TOPICS IN ANALYSIS

MANJUNATH KRISHNAPUR

Part-1: Questions of approximation

1. Weierstrass' approximation theorem

Unless we say otherwise, all our functions are allowed to be complex-valued. For e.g., $C[0,1]$ means the set of complex-valued continuous functions on $[0,1]$. When equipped with the sup-norm $\|f\|_{\sup} := \max\{|f(x)| : x \in [0,1]\}$, it becomes a Banach space. Weierstrass showed that polynomials are dense in $C[0,1]$.

Theorem 1 (Weierstrass). *If $f \in C[0,1]$ and $\epsilon > 0$ then there exists a polynomial P such that $\|f - P\|_{\sup} < \epsilon$. If f is real-valued, we may choose P to be real-valued.*

Bernstein's proof. Define $B_n^f(x) := \sum_{k=0}^n f(k/n)\binom{n}{k}x^k(1-x)^{n-k}$, called the Bernstein polynomial of degree n for the function f. Make the following observations about the coefficients $p_{n,x}(k) = \binom{n}{k}x^k(1-x)^{n-k}$.

$$\sum_{k=0}^n p_{n,x}(k) = 1, \qquad \sum_{k=0}^n kp_{n,x}(k) = nx, \qquad \sum_{k=0}^n (k-nx)^2 p_{n,x}(k) = nx(1-x),$$

all of which can be easily checked using the binomial theorem. In probabilistic language, $p_{n,x}$ is a probability distribution on $0,1,\ldots,n$ whose mean is nx and standard deviation is $nx(1-x)$. From these observations we immediately get

$$\sum_{k:|\frac{k}{n}-x|\ge\delta} p_{n,x}(k) \le \frac{1}{\delta^2 n^2}\sum_{k=0}^n (k-nx)^2 p_{n,x}(k) = \frac{x(1-x)}{n\delta^2}.$$

Thus, denoting $\omega_f(\delta) = \sup\limits_{|x-y|\le\delta} |f(x)-f(y)|$, we get

$$\begin{aligned}
|B_n^f(x) - f(x)| &\le \sum_{k\,:\,|\frac{k}{n}-x|<\delta} |f(x)-f(k/n)|p_{n,x}(k) + \sum_{k\,:\,|\frac{k}{n}-x|\ge\delta} |f(x)-f(k/n)|p_{n,x}(k) \\
&\le \omega_f(\delta)\sum_{k\,:\,|\frac{k}{n}-x|<\delta} p_{n,x}(k) + 2\|f\|_{\sup}\frac{x(1-x)}{n\delta^2} \\
&\le \omega_f(\delta) + \frac{1}{2n\delta^2}\|f\|_{\sup}.
\end{aligned}$$

First pick $\delta > 0$ so that $\omega_f(\delta) < \epsilon/2$ and then pick $n > \frac{\|f\|_{\sup}}{\epsilon\delta^2}$ to get $\|B_n^f - f\|_{\sup} < \epsilon$. ■

Here is another proof of Weierstrass' theorem, probably closer to the original proof. The idea is that real analytic functions (on an open neighbourhood of $[0, 1]$) are obviously uniformly approximable by polynomials (by truncating their power series), hence it suffices to show that any continuous function can be approximated uniformly by a real-analytic function. The key idea is of convolution.

Definition 2. For $f, g : \mathbb{R} \mapsto \mathbb{R}$, define $(f * g)(x) := \int_{\mathbb{R}} f(x-t)g(t)dt$, whenever the integral exists.

Exercise 3. (1) If f is bounded and measurable and g is (absolutely) integrable, then $f * g$ and $g * f$ are well-defined and equal.

(2) If f is bounded and measurable and g is smooth, then $f * g$ is smooth.

(3) If f is bounded and measurable and g is real-analytic and integrable, then $f * g$ is real-analytic.

Exercise 4. Complete the following steps. For definiteness, take all functions to be real-valued and defined on the whole of the real line.

(1) A real-analytic function can be uniformly approximated on compact sets by polynomials.

(2) If φ is a real-analytic probability density, then so is $\varphi_\sigma(x) := \frac{1}{\sigma}\varphi(x/\sigma)$.

(3) If f is a compactly supported continuous function, then $f * \varphi_\sigma$ is real analytic.

(4) As $\sigma \to 0$, we have $f * \varphi_\sigma \to f$ uniformly on compact sets.

(5) Deduce Weierstrass' theorem.

There are many examples of real-analytic probability densities. For example, (1) $\varphi(x) = \frac{1}{\pi(1+x^2)}$ (Cauchy density) and (2) $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ (normal density).

Remark 5. If the above densities are used, then the approximating functions $f * \varphi_\sigma$ used in the above exercise has a more special meaning.

(1) The Cauchy density $\varphi(x) = \frac{1}{\pi(1+x^2)}$. In this case, $(f * \varphi_y)(x) = u(x, y)$ where $u : \bar{\mathbb{H}} \to \mathbb{R}$ is the unique function that solves the Dirichlet problem on the upper-half plane $\mathbb{H} := \{(x, y) : y > 0\}$ with boundary condition f. What this means is that (a) u is continuous on $\bar{\mathbb{H}}$, (b) $u(\cdot, 0) = f(\cdot)$, (c) $\Delta u = 0$ on $\mathbb{H}$.

The point is that $(f * \varphi_y)$ is just u restricted to the line with y-co-ordinate equal to y and approaches f (at least pointwise) when $y \to 0$.

(2) The normal density $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. In this case $(f * \varphi_t)(x) = u(x, t)$ where u solves the heat equation with initial condition f. What this means is that (a) u is continuous on $\mathbb{R} \times \bar{\mathbb{R}}_+$, (b) $u(\cdot, 0) = f(\cdot)$, (c) $\frac{\partial}{\partial t}u(x, t) = \frac{1}{2}\frac{\partial^2}{\partial x^2}u(x, t)$ on $\mathbb{R} \times \mathbb{R}_+$.

Again, $(f * \varphi_{\sqrt{t}}) = u(\cdot, t)$ is the function ("temperature") at time t, and approaches the initial condition f (at least pointwise) as t approaches 0.

Some questions to think about. What about polynomials in m variables? Are they dense in the space $C(K)$ for $K \subseteq \mathbb{R}^m$? What about polynomials in one complex variable? Are they dense in the space $C(\bar{\mathbb{D}})$ where $\bar{\mathbb{D}}$ is the closed unit disk in the complex plane?

A somewhat challenging exercise.

Exercise 6. If $f : [0, \infty) \mapsto \mathbb{R}$ is continuous and $f(x) \to 0$ as $x \to +\infty$, then show that for any $\epsilon > 0$, there is polynomial p such that $|f(x) - p(x)e^{-x}| < \epsilon$ for all $x \geq 0$.

2. Fejér's theorem

Let S^1 denote the unit circle which we may identify with $[-\pi, \pi)$ using the map $\theta \mapsto e^{i\theta}$. Continuous functions on S^1 may be identified with continuous functions on $I = [-\pi, \pi]$ such that $f(-\pi) = f(\pi)$ or equivalently, with 2π-periodic continuous functions on $\mathbb{R}$.

Let $e_k(t) = e^{ikt}$ for $t \in [-\pi, \pi)$ (these are 2π-periodic as k is an integer). A supremely important fact is that e_k are orthonormal in $L^2(I, dt/2\pi)$, i.e., $\int_I e_k(t)\bar{e}_\ell(t)\frac{dt}{2\pi} = \delta_{k,\ell}$. The question of whether this is a complete orthonormal basis is answered to be "yes" by the following theorem. Note that the L^2 norm is dominated by the sup-norm, hence a dense subset of $C(S^1)$ is also dense in $L^2(S^1)$.

Theorem 1 (Fejér). *Given any $f \in C(S^1)$ and $\epsilon > 0$, there exists a trigonometric polynomial $P(e^{it}) = \sum_{k=-N}^{N} c_k e^{ikt}$ such that $\|f - P\|_{\sup} < \epsilon$.*

Proof. Define $\hat{f}(k) = \int_I f(t)e^{-ikt}\frac{dt}{2\pi}$ and set

$$\begin{aligned}\sigma_N f(t) &= \sum_{k=-N}^{N}\left(1 - \frac{|k|}{N+1}\right)\hat{f}(k)e^{ikt} \\ &= \sum_{k=-N}^{N}\left(1 - \frac{|k|}{N+1}\right)e^{ikt}\int_I f(s)e^{-iks}\frac{ds}{2\pi} \\ &= \int_I f(s)K_N(t-s)ds\end{aligned}$$

where the *Fejér kernel* K_N is defined as

$$K_N(u) = \sum_{k=-N}^{N}\left(1 - \frac{|k|}{N+1}\right)e^{iku} = \frac{1}{N+1}\frac{\sin^2\left(\frac{N+1}{2}u\right)}{\sin^2\left(\frac{u}{2}\right)}$$

The key observations about K_N (use the two forms of K_N whichever is convenient)

$$K_N(u) \geq 0 \text{ for all } u, \qquad \int\limits_I K_N(u)\frac{du}{2\pi} = 1, \qquad \int\limits_{I\setminus[-\delta,\delta]} K_N(u)\frac{du}{2\pi} \leq \frac{1}{N+1}\frac{1}{\sin^2(\delta/2)}.$$

In probabilistic language, $K_N(\cdot)$ is a probability density on I which puts most of its mass near 0 (for large N). Therefore,

$$\begin{aligned}|\sigma_N f(t) - f(t)| &\le \int_{-\delta}^{\delta} |f(t)-f(s)|K_N(t-s)ds + \int_{I\setminus[-\delta,\delta]} |f(t)-f(s)|K_N(t-s)ds \\ &\le \omega_f(\delta) + 2\|f\|_{\sup}\frac{1}{N+1}\frac{1}{\sin^2(\delta/2)}.\end{aligned}$$

Pick δ so that $\omega_f(\delta) < \epsilon/2$ and then pick $N+1 > \frac{4\|f\|_{\sup}}{\epsilon\sin^2(\delta/2)}$ to get $\|\sigma_N f - f\|_{\sup} < \epsilon$. ■

Some applications: In the following exercise, derive Weierstrass' theorem from Fejér's theorem.

Exercise 2. Let $f \in C_{\mathbb{R}}[0,1]$.

(1) Construct a function $g : [-\pi,\pi] \to \mathbb{R}$ such that (a) g is even, (b) $g = f$ on $[0,1]$ and (c) g vanishes outside $[-2,2]$.

(2) Invoke Fejér's theorem to get a trigonometric polynomials T such that $\|T-g\|_{\sup} < \epsilon$.

(3) Use the series $e^z = \sum_{k=0}^{\infty}\frac{1}{k!}z^k$ to replace the exponentials that appear in T by polynomials. Be clear about the uniform convergence issues.

(4) Conclude that there exists a polynomial P with *real* coefficients such that $\|f-P\|_{\sup} < 2\epsilon$..

A more interesting application is the theorem of Weyl that the set $\{nx \pmod 1\}$ is equidistributed in $[0,1]$ whenever x is irrational. You are guided to prove this statement in the following exercise.

Exercise 3. Let $x \in [0,1]$. Let $x_n = e^{2\pi i n x}$ and $S = \{x_1, x_2, \ldots\}$.

(1) Show that S is dense in S^1 if and only if x is irrational.

(2) If $f \in C(S^1)$, show that $\frac{1}{n}\sum_{k=1}^{n} f(x_k) \to \int_0^1 f(e^{it})\frac{dt}{2\pi}$. **[Hint:** First do it for $f(e^{it}) = e^{2\pi i p t}$ for some $p \in \mathbb{Z}$]

(3) For any arc $I = \{e^{it} : a < t < b\}$, show that as $n \to \infty$,

$$\frac{1}{n}\#\{k \le n : x_k \in I\} \to \frac{b-a}{2\pi}.$$

The point is that the points $x_1, x_2, \ldots$ spend the same amount of "time" in any arc of a given length. This is what we mean by *equidistribution*.

From Fejér's theorem, it follows that $\{e_n : n \in \mathbb{Z}\}$ is an orthonormal basis for $L^2(S^1)$, hence $S_n^f \to f$ in L^2 for every $f \in L^2$ and

$$\int_0^{2\pi} |f(t)|^2 \frac{dt}{2\pi} = \sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2 \qquad \text{(Plancherel identity)}.$$

But for $f \in C[0,1]$, the convergence need not be uniform (or even pointwise). But with extra smoothness assumption on f, one can achieve uniform convergence.

Exercise 4. Let $f \in C^2(S^1)$ (i.e., as a 2π-periodic function on $\mathbb{R}$, f is twice differentiable and f'' is continuous and 2π-periodic). Then, show that $S_n^f \to f$ uniformly. [**Hint:** Express Fourier coefficients of f' in terms of Fourier coefficients of f]

Question: Could we have proved this exercise first, and then used the density of $C^2(S^1)$ in $C(S^1)$ (in fact $C^\infty(S^1)$ is also dense in $C(S^1)$) to get an alternate proof of Fejér's theorem?

A brief history of Fejér's theorem: This is a cut-and-dried history, possibly inaccurate, but only meant to put things in perspective!

(1) The vibrating string problem is an important PDE that arose in mathematical physics, and asks for a function $u : [a,b]\times\bar{\mathbb{R}}_+ \to \mathbb{R}$ satisfying $\frac{\partial^2}{\partial t^2}u(x,t) = \frac{\partial^2}{\partial x^2}u(x,t)$ for $(x,t) \in (a,b)\times\mathbb{R}_+$ and satisfying the initial conditions $u(x,0) = f(x)$ and $\frac{\partial}{\partial t}u(x,t) \Big|_{\, t=0} = g(x)$, where f and g are specified initial conditions.

(2) Taking $[a,b] = [-\pi,\pi]$ (without loss of generality), it was observed that if $f(x) = e^{ikx}$ and $g(x) = e^{i\ell x}$, then $u(x,t) = \cos(kt)e^{ikx} + \frac{1}{\ell}\sin(\ell t)e^{i\ell x}$ solves the problem.

(3) Linearity of the system meant that if f and g are trigonometric polynomials, then by taking linear combinations of the above solution, one could obtain the solution to the vibrating string problem.

(4) Thus, the question arises, whether given f and g we can approximate them by trigonometric polynomials (and hopefully the corresponding solutions will be approximate solutions).

(5) Fourier made the fundamental observation that $e_k(\cdot)$ are orthonormal on $[-\pi,\pi]$ and deduced that if the notion of approximation is in mean-square sense (i.e., L^2 distance $\sqrt{\int |f-g|^2}$), then the best degree-n trigonometric polynomial approximation to f is

$$S_n f(x) := \sum_{k=-n}^{n} \hat{f}(k)e^{ikx}.$$

(6) Before Fejér, it was an open question whether $\|S_n f - f\|_{L^2} \to 0$ as $n \to \infty$. In other words, is $\{e_k\}_{k\in\mathbb{Z}}$ a complete orthonormal set for $L^2([-\pi,\pi])$?

(7) Since continuous functions are dense is $L^2[-\pi,\pi]$, it suffices to show that continuous functions can be uniformly approximated by trigonometric polynomials.

(8) In $C(S^1)$, it is no longer the case that $S_n f$ is the best approximation (in sup-norm sense). Fejér's innovative idea was to consider averages of $S_n f$, i.e., $\sigma_n f := \frac{1}{2n+1}\sum_{k=0}^{2n} S_k f$ (the same trigonometric polynomials that appeared in the proof) and show that they do converge to f uniformly.

References and further reading: Weierstrass' theorem is generalized to more abstract forms such as Stone-Weierstrass theorem.

Theorem 5 (Stone Weierstrass theorem). *Let X be a compact Hausdorff space and let $\mathcal{A} \subseteq C_{\mathbb{R}}(X)$. If $\mathcal{A}$ is (a) a real vector space, (b) closed under multiplication, (c) contains constant functions and (d) separates points of X. Then, $\mathcal{A}$ is dense in $C(X)$ in sup-norm.*

The separation condition means that for any distinct points $x, y \in X$, there is some $f \in \mathcal{A}$ such that $f(x) \neq f(y)$. If that were not the case, then no function that gives distinct values to x and y could be approximated by elements of $\mathcal{A}$. A proof based on Weierstrass' theorem can be found in most analysis books. Observe that the Stone-Weierstrass theorem aimplies Fejér's theorem too.

Weyl's equidistribution theorem mentioned here is the simplest one. Weyl showed also that for any real polynomial $p(\cdot)$, the sequence $\{e^{2\pi i p(n)} : n \geq 1\}$ is equidistribted in S^1 whenever at least one of the coefficients of p (other than the constant coefficient) is irrational.

Some references:

(1) B.Sury, *Weierstrass' theorem - leaving no stone unturned*, a nice expository article on Weierstrass' theorem available at `http://www.isibang.ac.in/~sury/hyderstone.pdf`.

(2) Rudin, *Principles of mathematical analysis* or Simmon's *Topology and modern analysis* for a proof of Stone-Weierstrass' theorem.

(3) Katznelson, *Harmonic analysis* or many other book on Fourier series for basics of Dirichlet and Fejér kernels.

3. MÜNTZ-SZASZ THEOREM IN L^2

Theorem 1. *Let $0 \leq n_1 < n_2 < \ldots$ and let $W = span\{x^{n_j} : j \geq 1\}$. Then, W is dense in $L^2[0,1]$ if and only if $\sum_j \frac{1}{n_j} = \infty$.*

Almost exactly the same criterion is necessary and sufficient for W to be dense in $C[0,1]$, except that for uniform approximation we must take $n_1 = 0$ (otherwise functions not vanishing at 0 cannot be approximated). In the above theorem, n_j are not required to be integers. From the above theorem, it is easy to deduce that if $\sum_j \frac{1}{n_j} < \infty$, then W cannot be dense in $C[0,1]$. This is simply beacause $\|f\|_{L^2} \leq \|f\|_{\text{sup}}$ for any $f \in C[0,1]$.

Some preliminaries in linear algebra: Let V be an inner product space over $\mathbb{R}$ and let $v_1, \ldots, v_k$ be elements of V. The Gram matrix of these vectors is the $k \times k$ matrix $A := (\langle v_i, v_j \rangle)_{i,j \leq k}$ whose entries are inner products of the given vectors. If $V = \mathbb{R}^k$ itself (the same k as the number of vectors), then $A = B^t B$ where $B = [v_1 \ldots v_k]$ is the $k \times k$ matrix whose columns are the given vectors. In this case, $\det(A) = \det(B)^2$ which is the squared volume of the parallelepiped formed by $v_1, \ldots, v_k$ (because $\det(B)$ is the signed volume of this parallelepiped). Convince yourself that even for general V, $\det(A)$ has the same meaning (but $\det(B)$ need not make sense, for example, if $V = \mathbb{R}^m$ with some $m > k$).

Now, let $u, v_1, \ldots, v_k$ be vectors in V. Let A be the Gram matrix of these $k+1$ vectors, and let B be the Gram matrix of $v_1, \ldots, v_k$. Using the above-mentioned volume interpretation of the determinants and the formula "volume = base volume × height" formula, we see that

$$\det(A) = \det(B) \times \text{dist.}^2(u, \text{span}\{v_1, \ldots, v_k\}).$$

Here the dist.^{2} term just means $\|P_W^\perp u\|^2$ where $P_W^\perp$ is the orthogonal projection to $W^\perp$ where $W = \text{span}\{v_1, \ldots, v_k\}$.

Example 2. Let $n_0, n_1, \ldots, n_k$ be distinct positive numbers and let $u = x^{n_0}, v_1 = x^{n_1}, \ldots, v_k = x^{n_k}$, all regarded as elements of $L^2[0,1]$. Let $W_k = \text{span}\{v_1, \ldots, v_k\}$. Then,

$$\text{dist.}^2(u, W_k) = \frac{\det(A)}{\det(B)}$$

where $A = \left(\frac{1}{n_i+n_j+1}\right)_{0\le i,j\le k}$ and $B = \left(\frac{1}{n_i+n_j+1}\right)_{1\le i,j\le k}$. The matrices here are called Hilbert matrices, and their determinants can be evaluated explicitly.

Cauchy determinant identity: Let $x_1, \ldots, x_k$ be distinct and $y_1, \ldots, y_k$ be distinct (we take them to be real numbers, but the same holds over any field). Then,

$$\det\left(\frac{1}{x_i+y_j}\right)_{1\le i,j\le k} = \frac{\prod\limits_{i<j}(x_i-x_j)(y_i-y_j)}{\prod\limits_{i,j}(x_i+y_j)}.$$

To see this, observe that

$$\prod_{i,j}(x_i+y_j)\det\left(\frac{1}{x_i+y_j}\right)_{1\le i,j\le k}$$

is a polynomial in x_is and y_js of degree at most n^2-n, and vanishes whenever two of the x_is are equal or two of the y_js are equal. Hence, the polynomial is divisible by $\prod\limits_{i<j}(x_i-x_j)(y_i-y_j)$. The latter is a polynomial of degree $n(n-1)$, hence we conclude that

$$\prod_{i,j}(x_i+y_j)\det\left(\frac{1}{x_i+y_j}\right)_{1\le i,j\le k} = C\prod_{i<j}(x_i-x_j)(y_i-y_j)$$

for some constant C. How to see that $C=1$?

Proof of Theorem 1. Let $0 \le n_0 \notin \{n_1, n_2, \ldots\}$ and set $u = x^{n_0}, v_1 = x^{n_1}, \ldots, v_k = x^{n_k}$, all regarded as elements of $L^2[0,1]$. As already explained,

$$\text{dist.}^2(u, W_k) = \frac{\det(A)}{\det(B)}$$

where $A = \left(\frac{1}{n_i+n_j+1}\right)_{0\le i,j\le k}$ and $B = \left(\frac{1}{n_i+n_j+1}\right)_{1\le i,j\le k}$. Cauchy's identity applies to both these determinants (take $x_i = y_i = n_i + \frac{1}{2}$) and hence, after canceling a lot of terms,

$$\begin{aligned}\text{dist.}^2(u, W_k) &= \frac{\prod_{i=1}^{k}(n_0-n_j)^2}{(2n_0+1)\prod_{j=1}^{k}(n_0+n_j+1)^2} = \frac{1}{2n_0+1}\prod_{j=1}^{k}\left(1-\frac{n_0}{n_j}\right)^2\left(1+\frac{n_0+1}{n_j}\right)^{-2} \\ &= \frac{1}{2n_0+1}\prod_{j=1}^{k}\left(1-\frac{A}{n_j}+O\left(\frac{1}{n_j^2}\right)\right)\end{aligned}$$

where $A \neq 0$. Recall that for $0 < x_j < 1$, the infinite product $\prod_{j=1}^{\infty}(1-x_j)$ is positive if $\sum_j x_j < \infty$ and zero if $\sum_j x_j = \infty$. From this, it immediately follows that

$$\lim_{k\to\infty} \text{dist.}^2(u, W_k) = 0 \text{ if and only if } \sum_j \frac{1}{n_j} = \infty.$$

Thus, if $\sum_j \frac{1}{n_j} < \infty$, then if we take $n_0 \notin \{n_1, n_2, \ldots\}$, it follows that x^{n_0} is not in the closed span of $W = \{x^{n_j} : j \ge 1\}$. In particular, W is not dense in $L^2[0,1]$.

Conversely, if $\sum_j \frac{1}{n_j} = \infty$, then for any $n_0 > 0$, we see that $x^{n_0} \in \bar{W}$. Thus all polynomials are in $\bar{W}$ which shows that $\bar{W} = L^2[0,1]$. ■

Hilbert arrived at the Hilbert matrix in studying the following question. How closely can x^n be approximated (in $L^2[0,1]$) by polynomials of lower degree? This just means finding

$$r_n = \text{dist.}(x^n, \text{span}\{1, x, \ldots, x^{n-1}\}).$$

The corresponding question in $C[0,1]$ is much deeper and was (asked and) answered by Chebyshev. We shall see it later.

Exercise 3. Find an explicit form of r_n. How big or small (i.e., the order of decay/growth) is it?

4. MÜNTZ-SZASZ THEOREM IN $C[0,1]$

Theorem 1. *Let $0 = n_0 < n_1 < n_2 < \ldots$ and let $W = span\{x^{n_j} : j \ge 0\}$. Then, W is dense in $C[0,1]$ if and only if $\sum_j \frac{1}{n_j} = \infty$.*

Sketch of the proof. Let $W = \text{span}\{x^{n_j} : j \ge 0\}$. Recall that W is not dense in $C[0,1]$ if and only if there is a non-zero bounded linear functional on $C[0,1]$ that vanishes on W (by Hahn-Banach theorem). We know that the dual of $C[0,1]$ is the space of all complex Borel measures on $[0,1]$, acting by $f \mapsto \int_{[0,1]} f d\mu$ (one of F. Riesz's many representation theorems). Thus, W is not dense if and only if we can find a complex Borel measure μ on $[0,1]$ such that $\int t^{n_j} d\mu(t) = 0$ for all j.

For any μ, consider the function $F_\mu(z) = \int t^z d\mu(t)$. This is a holomorphic function on the right half-plane. A question is whether it can vanish at n_j for all j, without μ being identically zero. A holomorphic function can have no accumulation points inside the domain of holomorphicity, but there is no restriction on vanishing at a sequence of points that go to the boundary (or infinity).

However, if there are some bounds on the growth of the holomorphic function, then its sequence of zeros must approach the boundary sufficiently fast.

We skip details for now, but what it amounts to is that when $\sum_j \frac{1}{n_j} = \infty$, such functions do not exist. Consequently W is dense in $C[0,1]$. ■

5. Mergelyan's theorem

On a compact subset K of the complex plane, what functions can be uniformly approximated by polynomials? Two examples to show what can go wrong.

Let $K = \bar{\mathbb{D}} = \{z : |z| \le 1\}$. Then $\bar{z}$ cannot be uniformly approximated by polynomials. This is because a uniform limit of polynomials must be holomorphic in the open disk $\mathbb{D}$. Thus not all continuous functions can be uniformly approximated by polynomials.

What about analytic functions? If $K = 2\bar{\mathbb{D}} \setminus \mathbb{D} = \{z : 1 \le |z| \le 2\}$, then the function $1/z$ cannot be uniformly approximated on K by polynomials. This is because polynomials integrate to zero on contours in the interior of K, but $\int_\gamma \frac{1}{z} dz \neq 0$ if γ has non-zero winding around 0.

Mergelyan's theorem gives the complete answer to the question. Let $\mathcal{A}(K)$ be the space of continuous functions on K that are holomorphic in the interior of K. Endow it with the sup-norm on K.

Theorem 1 (Mergelyan)**.** *Let K be a compact subset in the complex plane such that $\mathbb{C} \setminus K$ has finitely many connected components. Choose points $p_1, \ldots, p_m$, one in each of the bounded components of $\mathbb{C} \setminus K$. Let $\mathcal{R}$ be the collection of all rational functions whose poles are contained in $\{p_1, \ldots, p_m\}$.*

Then $\mathcal{R}$ is dense in $\mathcal{A}(K)$. In particular, if $\mathbb{C} \setminus K$ is connected, then polynomials are dense in $A(K)$.

For example, if K is the closure of a bounded simply connected region, then all continuous functions that are holmorphic in the interior can be approximated uniformly by polynomials. If $K = [0,1]$ (or any curve $\gamma : [0,1] \mapsto \mathbb{C}$ that is injective), then the interior is empty and $\mathbb{C} \setminus K$ is connected. Hence the analyticity condition is superfluous and all continuous functions are approximable by polynomials. If $K = S^1$, again the interior is empty but $\mathbb{C} \setminus K$ has one bounded component $\mathbb{D}$. Taking $p = 0$ and applying Mergelyan's theorem gives us Fejér's theorem.

As the proof of Mergelyan's theorem uses certain advanced theorems in complex analysis, we shall postpone it to later.

6. Chebyshev's approximation question

For $f \in C[0,1]$, $n \ge 1$ and $a < b$, let

$$\gamma(f, n, a, b) := \inf\{\|f - p\|_{\text{sup}} : p \text{ is a polynomial of degree at most } n\}.$$

Weierstrass' theorem is the statement that $\gamma(f, n) \to 0$ as $n \to \infty$. But what is the rate at which it goes to zero? Equivalently, for a given n, how good is the approximation? Try to work out the bound you get from the proofs we gave of Weierstrass' theorem. In a landmark paper, Chebyshev showed that $\gamma(x^n, n-1, -1, 1) = 2^{-n+1}$.

For instance, if we use x^{n-1} to approximate x^n, then,

$$\|x^n - x^{n-1}\|_{\sup[-1,1]} \geq \left(1 - \frac{1}{n}\right)^n - \left(1 - \frac{1}{n}\right)^{n-1} = \left(1 - \frac{1}{n}\right)^{n-1} \frac{1}{n} \sim e^{-1}n^{-1}.$$

But we shall see that it is possible to find degree $n-1$ polynomials that are exponentially close to x^n. To do this, he introduced an immortal class of polynomials, now known as *Chebyshev polynomials* (of the first kind).

In basic trigonometry, we see that

$$\cos(2\theta) = 2\cos^2(\theta) - 1, \quad \cos(3\theta) = 4\cos^2(\theta) - 3\cos(\theta).$$

It is not hard to see that in general, $\cos(n\theta)$ is a polynomial of degree n in $\cos(\theta)$. Thus, $\cos(n\theta) = T_n(\cos\theta)$, where T_n is defined to be the nth Chebysev polynomial (of the first kind). For instance, $T_2(x) = 2x^2 - 1$ and $T_3(x) = 4x^3 - 3x$.

By the identity $\cos((n+1)\theta) + \cos((n-1)\theta) = 2\cos(\theta)\cos(n\theta)$, we see that $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$. In fact, this recursion, together with the specification $T_0(x) = 1$ and $T_1(x) = x$, could be taken as an alternative definition of the Chebyshev polynomials.

Some easy properties: T_n has degree n. The coefficient of x^k in T_n is zero unless $n-k$ is even. The highest coefficient of T_n is 2^{n-1}. Lastly $\|T_n\|_{\sup[-1,1]} = 1$.

Consequently, $|x^n - p(x)| \leq 1$ for all $x \in [-1,1]$, where $p(x) = x^n - 2^{-n+1}T_n(x)$ is a polynomial of degree $n-1$. Therefore, $\gamma(x^n, n-1, -1, 1) \leq 2^{-n+1}$. Chebyshev's theorem is that $2^{-n+1}T_n$ is the best approximation to x^n among lower degree polynomials. We shall prove it shortly.

A less obvious looking property of Chebyshev polynomials is that they are orthogonal w.r.t. the arcsine measure $d\mu(x) = \frac{1}{\pi\sqrt{1-x^2}}dx$ on $[-1,1]$. That is,

$$\int_{-1}^{1} T_n(x)T_m(x)\frac{1}{\pi\sqrt{1-x^2}}dx = 0 \quad \text{if } m \neq n.$$

To do this without calculations, define the map $\varphi : S^1 \mapsto [-1,1]$ by $\varphi(e^{i\theta}) = \cos\theta$. The arcsine measure is precisely the push-forward of the normalized Lebesgue measure on S^1 under φ. Further, $T_k \circ \varphi = \mathrm{Re}\{e_k\}$. From this, it easily follows that

$$\int_{-1}^{1} T_m(x)T_n(x)\frac{1}{\pi\sqrt{1-x^2}}dx = \begin{cases} 1 & \text{if } m = n = 0 \\ \frac{1}{2} & \text{if } m = n \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Relevance to the approximation question: Any $x \in [-1,1]$ can be written as $x = \cos\theta$, hence $T_n(x) = \cos(n\theta) \in [-1,1]$. Thus, $\|T_n\|_{\sup} = 1$. The monic polynomial $2^{-n+1}T_n$ has sup-norm 2^{-n+1} in $[-1,1]$. Write $p(x) = x^n - 2^{-n+1}T_n(x)$, a polynomial of degree $n-2$. Then,

$$\|x^n - p_n(x)\|_{\sup[-1,1]} = 2^{-n+1}\|T_n\|_{\sup[-1,1]} = 2^{-n+1}.$$

This is much smaller than the $1/n$ that we got earlier by approximating x^n by x^{n-1}. We next show that this is the best possible.

The oscillation idea: Let $f \in C[-1,1]$, let p be a polynomial of degree n, and $\delta > 0$. Suppose $f - p$ oscillates between $\pm\delta$ many times, say m. By this we mean that there exist $x_1 < x_2 < \ldots < x_{m+1}$ in $[-1,1]$ such that $|f(x_j) - p(x_j)| \geq \delta$ for all j and $\text{sgn}(f(x_j) - p(x_j))$ alternates between $+1$ and -1 as j runs from 1 to $m+1$.

Now suppose q is another polynomial of degree n and $\|f - q\|_{\text{sup}} < \delta$. Then, $\text{sgn}(q(x_j) - p(x_j))$ alternates between $+1$ and -1 as j runs from 1 to $m+1$. Indeed, suppose $f(x_1) - p(x_1) \geq \delta$ and $f(x_2) - p(x_2) \leq -\delta$. Then $q(x_1) > f(x_1) - \delta \geq p(x_1)$ and $q(x_2) < f(x_2) + \delta \leq p(x_2)$. This shows that $q - p$ must have at least m roots, one in (x_j, x_{j+1}) for each $1 \leq j \leq m$.

If $m \geq n+1$, this is not possible, as $q - p$ has degree at most n. The way out of the contradiction is that $\|f - q\|_{\text{sup}[-1,1]} \geq \delta$ for every degree n polynomial q. We collect this conclusion as a lemma below.

Lemma 1. *If $f \in C[0,1]$, p is a polynomial of degree n, and there exist $n+2$ points $x_1 < x_2 < \ldots < x_{n+2}$ in $[-1,1]$ such that $|f(x_j) - p(x_j)| \geq \delta$ for all j and sgn$(f(x_j) - p(x_j))$ alternates between $+1$ and -1 as j runs from 1 to $n+2$. Then, for any polynomial q of degree n or less, $\|f - q\|_{\text{sup}[-1,1]} \geq \delta$.*

Chebyshev polynomial is the best approximation to the monomial: Let $f(x) = x^n$, $p(x) = x^n - 2^{-n+1}T_n(x)$ (a degree $n-1$ polynomial) and $\delta = 2^{-n+1}$. Note that $f(x) - p(x) = 2^{-n+1}T_n(x)$. Write $x = \cos\theta$ and let θ range over $[0,\pi]$ (so x runs through $[-1,1]$). Recall that $T_n(\cos\theta) = \cos(n\theta)$, take $\theta_k = k\pi/n$ for $k = 0, 1, \ldots, n$, and note that $T_n(\cos\theta_k) = (-1)^k$. Thus, $f - p$ alternates $n+1$ times between $\pm\delta$. From Lemma 1, we conclude that for any polynomial q of degree $n-1$ or less, $\|f - q\|_{\text{sup}[-1,1]} \geq 2^{-n+1}$.

An application: We have found the best way to approximate x^n by a polynomial of lower degree. In principle, replacing the highest power in the approximating polynomial by a lower degree Chebyshev polynomial, and continuing, it should be possible to reduce the degree of the approximating polynomial while keeping a reasonable level of approximation. This raises the question,[1] how small a degree m can we take and still approximate x^n (on $[-1,1]$) by a degree m polynomial?

First we find the expansion of x^n in terms of Chebyshev polynomials (this is obviously possible since T_k has degree k for each k). If

$$x^n = \sum_{k=0}^{n} c_{n,k} T_k(x),$$

[1] This part of the notes is taken from Nisheet Vishnoi's notes which contains more on these problems and their uses in algorithms. The derivation of expansion of x^n in terms of Chebyshev polynomials given here was suggested by Chaitanya Tappu, and is more natural than what we did in class.

then we must have

$$c_{n,0} = \int_{-1}^{1} x^n T_0(x) \frac{dx}{\pi\sqrt{1-x^2}} dx,$$

$$c_{n,k} = 2\int_{-1}^{1} x^n T_k(x) \frac{dx}{\pi\sqrt{1-x^2}} dx \quad \text{for } k \geq 1,$$

by the orthogonality of T_ks with respect to the arcsine measure. Make the change of variables $x = \cos\theta$ to write

$$\begin{aligned}\int_{-1}^{1} x^n T_k(x) \frac{dx}{\sqrt{\pi(1-x^2)}} dx &= \frac{1}{\pi}\int_0^{\pi} \cos^n\theta \times \cos(k\theta) d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{e^{i\theta}+e^{-i\theta}}{2}\right)^n \left(\frac{e^{ik\theta}+e^{-ik\theta}}{2}\right) d\theta \\ &= \frac{1}{2^n}\binom{n}{\frac{n+k}{2}}.\end{aligned}$$

To see the last inequality, expand $(e^{i\theta}+e^{-i\theta})^n$ and observe that only the terms with $e^{\pm ik\theta}$ give a non-zero integral. There are two of them, each with the binomial coefficient. If we write $p_{n,k} = 2^{-n}\binom{n}{(n+k)/2}$ for $k = n, n-2, \ldots, -n+2, -n$, then $c_{n,0} = p_{n,0}$ and $c_{n,k} = p_{n,k} + p_{n,-k}$.

The basic observation is that $(p_{n,k})_k$ is the Binomial probability distribution (the distribution of $\xi_1 + \ldots + \xi_n$, where ξ_k are independent and equal to ± 1 with probability $1/2$ each). Most of the mass of this distribution is concentrated in $|k| \lesssim \sqrt{n}$. For example, the Chebyshev inequality gives

$$\sum_{k:|k|\geq d} p_{n,k} \leq \frac{n}{4d^2}$$

which becomes small when $d >> \sqrt{n}$. A better bound is the following

$$\text{Bernstein/Chernoff bound:} \sum_{k:|k|\geq d} p_{n,k} \leq 2e^{-d^2/2n}.$$

Thus, if we set $P_{n,d}(x) = \sum_{k=0}^{d} c_{n,k}T_k(x)$, then

$$\begin{aligned}\|x^n - P_{n,d}\|_{\sup[-1,1]} &= \Big|\sum_{k=d+1}^{n} c_{n,k}T_k(x)\Big| \\ &\leq 2\sum_{k:k>d} p_{n,k} \qquad (\text{as } \|T_k\|_{\sup[-1,1]} = 1) \\ &\leq 2e^{-d^2/2n}\end{aligned}$$

by the Chernoff bound. Thus, we can approximate x^n well by polynomials of degree d provided d is much larger than $\sqrt{n}$. For example, if $d = \sqrt{2Bn\log n}$, then $\|x^n - P_{n,d}\|_{\sup[-1,1]} \leq n^{-B}$.

This finishes our discussion of approximation questions. Much more can be found in the references we mentioned earlier.

Exercise 2. Show that $T_n(x) = \frac{(-1)^n}{1\times 3\times\ldots\times(2n-1)}\sqrt{1-x^2}\frac{d^n}{dx^n}(1-x^2)^{n-\frac{1}{2}}$.

One can also take this as the definition and prove the other properties (recursion, orthogonality, etc.). The following exercise introduces Chebyshev polynomials of the second kind.

Exercise 3. Argue that $\frac{\sin((n+1)\theta)}{\sin\theta}$ is a polynomial of $\cos\theta$. Hence define the polynomials U_n, $n \geq 0$ by $U_n(\cos\theta) = \frac{\sin((n+1)\theta)}{\sin\theta}$. Show that (1) $U_n(x) = \frac{1}{n+1}T'_{n+1}(x)$, (2) $\int_{-1}^{1} U_n(x)U_m(x)\sqrt{1-x^2}dx = 0$ if $m \neq n$.

Part-2: Moment problems

1. Moment problems

If μ is a measure on $\mathbb{R}$, the number $\alpha_k = \int x^k d\mu(x)$ is said to be its kth moment, if it exists. Throughout this section, we work with measures for which all moments do exist. In particular, all measures will be finite, and often we normalize them to be probability measures.

The moment problem: Given a sequence $\alpha = (\alpha_0, \alpha_1, \ldots)$ of real numbers, does there exist a Borel measure on $\mathbb{R}$ whose nth moment is α_n? Is it unique? What if the measure is restricted to the half-line $[0, \infty)$ or to an interval?[2]

Necessary condition: The integral of a positive function against a measure is positive. Suppose α is the moment sequence of a measure μ whose support is the closed set $I \subseteq \mathbb{R}$. Then, for any (real) polynomial $p(x) = \sum_{j=0}^{n} c_j x^j$ such that $p(x) \geq 0$ for all $x \in I$, we must have $\int p(x) d\mu(x) \geq 0$. Since $\int p(x) d\mu(x) = \sum_{j=0}^{n} c_j \alpha_j$, writing $L(p) := \sum_{j=0}^{n} c_j \alpha_j$, we see that

$$L(p) \geq 0 \quad \text{whenever} \quad p(x) \geq 0 \text{ for all } x \in I. \tag{1}$$

The first main theorem is that this condition is also sufficient.

Theorem 1 (Existence part of the moment problem)**.** *Let I be a closed subset of $\mathbb{R}$ and let $\alpha = (\alpha_0, \alpha_1, \ldots)$ be a sequence of real numbers. There exists a measure μ on I such that $\int x^k d\mu(x) = \alpha_k$ for all $k \geq 0$ if and only if the positivity condition* (1) *holds.*

We shall prove this in the next section. For now, we take I to be an interval and find more tractable conditions for (1) to hold. The question is, what polynomials are positive on I?

The whole line: Let $I = \mathbb{R}$. Write

$$p(x) = a_n \prod_{j=1}^{k} (x - t_j) \prod_{j=1}^{\ell} (x - z_j)(x - \bar{z}_j)$$

where $k, \ell \geq 0$ and $t_j \in \mathbb{R}$ and $z_j \in \mathbb{C} \setminus \mathbb{R}$. Let $q(x) = \prod_{j=1}^{\ell} (x - z_j)$ (a complex polynomial), so that $p(x) = a_n |q(x)|^2 \prod_{j=1}^{k} (x - t_j)$. Thus p is positive on $\mathbb{R}$ if and only if $a_n > 0$ (take $x \to +\infty$ to see this) and each distinct real root of p occurs with even multiplicity. Then, letting $q = q_1 + i q_2$ and $r(x)^2 = \prod_{j=1}^{k} (x - t_j)$,

$$p = (\sqrt{a_n} r q_1)^2 + (\sqrt{a_n} r q_2)^2.$$

Conversely, any such polynomial is positive on $\mathbb{R}$.

[2] There are various names, such as the Hamburger moment problem, the Stieltjes' moment problem, Hausdorff moment problem, etc.

Thus, in condition (1) it suffices to take p to be the square of another polynomial and thus the condition becomes $L(p^2) \geq 0$ for all $p \in \mathcal{P}$. Writing $p(x) = \sum_{j=0}^{n} c_j x^j$, this can be written in terms of the sequence α as

$$0 \leq \sum_{j,k=0}^{n} c_j c_k \alpha_{j+k} \quad \text{for all } n \geq 1 \text{ and } c_0, \ldots, c_n \in \mathbb{R}. \tag{2}$$

This can also be phrased as saying that the infinite matrix $H_\alpha = (\alpha_{i+j})_{i,j \geq 0}$ is positive semi-definite (meaning that $\det[(H_\alpha(i,j))_{0 \leq i,j \leq n}] \geq 0$ for all $n \geq 0$).

Half-line: Let $I = [0, \infty)$. Going by the same logic as before, we see that if p is positive on $[0, \infty)$, then all its real roots in $(0, \infty)$ must have even multiplicity, but the negative roots are not restricted. Hence

$$p(x) = q(x) \prod_{j=1}^{m} (x + t_j)$$

where $q(x) \geq 0$ for all $x \in \mathbb{R}$. Expanding the product further, and writing $q = q_1^2 + q_2^2$, we see that $p(x)$ is a positive linear combination of polynomials of the form $x^k r(x)^2$ where r is a real polynomial and $k \geq 0$. Since even powers of x can be absorbed into r, we see that any polynomial positive on $[0, \infty)$ is a linear combination (with positive coefficients) of polynomials of the form r^2 and $xr^2(x)$. Thus the condition (1) is equivalent to

$$L(p^2) \geq 0 \text{ and } L(xp^2(x)) \geq 0 \quad \text{for all} \quad p \in \mathcal{P}.$$

Again, writing $p(x) = \sum_{j=0}^{n} c_j x^j$, we can write these conditions as

$$0 \leq \sum_{j,k=0}^{n} c_j c_k \alpha_{j+k} \quad \text{and} \quad 0 \leq \sum_{j,k=0}^{n} c_j c_k \alpha_{j+k+1} \quad \text{for all } n \geq 1 \text{ and } c_0, \ldots, c_n \in \mathbb{R}.$$

This is equivalent to positivity of the determinants of $(H_\alpha(i,j))_{0 \leq i,j \leq n}$ and $(H_\alpha(i,j))_{0 \leq i \leq n, 1 \leq j \leq n+1}$.

Compact interval: Let $I = [0, 1]$. We claim that if $p \geq 0$ on I, then it can be written as a positive linear combination of the polynomials $x^k(1-x)^\ell$ for $k, \ell \geq 0$. Accepting this claim, the condition (1) becomes equivalent to

$$L(x^k(1-x)^\ell) \geq 0 \quad \text{for all } k, \ell \geq 0.$$

Expanding $(1-x)^\ell$, this is the same as

$$\sum_{j=0}^{\ell} \binom{\ell}{j} (-1)^j \alpha_{k+j} \geq 0$$

There is a nice way to express this. Define the difference operator from sequences to sequences by $(\Delta c)_k = c_{k+1} - c_k$. Then the above conditions can also be written succinctly as $(-1)^p (\Delta^p \alpha)_k \geq 0$ for all p, k (the original sequence is positive, the differences are negative, second differences are positive, etc.).

Exercise 2. (1) Prove the claim. (2) Prove the equivalence of the derived condition to alternating signs of successive differences.

2. Some theorems similar in spirit to the moment problems

There are other theorems that one sees in analysis that are similar in spirit to the moment problem. We mention a couple of them in this section. These will not be required in future sections.

Theorem 3 (Riesz's representation theorem.). *Let X be a locally compact Hausdorff space. Let $L : C_c(X) \mapsto \mathbb{R}$ be a linear functional. Then, there exists a (regular) Borel measure μ on X such that $Lf = \int f d\mu$ for all $f \in C_c(X)$ if and only if L is positive (i.e., $L(f) \geq 0$ whenever $f \geq 0$).*

Presumably Riesz had the solutions to the moment problems in mind when he formulated this theorem. But the solutions to the moment problems cannot be deduced directly from Riesz's representation.

Other questions with the same flavour as the moment problem are as follows: We are given a linear functional on a subspace of continuous functions. The problem being to determine if it comes from a measure. Here are two concrete problems of interest.

Example 4. For a measure μ on S^1, we define its Fourier coefficients $\hat{\mu}(k) = \int e_{-k} d\mu$. The question "what sequences of complex numbers can arise as the Fourier coefficients of a measure?" is clearly very similar in spirit to the moment problem.

Given $\alpha = (\alpha_k)_{k \in \mathbb{Z}}$, a necessary condition for α to be the Fourier coefficients of a measure is that for any trigonometric polynomial $p = \sum_{k=-n}^{n} c_k e_k$,

$$0 \leq \sum_{j,k=-n}^{n} c_j \bar{c}_k \alpha_{j-k}.$$

We leave it to you to figure why. The non-trivial point is that these conditions are also sufficient. Here uniqueness of the measure comes for free!

Example 5. For a finite measure μ on $\mathbb{R}$, define its Fourier transform $\hat{\mu} : \mathbb{R} \mapsto \mathbb{C}$ by $\hat{\mu}(t) = \int_{\mathbb{R}} e_{-t} d\mu$ where $e_t(x) = e^{itx}$. Given a function $f : \mathbb{R} \mapsto \mathbb{C}$, is it the Fourier transform of a finite measure? Two necessary conditions are

(1) f is continuous

(2) $\sum_{j,k=1}^{n} c_j \bar{c}_k \hat{\mu}(t_j - t_k) \geq 0$ for all $n \geq 1$, all $c_1, \ldots, c_n \in \mathbb{C}$ and all $t_1, \ldots, t_n \in \mathbb{R}$.

Bochner's theorem asserts that these conditions are also sufficient. Again, uniqueness holds, in contrast to moment problem on the whole line.

Given the similarity between the moment problems, the theorems on Fourier transforms, and the Riesz representation theorem, one might prefer a more abstract statement that captures the general situation. We give one such theorem due to M. Riesz.

3. Marcel Riesz's extension theorem

Before we state the theorem, recall that a cone in a real vector space is a set that is closed under multiplication by positive scalars. A convex cone is a cone that is closed under convex combination of its elements. This is the same as a cone that is closed under addition of its elements. For example, the first quadrant is a convex cone in $\mathbb{R}^2$.

Theorem 6 (M. Riesz's extension theorem)**.** *Let W be a subspace of a real vector space X. Let K be a convex cone in X such that ~~span(K) + W = X~~ $K + W = X$. Let $L : W \mapsto \mathbb{R}$ be a linear functional such that $L(v) \geq 0$ for all $v \in W \cap K$. Then, there exists a linear functional $\tilde{L} : X \mapsto \mathbb{R}$ such that $\tilde{L}(v) = L(v)$ for all $v \in W$ and $\tilde{L}(v) \geq 0$ for all $v \in K$.*

The proof will be reminiscent of the proof of Hahn-Banach theorem that you have seen in Functional analysis class. Historically, perhaps both that proof and this were arrived at in stages, by polishing and making more abstract the solutions of the moment problems.

Proof. If $K \subseteq W$, then $W = X$ and there is nothing to prove. Otherwise, pick $u \in K \setminus W$ and let $W' = W + \mathbb{R}u$, a subspace strictly larger than W. We show that it is possible to extend L to W' so that it is positive on $K \cap W'$. There is no choice but to define the extension as $L'(w+\alpha u) = L(w)+\alpha t$ for some $t \in \mathbb{R}$. The only freedom is in t, and we must choose it so that $L(w) + \alpha t \geq 0$ whenever $w \in W$ and $w + \alpha u \in K$. It is enough to check this condition for $\alpha = \pm 1$, since K and W are both closed under multiplication by $|\alpha|$. Thus, the conditions for positivity of L' are precisely that

$$L(w) + t \geq 0 \text{ for } w \in W \cap (K - u), \quad \text{and} \quad L(w) - t \geq 0 \text{ for } w \in W \cap (K + u).$$

We may rewrite this as

$$-L(w_1) \leq t \leq L(w_2) \text{ for all } w_1 \in W \cap (K - u) \text{ and } w_2 \in W \cap (K + u).$$

Such a choice of t is possible if and only if $-L(w_1) \leq L(w_2)$ for all $w_1 \in W \cap (K + u)$ and $w_2 \in W \cap (K - u)$. But if $w_1 \in W \cap (K - u)$ and $w_2 \in W \cap (K + u)$, then $w_2 + w_1 \in K \cap W$ and hence $L(w_2) + L(w_1) = L(w_2 + w_1) \geq 0$ by the positivity of L. This completes the proof that a positive (on K) linear functional on a subspace can be extended to a subspace got by adding one new element.

Gap in the proof: We have not checked if $W \cap (K + u)$ is non-empty (note that $W \cap (K - u)$ cannot be empty as it contains the zero vector). If $W \cap (K + u)$ is empty, it may happen that the value of t obtained above is $+\infty$. This is possible if we only assume that $\text{span}(K) + W = X$ as we originally did. But if we assume that $K + W = X$, then $-u = w + k$ for some $w \in W$ and $k \in K$, and hence, $w \in W \cap (K + u)$, ensuring the non-emptyness of $W \cap (K + u)$.

The rest is the usual Zorn's lemma ritual. Consider all positive (on K) extensions of L, i.e., tuples $(\hat{W}, \hat{L})$ such that $\hat{W}$ is a subspace containing W, $\hat{L}$ is a positive (on K) linear functional on $\hat{W}$ that extends L. This set is partially ordered by the order $(W_1, L_1) \leq (W_2, L_2)$ if $W_1 \subseteq W_2$ and L_2 is an extension of L_1. Given a totally ordered subset (a chain) $\{(W_i, L_i)\}$, it is clear that $(\cup_i W_i, \vee_i L_i)$

(how are they defined?) is a maximal element of the chain. Applying Zorn's lemma, we get a maximal element (W_0, L_0). If $W_0 \neq X$, then as above, it is possible to extend L_0 to a strictly larger subspace while preserving positivity on K, contradicting the maximality of (W_0, L_0). Thus, $W_0 = X$ and the theorem is proved. ■

Remark 7. Here is a standard example to show that the hypothesis $K+W = X$ cannot be replaced by $\text{span}(K)+W = X$. Let $X = \mathbb{R}^2$, $W = \{(x,0) : x \in \mathbb{R}\}$, and $K = \{(x,y) : y > 0\}$. Let $L(x,0) = x$, a linear functional positive on K (tautologically, since $W \cap K = \emptyset$). Any extension must look like $\tilde{L}(x,y) = x + ty$ for some $t \in \mathbb{R}$. But then $\tilde{L}(-2t,1) = -t$ while $\tilde{L}(0,1) = t$, showing that both cannot be positive although $(0,1)$ and $(-2t,1)$ are both in K.

Remark 8. (For those who feel a pang of uneasiness when using Zorn's lemma). If there are countably many elements $u_1, u_2, \ldots$ in K such that $X = W + \text{span}\{u_1, u_2, \ldots\}$, then a simple induction argument may be used in place of Zorn's lemma. In many applications this suffices.

As a corollary, we derive solutions to the moment problems. Just to illustrate the idea, we first deal with the case when I is compact.

Theorem 9. *Let $I \subseteq \mathbb{R}$ be a non-empty compact set, and let $\alpha = (\alpha_0, \alpha_1, \ldots)$ be a sequence of real numbers such that if a polynomial $p(x) = \sum_{j=0}^{n} c_j x^j \geq 0$ for all $x \in I$, then $\sum_{j=0}^{n} \alpha_j c_j \geq 0$. Then, there is a Borel measure μ on I such that α_n is the nth moment of μ for every n.*

Proof. Let $X = C[0,1]$, $W = \mathcal{P}$, $K = \{f \in C[0,1] : f \geq 0\}$. It is clear that W is a subspace of X and K is a convex cone. To see that $W + K = X$, write any $f \in C[0,1]$ as $(f + \|f\|_{\sup(I)}) - \|f\|_{\sup(I)}$. The first summand is in K while the second is in $\mathcal{P}$ (being a constant!). Hence L extends to a positive linear functional on $C[0,1]$, which, by Riesz's representation theorem is represented by integration with respect to a Borel measure on $[0,1]$. ■

Remark 10. In the above theorem, uniqueness of the measure is easy to prove. This is because polynomials are dense in $C(I)$, by Weierstrass' theorem. Hence, the extension has to be unique (two bounded linear functionals on a Banach space that agree on a dense subset must agree everywhere). Uniqueness is not true in general, and not easy to prove when it is, for non-compact domains.

The use of Riesz's representation was a little extravagant, but employed to make the point quickly. We now give a direct argument that works for unbounded sets also.

Theorem 11. *If I be a closed subset of $\mathbb{R}$. Let $\alpha = (\alpha_0, \alpha_1, \ldots)$ be a real sequence. Define $L(p) = \sum_{j=0}^{n} c_j \alpha_j$ for any polynomial $p(x) = \sum_{j=0}^{n} c_j x^j$. If $L(p) \geq 0$ whenever $p \geq 0$ on I, then there exists a Borel measure μ such that $\alpha_n = \int x^n d\mu(x)$ for all n.*

Proof of the Theorem for the case $I = \mathbb{R}$. **Step 1:** To apply M. Riesz's extension theorem, let (here $\mathbf{1}_{(-\infty,\infty]}$ just means the constant function $\mathbf{1}$)

$$W = \mathcal{P}, \quad V = \text{span}\{\mathbf{1}_{(-\infty,b]} : b \in \mathbb{R} \cup \{\infty\}\}, \quad X = W + V, \quad K = \{f \in X : f \geq 0\},$$

and $L : W \mapsto \mathbb{R}$ as in the statement of the theorem. To apply the extension theorem, we need to check that $W + K = X$. If $f \in X$, write $f = p + g$ with $g = \sum_{i=1}^{m} a_i \mathbf{1}_{A_i}$ where A_i are disjoint (left-open, right-closed) intervals, possibly including intervals of the form $(-\infty, b]$ and (b, ∞). Let $a = \min\{a_1, \ldots, a_m\}$ and write $f = (p + a) + (g - a)$. Clearly $g - a \in K$ and $p + a \in W$. Thus $W + K = X$. Consequently, L extends to all of X as a positive linear functional. We continue to denote it by L.

Step 2: To get a measure, define $G(t) = L(\mathbf{1}_{(-\infty,t]})$. If $s < t$, then $0 \le \mathbf{1}_{(s,t]} = \mathbf{1}_{(-\infty,t]} - \mathbf{1}_{(-\infty,s]}$ and hence $G(s) \le G(t)$ by the positivity of L. Thus, G is an increasing function on $\mathbb{R}$. It is also clear that $0 \le G(t) \le \alpha_0$ for all t because $0 \le \mathbf{1}_{(-\infty,t]} \le 1$.

We claim that $G(-\infty) = 0$ and $G(+\infty) = \alpha_0$. To see this, use the Chebyshev-like idea and write $\mathbf{1}_{(-\infty,-b]}(t) \le t^2/b^2$ for $b > 0$. Applying L, we get $G(-b) \le \alpha_2/b^2$ which shows that $G(t) \to 0$ as $t \to -\infty$. Similarly, show that $\alpha_0 - G(b) = L(\mathbf{1}_{(b,\infty)}) \le \alpha_2/b^2$ to see that $G(b) \to \alpha_0$ as $b \to +\infty$. The claim is proved.

It would be clean if we could show that G is right-continuous, but I was not able to (is it false in general?). But we can easily modify it to be right continuous by defining $F : \mathbb{R} \mapsto \mathbb{R}_+$ by

$$F(t) = \inf\{G(s) : s \in \mathbb{Q}, s > t\}.$$

Clearly F is increasing and right-continuous. It also satisfies $F(+\infty) = \alpha_0$ and $F(-\infty) = 0$. Therefore, there exists a unique Borel measure[3] μ on $\mathbb{R}$ such that $\mu(a, b] = F(b) - F(a)$ for any $a < b$.

Let D be the set of continuity points of G. Then D^c is countable (since G is increasing) and hence D is dense. We note for future use that $F(t) = G(t)$ for all $t \in D$.

Step 3: We make some estimates on the tails of μ. Using $\mathbf{1}_{|x| \ge b} \le b^{-2k}|x|^{2k}$ and positivity of L, we get

$$L(\mathbf{1}_{(-\infty,-b]}) + L(\mathbf{1}_{[b,\infty)}) \le b^{-2k}\alpha_{2k}$$

for every $k \ge 1$. From this, it easily follows that (at least when $b \in D$)

$$\mu(-\infty, b] + \mu[b, \infty) \le \alpha_{2k} b^{-2k}.$$

From this we bound the tails of integrals with respect to μ as follows[4].

$$\int |u|^n \mathbf{1}_{|u|>b} d\mu(u) = \int_0^\infty \mu\{|u|^n \mathbf{1}_{|u| \ge b} \ge t\} dt.$$

[3] A quick proof if you have not seen this before. Define $H : (0, \alpha_0) \mapsto \mathbb{R}$ by $H(u) = \inf\{t \in \mathbb{R} : F(t) \ge u\}$. Then, if λ is the Lebesgue measure on $(0, \alpha_0)$, define $\mu = \lambda \circ H^{-1}$. Check that $H(u) \le t$ if and only if $u \le F(t)$ or in other notation $H^{-1}(-\infty, t] = (0, F(t)]$. Therefore, $\mu(-\infty, t] = \lambda(0, F(t)] = F(t)$.

[4] If (X, μ) is a measure space and $f : X \mapsto \mathbb{R}_+$ is a positive function, then $\int_X f(x) d\mu(x) = \int_0^\infty \mu\{f > t\} dt$ by a simple Fubini argument applied to the double integral $\iint_{X \times \mathbb{R}_+} \mathbf{1}_{0<t<f(x)} dt d\mu(x)$. Some people call this the "bath-tub principle". In probability it is often written in the form $\mathbf{E}[X] = \int_0^\infty \mathbf{P}\{X > t\} dt$ for a positive random variable X.

Observe that $|u|^n \mathbf{1}_{|u|\geq b} \geq t$ if and only if $|u|^n \geq t$ and $|u| \geq b$. For $t \leq b^n$ this means $|u| \geq b$ while for $t > b^n$ this means $|u| \geq t^{1/n}$. Therefore,

$$\begin{aligned}\int |u|^n \mathbf{1}_{|u|>b} d\mu(u) &= \int_0^{b^n} \mu\{u : |u| \geq b\} du + \int_0^{b^n} \mu\{u : |u| \geq t^{1/n}\} du \\ &= b^n \mu([-b^n, b^n]^c) + \int_{b^n}^{\infty} \mu([-t^{1/n}, t^{1/n}]^c) dt\end{aligned}$$

Using the usual Chebyshev idea, we write the bounds $\mu([-s,s]^c) \leq \alpha_{2m} s^{-2m}$ valid for any m, apply it with $m = 1$ for the first term and $m = n$ for the second term to get

$$\begin{aligned}\int |u|^n \mathbf{1}_{|u|>b} d\mu(u) &\leq \alpha_2 b^{-n} + \alpha_{2n} \int_{b^n}^{\infty} t^{-2} dt \\ &= (\alpha_2 + \alpha_{2n}) b^{-n}.\end{aligned} \tag{3}$$

Moral: When in distress, remember Chebyshev's inequality or the idea behind it: $\mathbf{1}_{[b,\infty)}(t) \leq t/b$ or more generally $\mathbf{1}_{[b,\infty)}(t) \leq f(t)/f(b)$ for an increasing function f.

Step 4: Now fix a large M and N and let $-M = t_0 < t_1 < \ldots < t_N = M$ be closely spaced points (quantification later). Let n be odd so that u^n is increasing on the whole line. Therefore,

$$u^n \mathbf{1}_{(-\infty,-M]}(u) + \sum_{j=0}^{N-1} t_j^n \mathbf{1}_{(t_j,t_{j+1}]}(u) \leq u^n \leq \sum_{j=0}^{N-1} t_j^n \mathbf{1}_{(t_j,t_{j+1}]}(u) + u^n \mathbf{1}_{[M,\infty)}(u) \tag{4}$$

Integrate w.r.t μ and use (3) to get

$$-\frac{\alpha_{2n}}{M} + \sum_{j=0}^{N-1} t_j^n (F(t_{j+1}) - F(t_j)) \leq \int u^n d\mu(u) \leq \sum_{j=0}^{N-1} t_{j+1}^n (F(t_{j+1}) - F(t_j)) + \frac{\alpha_{2n}}{M} \tag{5}$$

Similarly, we want to get an inequality by applying L to (4). But $u^n \mathbf{1}_{[M,\infty)}$ is not in X, hence we bound it by u^{n+1}/M. Similarly $u^n \mathbf{1}_{(-\infty,-M]}(u) \geq -u^{n+1}/M$. Thus,

$$-\frac{1}{M} u^{n+1} + \sum_{j=0}^{N-1} t_j^n \mathbf{1}_{(t_j,t_{j+1}]}(u) \leq u^n \leq \sum_{j=0}^{N-1} t_j^n \mathbf{1}_{(t_j,t_{j+1}]}(u) + \frac{1}{M} u^{n+1}$$

Now we can apply L and use positivity to get

$$-\frac{\alpha_{n+1}}{M} + \sum_{j=0}^{N-1} t_j^n (G(t_{j+1}) - G(t_j)) \leq \alpha_n \leq \sum_{j=0}^{N-1} t_{j+1}^n (G(t_{j+1}) - G(t_j)) + \frac{\alpha_{n+1}}{M}$$

Compare this with (5). By taking M large, we can make the $1/M$ terms as small as we like. Then by taking N large, we can make sure that $t_{j+1}^n - t_j^n$ are small. By perturbing the points slightly as needed, we may assume that $t_j \in D$ for all j, and hence $F(t_j) = G(t_j)$. Now it is clear that α_n and $\int u^n d\mu(u)$ are sandwiched between two numbers that are very close to each other, and hence must be equal.

So far we assumed that n was odd. For even n, a very similar argument can be given if one is not too tired by now. ■

We stated the last theorem only for intervals. What about general closed sets I? Observe that if $L(p) \geq 0$ for p that is positive on I, then it is certainly the case that $L(p) \geq 0$ for p that is positive on the whole line. From the above proof, we get a measure μ supported on $\mathbb{R}$ whose moments are α_n. To argue that it is supported on I is an exercise.

Exercise 12. Suppose $[a, b] \subseteq I^c$. Argue that in the above proof, when L is extended to X, the resulting functional satisfies $G(a) = G(b)$. Deduce that $\mu(I^c) = 0$.

If you understood the above proof, the following should be easier.

Exercise 13. Prove Riesz's representation theorem for $C_c(\mathbb{R})$: If L is a positive linear functional on $C_c(\mathbb{R})$, then there exists a Borel measure μ such that $L(f) = \int f d\mu$ for all $f \in C_c(\mathbb{R})$.

Remark 14. Can we prove Riesz's representation theorem for general locally compact Hausdorff spaces? Presumably it will work, by extending L from $C_c(X)$ to $C_c(X) + W$ where W is the span of indicators of all compact sets. Then we must define the measure μ by taking $\mu(A) = \sup\{L(\mathbf{1}_K) : K \subseteq A \text{ and } K \text{ is compact}\}$. But then one must show that μ is a measure, it is outer regular etc., and that it agrees with L on $C_c(X)$. This starts looking like the length proof in Rudin's *Real and complex analysis*. The proof is simpler for $X = \mathbb{R}$ (and in the moment problem above), because we assumed the existence of Lebesgue measure and that an increasing right continuous function is the CDF of a measure got by pushing forward the Lebesgue measure...

4. Measures, Sequences, Polynomials, Matrices

To be more precisely, we should have titled this section as "Measures on the line having all moments, positive semi-definite sequences, orthogonal polynomial sequences and Jacobi matrices". All these objects are intimately connected to each other and to the moment problem. This will also lead to the resolution of the uniqueness part of the moment problem, but we may not completely discuss it. Let us introduce all the four objects in the title.

(1) Measures. By this, in this section we shall mean positive Borel measures on the line whose moments are all finite. It is convenient to consider two cases separately. *Case 1:* The measure is has infinite support, *Case 2:* The measure is supported on finitely many points, i.e., $\mu = p_1\delta_{\lambda_1} + \ldots + p_n\delta_{\lambda_n}$, where λ_i are distinct real numbers and p_i are strictly positive numbers.

(2) Positive semi-definite sequences. By this we mean a sequence $\alpha = (\alpha_0, \alpha_1, \ldots)$ such that the infinite matrix $H_\alpha = (\alpha_{i+j})_{i,j \geq 0}$ is a positive semi-definite matrix. This just means that for any $n \geq 1$, and any real numbers $c_0, \ldots, c_n$,

$$\sum_{i,j=0}^{n} c_i c_j \alpha_{i+j} \geq 0$$

Case 1: The sequence is positive definite. That is, equality holds above if and only if all c_is vanish. *Case 2:* The sequence is positive semi-definite but not positive definite. There is a smallest n for which equality holds above for some c_is, not all zero.

(3) Orthogonal polynomial sequence. By this we mean a sequence of polynomial $\varphi_0, \varphi_1, \dots$ such that -

  (a) The degree of φ_j is exactly j for every $j \geq 0$.

  (b) If φ_j are declared to be an orthonormal set in $\mathcal{P}$, then in the resulting inner product space, the multiplication operator $M : \mathcal{P} \mapsto \mathcal{P}$ defined by $Mf(x) = xf(x)$ is symmetric: $\langle Mf, g\rangle = \langle f, Mg\rangle$ for all $f, g \in \mathcal{P}$.

For example, $\varphi_j(x) = x^j$ is not a valid choice for an orthogonal polynomial sequence, because $\langle M\varphi_1, \varphi_2\rangle = 1$ while $\langle \varphi_1, M\varphi_2\rangle = 0$.

What we describe so far is *Case 1*. *Case 2* is when we have a finite sequence $\varphi_0, \dots, \varphi_{n-1}$ such that ... (details later)

(4) Jacobi matrix. A tridiagonal matrix is a finite or infinite matrix whose (i, j) entry is zero unless $|j-i| \leq 1$ (only the main diagonal and the the diagonals immediately above and below it, can contain non-zero entries). A Jacobi matrix is a tridiagonal matrix that is symmetric and has strictly positive entries on the super-diagonal (hence also the sub-diagonal). The main diagonal entries will be labelled $a_0, a_1, \dots$ while the super-diagonal entries will be labelled $b_0, b_1, \dots$. *Case 1:* Infinite Jacobi matrix $T = T(a, b) = (t_{i,j})_{i,j\geq 0}$ such that $t_{i,i} = a_i$ and $t_{i,i+1} = t_{i+1,i} = b_i$ for $i \geq 0$. *Case 2:* Finite Jacobi matrix $T_{n\times n}$ whose main diagonal has $a_0, \dots, a_{n-1}$ and super-diagonal has $b_0, \dots, b_{n-1}$.

What we shall see is that these objects are very closely linked and almost (but not quite!) in one-one correspondence with each other. The objects in *Case 1* are related to each other and the objects in *Case 2* are related to each other. Rather than carrying the two cases all the time, let us first describe the connections in the first case. Later we shall discuss the second case.

5. Measures, Sequences, Polynomials, Matrices: Case 1

5.1. **Measure to sequence.** Given a measure μ whose support is not finite, let $\alpha_n = \int x^n d\mu(x)$ be the nth moment of μ. We claim that the moment sequence $\alpha = (\alpha_0, \alpha_1, \dots)$ is positive definite. This is because

$$\sum_{i,j=0}^{m} c_i c_j \alpha_{i+j} = \sum_{i,j=0}^{m} c_i c_j \int x^{i+j} d\mu(x) = \int \left(\sum_{i=1}^{m} c_i x^i\right)^2 d\mu(x) \geq 0.$$

Equality holds if and only if the polynomial $\sum_{i=0}^{n} c_i x^i$ vanishes on the support of μ. As the latter is an infinite set, this forces $c_i = 0$ for all i. Thus, α is positive definite.

An alternate way to say the same thing is that the matrix $H_\alpha = (\alpha_{i+j})_{i,j\geq 0}$ is positive definite, meaning that all finite principal submatrices of H_α have strictly positive determinant.

5.2. **Sequence to polynomial sequence.** Given a positive semi-definite α, we can define an inner product on $\mathcal{P}$ by defining $\langle x^i, x^j\rangle = \alpha_{i+j}$ for $i, j \geq 0$ and extending by linearity. That is

$$\left\langle \sum_{i=0}^{n} c_i x^i, \sum_{j=0}^{m} d_j x^j \right\rangle = \sum_{i=0}^{n}\sum_{j=0}^{m} c_i d_j \alpha_{i+j}.$$

The bilinearity and symmetry are clear while the positive definiteness of α ensures that $\langle p, p\rangle > 0$ for any $p \neq 0$.

Apply Gram-Shmidt process to $x^0, x^1, x^2, \ldots$ (in that order) to get $\varphi_0, \varphi_1, \ldots$, an orthonormal set that spans the whole space $\mathcal{P}$. It is also clear that φ_j is a polynomial of degree j and that it has positive leading coefficient. There is another property of this sequence

Observation: Let $M : \mathcal{P} \mapsto \mathcal{P}$ be defined by $(Mp)(x) = xp(x)$. Then,

$$\langle Mx^i, x^j\rangle = \langle x^{i+1}, x^j\rangle = \alpha_{i+1+j}, \quad \langle x^i, Mx^j\rangle = \langle x^i, x^{j+1}\rangle = \alpha_{i+j+1}$$

showing that M is symmetric: $\langle Mp, q\rangle = \langle p, Mq\rangle$ for all $p, q \in \mathcal{P}$.

By an *orthogonal polynomial sequence* we mean a sequence of polynomials $\varphi_0, \varphi_1, \varphi_2, \ldots$ such that φ_j has degree j, has positive leading coefficient, and such that if an inner product on $\mathcal{P}$ is defined by declaring φ_js to be orthonormal, then the multiplication operator is symmetric.

Remark 15. (Rameez) The symmetry of M can be equivalently stated as the condition that the Gram matrix of $x^0, x^1, x^2, \ldots$ is a Hankel matrix. This shows that $x^0, x^1, x^2, \ldots$ is not an orthogonal polynomial sequence (because the identity matrix is not Hankel!).

5.3. **Polynomial sequence to Jacobi matrix.** Let $\varphi_0, \varphi_1, \ldots$ be an orthogonal polynomial sequence. Let $\langle \star, \star\rangle$ denote the inner product on $\mathcal{P}$ got by declaring $\langle \varphi_j, \varphi_k\rangle = \delta_{j,k}$ and extending by linearity (possible since the span of $\{\varphi_j\}$ is all of $\mathcal{P}$). Two simple observations: $\langle Mp, q\rangle = \langle p, Mq\rangle$ (by definition of orthogonal polynomials) and $\langle \varphi_k, p\rangle = 0$ if p has degree less than k.

For $k \geq 0$, $M\varphi_k$ has degree $k+1$ and hence there is a unique way to write it as $M\varphi_k = \sum_{j=0}^{k+1} c_{k,j}\varphi_j$. For $j < k-1$, by the symmetry of M, we see that $c_{k,j} = \langle \varphi_k, M\varphi_j\rangle = 0$ since $M\varphi_j$ has degree less than than k. Further,

$$c_{k,k+1} = \langle M\varphi_k, \varphi_{k+1}\rangle = \langle \varphi_k, M\varphi_{k+1}\rangle = c_{k+1,k}.$$

Writing $a_k = c_{k,k}$ and $b_k = c_{k,k+1}$, we see that (with the convention that $b_{-1} = 0$)

$$b_{k-1}\varphi_{k-1} + a_k\varphi_k + b_k\varphi_{k+1} = M\varphi_k.$$

It will be convenient to collect the coefficients a_k, b_ks as an infinite tridiagonal matrix

$$T = T(a; b) = \begin{bmatrix} a_0 & b_0 & 0 & \ldots & \ldots \\ b_0 & a_1 & b_1 & \ldots & \ldots \\ 0 & b_1 & a_2 & b_2 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

To be a Jacobi matrix, we also require $b_k > 0$ for all k. To see this, let A_k be the leading coefficient of φ_k and observe that $\varphi_{k+1} - \frac{A_{k+1}}{A_k} M\varphi_k$ has degree less than k and hence is orthogonal to φ_{k+1}. Thus,

$$b_k = \langle M\varphi_k, \varphi_{k+1}\rangle = \frac{A_{k+1}}{A_k} > 0.$$

Remark 16. In terms of the Jacobi matrix, the three term recurrence can be written in the matrix form

$$T\varphi_\bullet(x) = x\varphi_\bullet(x)$$

where $\varphi_\bullet(x) = (\varphi_0(x), \varphi_1(x), \varphi_2(x), \dots)^t$. Formally, this looks like an eigenvalue equation. The appearance is more than skin deep.

5.4. **From Jacobi matrix to orthogonal polynomial sequence.** Let $T = T(a, b)$ be a finite or infinite tridiagonal matrix with $T_{i,i} = a_i \in \mathbb{R}$ and $T_{i,i+1} = b_i > 0$. We want to recover the orthogonal polynomial sequence. The short answer is that we solve the "eigenvalue equation" $T\mathbf{v} = \lambda\mathbf{v}$ for any $\lambda \in \mathbb{R}$ and write the eigenvector as $\mathbf{v} = (\varphi_0(\lambda), \varphi_1(\lambda), \dots)^t$. These φ_ks are the orthogonal polynomials.

Let us examine this in more detail. Fix any $\lambda \in \mathbb{R}$, set $v_0 = 1$ and recursively solving for $v_1, v_2, \dots$ from the equations

$$a_0v_0 + b_0v_1 = \lambda v_0$$

$$b_0v_0 + a_1v_1 + b_1v_2 = \lambda v_1$$

$$b_1v_1 + a_2v_2 + b_2v_3 = \lambda v_2 \quad \dots \quad \dots$$

As $b_k > 0$ for all k, this is possible and we get a vector $\mathbf{v} = (v_0, v_1, \dots)$ that satisfies $T\mathbf{v} = \lambda\mathbf{v}$. Now let us change notation and show the dependence on λ by writing v_k as $\varphi_k(\lambda)$, starting with $\varphi_0(\lambda) = 1$. It is clear from the recursions that φ_k is a polynomial of degree k and that it has positive leading coefficient. We must check one last point: If φ_k are declared to be orthonormal, then the multiplication M must be symmetric. That is indeed the case, as we can check that $\langle M\varphi_k, \varphi_\ell\rangle = \langle \varphi_k, M\varphi_\ell\rangle$ from the recursions

$$M\varphi_k = b_{k-1}\varphi_{k-1} + a_k\varphi_k + b_k\varphi_{k+1}.$$

Important observation: We want to say that this mapping from Jacobi matrices to OP-sequences and the mapping of the previous section from OP-sequences to Jacobi matrices are inverses of each other. This is almost correct in that we recover the orthogonal polynomial sequence up to an overall constant factor. Indeed, in the recovery, we always get $\varphi_0 = 1$. This is easily seen by staring for a minute at the three-term recurrence.

It would have been cleaner if we had assumed all measures to be probability measures, all positive definite sequences to have $\alpha_0 = 1$, all OP sequences to have $\varphi_0 = 1$. Then the above mapping would have been exactly the inverse of the one from OP sequences to Jacobi matrices. *Henceforth, let us adopt this convention.*

5.5. **From orthogonal polynomials to positive definite sequence.** Given an orthogonal polynomial sequence $\varphi_0, \varphi_1, \dots$ and the associated inner product, we construct a positive definite sequence as follows. There is a unique way to write $x^k = \sum_{j=0}^{k} c_{k,j}\varphi_j$ from which we get

$$\langle x^k, x^\ell \rangle = \sum_{j=0}^{k\wedge\ell} c_{k,j}c_{\ell,j}.$$

Since we already know that M is symmetric in this inner product, it follows that the above quantity must depend only on $k+\ell$. Denote this number by $\alpha_{k+\ell}$. This is a positive definite sequence because H_α is the Gram matrix of $x^0, x^1, x^2, \dots$ and these are linearly independent.

It is also a easy to see that this mapping is the inverse of the mapping that we gave earlier from positive definite sequences to orthogonal polynomial sequences.

It may look a little unsatisfactory that the mapping given here is not explicit. It can be made explicit. Fix $k \geq 0$ and write $\alpha_k = \langle x^k, x^0 \rangle = c_{k,0}$, the constant term in φ_k. This can be put in a more interesting form in terms of the Jacobi matrix (recall that we already know how to move between OP-sequences and Jacobi matrices).

5.6. **From Jacobi matrix to positive definite sequence.** Let $T = T(a,b)$ be a Jacobi matrix. Define $\beta_k = \langle T^k e_0, e_0 \rangle$ for $k \geq 0$. We claim that this is a positive semi-definite sequence and that this is the inverse of the mapping we have see from positive semi-definite sequence to tridiagonal matrices (via orthogonal polynomial sequence and three term recurrence).

As we have seen how to recover orthogonal polynomials from T, let us write

$$T\varphi_\bullet(x) = x\varphi_\bullet(x)$$

where $\varphi_\bullet(x) = (\varphi_0(x), \varphi_1(x), \varphi_2(x), \dots)^t$. Therefore, $T^k\varphi_\bullet(x) = x^k\varphi_\bullet(x)$ or in terms of the co-ordinate vectors $e_0, e_1, \dots$

$$\sum_{j=0}^{\infty} \varphi_j(x) T^k e_j = \sum_{j=0}^{\infty} x^k \varphi_j(x) e_j.$$

Take inner product with e_0 (this is inner product in ℓ^2) to get

$$\sum_{j=0}^{\infty} \varphi_j(x) \langle T^k e_j, e_0 \rangle = x^k.$$

This gives the expansion of x^k in terms of the orthogonal polynomials that we needed above (it should not worry you that the sum here is infinite, indeed $\langle T^k e_j, e_0 \rangle = 0$ as can be seen from the tridiagonal structure of T). In particular $c_{k,0} = \langle T^k e_0, e_0 \rangle$. Combining with the previous observation of how to recover the α_ks from $c_{k,0}$, this shows that $T \mapsto (\beta_0, \beta_1, \dots)$ mapping Jacobi matrices into positive definite sequence is the inverse of the mapping in the other direction that we have seen earlier (going through OP-sequences).

Remark 17. A better way as pointed out by Sayantan Khan in class. The mapping $\varphi_j \leftrightarrow e_j$, $j \geq 0$, is an isomorphism with $\mathcal{P}$ (with $\{\varphi_j\}$ as ONB) and $V = \text{span}\{e_0, e_1, \ldots\}$ where e_j is the vector with 1 at the jth place and 0s elsewhere. Under this isomorphism, $M : \mathcal{P} \mapsto \mathcal{P}$ becomes $T : V \mapsto V$. As we saw earlier, $\alpha_k = \langle x^k, x^0 \rangle = \langle M^k \varphi_0, \varphi_0 \rangle$ which, by the isomorphism, equals $\langle T^k e_0, e_0 \rangle$.

5.7. The picture so far.

$$\text{Measure} \to \text{PD sequence} \leftrightarrows \text{OP sequence} \leftrightarrows \text{Jacobi matrix}$$

The key question was whether a positive definite sequence is the moment sequence of a unique measure. We have not touched that question but introduced two other objects that are in one-one correspondence with positive definite sequences. We shall return to this question after talking about some nice consequences of the rich interactions between these objects in the next two sections.

5.8. **Exercises.** In these exercises, the relationship between the positive definite sequences, orthogonal polynomials and Jacobi matrices is further strengthened.

Exercise 18. Let α be a positive definite sequence. Let $D_m = \det\left(\alpha_{i+j}\right)_{0\leq i,j\leq m-1}$. Show that the corresponding orthogonal polynomials are given by $\varphi_0(x) = 1$ and for $m \geq 1$,

$$\varphi_m(x) = \frac{(-1)^{m-1}}{\sqrt{D_{m-1}D_m}} \det \begin{bmatrix} 1 & x & \ldots & x^m \\ \alpha_0 & \alpha_1 & \ldots & \alpha_m \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{m-1} & \alpha_m & \ldots & \alpha_{2m-1} \end{bmatrix}$$

Exercise 19. Let $T = T(a, b)$ be the Jacobi matrix corresponding to the positive definite sequence α. Let D_n be the determinant of $(\alpha_{i+j})_{0\leq i,j\leq n-1}$. Show that

$$b_k = \frac{\sqrt{D_{k-1}D_{k+1}}}{D_k}$$

for $k \geq 1$ and $b_0 = \frac{\sqrt{D_1}}{D_0}$. If $\alpha_n = 0$ for all odd n (for eg., if it is the moment sequence of a symmetric measure), show that $a_n = 0$ for all n. [**Remark:** There is also a formula for a_ns in general, but we skip it for now]

Exercise 20. Let the OP sequence $\varphi_0, \varphi_1, \ldots$ correspond to the Jacobi matrix $T = T(a, b)$. If T_n is the top $n \times n$ sub-matrix of T, show that φ_n is (up to a constant) the characteristic polynomial of T_n. Deduce that,

(1) The roots of φ_n are all real and distinct.

(2) The roots of φ_n and φ_{n-1} interlace.

6. Quadrature formulas

Let μ be a measure on the line with all moments finite. Assume that μ is not supported on finitely many points. Fix $n \geq 1$. We seek n distinct points $\lambda_1, \ldots, \lambda_n$ and poisitive weights $w_1, \ldots, w_n$ such that

$$\int Q(x)d\mu(x) = \sum_{k=1}^{n} w_k Q(\lambda_k)$$

for as many polynomials Q as possible. Since we have a choice of $2n$ parameters for the points and weights, we may expect that this can be done for all polynomials of degree $2n-1$ or less (it has $2n$ coefficients).

Why care? It has to do with numerical integration. Once we fix n and choose λ_is and w_is, given any $f : \mathbb{R} \mapsto \mathbb{R}$, we numerically compute its integral with respect to μ by

$$\int f(x)d\mu(x) \approx \sum_{k=1}^{n} w_k f(\lambda_k).$$

This gives the exact answer for polynomials of degree up to $2n-1$. Hence, if f is nice enough that it is well approximated by its Taylor expansion to order $2n-1$, then the above approximation gives a reasonably close answer to $\int f d\mu$.

How to find the points and weights? Note that what we are asking for is a measure $\mu_n = \sum_{k=1}^{n} w_k \delta_{\lambda_k}$ whose first $2n-1$ moments agree with those of μ.

Assume that μ is a probability measure, without loss of generality. From μ, we go to the infinite tridiagonal matrix $T = T(a,b)$ (via moments, orthogonal polynomials and the three-term recurrence). Let T_n be the top $n \times n$ principal submatrix of T. Let μ_n be the measure corresponding to T_n, i.e., the spectral measure of T_n at the vector e_0. Recall that this is given by

$$\mu_n = \sum_{k=1}^{n} w_k \delta_{\lambda_k}$$

where λ_k are the eigenvalues of T_n and $w_k = Q_{1,k}^2$, where $T_n = Q\Lambda Q^t$ is the spectral decomposition of T_n, with $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and Q an orthogonal matrix.

Recall that the moment sequence α can be recovered from the tridiagonal matrix μ by the equations $\alpha_k = \langle T^k e_0, e_0 \rangle$. This is just the $(0,0)$ entry of T^k, which can also be written as

$$\sum_{i_1, \ldots, i_{k-1} \geq 0} T_{0,i_1} T_{i_1,i_2} \ldots T_{i_{k-1},0}.$$

Since T is tridiagonal, the non-zero terms must have $i_1, \ldots, i_{k-1} \leq \lfloor k/2 \rfloor$. Hence,

$$\langle T^k e_0, e_0 \rangle = \langle (T_n)^k e_0, e_0 \rangle$$

for $k \leq n-1$. This shows that the first $2n-1$ moments of μ_n and μ are identical.

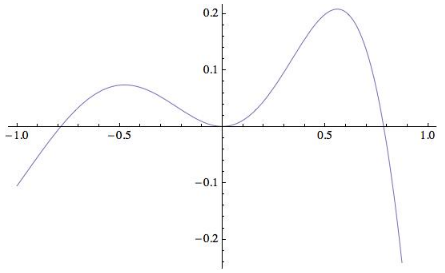

FIGURE 1. Plot of the function $e^x \cos(2x)\log(1+x^2)$

Remark 21. It is also possible to express the points and weights in terms of the orthogonal polynomials for μ. Indeed, $\lambda_1, \dots, \lambda_n$ are the roots of φ_n, and $(\varphi_0(\lambda_k) \dots \varphi_{n-1}(\lambda_k))^t$ is an eigenvector corresponding to the eigenvalue λ_k. After normalizing, this becomes the kth column of R. Hence,

$$w_k = \frac{\varphi_0(\lambda_k)^2}{\sum_{j=0}^{n-1} \varphi_j(\lambda_k)^2} = \frac{1}{\sum_{j=0}^{n-1} \varphi_j(\lambda_k)^2}.$$

An example - Lebesgue measure on $[-1,1]$**:** If μ is the Lebesgue measure on $[-1,1]$, then the corresponding orthogonal polynomials are called Legendre polynomials. There are explicit formulas to express them. The zeros of the Legendre polynomial can be computed and so can the weights. This gives us a way to numerically integrate functions over $[-1,1]$. Just to illustrate, here is an example:

If we want four points, the points and weights are given by (computations on Mathematica)

$$\lambda = (-0.861136, -0.339981, 0.339981, 0.861136), \quad w = (0.347855, 0.652145, 0.652145, 0.347855).$$

The function $f(x) = e^x \cos(2x)\log(1+x^2)$ (chosen without fear or favour) has integral 0.0350451 and the numerical approximation using the above points and weights gives 0.036205. With five points, it improves to 0.0348706 and with 10 points, the agreement is up to 7 decimal places! In contrast, with equispaced points and equal weights, the approximations are 0.0255956 for 100 points, 0.0327198 for 1000 points and 0.0349526 for 10000 points. In general, what is the error like? Let $f \in C^{(n)}$ (on an open set containing $[-1,1]$) and write $f(x) = Q_n(x) + R_n(x)$, where Q_n is the $2n-1$ order Taylor expansion of f. The remainder term R_n can be estimated by

$$\sup_{x\in[-1,1]} |R_n(x)| \le \frac{1}{(2n)!}\|f^{(2n)}\|_{\sup[-1,1]}.$$

Since $\sum_{j=1}^{n} w_j Q_n(\lambda_j) = \int_{-1}^{1} Q_n(x)dx$, we get

$$\begin{aligned}|\int_{-1}^{1} f(x)dx - \sum_{k=1}^{n} w_k f(\lambda_k)| &\le \int_{-1}^{1} |R_n(x)|dx + \frac{1}{n}\sum_{k=1}^{n} w_k R_n(\lambda_k) \\ &\le \frac{2}{(2n)!}\|f^{(2n)}\|_{\sup[-1,1]}\end{aligned}$$

For instance, if the derivatives are uniformly bounded in $[-1,1]$ (or grow at most exponentially etc.) then the error term is $O(e^{-cn\log n})$. In contrast, for n equispaced points, the error goes down like $O(1/n)$ and for n randomly chosen points the error goes down like $1/\sqrt{n}$.

Similarly, one uses zeros of Chebyshev polynomials, Hermite polynomials (OPs for Gaussian measure), Laguerre polynomials (OPs for $e^{-x}dx$ on $\mathbb{R}_+$), etc., to integrate against $\frac{1}{\sqrt{1-x^2}}$, e^{-x^2}, e^{-x}, respectively. They carry names such as Chebyshev quadrature, Gaussian quadrature etc.

This may be a good occasion to say something explicit about orthogonal polynomials for special measures. The few examples are, the uniform measure (Legendre polynomials), the Gaussian measure (Hermite polynomials), Exponential measure (Laguerre polynomials), arcsine measure (Chebyshev polynomials). The uniform and arcsine fall into the family of Beta measures (whose orthogonal polynomials are called Jacobi polynomials) and the exponential is part of the Gamma family of distributions.

Exercise 22. Define $P_n(x) = \frac{d^n}{dx^n}(1-x^2)^{2n}$. Show that P_n are orthogonal on $[-1,1]$ with respect to Lebesgue measure. Find c_n so that c_nP_n become orthonormal. These are the Legendre polynomials.

The expression for Legendre polynomials in the exercise is called Rodrigues' formula. Similarly, one can show that

$$H_n(x) := e^{x^2/2}\frac{d^n}{dx^n}e^{-x^2/2}$$

are orthogonal with respect to the standard Gaussian measure on $\mathbb{R}$.

7. Another proof that positive semi-definite sequences are moment sequences

Let α be a positive semi-definite sequence with $\alpha_0 = 1$ (without loss of generality). We want to show that there is a measure μ such that $\alpha_n = \int x^n d\mu(x)$ for all n.

The idea of this proof is to solve a sequence of problems approximating our problem, and then extract a limit solution that will solve the actual problem. At this level of generality, this is a very repeatable (and natural) idea. In addition, it will illustrate one of the theorems we learned in functional analysis class.

If α is positive semi-definite but not positive definite, we shall see a simple proof that it is the moment sequence of a unique measure which in fact has finite support. Hence, let us assume that α is positive definite below.

Step-1: For any n, there exists a measure μ_n such that $\alpha_k = \int x^k d\mu_n(x)$ for $0 \le k \le n-1$.

We saw this in the previous section. From α, construct the OP sequence and then the Jacobi matrix T. Let μ_n be the spectral measure at e_0 of T_n, where T_n is the top $n \times n$ principal submatrix of T, that is, $\mu_n = \sum_{j=1}^n Q_{1,j}^2 \delta_{\lambda_j}$ where $T_n = Q\Lambda Q^t$ is the spectral decomposition of T_n with $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. Then, μ_n has the first $2n-1$ moments equal to those of μ.

Step-2: There is a subsequence n_k such that μ_{n_k} converges weakly to a probability measure μ.

This is a direct consequence of Helly's theorem,[5] since $\mu_n(\mathbb{R}) = \alpha_0$ for all n.

Step-3: We claim that μ has the moment sequence α.

Fix an even number $2k$ and write

$$\begin{aligned}\int x^{2k} d\mu_n(x) &= \int_0^\infty \mu_n\{x : x^{2k} > t\} dt \\ &= \int_0^\infty \mu_n(-\infty, -t^{1/2k}) dt + \int_0^\infty \mu_n(t^{1/2k}, \infty) dt.\end{aligned}$$

Consider the first integral, take $n = n_j$ and let $j \to \infty$. For a.e. x (according to Lebesgue measure), the integrand converges to $\mu(-\infty, -t^{1/2k})$. If we can justify the hypothesis of DCT, it follows that the integral converges to $\int_0^\infty \mu(-\infty, -t^{1/2k}) dt$. Similarly for the second integral. Taking the sum, we get $\int x^{2k} d\mu(x)$, showing that the even moments of μ_{n_j} converge to those of μ. But for every k, the kth moment of μ_{n_j} is α_k for large enough j. Therefore, the $2k$th moment of μ is α_{2k}. Argue similarly for odd moments.

To justify DCT, use the bounds (Chebyshev again!)

$$\mu_n(-\infty, -t^{1/2k}) \leq \frac{1}{t^2} \int x^{4k} d\mu_n(x) = \alpha_{4k} t^{-2},$$

the last inequality being for large enough n. Of course we also have the bound α_0 for the left side, which we use for $t < 1$. Thus, the integrand is dominated by $\alpha_0 + \alpha_{4k} t^{-2} \mathbf{1}_{t \geq 1}$ which is integrable. This completes the proof.

Remark 23. We shall have occasion to use Helly's theorem again. It is a compactness criterion for measures on the line (with the topology of weak convergence). It is instructive to compare it and its proof with other compactness theorems that you have seen, like the Arzela-Ascoli theorem or Montel's theorem in complex analysis.

Helly's theorem can be seen as a special case of Banach-Aloglu theorem as follows: The space $C_0(\mathbb{R})$ of continuous functions vanishing at infinity is a Banach space under the sup-norm, and its dual is the space of all signed measures that are Radon. The weak-* topology on the dual is precisely the topology of weak convergence. Thus, a sequence $\{\mu_n\}$ as in Helly's theorem is contained in a ball in $(C_0(\mathbb{R}))^*$ and hence pre-compact.

In general, compactness does not imply sequential compactness (note that the weak-* topology is not metrizable in general), but the separability of $C_0(\mathbb{R})$ can be used to show the sequential

[5] *Helly's theorem:* If μ_n is a sequence of Borel measures on $\mathbb{R}$ such that $\mu_n(\mathbb{R}) \leq A$ for some A for all n, then there is a subsequence n_k and a measure μ such that $\mu_n[a, b] \to \mu[a, b]$ for all a, b such that $\mu\{a, b\} = 0$.

Proof: For each x, the sequence $\mu_n(-\infty, x]$ has a subsequential limit. Enumerate rationals in a sequence, take subsequences of subsequences etc., and use a diagonal argument to get a single subsequence along which $G(x) := \lim_{k\to\infty} \mu_{n_k}(-\infty, x]$ exists for all $x \in \mathbb{Q}$. Now define $F(x) = \inf\{G(y) : y > x\}$, an increasing, right-continuous, bounded function. Let μ be the measure such that $\mu(-\infty, x] = F(x)$ for all x. Check that $\mu_{n_k}[a, b] \to \mu[a, b]$ at least if $\mu\{a\} = \mu\{b\} = 0$.

compactness as required in Helly's theorem. The best way to understand this is to assume that a Banach space is separable and prove Banach-Alaoglu theorem for its dual by imitating the proof of Helly's theorem (take a countable dense set in X...).

8. Some special orthogonal polynomials

We have talked about general measures and not actually worked out any examples. Here we present a few.

Gaussian measure: Let $d\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}dx$ on the line. The odd moments are zero while the even moments are

$$\begin{aligned}\alpha_{2n} &= \frac{2}{\sqrt{2\pi}}\int_0^\infty x^{2n}e^{-x^2/2}dx = \frac{2}{\sqrt{2\pi}}\int_0^\infty (2t)^n e^{-t}\frac{dt}{\sqrt{2t}} = \frac{2^n}{\sqrt{\pi}}\Gamma(n+\frac{1}{2})\\ &= 2^n(n-\frac{1}{2})(n-1-\frac{1}{2})\dots(1-\frac{1}{2}) = (2n-1)\times(2n-3)\times\dots\times 3\times 1.\end{aligned}$$

This has the nice interpretation as the number of matchings of the set $\{1,2,\dots,2n\}$ into n pairs[6].

I do not know how to derive the orthogonal polynomials using Gram-Schmidt or the determinant formula that we gave in an earlier exercise. We simply define for $n \ge 0$,

$$H_n(x) = (-1)^n e^{x^2/2}\frac{d^n}{dx^n}e^{-x^2/2}$$

which is clearly a polynomial of degree n. Also,

$$\begin{aligned}\int H_nH_m d\mu &= \frac{(-1)^{m+n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left[\frac{d^n}{dx^n}e^{-x^2/2}\right]H_m(x)dx\\ &= \frac{(-1)^{m+2n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x^2/2}\left[\frac{d^n}{dx^n}H_m(x)\right]dx\end{aligned}$$

by integrating by parts n times (the boundary terms vanish because of the rapid decay of $e^{-x^2/2}$). If $m < n$, the integrand is zero (since H_m has degree m) and if $m = n$, we observe that $H_n(x) = x^n + \dots$ to see that $\frac{d^n}{dx^n}H_n(x) = n!$. The rest of the integral is one, and we arrive at $\int H_n^2(x)d\mu(x) = n!$, from which we get the OPs as

$$\varphi_n(x) = \frac{1}{\sqrt{n!}}H_n(x).$$

These form an ONB for $L^2(\mathbb{R},\mu)$. As a corollary, $\frac{1}{\sqrt[4]{2\pi}\sqrt{n!}}\varphi_n(x)e^{-x^2/4}$, $n \ge 0$, form an ONB for $L^2(\mathbb{R})$. Completeness may require an argument.

[6] This is not mere numerology. If $(X_1,\dots,X_{2n})$ are jointly Gaussian with zero means and covariance $\mathbf{E}[X_iX_j] = \sigma_{i,j}$, then $\mathbf{E}[X_1\dots X_{2n}]$ is equal to $\sum_M w(M)$, where the sum is over all matchings of $\{1,2\dots,2n\}$ and the weight of a matching $M = \{\{i_1,j_1\},\dots,\{i_n,j_n\}\}$ is given by $w(M) = \sigma_{i_1,j_1}\sigma_{i_2j_2}\dots\sigma_{i_nj_n}$. This is sometimes called *Wick formula* or *Feynman diagram formula*.

The Jacobi matrix corresponding to this is given by $T = T(a,b)$ where

$$a_n = \int x\varphi_n(x)^2 d\mu(x), \quad b_n = \int x\varphi_n(x)\varphi_{n+1}(x)d\mu(x).$$

It is easy to see that H_n (and hence φ_n) is even or odd according as n is even or odd. Hence $a_n = 0$, being the integral of an odd function. Further, if we write $x\varphi_n(x) = C_n\varphi_{n+1}(x) +$ [lower order terms], then it is clear that $b_n = C_n$. But it is easy to work out that

$$C_n = \frac{[x^n]\varphi_n(x)}{[x^{n+1}]\varphi_{n+1}(x)} = \frac{1/\sqrt{n!}}{1/\sqrt{(n+1)!}} = \sqrt{n+1}.$$

Thus, the Jacobi matrix for this measure is

$$T = \begin{bmatrix} 0 & \sqrt{1} & 0 & \dots & \dots \\ \sqrt{1} & 0 & \sqrt{2} & \dots & \dots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Uniform measure: Let μ be the uniform probability measure on $[-1,1]$. The moments are

$$\alpha_n = \begin{cases} 0 & \text{if } n \text{ is odd,} \\ \frac{1}{n+1} & \text{if } n \text{ is even.} \end{cases}$$

Observe that the Hankel matrix H_α is very similar to the Hilbert matrix. Again, rather than working out the orthogonal polynomials, we simply present the answer. Define the *Legendre polynomials*

$$P_n(x) := \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n, \text{ for } n \geq 0.$$

Clearly P_n is a polynomial of degree n and $[x^n]P_n(x) = \frac{(2n)!}{2^n(n!)^2}$. We leave it as an exercise to check that

$$\int_{-1}^{1} P_n(x)P_m(x)\frac{dx}{2} = \frac{1}{2n+1}\delta_{n,m}.$$

Thus, $\varphi_n(x) = \sqrt{2n+1}P_n(x)$, $n \geq 0$, are the orthogonal polynomials.

To get the Jacobi matrix, we compute a_n and b_n as before. Again, φ_n are alternately odd and even, hence $a_n = 0$. Further,

$$b_n = \int_{-1}^{n} x\varphi_n(x)\varphi_{n+1}(x)\frac{dx}{2} = \frac{[x^n]\varphi_n(x)}{[x^{n+1}]\varphi_{n+1}(x)} = \frac{n+1}{\sqrt{2n+1}\sqrt{2n+3}}.$$

Exercise 24. Let $d\mu(x) = e^{-x}dx$ on $\mathbb{R}_+$. Find the moments, orthogonal polynomials and the Jacobi matrix corresponding to this measure.

Hint: Consider the Laguerre polynomials

$$L_n(x) = \frac{1}{n!}e^x\frac{d^n}{dx^n}[x^n e^{-x}].$$

Other special orthogonal polynomials: In a similar fashion, it is possible to obtain explicitly the orthogonal polynomials and the Jacobi matrix for the Beta family of distributions (that includes the uniform measure and also the arcsine measure) and the Gamma family of distributions (a special case being the exponential measure $e^{-x}dx$ on $\mathbb{R}_+$). The corresponding orthogonal polynomials are called Jacobi polynomials and generalized Laguerre polynomials. In addition to the general properties shared by all orthogonal polynomials, these special ones also satisfy differential equations, recursions involving the polynomials and the derivatives etc. They arise in a variety of problems. For example, the Legendre polynomials arise naturally in the representation theory of the orthogonal group.

9. The uniqueness question: some sufficient conditions

Now suppose we have a positive definite sequence α. We also have the associated OP sequence $\varphi_0, \varphi_1, \ldots$ and the Jacobi matrix $T = T(a, b)$. The question is whether there is a unique measure having moment sequence α? If not, what are all the measures that have this moment sequence?

First we give examples to show that uniqueness need not always hold. A standard example is the measure $d\mu(t) = f(t)dt$ where

$$f(t) = \frac{1}{\sqrt{2\pi}t} e^{-\frac{1}{2}(\log t)^2} dt \quad \text{for } t > 0.$$

In probabilistic language, if X has $N(0, 1)$ distribution, then e^X has density f. The moments of μ are given by

$$\alpha_n = \int_0^\infty t^n f(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{nx} e^{-\frac{1}{2}x^2} dx = e^{\frac{1}{2}n^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{1}{2}(x-n)^2} dx = e^{\frac{1}{2}n^2}.$$

To get some other measures, consider the sum

$$\sum_{k\in\mathbb{Z}} e^{-\frac{1}{2}(k-n)^2} = e^{-\frac{1}{2}n^2} \sum_{k\in\mathbb{Z}} e^{-\frac{1}{2}k^2} e^{kn}.$$

The left hand side does not depend on n (index the sum by $k - n$ instead of k)! Denoting it as Z and $p_k = e^{-\frac{1}{2}k^2}/Z$, we see that

$$\sum_{k\in\mathbb{Z}} p_k e^{kn} = \alpha_n.$$

Thus, the discrete measure $\nu = \sum_{k\in\mathbb{Z}} p_k \delta_{e^k}$ and μ have the same moment sequence. Instead of summing k over integers, if we sum over $\mathbb{Z} + t$ for some $t \in (0, 1)$, we would get other measures with the same moment sequence.

Here is another kind of example. Observe that

$$\begin{aligned}\int_0 t^n f(t)\sin(2\pi\log t)dt &= \int_{-\infty}^{\infty} e^{nx}e^{-\frac{1}{2}x^2}\sin(2\pi x)dx \\ &= e^{\frac{1}{2}n^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-n)^2}\sin(2\pi x)dx \\ &= e^{\frac{1}{2}n^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\sin(2\pi x)dx\end{aligned}$$

where the last line used $\sin(2\pi(x+n)) = \sin(2\pi x)$. The above integral is zero because the integrand is an odd function. Thus if $g_c(t) = f(t)(1 + c\sin(2\pi t))$, with $|c| \le 1$, then $g_c \ge 0$ and

$$\int t^n g_c(t)dt = \int t^n f(t)dt \quad \text{for all } f.$$

Exercise 25. Fix $0 < \lambda < 1$. For a suitable choice of β, show that $\int x^n e^{-|x|^\lambda}\sin(\beta|x|^\lambda \text{sgn}(x))dx = 0$ for all n. Produce many measures having a common moment sequence.

Sufficient conditions for uniqueness: For practical purposes, it is useful to have sufficient conditions for recovery. Here are three (progressively stronger) sufficient conditions that are sufficient for most purposes.

(1) If μ is compactly supported, it is determined by its moment sequence. In terms of the moment sequence, this is equivalent to $\alpha_{2n} \le C^n$ for some $C < \infty$ (i.e., $\limsup_{n\to\infty} \alpha_{2n}^{1/2n} < \infty$).

(2) If μ has finite Laplace transform in a neighbourhood of zero, i.e., if $\mathcal{L}_\mu(t) = \int e^{tx}d\mu(x) < \infty$ for $t \in (-\delta, \delta)$ for some $\delta > 0$, then μ is determined by its moment sequence. This condition is equivalent to $\alpha_{2n} \le (Cn)^n$ for some $C < \infty$ (i.e., $\limsup_{n\to\infty} \frac{1}{n}\alpha_{2n}^{1/2n} < \infty$).

(3) If $\sum_{n=1}^{\infty} \alpha_{2n}^{-1/2n} = \infty$ (and α is positive definite), there is a unique measure whose moment sequence is α. This is known as *Carleman's condition*.

We just justify the first condition. First, observe that if μ is supported on $[-M, M]$, then $\alpha_{2n} \le M^{2n}$. Conversely, if $\alpha_{2n} \le C^{2n}$, observe that $\mu([-M, M]^c) \le M^{-2n}\alpha_{2n}$ which goes to zero as $n \to \infty$, provided $M > C$. Thus μ is supported on $[-C, C]$.

Now if a moment sequence α satisfying $\alpha_{2n} \le M^{2n}$ is given, and μ and ν are two measures on $[-M, M]$ having the moment sequence α, we see that $\int p(x)d\mu(x) = \int p(x)d\nu(x)$ for all polynomials p. Use Weierstrass' approximation to conclude that $\int f d\mu = \int f d\nu$ for all $f \in C[-M, M]$. For any $[a, b] \subseteq [-M, M]$, it is easy to find continuous functions that decrease to $\mathbf{1}_{[a,b]}$. Monotone convergence theorem implies that $\mu[a, b] = \nu[a, b]$ and thus $\mu = \nu$.

10. The uniqueness question: Finitely supported measures

In this section, we consider finitely supported measures, positive semi-definite sequences (that are not positive definite), finite sequences of orthogonal polynomials, and finite Jacobi matrices.

As before, we show how to go from one to the next, but crucially, we can also go back from Jacobi matrices to finitely supported measures, completing the cycle. This will also motivate our next discussion on the importance of the spectral theorem in going from a positive semi-definite sequence to a measure. Since the steps are analogous, we keep this account brief.

Let $\mu = p_1\delta_{\lambda_1} + \ldots + p_n\delta_{\lambda_n}$ where $n \geq 1$, $\lambda_1 < \ldots < \lambda_n$ and $p_i > 0$ with $p_1 + \ldots + p_n = 1$ be a measure supported on finitely many points of the real line.

The kth moment of μ is $\alpha_k = \int x^k d\mu(x) = \sum_{j=1}^{n} p_j \lambda_j^k$. Clearly $\alpha_0 = 1$. As before, the matrix $H_\alpha = (\alpha_{i+j})_{i,j\geq 0}$ is positive semi-definite, because for any $m \geq 0$ and $c_0, \ldots, c_m \in \mathbb{R}$,

$$0 \leq \int \Big|\sum_{i=0}^{m} c_i x^i\Big|^2 d\mu(x) = \sum_{i,j=1}^{N} c_i c_j \int x^{i+j} d\mu(x) = \sum_{i,j=1}^{N} c_i c_j \alpha_{i+j}.$$

Equality holds in the above inequality if and only if $\sum_{i=0}^{m} c_i x^i = 0 \; a.e.[\mu]$ which is the same as saying that $\sum_{i=0}^{m} c_i \lambda_k^i = 0$ for $1 \leq k \leq n$. Writing in matrix form, this is equivalent to

$$\begin{bmatrix} \lambda_1^0 & \lambda_1^1 & \ldots & \lambda_1^m \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_n^0 & \lambda_1^1 & \ldots & \lambda_n^m \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_m \end{bmatrix} = \mathbf{0}.$$

If $m = n-1$, the matrix on the left is square and has determinant $\prod_{i<j}(\lambda_j - \lambda_i)$ which is non-zero as the λ_is are distinct. For $m \geq n$, clearly there exist c_js such that the equation is satisfied, since the matrix has rank at most n. Thus, H_α has rank n, and more specifically, its top $k \times k$ principal sub-matrix is non-singular for $k \leq n-1$ and singular for $k \geq n$. Thus, α is positive semi-definite but not positive definite.

Applying Gram-Schmidt to $1, x, x^2, \ldots$, we get polynomials $\varphi_0, \varphi_1, \ldots, \varphi_{n-1}$, where φ_j has degree j. We cannot proceed further, as x^n is linearly dependent on $1, x, \ldots, x^{n-1}$ in the given inner product (i.e., in $L^2(\mu)$). This is the orthogonal polynomial sequence.

To get the three term recurrence, we again write, for $k \leq n-2$,

$$x\varphi_k(x) = c_{k,k+1}\varphi_{k+1}(x) + \ldots + c_{k,0}\varphi_0(x).$$

Using the inner product of $L^2(\mu)$ (since $M : L^2(\mu) \mapsto L^2(\mu)$ is symmetric), we reason as before that $c_{k,j} = 0$ for $j \leq k-2$, $c_{k,k+1} = c_{k+1,k}$ and writing $a_k = c_{k,k}$ and $b_k = c_{k,k+1}$ (this is positive, why?) thus get the three term recurrence

$$\begin{aligned} x\varphi_0(x) &= a_0\varphi_0(x) + b_0\varphi_1(x), \\ x\varphi_k(x) &= b_{k-1}\varphi_{k-1}(x) + a_k\varphi_k(x) + b_k\varphi_{k+1}(x) \quad \text{for } 1 \leq k \leq n-2. \end{aligned}$$

Lastly, it also holds that (we leave the reasoning to you)

$$x\varphi_{n-1}(x) \overset{L^2(\mu)}{=} b_{n-2}\varphi_{n-2}(x) + a_{n-1}\varphi_{n-1}(x).$$

Equality in $L^2(\mu)$ means that the difference has zero norm in $L^2(\mu)$, or equivalently, equality holds for $x \in \{\lambda_1, \dots, \lambda_n\}$. The equality cannot be for all x as the left side is a polynomial of degree n but the right side has lower degree.

The three term recurrences can be written in matrix form as

$$\begin{bmatrix} a_0 & b_0 & 0 & \dots & 0 \\ b_0 & a_1 & b_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & b_{n-2} & a_{n-1} \end{bmatrix} \begin{bmatrix} \varphi_0(x) \\ \vdots \\ \varphi_{n-1}(x) \end{bmatrix} \overset{L^2(\mu)}{=} x \begin{bmatrix} \varphi_0(x) \\ \vdots \\ \varphi_{n-1}(x) \end{bmatrix}$$

The equality is in L^2 because the very last equation holds only when $x \in \{\lambda_1, \dots, \lambda_n\}$. Let T_n denote the Jacobi matrix on the left.

The previous equality holds for $x \in \{\lambda_1, \dots, \lambda_n\}$, showing that λ_k is an eigenvalue of T_n with eigenvector $(\varphi_0(\lambda_k), \dots, \varphi_{n-1}(\lambda_k))^t$. Thus, if we are given the Jacobi matrix, we recover the support of μ, it is precisely the spectrum of T_n. We can also recover the weights as follows (we have seen very similar reasoning earlier). Observe that T is the matrix for the multiplication operator on $L^2(\mu)$. Therefore,

$$\langle T^m e_0, e_0 \rangle = \langle x^m, x^0 \rangle_{L^2(\mu)} = \alpha_m = \sum_{k=1}^{n} \lambda_k^m p_k.$$

On the other hand, the spectral decomposition of the Jacobi matrix is $T_n = Q \Lambda Q^t$ where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ and

$$Q_{i,j} = \frac{\varphi_i(\lambda_j)^2}{\sum_{\ell=0}^{n-1} \varphi_i(\lambda_j)^2}.$$

Therefore, it is clear that

$$\langle T_n^m e_0, e_0 \rangle = \langle Q \Lambda^m Q^t e_0, e_0 \rangle = \sum_{j=1}^{n} \lambda_j^m Q_{0,j}^2.$$

Equating with the earlier identity, we have recovered the measure as

$$\mu = \sum_{k=1}^{n} p_k \delta_{\lambda_k}$$

where λ_k are the eigenvalues of T_n and $p_k = Q_{0,k}^2$ are the squared entries of the first row of the eigenvector matrix of T_n.

Conclusion: From a finitely supported measure, we can compute its moments. From the moment sequence we can recover the measure by going first to the orthogonal polynomials, then to the Jacobi matrix describing the three-term recurrence, and from there to the measure, via the *spectral decomposition of the Jacobi matrix*. In summary, the measure is just the *spectral measure of the Jacobi matrix at the vector* e_0.

11. THE UNIQUENESS QUESTION: CONNECTION TO SPECTRAL THEOREM

If α is a positive definite sequence, we construct its Jacobi matrix $T = T(a, b)$. Going by the finite support case, we may expect that the measure (or one measure) with moment sequence α can be recovered from T by taking spectral measure at e_0. There are many subtleties on the way.

First w regard T as an operator on sequences by mapping $(Tx)_n = b_{n-1}x_{n-1} + a_n x_n + b_n x_{n+1}$. While this is well-defined for any $x \in \mathbb{R}^{\mathbb{N}}$, to talk about spectral theorem, we must work inside a Hilbert space. Here the natural Hilbert space is $\ell^2 = \{(x_0, x_1, x_2, \ldots) : \sum_n x_n^2 < \infty\}$.

A special case is when the entires of T are bounded. In this case, by Cauchy-Schwarz inequality

$$|(Tx)_n|^2 \le (a_n^2 + b_{n-1}^2 + b_n^2)(x_{n-1}^2 + x_n^2 + x_{n+1}^2)$$

and hence $\|Tx\|_{\ell^2}^2 \le 3M^2\|x\|_{\ell^2}^2$ where M is a bound for $|a_n|$s and $|b_n|$s. Thus $T : \ell^2 \mapsto \ell^2$ is a bounded operator. It also satisfies $\langle Tx, y\rangle = \langle x, Ty\rangle$ for all $x, y \in \ell^2$, which makes it self-adjoint.

The spectral theorem for bounded self-adjoint operators tells us that we can write $T = \int \lambda dE(\lambda)$ where E is a *projection valued measure*. This representation is also unique etc. As a consequence, there is a measure μ (it is defined by $\mu(A) = \langle E(A)e_0, e_0\rangle$ for $A \in \mathcal{B}_{\mathbb{R}}$) such that or any m,

$$\langle T^m e_0, e_0\rangle = \int x^m d\mu(x).$$

Recall that the positive definite sequence can be recovered from the Jacobi matrix as $\alpha_n = \langle T^n e_0, e_0\rangle$ to see that μ has the moment sequence α.

If the entries of T are not bounded, it is no longer the case that T defines a bounded linear operator on ℓ^2. By restricting the domain to $D = \{x \in \ell^2 : x_n = 0 \text{ eventually}\}$, we see that $T : D \mapsto \ell^2$ is linear. Since D is dense in ℓ^2, if we had $\|Tx\|_{\ell^2} \le C\|x\|_{\ell^2}$ for $x \in D$, then it would extend to all of ℓ^2 as a bounded linear operator. That is not the case when the entries of T are unbounded. These operators are called *unbounded operators*

Example 26. Let $a_n = n$ and $b_n = 0$. Then of T acts on D by $(Tx)_n = nx_n$. We could also have defined T on a larger domain $D_1 = \{x : \sum nx_n^2 < \infty\}$ because then T clearly maps D_1 into ℓ^2. It is better to denote tis operator as T_1 and regard it as an extension of T.

In functional analysis class one learns how to associate an adjoint operator T^* which is defined on another proper subspace D^*. For our T, the symmetry of the Jacobi matrix forces that $D^* \supseteq D$ and that $T^*\big|_D = T$. We say that T is *symmetric*. This is not sufficient to get a spectral decomposition. What one needs is *self-adjointness*, i.e., for D and D^* to coincide and T and T^* to coincide. Once T is self-adjoint, spectral theorem can be proved in full force (the only difference is that when T is bounded, the projection valued measure E and the spectral measure μ are both compactly supported, while they need no be so now).

Example 27. Take $T : D \mapsto \ell^2$ and $T_1 : D_1 \mapsto \ell^2$ as in the previous example. It is easy to work out that $D^* = D_1^* = D_1$. Further, T^*, T_1^* and T coincide on D_1. Thus (in the language introduced next), T is symmetric, while T_1 is self-adjoint.

To achieve this one tries to extend T to a larger domain $D_1 \supseteq D$ and get an operator $T_1 : D_1 \mapsto \ell^2$. Then it turns out that $D_1^* \subseteq D^*$ and $T_1^* : D_1^* \mapsto \ell^2$ is the adjoint of T_1. General theorems assert the existence of self-adjoint extensions (at least for our Jacobi matrices), but there can be several self-adjoint extensions. This has repercussions in the moment problem.

Theorem 28. *Let T be the Jacobi matrix of a positive definite sequence α (with $\alpha_0 = 1$). Regard it as an operator $T : D \mapsto \ell^2$ where D is the set of sequences that are eventually 0.*

(1) If $\tilde{T} : \tilde{D} \mapsto \ell^2$ is a self-adjoint extension of T, then the spectral measure of $\tilde{T}$ at the vector e_0 is a probability measure whose moment sequence is α. If the self-adjoint extension is unique, this is the unique measure having this moment sequence.

(2) If there are distinct self-adjoint extensions, then they have distinct spectral measures at e_0, each having the same moment sequence α.

Part 3: Isoperimetric inequality

1. Isoperimetric inequality

Isoperimetric nequality is a well-known statement in the following form: *Among all bodies in space (in plane) with a given volume (given area), the one with the least surface area (least perimeter) is the ball (the disk).*

Several things need to be made precise. The notion of volume in space or area in the plane are understood to mean Lebesgue measure on $\mathbb{R}^3$ or $\mathbb{R}^2$ or more generally on $\mathbb{R}^d$ (we denote it by $m_d(A)$). Then of course we restrict the notion of "bodies" to Borel sets (or Lebesgue measurable sets).

Still, in measure theory class we (probably!) did not study the notion of surface area of a Borel set in $\mathbb{R}^3$ or the perimeter of a Borel set in $\mathbb{R}^2$. We first need to fix this notion. And then state a precise theorem. First we state a form of the isoperimetric inequality which completely avoids the notion of surface area or perimeter.

Theorem 1 (Isoperimetric inequality)**.** *Let A be Borel subsets of $\mathbb{R}^d$ and let B be a closed ball such that $m_d(A) = m_d(B)$. Then, for any $\epsilon > 0$, we have $m_d(A_\epsilon) \geq m_d(B_\epsilon)$ where $A_\epsilon = \{x \in \mathbb{R}^d : d(x, y) \leq \epsilon \text{ for some } y \in A\}$.*

How does this relate to the informally stated version above? If at all we can define the surface area of A, it must be the limit (or $\limsup$ or $\liminf$) of $(m_d(A_\epsilon) - m_d(A))/\epsilon$ as $\epsilon \to 0$. For simplicity, let us define the surface area (or "perimeter") of a Borel set $A \subseteq \mathbb{R}^d$ as

$$\sigma_d(A) := \limsup_{\epsilon \to 0} \frac{m_d(A_\epsilon) - m_d(A)}{\epsilon}$$

which is either a non-negative real number or $+\infty$. If A is a bounded set with smooth boundary, then the above definition agrees with our usual understanding of perimeter/surface area.

Theorem 1 clearly gives the following theorem as a corollary.

Theorem 2 (Isoperimetric inequality - standard form)**.** *Let A be Borel subsets of $\mathbb{R}^d$ and let B be a closed ball such that $m_d(A) = m_d(B)$. Then, $\sigma_d(A) \geq \sigma_d(B)$.*

In this sense, we are justified in saying that Theorem 1 is stronger than Theorem 2. In addition, note the great advantage of the former being easy to state for all Borel sets without having to define the notion of surface area. However, we have omitted a key point in the isoperimetric inequality which is the uniqueness of the surface-area-minimizing set.

Theorem 3 (Equality in isoperimetric inequality)**.** *In the setting of Theorem 1 assume that A is closed. If $m_d(A_\epsilon) = m_d(B_\epsilon)$ for some $\epsilon > 0$, then $A = B(x, r)$ for some $x \in \mathbb{R}^d$.*

However, the analogous statement for Theorem 1 is false without further qualifications. For example, if A is the disjoint union of a closed disk and a closed line segment, then it has the same area and the same perimeter as the ball. But the uniqueness is "essentially true", for example, if one restricts to sets with smooth boundary or alternately by taking a more general notion of perimeter (which does distinguish a disk from a union of a disk and a line segment). We shall present two proofs of Theorem 1. A short one using the Brunn-Minkowski inequality and a longer but more natural one by Steiner symmetrization.

Exercise 4. Show that the isoperimetric inequality is equivalent to the following statement: If $A \subseteq \mathbb{R}^d$ is measurable, then $|A|^{\frac{d-1}{d}} \leq C_d \sigma_d(A)$ where $C_d^{-1} = d^{1-\frac{1}{d}} \tau_d^{1/d}$ and $\tau_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ is the surface area of the unit sphere S^{d-1}.

Here is a proof of isoperimetric inequality in the plane under some restrictions.

Exercise 5. Let $\gamma(t) = (x(t), y(t))$, $0 \leq t \leq L$ be a simple smooth curve in the plane, parameterized by its arc length, i.e., $\|\dot{\gamma}(t)\| = 1$ for all $t \in [0, 2\pi]$. Let A be the area enclosed by γ and let L be the length of γ.

(1) Show that the length of the curve is given by $L^2 = \int_0^{2\pi} |\dot{\gamma}(t)|^2 dt$ and $A = -\int_0^{2\pi} y(t)\dot{x}(t)dt$.

(2) WLOG assume that $\int_0^{2\pi} y(t)dt = 0$ and show that $\int_0^{2\pi} y(t)^2 dt \leq \int_0^{2\pi} \dot{y}(t)^2 dt$. [**Hint:** Assume that the Fourier series $y(t) = \sum_{n\in\mathbb{Z}} \hat{y}_n e^{int}$ converges nicely and uniformly]

Presentation topic: Proof of isoperimetric inequality via symmetrization and induction on the dimension.

2. Brunn-Minkowski inequality and a first proof of isoperimetric inequality

For simplicity write $|A|$ for $m_d(A)$, the d-dimensional Lebesgue measure. For nonempty sets $A, B \subseteq \mathbb{R}^d$, define their Minkowski sum $A + B := \{a + b : a \in A, b \in B\}$.

Theorem 6 (Brunn-Minkowski inequality). *If A, B are non-empty Lebesgue measurable subsets of $\mathbb{R}^d$, and if $A + B$ is also Lebesgue measurable, then,*

$$|A + B|^{1/d} \geq |A|^{1/d} + |B|^{1/d}.$$

The proof is very easy in one dimension. In fact, it is a continuous analogue of the following inequality that we leave as an exercise.

Exercise 7 (Cauchy-Davenport inequality). Let A, B be non-empty finite subsets of $\mathbb{Z}$. Then $|A + B| \geq |A| + |B| - 1$ and the inequality cannot be improved (here $|A|$ denotes the cardinality of A).

Use the same idea to prove Brunn-Minkowski inequality for $d = 1$.

Proof of Theorem 1 using Brunn-Minkowski inequality. Assume $|A| = |rB|$ where B is the unit ball and $r > 0$. Then $A_\epsilon = A + \epsilon B$ and hence by Brunn-Minkowski

$$\begin{aligned}|A_\epsilon|^{1/d} &\geq |A|^{1/d} + \epsilon|B|^{1/d} \\ &= r|B|^{1/d} + \epsilon|B|^{1/d} \\ &= |(r+\epsilon)B|^{1/d}.\end{aligned}$$

Since $(rB)_\epsilon = (r+\epsilon)B$, we have proved that $|A_\epsilon| \geq |(rB)_\epsilon|$ as required. ■

Proof of Brunn-Minkowski inequality. The proof will proceed by proving it when the two sets are rectangles (parallelepipeds) with sides parallel to the co-ordinate, then for finite unions of rectangles, and finally

Step 1: Suppose $A = \mathbf{x} + [0, a_1] \times \ldots \times [0, a_d]$ and $B = \mathbf{y} + [0, b_1] \times \ldots \times [0, b_d]$ are any two closed parallelepipeds with sides parallel to the axes (we shall refer to them as standard parallelepipeds). Then $A + B = \mathbf{x} + \mathbf{y} + [0, a_1 + b_1] \times \ldots \times [0, a_d + b_d]$. Thus,

$$\begin{aligned}\frac{|A|^{1/d} + |B|^{1/d}}{|A+B|^{1/d}} &= \left(\prod_{k=1}^{d} \frac{a_k}{a_k + b_k}\right)^{1/d} + \left(\prod_{k=1}^{d} \frac{b_k}{a_k + b_k}\right)^{1/d} \\ &\leq \frac{1}{d}\sum_{k=1}^{d} \frac{a_k}{a_k + b_k} + \frac{1}{d}\sum_{k=1}^{d} \frac{b_k}{a_k + b_k} \qquad \text{(AM-GM inequality)} \\ &= 1.\end{aligned}$$

Step 2: Suppose $A = A_1 \sqcup \ldots \sqcup A_m$ and $B = B_1 \sqcup \ldots \sqcup B_n$ are finite unions of standard closed parallelepipeds with pairwise disjoint interiors. When $m = n = 1$ we have already proved the theorem. By induction on $m + n$, we shall prove it for all $m, n \geq 1$. This is the cleverest part of the proof.

Translating A or B does not change any of the quantities in the inequality, hence we may freely do so. Assume $m \geq 2$ without loss of generality (else interchange A and B).

Claim: There is at least one axis direction $j \leq d$ and a number $t \in \mathbb{R}$ such that each of the sets $A' := A \cup \{x : x_j \leq t\}$ and $A'' := A \cap \{x : x_j < t\}$ are both unions of atmost $m - 1$ standard parallelepipeds with pairwise disjoint interiors.

Proof of the claim: Let $R_1 = [a_1, b_1] \times \ldots \times [a_d, b_d]$ and $R_2 = [p_1, q_1] \times \ldots \times [p_d, q_d]$ be two among the parallelepipeds that comprise A. If $I_j = [a_j, b_j] \cap [p_j, q_j]$, then $I_1 \times \ldots \times I_d \subseteq R_1 \cap R_2$. But R_1 and R_2 have disjoint interiors, hence I_j must be empty or be a singleton for some j. This means $b_j \leq t \leq p_j$ or $q_j \leq t \leq a_j$, and we set $t = b_j$ or $t = q_j$ accordingly. The hyperplane $\{x : x_j = t\}$ will do the job, since R_1 will lie on one side of it and R_2 on the other (the boundary of both may intersect the hyperplane). The claim is proved.

Set $\lambda = |A'|/|A|$. By the above claim, $0 < \lambda < 1$ and each of A' and A'' is a disjoint union of at most $m-1$ parallelepipeds (with sides parallel to the axes). Now translate B along the jth direction, i.e., for each s consider $B_s := B + s\mathbf{e_j}$ and let $B'_s = B_s \cap \{x : x_j \leq t\}$ and $B''_s = B_s \cap \{x : x_j \geq t\}$. Choose a value of s such that $|B'_s| = \lambda|B|$ and set $B' = B'_s$ and $B'' = B''_s$.

By the induction hypothesis,

$$|A' + B'| \geq \left(|A'|^{1/d} + |B'|^{1/d}\right)^d = \lambda \left(|A|^{1/d} + |B|^{1/d}\right)^d,$$

$$|A'' + B''| \geq \left(|A''|^{1/d} + |B''|^{1/d}\right)^d = (1-\lambda) \left(|A|^{1/d} + |B|^{1/d}\right)^d.$$

Further, observe that $A' + B' \subseteq \{x : x_j \leq 2t\}$ and $A'' + B'' \subseteq \{x : x_j \geq 2t\}$ and hence $|(A' + B') \cap (A' + B')| = 0$, the intersection being contained in the hyperplane $\{x : x_j = t\}$. Therefore,

$$\begin{aligned} |A + B| &= |A' + B'| + |A'' + B''| \\ &= \lambda \left(|A|^{1/d} + |B|^{1/d}\right)^d + (1-\lambda) \left(|A|^{1/d} + |B|^{1/d}\right)^d \\ &= \left(|A|^{1/d} + |B|^{1/d}\right)^d. \end{aligned}$$

This completes the proof when A, B are finite unions of standard parallelepipeds.

Step 3: Let A and B be compact sets. Let $Q = [-1,1]^d$ and fix $\epsilon > 0$. Observe that compactness of A implies that there exist $x_1, \ldots, x_n \in A$ (for some n) such that $A \subseteq A''$ where $A'' = \cup_{i=1}^n (x_i + \epsilon Q)$. It is easy to see that $A'' \subseteq A_{\epsilon\sqrt{d}}$ and that A'' may be written as a finite union of standard rectangles whose interiors are pairwise disjoint. Similarly find $B'' = \cup_{j=1}^m (y_j + \epsilon Q)$ that is a union of standard rectangles whose interiors are pairwise disjoint and such that $B \subseteq B'' \subseteq B_{\epsilon\sqrt{d}}$.

Then, observe that $A'' + B'' \subseteq (A+B)_{2\sqrt{d}\epsilon}$. Since A'' and B'' are finite unions of standard parallelepipeds, by the previous case, we know that Brunn-Minkowski inequality applies to them. Thus,

$$\begin{aligned} |(A+B)_{2\sqrt{d}\epsilon}| &\geq |A'' + B''| \\ &\geq (|A''|^{1/d} + |B''|^{1/d})^d \\ &\geq (|A|^{1/d} + |B|^{1/d})^d. \end{aligned}$$

This is true for every $\epsilon > 0$. As $A + B$ is compact we see that $\cap_{\epsilon > 0}(A+B)_{2\sqrt{d}\epsilon} = A + B$ and hence $|(A+B)_{2\sqrt{d}\epsilon}| \downarrow |A + B|$ as $\epsilon \downarrow 0$. Therefore, Brunn-Minkowski inequality holds true when A and B are compact.

Step 4: Let A and B be general Borel sets. If either of A or B has infinite Lebesgue measure, there is nothing to prove. Otherwise, by regularity of Lebesgue measure, there are compact sets $A' \subseteq A$ and $B' \subseteq B$ such that $|A \setminus A'| < \epsilon$ and $|B \setminus B'| < \epsilon$. Then of course $A + B \supseteq A' + B'$ and hence

$$|A + B|^{1/d} \geq |A' + B'|^{1/d} \geq |A'|^{1/d} + |B'|^{1/d} \geq (|A| - \epsilon)^{1/d} + (|B| - \epsilon)^{1/d}.$$

Letting $\epsilon \to 0$ we get the inequality for A and B. ■

Remark 8. If we do not assume that $A+B$ is measurable, then (see the last step) we still get

$$m_*(A+B)^{1/d} \geq |A|^{1/d} + |B|^{1/d}$$

where m_* is the inner Lebesgue measure, $m_*(C) := \sup\{m(K) : K \subseteq B,\ K \text{ compact}\}$. We shall see in the next section that $A+B$ is not necessarily measurable.

Exercise 9. Let K be a bounded convex set in $\mathbb{R}^d$. Fix a unit vector $u \in \mathbb{R}^d$ and let $K^t := \{x \in K : \langle x, u\rangle = t\}$ denote the sections of K for any $t \in \mathbb{R}$. Let $I = \{t : |K_t| > 0\}$ and let $f : I \mapsto \mathbb{R}$ be defined by $f(t) = |K_t|^{1/(n-1)}$. Show that I is an interval and that f is concave. **[Note:** Here $|K_t|$ denotes the $(d-1)$-dimensional Lebesgue measure of K_t in the hyperplane $\{x \in \mathbb{R}^d : \langle x, u\rangle = t\}$.]

3. Measurability questions

We want to exhibit measurable sets $A, B \subseteq \mathbb{R}$ such that $A+B$ is not measurable. In fact we shall produce an example with $B = A$. This construction is due to Sierpinski[7] and may also be taken simply as a construction of a non-measurable set (quite different from the one usually presented in measure theory class).

Step 1: Let $K \subseteq [0,1]$ be the usual $1/3$-set of Cantor. Then $K + K \supseteq [0,2]$.

To see this, recall that Cantor set consists of numbers whose ternary expansion has digits 0 and 2 (but not 1). Hence, if $x, y \in \frac{1}{2}\cdot K = \{u/2 : u \in K\}$, then $x = \sum_{i=1}^{\infty} \frac{x_i}{3^i}$ and $y = \sum_{i=1}^{\infty} \frac{y_i}{3^i}$ with $x_i, y_i \in \{0,1\}$. Now consider any $t \in [0,1]$ and write $t = \sum_{i=1}^{\infty} \frac{t_i}{3^i}$ where $t_i \in \{0,1,2\}$. Clearly, we can find $x_i, y_i \in \{0,1\}$ such that $x_i + y_i = t_i$ for each i. Thus, a given $t \in [0,1]$ can be written as $x+y$ with $x, y \in \frac{1}{2}\cdot K$ and hence a number in $[0,2]$ can be written as a sum of two elements of K.

Step 2: Regard $\mathbb{R}$ as a vector space over $\mathbb{Q}$. Then the first step says that the span of K is $\mathbb{R}$. Hence, by a standard application of Zorn's lemma, there exists a basis $B \subseteq K$ for the vector space.

Step 4: Define $E_0 = B \sqcup (-B) \sqcup \{0\}$ and $E_n = E_{n-1} + E_{n-1}$ for $n \geq 1$. From the previous step, it follows that

$$\bigcup_{n\geq 0} \bigcup_{q \geq 1} \frac{1}{q} E_n = \mathbb{R}.$$

Indeed, given $x \in \mathbb{R}$, write it as $x = r_1 b_1 + \ldots + r_n b_n$ with $n \geq 1$, $r_i \in \mathbb{Q}$, $b_i \in B$. Taking q to be the product of the denominator of r_is, we get $x = \frac{1}{q}(p_1 b_1 + \ldots + p_n b_n)$ with $p_i \in \mathbb{Z}$. Negating b_i if necessary (it will still be in E_0), we may assume $q \geq 1$ and $p_i \geq 1$.

Step 5: Let m be the smallest n for which $m^*(E_n) > 0$. Since E_0 is a subset of a set of zero measure, $m \geq 1$. Hence it makes sense to set $A = E_{m-1}$. Then A is Lebesgue measurable (since its outer measure is zero). We claim that $A + A = E_m$ is not Lebesgue measurable.

[7] We have taken this presentation from Rubel's paper *A pathological Lebesue measurable function*.

Indeed, if E_m was measurable, by Steinhaus' lemma, E_m+E_m contains an open interval around 0 (since E_m is symmetric, E_m+E_m is the same as E_m-E_m). But then E_{m+1} contains an interval around 0. Thus, given any $x \in \mathbb{R}$, we can find $q \geq 1$ so that $x/q \in E_{m+1}$. The conclusion is that every element of $\mathbb{R}$ can be written as a linear combination of at most 2^{m+1} distinct elements of B. But B is an infinite set (i.e., $\mathbb{R}$ has infinite dimensions over $\mathbb{Q}$), and hence $b_1+\ldots+b_k \notin E_{m+1}$ if $k > 2^{m+1}$ and b_i are distinct elements of B. This contradiction can only be resolved by accepting that E_m cannot be measurable. ■

Remark 10. There is also an example to show that the Minkowski sum of Borel sets need not be Borel. However, the sum-set will necessarily be Lebesgue measurable.

Here is two simpler facts, one of which was used in the proof of Brunn-Minkowski inequality.

Exercise 11. Show that the Minkowski sum of two compact sets is $\mathbb{R}$ is necessarily compact. Show that the Minkowski sum of two closed sets in $\mathbb{R}$ need not be closed.

4. Functional form of isoperimetric inequality

It is a general idea that a statement about sets must have an analogous statement for functions and vice versa. When the function is taken to be the indicator of a set the functional inequality should reduce to the inequality for sets. This may not make sense immediately as there may be assumptions of smoothness etc., that are not satisfied by indicator functions, but in some approximate sense this should hold. Here is the functional analogue of the isoperimetric inequality.

Theorem 12 (Sobolev inequality). *Let $d \geq 2$ and $p = \frac{d}{d-1}$. Then, for every $f \in C_c^1(\mathbb{R}^d)$, we have*

$$\|f\|_{L^p} \leq \|\nabla f\|_{L^1}.$$

The idea is explained easily when $d=2$.

Proof for $d=2$. Let f_i denote the ith partial derivative. Then, for any $(x,y) \in \mathbb{R}^2$, we have

$$f(x,y) = \int_{-\infty}^{x} f_1(s,y)ds \implies |f(x,y)| \leq \int_{\mathbb{R}} |f_1(s,y)|ds.$$

$$f(x,y) = \int_{-\infty}^{y} f_2(x,t)dt \implies |f(x,y)| \leq \int_{\mathbb{R}} |f_1(x,t)|dt.$$

Multiplying the two inequalities, we get

$$|f(x,y)|^2 \leq \left(\int_{\mathbb{R}} |f_1(s,y)|ds\right)\left(\int_{\mathbb{R}} |f_2(x,t)|dt\right).$$

Integrate over x and y and observe that the right hand side factors

$$\int_{\mathbb{R}^2} |f(x,y)|^2 dxdy \leq \left(\int_{\mathbb{R}^2} |f_1(s,y)|dsdy\right)\left(\int_{\mathbb{R}^2} |f_2(x,t)|dtdx\right)$$

Since $\|\nabla f\| = \sqrt{f_1^2 + f_2^2}$, we have $|f_1| \le \|\nabla f\|$ and $|f_2| \le \|\nabla f\|$, and therefore

$$\int_{\mathbb{R}^2} |f|^2 \le \left(\int_{\mathbb{R}^2} \|\nabla f\| \right)^2$$

which is precisely the claim of Sobolev inequality for $d = 2$. ■

The proof for $d \ge 3$ needs a little more work.

Proof for $d = 3$. We have as before

$$|f(x_1, x_2, x_3)| \le \int |f_1(s_1, x_2, x_3)| ds_1 =: I_1(x_2, x_3)$$
$$|f(x_1, x_2, x_3)| \le \int |f_2(x_1, s_2, x_3)| ds_2 =: I_2(x_1, x_3)$$
$$|f(x_1, x_2, x_3)| \le \int |f_3(x_1, x_2, s_3)| ds_3 =: I_3(x_1, x_2).$$

We again multiply them and write

$$|f(x_1, x_2, x_3)|^3 \le I_1(x_2, x_3) \times I_2(x_1, x_3) \times I_3(x_1, x_2).$$

But if we integrate with respect to x_1, x_2, x_3, the right side does not split as a product of integrals - this is the difference from the case $d = 2$. Note also that the power of f on the left is 3 while the desired inequality should have $3/2$. Take square roots and integrate over x_1 alone. We get

$$\begin{aligned}\int_{\mathbb{R}} |f(x_1, x_2, x_3)|^{3/2} dx_1 &\le I_1(x_2, x_3)^{1/2} \int_{\mathbb{R}} I_2(x_1, x_3)^{1/2} I_3(x_1, x_2)^{1/2} dx_1 \\ &\le I_1(x_2, x_3)^{1/2} \left(\int_{\mathbb{R}} I_2(x_1, x_3) dx_1 \right)^{1/2} \left(\int_{\mathbb{R}} I_3(x_1, x_2) dx_1 \right)^{1/2} \\ &=: I_1(x_2, x_3)^{1/2} J(x_3)^{1/2} K(x_2)^{1/2}\end{aligned}$$

by Cauchy-Schwarz. Now integrate over x_2. The second factor is independent of x_2.We apply Cauchy-Schwarz again to get

$$\begin{aligned}\int_{\mathbb{R}^2} |f(x_1, x_2, x_3)|^{3/2} dx_1 dx_2 &\le J(x_3)^{1/2} \left(\int_{\mathbb{R}} I_1(x_2, x_3) dx_2 \right)^{1/2} \left(\int_{\mathbb{R}} K(x_2) dx_2 \right)^{1/2} \\ &=: J(x_3)^{1/2} L(x_3)^{1/2} \left(\int_{\mathbb{R}} K(x_2) dx_2 \right)^{1/2}\end{aligned}$$

Now integrate over x_3 and apply Cauchy-Schwarz again to the first two factors to get

$$\int_{\mathbb{R}^3} |f(x_1, x_2, x_3)|^{3/2} dx_1 dx_2 dx_3 \le \left(\int K(x_2) dx_2 \right)^{1/2} \left(\int_{\mathbb{R}} J(x_3) dx_3 \right)^{1/2} \left(\int_{\mathbb{R}} L(x_3) dx_3 \right)^{1/2}.$$

It remains to observe that

$$\int_{\mathbb{R}} K(x_2) dx_2 = \int_{\mathbb{R}^3} |f_3|, \quad \int_{\mathbb{R}} J(x_3) dx_3 = \int_{\mathbb{R}^3} |f_2|, \quad \int_{\mathbb{R}} L(x_3) dx_3 = \int_{\mathbb{R}^3} |f_1|.$$

Each of the three is bounded by $\|\nabla f\|_{L^1}$ and hence,

$$\int_{\mathbb{R}^3} |f(x_1,x_2,x_3)|^{3/2} dx_1 dx_2 dx_3 \le \left(\int_{\mathbb{R}^3} \|\nabla f(x_1,x_2,x_3)\| dx_1 dx_2 dx_3 \right)^{3/2}$$

This is the Sobolev inequality for $d = 3$. ■

Exercise 13. Write the proof for general d. **[Hint:** Use Hölder's inequality in place of Cauchy-Schwarz where necessary.]

5. Proof of isoperimetric inequality by symmetrization

Using symmetrization techniques introduced by Steiner and induction on the dimension, we give a proof of the isoperimetric inequality[8].

Theorem 14. *Let A be a compact subset of $\mathbb{R}^d$ and let B be a closed ball with $|A| = |B|$. Then, $|A_\epsilon| \ge |B_\epsilon|$ for all $\epsilon > 0$.*

The theorem is obvious in one dimension. Indeed, if $M = \max A$ and $m = \min A$, then $A_\epsilon \setminus A$ contains the intervals $(M, M+\epsilon)$ and $(m-\epsilon, m)$ and hence has measure at least 2ϵ. But B is an interval and clearly $|B_\epsilon| = |B| + 2\epsilon$. Thus, $|A_\epsilon| \ge |B_\epsilon|$ for all $\epsilon > 0$.

Next we introduce the key notion of symmetrization. Given a unit vector $\mathbf{e} \in \mathbb{R}^{\mathbf{d}}$, let $\ell = \mathbb{R}\mathbf{e}$ (a line) and a set A, we define the sections of A as $A^t := A \cap (\ell^\perp + t\mathbf{e})$ for $t \in \mathbb{R}$ (the intersection of A with the affine hyperplane that orthogonal to ℓ and at a distance of t from the origin).

Definition 15. Given a line $\ell = \mathbb{R}\mathbf{e}$ and a compact set A, define the symmetrization of A with respect to ℓ as the set C such that: For any $t \in \mathbb{R}$, the section $C^t := C \cap (t + \ell^\perp)$ is the closed $(d-1)$-dimensional disk in the affine hyperplane $t + \ell^\perp$. Further, the center of C^t is on ℓ and the $(d-1)$-dimensional volume of C^t is the same as that of $A^t := A \cap (t + \ell^\perp)$. To be unambiguous, we adopt the following convention: If A^t is empty, then C^t is defined to be empty. If A^t is non-empty but has zero $(d-1)$-dimnesional Lebesgue measure, then C^t is defined to be a singleton. The resulting set C is denoted as $\sigma_\ell(A)$.

Exercise 16. Show that $\sigma_\ell(A)$ is compact for any compact A.

For compact A, let

$$\mathcal{M}(A) = \left\{ C \subseteq \mathbb{R}^d : C \text{ is compact, } |C| = |A|,\ |C_\epsilon| \ge |A_\epsilon| \text{ for all } \epsilon > 0 \right\}.$$

These are the sets that are better than A in isoperimetric sense. Theorem 14 is equivalent to saying that $\mathcal{M}(A)$ contains a ball. The main idea of the proof is that the symmetrization of a set has better isoperimetric profile than the original set.

[8] This proof is taken from the appendix to a paper of Figiel, Lindestrauss and Milman, where they prove the isoperimetric inequality on the sphere. They modeled it on a well-known proof for the Euclidean case which is written in many books but since I did not find one, I translated back their proof to the Euclidean case. Hence, there may be avoidable complications in the proof.

Lemma 17. *Let A be a compact subset of $\mathbb{R}^d$. Then, $\sigma_\ell(A) \in \mathcal{M}(A)$ for any line ℓ.*

Observe that $\sigma_\ell(A)$ is a set which is symmetric about the axis ℓ. Thus one may expect that by symmetrizing about various lines, the set becomes rounder and rounder, and approach a ball. The lemma assures us that the isoperimetric profile only gets better in the process. But a finite number of operations may not get to a ball. For a rigorous argument, we use an auxiliary functional on sets. Let the radius of a compact set be defined by

$$r(A) = \inf\{r > 0 : B(x, r) \supseteq A \text{ for some } x \in \mathbb{R}^d\}.$$

This will be used as follows.

Lemma 18. *If A is a compact subset that is not a ball, then there exist lines $\ell_1, \ldots, \ell_m$ (for some m) such that $\sigma_{\ell_1} \circ \ldots \circ \sigma_{\ell_m}(A)$ has strictly smaller radius than A.*

The isoperimetric inequality is an easy consequence of the previous two lemmas together with the next one.

Lemma 19. *Let A be a compact set. Then r attains its minimum on $\mathcal{M}(A)$.*

Proof of Theorem 14. Fix A and let B be a minimizer of r on $\mathcal{M}(A)$ (by Lemma 19). If B is not a ball, by Lemma 18 there is a sequence of symmetrizations that strictly reduce the radius. By Lemma 17, the resulting set is still in $\mathcal{M}(A)$, contradicting that B is a minimum of $r(\cdot)$ inside $\mathcal{M}(A)$. ■

It remains to prove the lemmas.

Proof of Lemma 18. Fix A and let $B = B(x, r(A))$ contain A. Take any line ℓ passing through x and symmetrize to get A_1. The ball remains fixed under symmetrization. Since $B \setminus A$ contains an open ball, $\partial B \setminus A_1$ contains a cap $C \subseteq \partial B$ (by cap, we mean a ball inside ∂B in spherical metric).Now pick a line ℓ_1 passing through x and a boundary point of C and symmetrize to get A_2. Then, draw a picture and convince yourself that $\partial B \setminus A_2$ contains a cap C' with radius double that of C. Continuing to reflect on further lines $\ell_2, \ell_3, \ldots$, in a finite number of steps we get to a set A_m such that $A_m \cap \partial B = \emptyset$. Then $r(A_m) < r(A)$. ■

Proof of Lemma 19. First we claim that r is continuous. In fact it is Lipschitz, i.e., $|r(A_1) - r(A_2)| \leq d_H(A_1, A_2)$. This is because $\epsilon > d_H(A_1, A_2)$ and $B(x, r) \supseteq A_1$ implies that $B(x, r + \epsilon) \supseteq A_2$.

We next claim that $\mathcal{M}(A)$ is closed. To see this, let $C_n \in \mathcal{M}(A)$ and $C_n \to C$ in Hausdorff metric. Then $C_\epsilon \supseteq C_n$ for large n showing that $|C_\epsilon| \geq \limsup\limits_{n\to\infty} |C_n| = |A|$. Put $\epsilon = 1/k$ and note that $\cap_{k\geq 1} C_{1/k} = C$ (as C is compact) to get $|C| = \lim\limits_{k\to\infty} |C_{1/k}| \geq |A|$. Further, for any $\delta > 0$ we have $C \subseteq (C_n)_\delta$ for large n and hence $|C| \leq |(C_n)_\delta| \leq |A_\delta|$. Now put $\delta = 1/k$ and use $A = \cap_k A_{1/k}$ to get $|C| \leq |A|$. We have now proved that $|C| = |A|$. Next fix $\epsilon > 0$ and $\delta > 0$ and observe that $|C_\epsilon| \leq \liminf |(C_n)_{\epsilon+\delta}| \leq |A_{\epsilon+\delta}|$ since $C \subseteq (C_n)_\delta$ for large n. Put $\delta = 1/k$ and let $k \to \infty$ to get $|C_\epsilon| \leq |\bar{A}_\epsilon|$, since $\cap_k A_{\epsilon+1/k} = \bar{A}_\epsilon$. Thus, $|C_\epsilon| \leq |\bar{A}_\epsilon|$ for every $\epsilon > 0$. Use this for $\epsilon - 1/k$

and take union over k. Since $C_{\epsilon-k^{-1}}$ increase to C_ϵ and $\overline{A_{\epsilon-k^{-1}}}$ increase to A_ϵ, taking limits we get $|C_\epsilon| \leq |A_\epsilon|$ for any $\epsilon > 0$.

Since $A \in \mathcal{M}(A)$, in minimizing r we may restrict ourselves to $\{C \in \mathcal{M}(A) : r(C) \leq r(A)\}$. Translation does not change isoperimetric profile, hence it suffices to $\mathcal{M}_0(A) = \{C \in \mathcal{M}(A) : C \subseteq B(0, r(A))\}$. But $\mathcal{M}_0(A)$ is a compact set (see Exercise 20) and r is continuous, there must be a minimum. ■

Exercise 20. Let (X, d) be a compact metric space. Then $(\mathcal{C}, d_H)$, the space of closed subsets endowed with Hausdorff metric, is also compact.

The following proof is easier understood with pictures, but I don't have time to draw some.

Some notation used in the following proof: Without loss of generality we shall take the line to be $\ell = \mathbb{R}e_d$ (where $e_d = (0, \ldots, 0, 1)$). For $t \in \mathbb{R}$, let $\tau_t(A) = A + te_d$ (translation in "vertical" direction). We use λ_d to denote the d-dimensional Lebesgue measure λ_{d-1} to denote the lower dimensional Lebesgue measure on any affine hyperplane in $\mathbb{R}^d$ (particularly on the hyperplane $\ell^\perp + te_d = \{x \in \mathbb{R}^d : x_d = t\}$).

Proof of Lemma 17. Fix A and ℓ and let $C = \sigma_\ell(A)$. Since $\lambda_{d-1}(C^t) = \lambda_{d-1}(A^t)$ for all t, it follows that $\lambda_d(A) = \lambda_d(C)$. This is because $\lambda_d(A) = \int_{\mathbb{R}} \lambda_{d-1}(A_t)dt$ and similarly for C.

It remains to compare $\lambda_d(C_\epsilon)$ with $\lambda_d(A_\epsilon)$. The sections of A_ϵ get contributions from many different sections of A. In fact,

$$(A_\epsilon)^t = \bigcup_{s:|s-t|\leq\epsilon} (\tau_{t-s}[A^s])_{\sqrt{\epsilon^2-(t-s)^2}}. \tag{1}$$

The notation does not show this, but the neighbourhoods on the right are taken inside the hyperplane $\ell^\perp + te_d$. Analogously,

$$(C_\epsilon)^t = \bigcup_{s:|s-t|\leq\epsilon} (\tau_{t-s}[C^s])_{\sqrt{\epsilon^2-(t-s)^2}}.$$

A key observation is that for fixed t, the sets on the right are concentric balls in H_t, hence there is at least one s for which $(\tau_{t-s}[C^s])_{\sqrt{\epsilon^2-(t-s)^2}}$ is equal to the whole set of $(C_\epsilon)^t$.

For that s, we use the inequality

$$\lambda_{d-1}((\tau_{t-s}[C^s])_{\sqrt{\epsilon^2-(t-s)^2}}) \leq \lambda_{d-1}((\tau_{t-s}[A^s])_{\sqrt{\epsilon^2-(t-s)^2}}). \tag{2}$$

This inequality follows inductively (we assume the validity of Theorem 14 for dimension $d-1$) and using $|C^s| = |A^s|$ (which implies $|\tau_{t-s}[C^s]| = |\tau_{t-s}[A^s]|$, of course). In (2), the left side is equal to $\lambda_{d-1}((C_\epsilon)^t)$ by the choice of s, while the right hand side is at most $\lambda_{d-1}((A_\epsilon)^t)$ by (1). Thus,

$$\lambda_{d-1}((C_\epsilon)^t) \leq \lambda_{d-1}((A_\epsilon)^t).$$

Integrate over t to get $\lambda_d(C_\epsilon) \leq \lambda_d(A_\epsilon)$. Thus, we have proved that $C \in \mathcal{M}(A)$. ■

Remark 21. As remarked earlier, this proof is taken from a paper of Figiel, Lendenstrauss and Milman where they prove isoperimetric inequality in the sphere $\mathbb{S}^{n-1}$. Brunn-Minkowski inequality does not make sense in the sphere (there is no addition operation) but the above proof by symmetrization goes though virtually identically, with spherical metric replacing the Euclidean metric and symmetrization done w.r.t. great circles in place of straight lines. One difference is in the proof of Lemma 17, where $\sqrt{\epsilon^2 - (t-s)^2}$ has to be replaced by some function of ϵ, t, s (it is not required to know what this function precisely is!). A lesser point is that in the proof of Lemma 19, the whole collection $\mathcal{M}(A)$ is compact (since the sphere is itself compact), and there is no need to bring in $\mathcal{M}_0$.

Part 4: Matching theorem and its applications

1. Three theorems in combinatorics

1.1. **Hall's matching theorem.** We recall some basic notions. A *graph* $G = (V, E)$ is a set V ("vertex set") together with an edge-set E where E is a symmetric relation on the set V (i.e., $E \subseteq V \times V$ and $(u, v) \in E$ implies $(v, u) \in E$). We shall also assume that the relation is anti-reflexive, i.e., $(u, u) \notin E$ for any $u \in E$. In such a case, we shall be loose in our language and say that $\{u, v\}$ is an edge or write $u \sim v$ and say u is adjacent to v. Also, we say that the edge $\{u, v\}$ is incident to the vertices u and v.

When we talk about a *directed graph*, the symmetry condition on E is dropped (and the convention is to say that the edge (u, v) is directed from u towards v) but we shall still require it to be anti-reflexive. Clearly, any undirected graph can also be considered as a directed graph.

A *bipartite graph* is an undirected graph whose vertex set V can be partitioned into V_1 and V_2 such that if $u \sim v$, then $u \in V_1, v \in V_2$ or $u \in V_2, v \in V_1$.

A (complete) *matching* of a graph is a collection of edges such that every vertex is adjacent to exactly one edge in the collection.

In a graph, $G = (V, E)$, let $N(A) = \{v \in V : v \sim u \text{ for some } u \in A\}$ be the neighbourhood of A.

Theorem 1 (Hall's marriage theorem). *Let $G = (V, E)$ be a finite bipartite graph with parts V_1, V_2 of equal cardinality. Then G has a complete matching if and only if $|N(A)| \geq |A|$ for all $A \subseteq V_1$.*

We shall derive Hall's theorem from a more general theorem of Dilworth on posets.

1.2. **Dilworth's theorem.** Let $\mathcal{P}$ be a partially ordered set. Recall that this means that there is a relation (denoted "$\leq$") on $\mathcal{P}$ that is reflexive ($x \leq x$ for all $x \in \mathcal{P}$), anti-symmetric ($x \leq y$ and $y \leq x$ imply $x = y$) and transitive ($x \leq y$ and $y \leq z$ imply $x \leq z$). It will be convenient to write the reversed relation as "$\geq$" (i.e., $x \geq y$ if $y \leq x$).

A *chain* is a totally ordered subset in $\mathcal{P}$. An *anti-chain* (also called *independent set*) is a subset in which no two distinct elements are comparable. Suppose we write the poset as a union of chains C_j for $j \in J$ (some index set). If A is any anti-chain in $\mathcal{P}$, it can put at most one point in each of the chains C_j. Therefore, $|A| \leq |J|$ (in these sections $|A|$ will denote the cardinality of A).

Example 2. The collection of all subsets of a given set is a poset with the order given by set inclusion is a poset. For example, if the given set is $\{1, 2, 3\}$, then $C_1 = \{\emptyset, \{1\}, \{1, 2\}, \{1, 2, 3\}\}$, $C_2 = \{\{2\}, \{2, 3\}\}$ and $C_3 = \{\{3\}, \{1, 3\}\}$ are all chains and $C_1 \cup C_2 \cup C_3 = \mathcal{P}$.

Theorem 3. *If m is the maximal size of an anti-chain in a finite poset, then the poset can be written as a union of m chains.*

Use Dilworth's theorem to solve a famous problem first posed by Erdös and Szekeres.

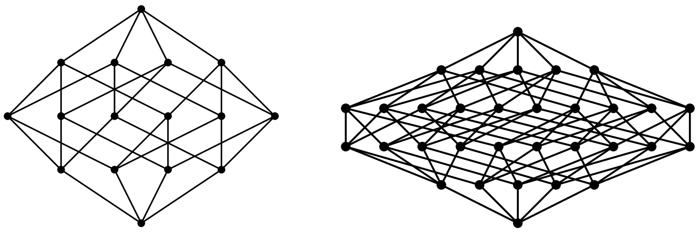

FIGURE 2. Poset of subsets of $\{1,2,3,4\}$ and of $\{1,2\ldots,5\}$.

Exercise 4. Let $N = mn+1$ and $a_1,\ldots,a_N$ be distinct real numbers. Then, there exists an increasing subsequence of cardinality $n+1$ or a decreasing subsequence of cardinality $m+1$ (or both).

In many interesting posets, it is hard to find the size of the maximal anti-chain. A very beautiful example is that of the Boolean poset consisting of subsets of $\{1,2,\ldots,n\}$, with ordering given by set-inclusion. It is clear that A_k, the collection of all subsets with a given cardinality k, is an anti-chain. Since $|A_k| = \binom{n}{k}$, among these anti-chains, the maximal size if $\binom{n}{\lfloor n/2 \rfloor}$. It is a beautiful result of Sperner that in fact this is the maximal size among all anti-chains of the Boolean poset. We outline it as an exercise.

The language used will be simpler if you imagine the Hasse diagram of the Boolean poset as shown in the figures above. At the bottom is the empty set and at the top is the whole set.

Exercise 5. Let an ant start at the empty set and moves upward by picking an element uniformly at random (among all elements in the layer immediately above that are connected by an edge in the Hasse diagram). At each step the picks are made independently. After n steps, the ant is at the top.

For any set $A \subseteq \{1,2,\ldots,n\}$, let $p(A)$ be the probability that the ant passes through the vertex A. Calculate $p(A)$. What can you say about $p(A_i)$s if $A_1,\ldots,A_m$ is an anti-chain?

Put these together to prove Sperner's lemma.

Here is another standard application of Dilworth's theorem. Let G be any graph. A *vertex cover* is any subset of vertices such that every edge is incident to at least one of the vertices in the subset. A *matching* (now we use it to mean incomplete matchings) is a collection of edges such that no vertex of the graph is incident to more than one edge in the collection.

Exercise 6. In a finite bipartite graph, show that the maximal number of edges in any matching is equal to the minimal number of vertices in any vertex cover. This is known as König's theorem.

1.3. **Proof of Hall's theorem and some consequences.** Here is how Dilworth's theorem implies Hall's theorem.

Proof of Hall's matching theorem. Given a bipartite graph as in the statement of Hall's theorem, define a partial order of V by declaring $u \leq v$ if $u \in V_1$, $v \in V_2$ and $u \sim v$ in the graph. We claim that the maximal size of an anti-chain is $n := |V_1| = |V_2|$. Indeed, if A is an anti-chain, $N(A \cap V_1)$ is disjoint from $A \cap V_2$ (if not, there is a vertex in $A \cap V_1$ that is adjacent to a vertex in $A \cap V_2$, contradicting the anti-chain property). Their cardinalities sum to at most $|V_2| = n$. Thus, using Hall's condition, $|A \cap V_1| \leq N(A \cap V_1)|$, we have

$$|A| = |A \cap V_1| + |A \cap V_2| \leq |N(A \cap V_1)| + |A \cap V_2| \leq n.$$

On the other hand, we do have anti-chains of cardinality n (eg., V_1 or V_2). Thus the size of a maximal anti-chain is precisely n.

By Dilworth's theorem, we can write V as a union of n chains. As $|V| = 2n$ and each chain has cardinality at most 2, this means that V is a union of n pairs $\{u, v\}$ with $u \leq v$ (i.e., $u \in V_1$, $v \in V_2$ and $u \sim v$). That is precisely the matching that we want. ■

As a useful consequence of Hall's theorem, we derive a theorem of Birkoff and von Neumann that every doubly stochastic matrix is a convex combination of permutation matrices. Recall that a doubly stochastic matrix is a square matrix having non-negative entries and whose row and column sums are all equal to 1. The space DS_n of all $n \times n$ doubly stochastic matrices is easily seen to be a convex set. It is also compact (as a subset of $\mathbb{R}^{n^2}$). If K is a compact convex set in $\mathbb{R}^d$, then a point is said to be an extreme point of K if it cannot be written as a strict convex combination of two distinct points in K and the set of all extreme points of K is denoted by $E(K)$. In other words, $x \in E(K)$ if and only if $x \in K$ and $x = \alpha y + (1 - \alpha)z$ for some $0 < \alpha < 1$ and $y, z \in K$ implies that $y = z$.

A well-known theorem of Krein and Milman states that for any non-empty compact convex set K in $\mathbb{R}^d$ is the convex hull of its extreme points. That is

$$K = \text{conv}(E(K)) := \left\{ \sum_{i=1}^{n} \alpha_i x_i : n \geq 1, x_i \in E(K), \alpha_i \geq 0 \text{ and } \sum_{i=1}^{n} \alpha_i = 1 \right\}.$$

In fact, Krein-Milman theorem is valid in general locally convex spaces, except that we must take the closure on the right. That is $K = \overline{\text{conv}(E(K))}$.

Example 7. The space of probability measures on $\mathbb{R}$ is a convex set whose extreme points are δ_a, $a \in \mathbb{R}$. The space of probability measure whose mean exists and is equal to 0 is also a convex set. Its extreme points are δ_0 and $\frac{b}{a+b}\delta_{-a} + \frac{a}{a+b}\delta_b$ for some positive a, b. In general, the set of measures with specified m moments will form a convex set. What are its extreme points?

The set DS_n is convex and compact. Its extreme points are precisely the set of permutation matrices (we had trouble justifying this in class, but it follows from the proof below) and then Krein-Milman would imply that all such matrices are convex combinations of permutation matrices. We show this directly, invoking Hall's theorem.

Theorem 8 (Birkoff-von Neumann theorem). *Every doubly stochastic matrix is a convex combination of permutation matrices.*

Proof. Let $A \in DS_n$. Define a bipartite graph with V_1 being the set of rows of A and V_2 being the set of columns of A and put an edge between ith row and jth column if and only if $a_{i,j} > 0$. If $R_1, \ldots, R_k$ are any k rows, the $k \times n$ matrix formed by these rows has a total sum of k (each row sums to 1) and hence the sum of all the column sums is k. Since each column sum is at most 1, there must be at least k con-zero columns. Therefore, $|N(S)| \geq |S|$ for $S = \{R_1, \ldots, R_k\}$. This shows the validity of Hall's conditions, and hence there is a matching of rows and columns in this bipartite graph.

Denote the matching by $i \sim \pi(i)$ where π is a permutation. Let $\alpha = \min\{a_{i,\pi(i)} : i \leq n\}$ which is positive. If P_π denotes the permutation matrix with 1s at $(i, \pi(i))$, then the matrix $A - \alpha P$ has row and column sums equal to $1 - \alpha$. If $\alpha = 1$, then $A = P$ and we are done. If $\alpha < 1$, we can rescale it to a doubly stochastic matrix and write $A = \alpha P + (1 - \alpha)B$ where $B \in DS_n$. Note that B has at least one more zero entry than A. Continue to write B as $\beta Q + (1 - \beta)C$ where Q is a permutation and C is a doubly stochastic matrix, etc. The process must terminate as the number of zeros in the doubly stochastic matrix increases by at least 1 in each step. We end with a representation of A as a convex combination of permutation matrices. ■

1.4. **Proof of Dilworth's theorem.** The proof will be by induction on the cardinality of the poset. Check the base case yourself.

Let $\mathcal{P}$ be a finite poset and let $a_1, \ldots, a_m$ be an anti-chain of maximal cardinality in $\mathcal{P}$. Then, for any $x \in \mathcal{P}$, there is some i such that $x \leq a_i$ or $a_i \leq x$ (otherwise $\{a_1, \ldots, a_m, x\}$ would be a larger anti-chain). Hence, if we define

$$\mathcal{P}_- = \{x : x \leq a_i \text{ for some } i \leq m\}, \quad \mathcal{P}_+ = \{x : a_i \leq x \text{ for some } i \leq m\},$$

then $\mathcal{P} = \mathcal{P}_- \cup \mathcal{P}_+$. Both $\mathcal{P}_-$ and $\mathcal{P}_+$ are posets and $\{a_1, \ldots, a_m\}$ is an anti-chain in both. If we could argue that these two posets had strictly smaller cardinality than $\mathcal{P}$, then inductively we could write them as unions of m chains:

$$\mathcal{P}_+ = C_1^+ \cup \ldots \cup C_m^+, \quad \mathcal{P}_- = C_1^- \cup \ldots \cup C_m^-$$

where each $C_i^\pm$ is a chain and $a_i \in C_i^\pm$. Since a_i is a maximal element in $\mathcal{P}_-$ (and hence in C_i^-) and a minimal element in $\mathcal{P}_+$ (and hence in C_i^+), it follows that $C_i = C_i^+ \cup C_i^-$ is a chain in $\mathcal{P}$. This gives the decomposition $\mathcal{P} = C_1 \cup \ldots \cup C_m$ of the given poset into chains.

The gap in the proof is that $\mathcal{P}_+$ or $\mathcal{P}_-$ could be all of $\mathcal{P}$ (clear, but give an explicit example) and hence induction does not help.

To fix this problem, we first take a maximal chain C_0 in $\mathcal{P}$ and set $\mathcal{Q} = \mathcal{P} \setminus C_0$. Then $\mathcal{Q}$ is strictly smaller than $\mathcal{P}$.

Case 1: Suppose $\mathcal{Q}$ has an anti-chain of size m, say $\{a_1, \ldots, a_m\}$. Now take this anti-chain in the argument outlined earlier (for the poset $\mathcal{P}$, now we may forget $\mathcal{Q}$). The proof is now legitimate because $\mathcal{P}_+$ does not contain the minimal element of C_0 (else $C_0 \cup a_i$ would be a chain for some i and $a_i \notin C_0$). Similarly $\mathcal{P}_-$ does not contain the maximal element of C_0. Both $\mathcal{P}_+$ and $\mathcal{P}_-$ have strictly smaller cardinality and hence the induction hypothesis applies. We get a chain decomposition of $\mathcal{P}$ as above.

Case 2: Suppose that the maximal cardinality of a chain in $\mathcal{Q}$ is m' which is strictly smaller than m (then $m' = m - 1$ in fact). Then write $\mathcal{Q}$ (by induction hypothesis) as a union of m' chains. Together with C_0 this decomposes $\mathcal{P}$ into m chains.

This completes the proof.

2. Haar measure on topological groups

Topological groups: A *topological group* is a group G endowed with a Hausdorff topology such that the operations $(xy) \mapsto xy$ (from $G \times G$ to G) and $x \mapsto x^{-1}$ (from G to G) are continuous.

As examples, we may take any finite or countable group (with discrete topology), the group $(\mathbb{R}^n, +)$, the group $GL(n, \mathbb{R})$ of $n \times n$ invertible matrices with real entries, similarly $GL(n, \mathbb{C})$, the unitary group $\mathcal{U}(n)$, the orthogonal group $O(n)$, various other subgroups of matrices (all with topology inherited from $\mathbb{R}^{n^2}$ or $\mathbb{C}^{n^2}$), the group $M_n(\mathbb{R})$ of isometries of $\mathbb{R}^n$ (which can be built from the translation group $\mathbb{R}^n$, the "rotation group" $O(n)$ and reflections $x \mapsto -x$), group of isometries of Hyperbolic space, groups constructed by taking products such as $(\mathbb{Z}/(2))^J$ for an arbitrary index set J, etc.

Another kind of example (for the sole purpose of giving an exercise): For a graph $G = (V, E)$, by an automorphism of G we mean a bijection $f : V \mapsto V$ such that $u \sim v$ if and only if $f(u) \sim f(v)$. The set of all such automorphisms $\text{Aut}(G)$ forms a group under composition. The graph is said to be *transitive* if for any $u, v \in V$, there exists $f \in \text{Aut}(G)$ such that $f(u) = v$. Examples of transitive graphs are $\mathbb{Z}^d$, lattices, regular trees, etc. Give the topology of pointwise convergence on $\text{Aut}(G)$. If the graph is rooted (i.e., one vertex is distinguished), then the automorphism is required to fix the root. They too form a group (an obvious subgroup of $\text{Aut}(G)$).

Exercise 9. If G is any transitive group with a countable vertex set where each vertex has finite degree , show that the group of automorphisms fixing the root vertex is compact.

Exercise 10. Identify the automorphism group of the rooted infinite binary tree shown in Figure 2.

Invariant measures: On a topological group we may talk of the Borel sigma-algebra and measures on it. We have seen some of these.

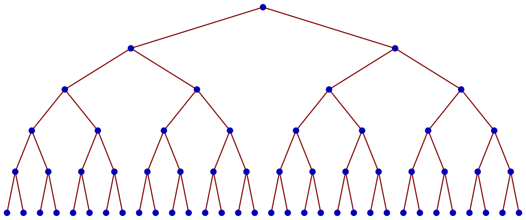

FIGURE 3. The rooted binary tree shown up to four levels. At the top is the root.

▶ On $\mathbb{R}^n$ we have the Lebesgue measure λ_n with the property that $\lambda_n(A+x) = \lambda_n(A)$ for all $A \in \mathcal{B}(\mathbb{R}^n)$ and for all $x \in \mathbb{R}^n$. Any constant multiple of λ_n also has this property of *translation-invariance*, and no other measure does.

▶ On $\mathbb{R}_+ = (0, \infty)$ with multiplication, define the measure $d\mu(x) = \frac{dx}{x}$. Check that $\mu(xA) = \mu(A)$ for all $A \in \mathcal{B}(\mathbb{R}_+)$ and for all $x \in \mathbb{R}_+$. For example,

$$\mu(a,b) = \int_a^b \frac{1}{x} dx = \log(b/a)$$

which is clearly the same as $\mu(2a, 2b)$.

▶ On $GL(n, \mathbb{R})$, define the measure $d\mu(X) = |\det X|^{-n} dX$ where dX denotes Lebesgue measure on $\mathbb{R}^{n^2}$ (of which $GL(n\mathbb{R})$ is an open set). Then if $A \in GL(n, \mathbb{R})$, the map $X \mapsto A.X$ has Jacobian determinant equal to $\det(A)^n$ (why?). Therefore, for a Borel set $\mathcal{S} \subseteq GL(n, \mathbb{R})$, we have

$$\begin{aligned} \mu(A\mathcal{S}) &= \int_{A\mathcal{S}} |\det X|^{-n} dX = \int_{\mathcal{S}} |\det(AX)|^{-n} d(AX) \\ &= \int_{\mathcal{S}} |\det(AX)|^{-n} \, |\det(A)|^n \, dX = \int_{\mathcal{S}} |\det(X)|^{-n} dX = \mu(\mathcal{S}). \end{aligned}$$

Thus μ is invariant under left multiplication. Check that it is also invariant under right multiplication.

▶ Let $G = (\mathbb{Z}/(2))^J$ where J is an arbitrary index set. Let $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ be the unique invariant measure on $\mathbb{Z}/(2)$. A basic theorem in probability theory (Kolmogorov's existence theorem) assures us that there is a unique "product measure" $\mu^{\otimes J}$ on G such that its projection to any finite number of co-ordinates $j_1, \ldots, j_n$ is precisely the n-fold product $\mu \otimes \ldots \otimes \mu$. This $\mu^{\otimes J}$ is an invariant measure on G (check!).

The general question is whether every topological group has an invariant measure.

Definition 11. Let G be a topological group. A nonzero, regular Borel measure μ is said to be a *left Haar measure* on G if $\mu(gA) = \mu(A)$ for all $g \in G$, $A \in \mathcal{B}(G)$. Similarly, a *right Haar measure* is one that satisfies $\mu(Ag) = \mu(A)$. If a measure is both left and right invariant, we call it a *Haar measure*.

Recall that regularity means that for any Borel set A,

$$\mu(A) = \inf\{\mu(U) : U \supseteq A, U \text{ open}\} = \sup\{\mu(K) : A \supseteq K, K \text{ compact}\}.$$

Some situations are problematic.

Example 12. Consider $\mathbb{Q}$ under addition. If μ is invariant, and the singleton $\{0\}$ has mass p, then every singleton must have mass p. If $p = 0$, the measure is identically zero, so we must take $p > 0$. Then μ is basically counting measure on $\mathbb{Q}$, and the only sets with finite measure are finite sets. It does not appear to be of much use, for instance, all nonempty open sets have infinite measure. Alternately, observe that the same measure would be an invariant measure for $\mathbb{Q}$ with discrete topology. The fact that we are taking a more interesting topology that respects addition on $\mathbb{Q}$ is not getting us a more interesting measure.

Example 13. Take any infinite dimensional normed space X, eg., ℓ^2 or $C[0,1]$. Addition is the group operation. If μ is a translation-invariant measure on X, that would be like Lebesgue measure in infinite dimensions - something looks suspicious! Here is one issue. Any such μ must give infinite measure to all open sets. To see this, observe that the unit ball contains countably many balls of identical radius $r > 0$ (intersection of the unit ball with each orthant is open). Since each of these smaller balls must have equal measure, either the unit ball has zero measure or infinite measure.

Example 14. Consider affine transformations on the real line, $f_{a,b}(x) = ax + b$, where $a > 0$ and $b \in \mathbb{R}$. These form a group under composition with the multiplication: $f_{a,b} \circ f_{c,d} = f_{ac,ad+b}$.

In searching for an invariant measure, we try $d\mu(a,b) = h(a,b)da\, db$. Push forward under left multiplication by $f_{A,B}$ to get $A^{-2}h(a/A,(b-B)/a)da\, db$. Invariance requires $h(a,b) = A^{-2}h(a/A,(b-B)/a)$ for almost all a, b and any A, B, which implies that $h(a,b) = a^{-2}$ (up to a constant).

Similarly, if we consider right multiplication by $f_{A,B}$, then the measure $h(a,b)da\ db$ pushes forward to $A^{-1}h(a/A, b - aB)db\, db$. Deduce that right invariance forces $h(a,b) = 1/a$ (again, up to constant factor).

This example shows that the right Haar measure and left Haar measure can both exist and be distinct.

Exercise 15. Consider the group of affine transformation $f_{A,b} : \mathbb{R}^n \mapsto \mathbb{R}^n$ defined by $f_{A,b}(x) = Ax + b$. Here $A \in GL(n,\mathbb{R})$ and $b \in \mathbb{R}^n$. Show that they form a group under composition and find the left and right Haar measures.

Now we are ready to state the results on existence and uniqueness of Haar measures.

Theorem 16 (André Weil)**.** *If G is a locally compact topological group, then it has a unique (up to multiplication by constants) left Haar measure. Similarly for right Haar measure.*

We shall not prove this. But we shall prove the theorem for compact groups.

Theorem 17 (Haar, von Neumann)**.** *If G is a compact topological group, then it has a unique (up to multiplication by constants) Haar measure.*

One can phrase invariance in terms of integrals instead of functions.

Exercise 18. Let G be a locally compact topological group. Let μ be a regular Borel measure on G. Show that the following are equivalent.

(1) μ is a left-Haar measure on G.

(2) For any $f \in C_c(G)$, we have $\int f(x)d\mu(x) = \int f(gx)d\mu(x)$ for all $g \in G$.

3. Proof of existence of Haar measure on compact groups

Let G be a compact group. We want to show the existence of a unique probability measure μ on G such that for any $f \in C(G)$ and any $y \in G$,

$$\int f(yx)d\mu(x) = \int f(xy)d\mu(x) = \int f(x)d\mu(x).$$

This μ is then the unique Haar measure on G.

The key idea: Distribute n points as spread out regularly as possible on G. Then the probability measure that puts mass $1/n$ at each of these points converges to a measure on G that is the Haar measure. For example, if $G = S^1$, it is clear that the most regular distribution of points is to take the nth roots of 1 (or rotate them all by one element of S^1).

There is a starting issue with this plan - what is the meaning of a well-distributed set of points? For simplicity of presentation of the key ideas, we first make the following assumption and remove it later.

Assumption: The topology of G is induced by an invariant metric d, i.e., $d(zx, zy) = d(x, y)$ for all $x, y \in G$.

Once we have a metric, we can talk about ϵ-nets. Recall that an ϵ-net is a set $A \subseteq G$ such that every point of G is at distance less than ϵ of some point in A. Since G is compact, finite ϵ-nets exists for every $\epsilon > 0$. Let N_ϵ be the smallest cardinality of any ϵ-net. The following lemma has the key idea which makes the proof work.

Lemma 19. *If $A = \{x_1, \ldots, x_{N_\epsilon}\}$ and $B = \{y_1, \ldots, y_{N_\epsilon}\}$ are two ϵ-nets of minimal cardinality for G, then there is a permutation π such that $d(x_i, y_{\pi(i)}) < 2\epsilon$ for every $i \leq N_\epsilon$.*

Proof. Define a bipartite graph with parts A and B (even if a point is common to A and B, it corresponds to two vertices in this graph) with edges $x_i \sim y_j$ if $d(x_i, y_j) < 2\epsilon$.

Suppose $A' \subseteq A$ and let $N(A') \subseteq B$ be its neighbourhood in the graph. Let $C = (A \setminus A') \cup N(A')$. We claim that C is an ϵ-net. To show this, take any $z \in G$, and find i, j such that $d(x_i, z) < \epsilon$ and $d(y_j, z) < \epsilon$. Then $d(x_i, y_j) < 2\epsilon$, hence $x_i \sim y_j$. Therefore, either (1) $x_i \in A \setminus A'$ in which case $x_i \in C$ or (2) $x_i \in A'$ in which case $y_j \in N(A') \subseteq C$. Thus every point of G s within ϵ of a point of C, showing that C is an ϵ-net. Therefore $N_\epsilon \leq |C| = N_\epsilon - |A'| + |N(A')|$. In other words, $|N(A')| \geq |A'|$.

Thus, Hall's conditions are satisfied, and we get a matching of the bipartite graph. That is precisely the permutation π. ■

For a finite set $A = \{x_1, \ldots, x_n\}$, let $L_A : C(G) \mapsto \mathbb{R}$ be defined by

$$L_A f = \frac{1}{n} \sum_{k=1}^{N} f(x_k) = \int f d\mu_A$$

where $\mu_A = \frac{1}{n} \sum_{k=1}^{n} \delta_{x_k}$. For any $f \in C(G)$, we define its modulus of continuity $\omega_f(\epsilon) = sup\{|f(x) - f(y)| : d(x,y) \leq \epsilon\}$. Then $\omega_f(\epsilon) \to 0$ as $\epsilon \to 0$. The above lemma easily implies that if A and B are two minimal cardinality ϵ-nets, then $|L_A f - L_B f| \leq \omega_f(2\epsilon)$. We now extend this comparision to nets for different ϵ.

Lemma 20. *Let A (and B) be minimal cardinality ϵ-net (respectively δ-net) for G. Then for any $f \in C(G)$, we have $|L_A f - L_B f| \leq \omega_f(2\epsilon) + \omega_f(2\delta)$.*

Proof. Let $A = \{x_1, \ldots, x_n\}$ and $B = \{y_1, \ldots, y_m\}$. Let $C = A.B = \{x_i y_j : i \leq n, j \leq n\}$. We can write $C = \bigcup_{i \leq m} x_i B = \bigcup_{j \leq m} A y_j$. Thus,

$$L_C f = \frac{1}{n} \sum_{i=1}^{n} L_{x_i B} f = \frac{1}{n} \sum_{j=1}^{m} L_{A y_j} f.$$

But Ay_j is a minimal cardinality ϵ net for each $j \leq m$, hence the numbers $L_{Ay_j} f$ are all within $\omega_f(2\epsilon)$ of $L_A f$. Therefore $L_C f$ (being an average of $L_{Ay_j} f$, $j \leq m$, is also within $\omega_f(2\epsilon)$ of $L_A f$. By an analogous argument, $|L_C f - L_B f| \leq \omega_f(2\delta)$. Putting these together, we see that $|L_A f - L_B f| \leq \omega_f(2\delta) + \omega_f(2\epsilon)$. ■

Lemma 21. *For each $\epsilon > 0$, fix a minimal cardinality ϵ-net A_ϵ. Then, $\lim_{\epsilon \to 0} L_{A_\epsilon} f$ exists for every $f \in C(G)$. The limit does not depend on the choice of the nets A_ϵ.*

Proof. For $f \in C(G)$. For $\epsilon > 0$ let K_ϵ be the collection of all numbers $L_A f$, where A varies over all minimal-cardinality δ-nets for any $\delta < \epsilon$. Clearly $K_\epsilon \subseteq [-\|f\|_{\sup}, \|f\|_{\sup}]$. Further, $\text{dia}(K_\epsilon) \leq 2\omega_f(2\epsilon)$. Therefore, it follows that $\cap \bar{K}_\epsilon$ is a singleton $\{c\}$, and that number is the limit of $L_{A_\epsilon} f$ along any sequence of minimal-cardinality ϵ-nets (as $\epsilon \to 0$). ■

Proof of Theorem 17 under Assumption 3. For each $f \in C(G)$, let Lf be the number given by the previous lemma, i.e., $Lf = \lim_{\epsilon \to 0} L_{A_\epsilon} f$ along any sequence of minimal cardinality ϵ-nets A_ϵ. Linearity and positivity of L is obvious. Also $L(\mathbf{1}) = 1$.

For any $g \in G$, let $\tau_g f(x) = f(gx)$. Then,

$$L(\tau_g f) = \lim_{\epsilon \to 0} L_{A_\epsilon}(\tau_g f) = \lim_{\epsilon \to 0} L_{gA_\epsilon} f = Lf$$

where the last equality follows from the fact that gA_ϵ is also a minimal cardinality ϵ-net. Similarly, $L(\sigma_g f) = Lf$ where $\sigma_g f(x) = f(xg)$.

By Riesz's representation theorem, $Lf = \int_G f d\mu$ for a probability measure μ. Invariance of L implies that this measure satisfies the second condition in Exercise 18. Hence, it is a bi-invariant probability measure on G. ■

Removing the assumption 3: If we don't assume that an invariant metric exists, then we cannot talk of ϵ-nets, but we shall simply consider the net of neighbourhoods of the identity[9].

Given a neighbourhood V of identity, xVy, $x, y \in G$, is an open cover for G. Hence it has a finite sub cover. A *blocking set* for V is a set of minimal cardinality that intersects every one of the sets xAy for $x, y \in G$. Write $a \sim b$ (w.r.t. V) if there is some x, y such that xVy contains both a and b. A *blocking set* is a set that intersects each of the sets xVy for $x, y \in G$. Minimum cardinality blocking sets will replace minimal cardinality ϵ-nets in our proof. We prove the analogous lemmas.

Note added later: An important missing point in this discussion was pointed out by Abu Sufian. We must show that finite blocking sets exist. When you consider a metric space and ϵ balls, in finding a blocking set we would need to consider $\epsilon/2$ balls. The analogue of this without the metric is the following.

Fact: Let V be a neighbourhood of the identity in a topological group G. Then, there exists a neighbourhood W of the identity such that $W.W.W := \{xyz : x, y, z \in W\}$ is contained in V.

It is easy to see this from the fact that the map $(x, y, z) \mapsto xyz$ from $G \times G \times G$ to G is continuous, hence the pull back of V is an open set containing (e, e, e), where e is the identity of the group.

We leave it as an exercise to work out the existence of a finite blocking set using this observation.

For $f \in C(G)$, we define $\omega_f(V) = \sup\{|f(x) - f(y)| : x \sim y\}$.

Lemma 22. *If A and B are blocking sets of minimal cardinality (w.r.t V), then $|L_A f - L_B f| \leq \omega_f(V)$.*

Proof. We shall apply Hall's marriage theorem to say that there is a bijection π between $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$ such that $a_i \sim b_{\pi(i)}$ (w.r.t. V) for all $i \leq n$. Once that is done,

$$|L_A f - L_B f| \leq \frac{1}{n} \sum_{k=1}^{n} |f(a_k) - f(b_{\pi(k)})| \leq \omega_f(V).$$

To check Hall's condition, let $A' \subseteq A$ and $N(A') = \{b \in B : b \sim a \text{ for some } a \in A'\}$. Set $C = (A \setminus A') \cup N(A')$. Show that C is a blocking set. But its cardinality is $|A| - |A'| + |N(A')|$ which shows that $|N(A')| \geq |A'|$. ■

Lemma 23. *Let V, W be two neighbourhoods of identity in G. Let A and B be minimal cardinality blocking sets w.r.t. V and W, respectively. Then $|L_A f - L_B f| \leq \omega_f(V) + \omega_f(W)$.*

[9] A *net* is a partially ordered set in which given any two elements, there is a common element greater than or equal to both. For example, the collection of all neighbourhoods of a point x_0 in a topological space, endowed with the reverse inclusion, is a net. Given U, V, we have $U \cap V$ lying above U and above V.

Proof. Let $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$. Compare $L_A f$ and $L_B f$ with

$$\frac{1}{mn} \sum_{i \leq n} \sum_{j \leq m} f(a_i b_j). \tag{1}$$

This can be written alternately as $\frac{1}{m} \sum_{j=1}^{m} L_{Ab_j}$ or as $\frac{1}{n} \sum_{i=1}^{n} L_{a_i B}$. Since each Ab_j is a minimal cardinality blocking set, $|L_{Ab_j} f - L_A f| \leq \omega_f(V)$ for all $j \leq m$. Similarly, $|L_{a_i B} f - L_B f| \leq \omega_f(W)$. This shows that the quantity in (1) is with $\omega_f(V)$ of $L_A f$ and within $\omega_f(W)$ of $L_B f$. Therefore, $|L_A f - L_B f| \leq \omega_f(V) + \omega_f(W)$. ■

Exercise 24. Let $f \in C(G)$. Given $\epsilon > 0$, show that there exists a neighbourhood V of identity such that for all neighbourhoods $e \in W \subset V$ and all minimal blocking sets A, we have $\omega_f(W) \leq \epsilon$.

We put all these to prove the existence of Haar measure.

Proof of Theorem 17. Fix $f \in C(G)$ and for V, a neighbourhood of identity, define

$$K_V = \{L_A f : A \text{ is a minimal cardinality blocking set w.r.t. } W \text{ for some } W \subseteq V,\ W \ni e\}.$$

Then, all elements of K_V are within $2\omega_f(V)$ of each other, hence dia$(K_V) \leq 2\omega_f(V)$ which goes to zero by the exercise above. Further, $K_V \supseteq K_W$ if $V \supseteq W$. Hence, the sets $\bar{K_V}$ have finite intersection property since $K_{V_1} \cap \ldots \cap K_{V_m} \supseteq K_W$ where $W = V_1 \cap \ldots \cap V_m$. From this, it follows that $\bigcap_V \bar{K_V}$ is a singleton that we denote as $\{Lf\}$. Another way to say this is that if V_i is a net of neighbourhoods that converge to $\{e\}$, then $\lim L_{A_i} f$ exists and is independent of the choice of the minimal cardinality blocking sets A_i chosen.

The mapping $L : C(G) \mapsto \mathbb{R}$ is linear, positive and $L(\mathbf{1}) = 1$. Hence $Lf = \int f d\mu$ for some probabilit measure μ (Riesz's representation theorem). Fix any $x_0 \in G$ and consider $g(x) = f(x_0 x)$. Then, for any blocking set A, it is clear that $L_A g = L_{x_0 A} f$. Since A is a minimal cardinality blocking set w.r.t. V if and only is $x_0 A$ is, it follows that $Lf = Lg$. In other words, L is left-invariant. Similarly it is also right invariant. That is, for any $f \in C_c(G)$, we have $\int f(x) d\mu(x) = \int f(gx) d\mu(x)$ for all $g \in G$. Hence μ is a Haar measure. ■

The uniqueness question: We have constructed a bi-invariant probability measure μ. Suppose ν is another left-invariant probability measure on G. Define the measure $\theta = \mu \star \nu$ by (the right hand side is a positive linear functional of f, hence represented by a measure)

$$\int f(x) d\theta(x) = \iint f(xy) d\mu(x) d\nu(y)$$

for $f \in C(G)$. By the right-invariance of μ, the inner integral is $\int f d\mu$ for every $y \in G$ (a constant independent of y). Integrating w.r.t ν gives us that $\int f d\theta = \int f d\mu$.

Apply Fubini's theorem (applicable since the integrand is bounded) to write

$$\int f(x) d\theta(x) = \iint f(xy) d\nu(y) d\mu(x) = \iint f(y) d\nu(y) d\mu(x)$$

by the left-invariance of ν. The inner integral is independent of x and we simply get $\int f d\theta = \int f d\nu$.

Thus, $\int f d\mu = \int f d\nu$ for all $f \in C(G)$, whence it follows that $\mu = \nu$.

Remark 25. What was all this? If you are comfortable with probability language, let X and Y be independent random variables with distribution μ and ν, respectively. Bi-invariance of μ means that gX, Xg and X all have the same distribution μ, for any fixed $g \in G$. Left-invariance of ν means that gY has the same distribution as Y, for any $g \in G$.

Now consider $Z = XY$. Using independence, we can argue that Z has the same distribution as X (condition on Y) and that Z has the same distribution as Y (condition on X). Hence X and Y have the same distribution, i.e., $\mu = \nu$.

Part 5: Asymptotics of integrals

1. Some questions

Consider the sequence $n!$, which, by definition, it is the product of the first n positive integers. Do we understand how large it is? For example, it is easy to see that $2^{n-1} \le n! \le n^{n-1}$. Both sides of this inequality are quantities we are more familiar with and can work with easily. However, they are quite far from each other. We can sharpen the bounds as follows. Write $\log n! = \sum_{k=1}^{n} \log k$ and hence $\int_{k-1}^{k} \log x dx \le \log k \le \int_{k}^{k+1} \log x dx$. Therefore,

$$\int_0^n \log x \, dx \le \log n! \le \int_1^{n+1} \log x \, dx$$

giving $n \log n - n \le \log n! \le n \log(n+1) - n + \log(n+1)$. Thus,

$$n^n e^{-n} \le n! \le n^{n+1} e^{-n}(n+1).$$

The ratio of the upper and lower bounds is only of order n^2 now. Can we sharpen it further and get an elementary expression $f(n)$ such that[10] $n! \sim f(n)$? Stirling's formula asserts that $n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$. We shall prove this later.

Similarly, one is often interested in the magnitudes of various quantities such as

(1) Asymptotics of the Bell numbers B_n, the number of ways to partition the set $\{1, 2, \dots, n\}$.

(2) Asymptotics of $p(n)$, the number of ways to partition the number n.

(3) Asymptotics of $H_n(x)$ (fixed x, large n), where H_n is the nth Hermite polynomial.

One could list many more. We shall see some basic techniques to get the asymptotics of such quantities. We shall restrict ourselves to quantities that can be expressed as integrals of certain kinds. Fortunately, this covers many examples.

(1) $n! = \int_0^\infty x^n e^{-x} dx$.

(2) $B_n = n! \frac{1}{2\pi i} \int_C (e^{e^z - 1}) \frac{dz}{z^{n+1}}$ where C is a simple closed contour enclosing the origin in the complex plane.

(3) $H_n(x) = (-1)^n n! e^{x^2} \frac{1}{2\pi i} \int_\gamma \frac{1}{(z-x)^{n+1}} e^{-z^2} dz$.

In general, if we have a sequence (a_n), and its generating function $F(z) = \sum_{n=0}^{\infty} a_n z^n$ or exponential generating function $G(z) = \sum_{n=0}^{\infty} a_n z^n / n!$ has a positive radius of convergence, we can write a_n as

$$a_n = \frac{1}{2\pi i} \int_\gamma F(z) z^{-n-1} dz, \quad a_n = n! \frac{1}{2\pi i} \int_\gamma G(z) z^{-n-1} dz.$$

We will work out a few examples in the next sections[11]. The method is important, more than than the statements of the results.

[10] Common notation: (1) $a_n \sim b_n$ means $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$, (2) $a_n \asymp b_n$ means that $cb_n \le a_n \le Cb_n$ for some constants c and C, (3) $a_n \approx b_n$ means $\log a_n \sim \log b_n$.Similar interpretation for $f(x) \sim g(x)$ etc.

[11] Much of this material is taken from a very well-written old book of de Bruijn titled *Asymptotic methods in analysis*.

2. Laplace's method

Let us return to the factorial function. We shall derive its asymptotics (Stirling's formula) using the integral representation

$$n! = \int_0^\infty e^{-x}x^n dx.$$

Here is a quick sketch of the idea. The integrand is $\exp\{-x+n\log x\}$. The exponent (and hence the integrand) is maximized at $x=n$. Near this point, the second order Taylor expansion of the exponent is (derivative term vanishes because we are at a maximum)

$$-x+n\log x = (-n+n\log n) - \frac{1}{2n}(x-n)^2.$$

If we blindly replace the exponent by this, we get

$$e^{-n+n\log n}\int_0^\infty e^{-\frac{1}{2n}(x-n)^2}dx = n^n e^{-n}\int_{-\sqrt{n}}^\infty e^{-\frac{1}{2}t^2}dt.$$

For large n, the integral can be extended to the whole line without affecting the value significantly, hence we get

$$n^n e^{-n}\int_{-\infty}^\infty e^{-\frac{1}{2}t^2}dt = n^n e^{-n}\sqrt{2\pi n}$$

which is precisely Stirling's approximation! We have not yet justified the steps, or shown in what precise sense this approximates $n!$, but the idea described here is general: The contribution to the integral comes from a certain neighbourhood (here of order $\sqrt{n}$ in length) of the point where the integrand is maximized (here n).

A general theorem: We now try a general integral of the form $I(\lambda) = \int_{\mathbb{R}} e^{-\lambda f(x)}g(x)dx$. We shall show that under appropriate assumptions on f and g, we have (as $\lambda \to \infty$)

$$I(\lambda) \sim \frac{\sqrt{2\pi}g(0)}{\sqrt{f''(0)}\sqrt{\lambda}}.$$

Assumptions:

(1) Let $f:\mathbb{R}\mapsto\mathbb{R}_+$ be C^2, with a unique minimum at 0. Assume that $f(0)=0$ (without loss of generality) and that $f''(0)>0$. For $\delta>0$, assume that $m_\delta = \inf_{|x|\ge\delta} f(x)$ is strictly positive.

(2) Let $g:\mathbb{R}\mapsto\mathbb{R}$ be continuous and assume that $g(0)>0$ (if $g(0)<0$, replace g by $-g$).

(3) Assume that the integral defining $I(\lambda)$ converges absolutely for all λ (or all sufficiently large λ).

With these assumptions, fix $\delta>0$ and write $I(\lambda)=I_1(\lambda)+I_2(\lambda)+I_3(\lambda)$ where

$$I_1(\lambda)=\int_{-\delta}^{\delta} e^{-\lambda f(x)}g(x)dx,\quad I_2(\lambda)=\int_{\delta}^{\infty} e^{-\lambda f(x)}g(x)dx,\quad I_3(\lambda)=\int_{-\infty}^{-\delta} e^{-\lambda f(x)}g(x)dx.$$

The contribution from I_2 and I_3 are small. Indeed, we can write

$$|I_2(\lambda)| \leq \int_\delta^\infty e^{-\lambda f(x)}|g(x)|dx \leq e^{-(\lambda-1)m_\delta} \int_{\mathbb{R}} e^{-f(x)}|g(x)|dx.$$

The same bound holds for $I_3(\lambda)$ and we summarize the two estimates as:

$$|I_1(\lambda)| \leq Be^{-c_\delta \lambda}, \quad |I_3(\lambda)| \leq Be^{-c_\delta \lambda} \tag{1}$$

for some (large) constant B and small constant c_δ. The same bound holds for $I_3(\lambda)$.

Now we turn to $I_1(\lambda)$. Since $f''(0) = \lim_{x\to 0} \frac{f(x)-f(0)-f'(0)x}{x^2}$, we have $\epsilon(\delta)$ that goes to zero as δ goes to zero, such that (from our assumptions $f(0) = f'(0) = 0$) for all $x \in [-\delta, \delta]$,

$$\frac{1}{2}(f''(0) - \epsilon)x^2 \leq f(x) \leq \frac{1}{2}(f''(0) + \epsilon)x^2$$

where we have written ϵ for $\epsilon(\delta)$ so as to not add to the ugliness in this world. We may assume that for the same ϵ and any $x \in [-\delta, \delta]$,

$$g(0) - \epsilon \leq g(x) \leq g(0) + \epsilon.$$

Let us take δ small enough that $g(0) - \epsilon > 0$ (then we also have $g(x) > 0$ for $|x| \leq \delta$). Then,

$$(g(0) - \epsilon)e^{-\frac{1}{2}\lambda(f''(0)+\epsilon)x^2} \leq g(x)e^{-\lambda f(x)} \leq (g(0) + \epsilon)e^{-\frac{1}{2}\lambda(f''(0)-\epsilon)x^2}, \quad \text{for } |x| \leq \delta. \tag{2}$$

We now integrate over $[\delta, \delta]$ and write

$$\begin{aligned} I_1(\lambda) &\leq (g(0) + \epsilon)\left\{\int_{\mathbb{R}} e^{-\frac{1}{2}\lambda(f''(0)-\epsilon)x^2}dx - \int_{[-\delta,\delta]^c} e^{-\frac{1}{2}\lambda(f''(0)-\epsilon)x^2}dx\right\} \\ &= (g(0) + \epsilon)\frac{\sqrt{2\pi}}{\sqrt{\lambda(f''(0) - \epsilon)}} - (g(0) + \epsilon)\frac{2}{\delta}\sqrt{\lambda(f''(0) - \epsilon)}e^{-\frac{1}{2}\lambda\delta^2(f''(0)-\epsilon)} \end{aligned}$$

where the last line follows from the Gaussian integral $\int e^{-x^2/2\sigma^2}dx = \sigma\sqrt{2\pi}$ and the simple estimate[12]

$$\int_a^\infty e^{-x^2/2\sigma^2}dx \leq \frac{1}{\sigma a}e^{-a^2/2\sigma^2}, \quad \text{for } a > 0.$$

Analogously, using the left inequality in (2) we get the estimate

$$I_1(\lambda) \geq (g(0) - \epsilon)\frac{\sqrt{2\pi}}{\sqrt{\lambda(f''(0) + \epsilon)}} - (g(0) - \epsilon)\frac{2}{\delta}\sqrt{\lambda(f''(0) + \epsilon)}e^{-\frac{1}{2}\lambda\delta^2(f''(0)+\epsilon)}$$

We can summarize these to bounds as

$$\frac{\sqrt{2\pi}(g(0) - \epsilon)}{\sqrt{\lambda(f''(0) + \epsilon)}} - C_\delta\sqrt{\lambda}e^{-c_\delta\lambda} \leq I_1(\lambda) \leq \frac{\sqrt{2\pi}(g(0) + \epsilon)}{\sqrt{\lambda(f''(0) - \epsilon)}} + C_\delta\sqrt{\lambda}e^{-c_\delta\lambda}$$

where C_δ, c_δ are constants that depend on δ (and on f and g). Combining this with (1), we get

$$\frac{\sqrt{2\pi}(g(0) - \epsilon)}{\sqrt{\lambda(f''(0) + \epsilon)}} - C_\delta\sqrt{\lambda}e^{-c_\delta\lambda} \leq I(\lambda) \leq \frac{\sqrt{2\pi}(g(0) + \epsilon)}{\sqrt{\lambda(f''(0) - \epsilon)}} + C_\delta\sqrt{\lambda}e^{-c_\delta\lambda}$$

[12] *Proof of the estimate:* $\int_a^\infty e^{-t^2/2\sigma^2}dt \leq \frac{1}{a}\int_a^\infty xe^{-x^2/2}dx = \frac{1}{a}e^{-a^2/2}$.

Divide by $\frac{\sqrt{2\pi}g(0)}{\sqrt{\lambda f''(0)}}$ and let $\lambda \to \infty$ to get

$$\frac{g(0)-\epsilon}{g(0)} \cdot \frac{\sqrt{f''(0)}}{\sqrt{f''(0)+\epsilon}} \le \limsup_{\lambda\to\infty} \frac{I(\lambda)}{\frac{\sqrt{2\pi}g(0)}{\sqrt{\lambda f''(0)}}} \le \frac{g(0)+\epsilon}{g(0)} \cdot \frac{\sqrt{f''(0)}}{\sqrt{f''(0)-\epsilon}}.$$

Now let $\delta \downarrow 0$ (and recall that $\epsilon = \epsilon(\delta) \to 0$) to get

$$I(\lambda) \sim \frac{\sqrt{2\pi}g(0)}{\sqrt{\lambda f''(0)}}.$$

Exercise 1. Show that $\int_0^\pi x^n \sin x dx \sim \frac{\pi^{n+2}}{n^2}$.

When the maximum of the integral occurs at and end of the interval of integration, the same methods can be followed to get a different answer.

Exercise 2. Let $I(\lambda) = \int_0^\infty e^{-\lambda f(x)} dx$ where $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$ has a unique minimum at zero. Make appropriate assumptions and show that $I(\lambda) \sim \frac{-1}{f'(0)\lambda} e^{\lambda f(0)}$.

3. Saddle point method

Here we are interested in evaluating integrals of the form $I(\lambda) = \int_{[A,B]} g(z)e^{\lambda f(z)}dz$ where f, g are holomorphic functions on some region and A, B are points in the region. The parameter λ is real and will go to infinity. We want the asymptotic behaviour of $I(\lambda)$.

In our examples, we shall take f and g to be entire functions. The idea consists of two steps:

(1) By holomorphicity, the integral does not change if we deform the contour of integration (keeping end points fixed at A, B). The first step is to choose the right contour, by which we mean whatever will make the second step work!

(2) Once the contour is chosen, write $I(\lambda)$ as an integral over an interval in the real line, $I(\lambda) = \int_a^b g(\gamma(t))e^{\lambda f(\gamma(t))}\dot{\gamma}(t)dt$. If the contour is well-chosen, Laplace's method (or some other) could apply to this integral and we could calculate the asymptotics.

How do we choose a good contour? Here are some guidelines. They are not guaranteed to work, but often do.

Guidelines: Consider the absolute value of $e^{\lambda f(z)}$ which is $e^{\lambda u(z)}$ where $u = \operatorname{Re} f$. For Laplace method to apply in the second step, we would like $e^{\lambda u(\gamma(t))}$ to be peaked at one point t_0 so that the entire contribution to the integral comes from a neighbourhood of t_0 (for large λ). Then $u(\gamma(t))$ should achieve a maximum at t_0. Let us assume t_0 is not an endpoint, then u achieves its maximum on γ at $\gamma(t_0)$.

But since u is harmonic, it has no maxima (or minima) in the plane. Therefore, $\gamma(t)$ must be a saddle point of u. Working backwards, we see that a good choice of γ is one that passes through one of the saddle points of u, and it should pass through the saddle point in such a way that the maximum of u on the curve is attained at this saddle point.

Example 3. If $f(z) = z^2$, then $u(x, y) = x^2 - y^2$ (where $z = x + iy$) and $\nabla u(z) = (2x, -2y)$. The only saddle point is $(0, 0)$. Along the x-axis, this is a minimum of u, and along the y-axis, this is a maximum of u. We would want our curve to pass through 0 in the direction of the y-axis.

In general, if $f = u + iv$ is holomorphic, then $f' = u_x + iv_x = u_x - iu_y$ (Cauchy-Riemann equations). Thus, saddles of u are precisely the zeros of f'. Further, the Taylor expansion of f near ζ looks like $f(z) = f(\zeta) + \frac{1}{2}(z - \zeta)^2 f''(\zeta) + \ldots$. Hence, if $z = \zeta + re^{i\theta}$ (small r) and $f''(\zeta) = Re^{i\alpha}$, then

$$u(z) = u(\zeta) + \frac{1}{2}r^2 R\cos(2\theta + \alpha) + \ldots$$

The direction of steepest descent (respectively ascent) is the θ for which $\cos(2\theta + \alpha) = -1$ (respectively $+1$). We define the direction of the steepest descent to be the *axis of the saddle*. In other words, it is the line of z such that $(z - \zeta)^2 f''(\zeta)$ is real and negative.

Exercise 4. Show the same in an alternative way by going through the Hessian of u given by

$$Hu(\zeta) = \begin{bmatrix} u_{x,x} & u_{x,y} \\ u_{y,x} & u_{y,y} \end{bmatrix}$$

and the fact that the direction of the steepest descent is the direction of the eigenvector corresponding to the negative eigenvalue of Hu.

Part-6: Asymptotics of the eigenvalues of the Laplacian

1. Weyl's law

The Laplacian, $\Delta := \sum_{i=1}^{n} \partial_i^2$, where $\partial_i = \frac{\partial}{\partial x_i}$, is perhaps the most important linear operator in mathematics. It shows up in many contexts in physics. For example, the law connection the charge distribution $\rho(\cdot)$ to the electirc potential generated by it is $\Delta\varphi = \rho$, the same if ρ is interpreted as mass distribution and φ as the gravitational potential. One could give many other examples. Just to mention two-

(1) Wave equation: $\frac{\partial^2}{\partial t^2}u(x,t) = \Delta u(x,t)$. Here $u(x,t)$ represents the displacement of a stretched membrane (say $x \in \Omega$, a domain in $\mathbb{R}^2$), eg. a drum, where the ends are tied down, $u(x,t) = 0$ for $x \in \partial\Omega$.

(2) Heat equation: $\frac{\partial}{\partial t}u(x,t) = \Delta u(x,t)$, where $u(x,t)$ is the temperature at location x at time t.

In mathematics, the importance of the Laplacian comes from its symmetry with respect to rotations and translations. If $f, g : \mathbb{R}^n \mapsto \mathbb{R}$ are smooth functions such that $g(x) = f(Ax+b)$ where $A_{n\times n}$ is an orthogonal matrix and $b \in \mathbb{R}^n$, then

$$(\Delta g)(x) = (\Delta f)(Ax+b).$$

In other words, the Laplacian commutes with isometries of $\mathbb{R}^n$. It is only natural when a system is described by second order derivatives, and there is symmetry of translation and rotation, that the Laplacian should make an appearance.

Here we are interested in eigenvalues and eigenfunctions of the Laplacian on bounded regions of the Euclidean space. The setting is that we have a nice bounded region $\Omega \subseteq \mathbb{R}^n$ with piecewise smooth boundary, and we consider functions $f : \Omega \mapsto \mathbb{R}$ satisfying $f\big|_{\partial\Omega} = 0$ and some smoothness requirements (eg., C^2) inside Ω. Let us see some examples.

Example 1. $\Omega = (0, L)$ in $\mathbb{R}$. Here $\Delta = \frac{d^2}{dx^2}$. Clearly, $\varphi_n(x) = \sin(\pi n x/L)$ satisfy $\Delta\varphi_n = -\pi^2 L^{-2} n^2 \varphi_n$ and also $\varphi_n(0) = \varphi_n(1) = 0$. We could say that φ_n, $n \geq 1$, are eigenfunctions of the Laplacian on $[0,1]$ with Dirichlet boundary conditions.

There are no other eigenfunctions (we are not saying why, as yet). We see that the nth largest eigenvalue of $-\Delta$ is $\pi^2 n^2 L^{-2}$. Equivalently, if $N(\lambda)$ is the number of eigenvalues not exceeding λ, then $N(\lambda) \sim (L/\pi)\sqrt{\lambda}$.

Example 2. $\Omega = (0,a) \times (0,b)$ in $\mathbb{R}^2$. It is clear that $\varphi_{n,m}(x,y) = \sin(\pi n x/a)\sin(\pi m y/b)$ satisfies $\Delta\varphi_{n,m} = -\pi^2(\frac{n^2}{a^2} + \frac{m^2}{b^2})\varphi_{n,m}$.

What is $N(\lambda)$? It is equal to the number of lattice points (m,n), $m,n \geq 1$, that lie inside the ellipse

$$\frac{x^2}{a^2/\pi^2} + \frac{y^2}{b^2/\pi^2} = \lambda.$$

Hence, $N(\lambda)$ is close to one-quarter of the area of the ellipse, which is $(ab/\pi)\lambda = (|\Omega|/\pi)\lambda$. More precisely, $N(\lambda) \sim (ab/4\pi)\lambda = \frac{1}{(2\pi)^2}|\Omega|\lambda$.

Based on such calculations, and perhaps a few more explicit examples, physicists conjectured (late 1800s) that the asymptotics of eigenvalues depends only on the volume of the domain (and the dimension). That was proved by Weyl. Let ω_d denote the volume of the unit ball in $\mathbb{R}^d$.

Theorem 3 (Weyl)**.** *Let Ω be a domain in $\mathbb{R}^d$ with piecewise smooth boundary. Let $N(\lambda)$ be the number of eigenvalues of $-\Delta$ on Ω, with Dirichlet boundary conditions. Then,*

$$N(\lambda) \sim (2\pi)^{-d}\omega_d|\Omega|\lambda^{d/2}.$$

This is only the most basic version of the theorem, in one setting. It can be extended by finding further corrections. And similar theorems exist for other boundary conditions (eg., Neumann boundary conditions), to related operators (eg., the Schrodinger operator $-\Delta + V$), to the Laplace-Beltrami operator on closed Riemannian manifolds, etc.

2. The spectrum of the Laplacian

There are three ingredients: the domain, the operator and the boundary condition[13].

The domain: We shall assume that Ω is a bounded, open set in $\mathbb{R}^d$ whose boundary is a union of finitely many piecewise smooth closed curves. Let $B = \partial\Omega$. Let $n(x)$ denote the unit outward normal to Ω at $x \in B$ (it exists except at finitely many points).

Boundary conditions: Standard boundary conditions are as follows

(1) Dirichlet: $u = 0$ on B.

(2) Neumann: $\frac{\partial}{\partial n}u = 0$ on B (except at the finitely many points where the normal is not well-defined).

(3) Mixed: Fix a nice (continuous/smooth) function $\sigma : B \mapsto \mathbb{R}$ and ask for $\frac{\partial u}{\partial n} + \sigma u = 0$ on B.

The operator: For $u \in C^2(\Omega)$, we define $\Delta u(x) = \sum_{i=1}^{d} \partial_i^2 u(x)$ for $x \in \Omega$.

What we want are eigenvalues and eigenfunctions of $-\Delta$. In principle, this must simply be a function $u \in C_c^2(\bar{\Omega})$ (continuous on $\bar{\Omega}$ and smooth in Ω) and a number $\lambda \in \mathbb{R}$ such that $-\Delta u = \lambda u$ inside Ω and also require that u is not identically zero and that satisfies the boundary conditions imposed.

But as we know, to talk about spectral theorem, we require the setting of a Hilbert space, although the operator need not be defined on all of the space. Also, a reasonable spectral theorem

[13] This section will be very sketchy, but gives an overview of many important ideas required to make sense of the eigenvalues and eigenfunctions of the Laplacian.

exists only for self-adjoint, or at least normal, operators. What is this Hilbert space for the Laplacian? Although all this can be made sense of, we shall change the setting slightly and work with quadratic forms. First, we introduce the required Hilbert spaces.

A simple integration by parts shows that for $f, g \in C_c(\mathbb{R}^2)$, we have

$$\int_\Omega (-\Delta f)(x)g(x)dx = \int_\Omega \nabla f(x).\nabla g(x)dx.$$

The right side is the required quadratic form, or more precisely, $\int_\Omega |\nabla f|^2$. When we work in a bounded open set Ω, then for $f, g \in C_c^2(\Omega)$, the above identity is still valid. However, if f, g are merely smooth (say on a neighbourhood of $\bar{\Omega}$), then we must be careful about the boundary terms and the identity changes to

$$\int_\Omega (-\Delta f)(x)g(x)dx = \int_\Omega \nabla f(x).\nabla g(x)dx - \int_{\partial\Omega} g\frac{\partial f}{\partial n}.$$

For simplicity of language, let us stick to 2-dimensions. If f satisfies, Neumann boundary condition, then the second term above vanishes. Thus, if f, g are smooth and f satisfies either the Dirichlet or the Neumann boundary condition, then $\int(-\Delta f)g = \int\langle\nabla f, \nabla g\rangle$. This leads us to study the quadratic form

$$Q[f, g] = \int_\Omega \nabla f.\nabla g.$$

What is the right class of functions for which this makes sense? It looks like we must require ∇f to be in L^2. This can be made sense of by the notion of weak derivative.

Weak derivative: If $f, g_i : \mathbb{R}^d \mapsto \mathbb{R}$ are locally integrable functions such that

$$\int f\partial_i g = -\int g_i\varphi$$

for all $\varphi \in C_c^\infty(\mathbb{R}^d)$, then we say that g_i is the weak ith partial derivative of f. If $f \in C^1(\mathbb{R}^d)$, then this is satisfied with $g_i = \partial_i f$, the usual definition of derivative (integration by parts formula). In general, if it exists, it is well defined $a.e.$ If all the weak partial derivative $g_1, \ldots, g_d$ exist, we say that $(g_1, \ldots, g_d)$ is the weak gradient of f.

Now we are ready to define the spaces that we want. Let Ω be a bounded open set in $\mathbb{R}^2$ (for simplicity, stick to $d = 2$ henceforth).

$$H^1(\Omega) = \{f \in L^2(\Omega) : \nabla f \text{ exists in the weak sense and belongs to } L^2(\Omega)\}.$$

On $H^1(\Omega)$, define the inner product $(f, g) = \langle f, g\rangle + \langle\nabla f, \nabla g\rangle$.

Fact: $H^1(\Omega)$ is complete under this inner product.

Define $H_0^1(\Omega)$ to be the closure of $C_c^\infty(\Omega)$ in $H^1(\Omega)$. Then, $H_0^1(\Omega)$ is also a Hilbert space with the same inner product. Functions in H_0^1 are the ones that are meant to satisfy Dirichlet boundary condition.

By the earlier discussion, once we move to the level of quadratic forms (instead of the Laplacian), the boundary condition is no longer required in the Neumann problem. In short, the quadratic form Q, when restricted to $H^1(\Omega)$ and to $H^1_0(\Omega)$, represent the quadratic forms induced by the Laplacian with the Neumann and Dirichlet boundary conditions, respectively.

Definition of eigenvalues and eigenvectors: Define

$$\mu_1 = \min_{f \in H^1, \|f\|=1} Q[f,f].$$

It is clear that $\mu_1 = 0$ (since $Q[f,f] \geq 0$ for all f and $Q[\mathbf{1},\mathbf{1}] = 0$). The minimum is attained by constant functions. Let ψ_1 be one such, normalized in $L^2(\Omega)$. We refer to μ_1 and ψ_1 as the first eigenvalue and the first eigenfunction of the Neumann-Lapacian, respectively.

For $k \geq 1$, let

$$\mu_k = \min_{\substack{f \in H^1, \|f\|=1 \\ f \perp \psi_1, \ldots, \psi_{k-1}}} Q[f,f].$$

It is true, but no longer obvious, that the minimum is attained. Let ψ_k be a minimizer (normalize it in $L^2(\Omega)$). We refer to μ_k and ψ_k as an eigenvalue-eigenfunction pair. Assuming the existence of minimizers, we proceed inductively and obtain $\mu_1 \leq \mu_2 \leq \ldots$ and $\psi_1, \psi_2, \ldots$. By definition, $\{\psi_1, \psi_2, \ldots\}$ is an orthonormal set in $L^2(\Omega)$. Observe also that

$$Q[\psi_k, \psi_j] = \begin{cases} 0 & \text{if } k \neq j, \\ \mu_k & \text{if } k = j. \end{cases}$$

The second is clear by definition of ψ_k and μ_k. The first is also easy (if $j > k$, observe that $Q[\psi_k + t\psi_j, \psi_k + t\psi_j] \leq Q[\psi_k, \psi_k]$ for all t, since $\psi_k + t\psi_j$ is also considered in the minimum. Use that to show that $Q[\psi_k, \psi_j] = 0$).

Another important fact (requires proof) is that $\mu_k \to \infty$ as $k \to \infty$. This ensures that the eigenfunctions form an orthonormal basis for $L^2(\Omega)$ (why?).

We have left two facts unproved: (a) Existence of minimizers and (b) That eigenvalues increase without bound.

In a similar fashion, one can work with the same quadratic form on $H^1_0(\Omega)$ and define $0 < \lambda_1 \leq \lambda_2 \ldots$ and an orthonormal basis $\{\varphi_1, \varphi_2, \ldots\}$ for $L^2(\Omega)$ such that

$$\lambda_k = \min_{\substack{f \in H^1_0, \|f\|=1 \\ f \perp \varphi_1, \ldots, \varphi_{k-1}}} Q[f,f].$$

These are defined to be the eigenvalues of the Dirichlet-Laplacian and the minimizers are the eigenfunctions.

The above definition is essentially the Rayleigh-Ritz formulas that we are familiar with in the case of symmetric matrices. We shall need the min-max theorem (actually max-min theorem, but that sounds odd!) for these eigenvalues.

Theorem 4 (Min-Max theorem)**.** *Let Ω be as above. Then*

$$\mu_k = \max_{\substack{W\subseteq H^1 \\ dim(W)\le k-1}} \min_{\substack{f\in H^1,\|f\|=1 \\ f\perp W}} Q[f,f] \quad \textit{and} \quad \lambda_k = \max_{\substack{W\subseteq H^1_0 \\ dim(W)\le k-1}} \min_{\substack{f\in H^1_0,\|f\|=1 \\ f\perp W}} Q[f,f].$$

3. Proof of Weyl's law using the min-max theorem

Let Ω be as before. Let $N_0(\lambda)$ be the number of Dirichlet eigenvalues in the interval $[0,\lambda]$ and let $N'(\lambda)$ be the number of Neumann eigenvalues in the same interval. Now we are ready to prove Weyl's law. We shall stick to the simplest version of it only.

Theorem 5. $N(\lambda) \sim (2\pi)^{-d/2}\omega_d|\Omega|\lambda^{d/2}$ *and as* $\lambda \to \infty$ *where* $N(\lambda) = N_0(\lambda)$ *or* $N'(\lambda)$*.*

The proof consists of three steps.

(1) Show the theorem for rectangles. This can be done because the eigenvalues of the Laplacian under both Dirichlet and Neumann conditions can be computed explicitly.

(2) Show the theorem for a finite union of standard rectangles. This can be done by comparison theorems using the min-max criteria. The essential point is to show that $N(\lambda)$ is nearly additive in the domain, i.e., $N_{\Omega_1\sqcup\Omega_2}(\lambda) \approx N_{\Omega}(\lambda) + N_{\Omega_2}(\lambda)$. That makes the appearance of $|\Omega|$ transparent.

(3) For a general Ω, sandwich it from inside and outside by regions that are finite unions of standard rectangles. Again invoke comparison theorems.

Remaining notes to be written. No time now!