Princeton University

Spring 2025 MAT425: Measure Theory

HW8 Sample Solutions

Apr 20 2025

Chayim Lowen (only Q5) and Serban Cicortas (all other Questions)

April 20, 2025

Problem 1.

Solution. See proof of Theorem 4.4 in Chapter 1 of Stein and Shakarchi: Real analysis. □

Problem 2.

Solution. We find a counter-example for $d = 1$. We define $f_n : \mathbb{R} \to \mathbb{R}$ by $f_n = \mathbf{1}_{[n,n+1]}$ and $f : \mathbb{R} \to \mathbb{R}$ as $f = 0$. Then f_n converges to f everywhere on $\mathbb{R}$. For $\varepsilon > 0$ small, we suppose there exists $M \subset \mathbb{R}$ with $m(M) < \varepsilon$ such that $f_n \to f$ uniformly on $\mathbb{R} \backslash M$. Then for any $\epsilon > 0$ there exists $N(\epsilon) > 0$ large enough such that $|f_n(x) - f(x)| < \epsilon$ for all $n > N(\epsilon)$ and $x \in \mathbb{R} \backslash M$, which is of course a contradiction. □

Problem 3.

Solution. See proof of Theorem 7.10 in Chapter 7.2 of Folland. □

Problem 4.

Solution. Since μ is concentrated on S, by the definition of the total variation measure it follows that $|\mu|$ is also concentrated on S. Similarly, we get for each $x \in S$ that $|\mu|(x) = |\mu(x)| = |c_x|$, since the partitions of $\{x\}$ are simply $\big\{\{x\}, \emptyset\big\}$. We conclude by countable additivity that:

$$|\mu| = \sum_{x \in S} |c_x| \delta_x$$

□

Problem 5. Let $N \in \mathbb{N}$ be an integer. The set of Hermitian matrices of order N is given by

$$\mathrm{Herm}_N(\mathbb{C}) := \{A \in \mathrm{Mat}_{N\times N}(\mathbb{C}) \mid A = A^*\}$$

where * is the conjugate-transpose operator. The set of unitary matrices of order N is given by

$$\mathcal{U}(N) := \{U \in \mathrm{Mat}_{N\times N}(\mathbb{C}) \mid UU^* = I_N\}$$

where I_N is the identity matrix of order N.

We write $\mathbb{T}^N$ for the set of $N \times N$ diagonal unitary matrices. These are precisely the matrices of the form $\mathrm{diag}(e^{i\theta_1}, \ldots, e^{i\theta_N})$ for $\theta_1, \ldots, \theta_N \in \mathbb{R}$. Write also D_N for the set of diagonal matrices in $\mathrm{Mat}_{N\times N}(\mathbb{R})$ and S_N for the set of permutation matrices in $\mathrm{Mat}_{N\times N}(\mathbb{R}) \subset \mathrm{Mat}_{N\times N}(\mathbb{C})$.

Note the following

- $\mathrm{Herm}_N(\mathbb{C})$ is an $\mathbb{R}$-vector subspace[1] of $\mathrm{Mat}_{N\times N}(\mathbb{C})$ of dimension N^2. The Euclidean norm on $\mathrm{Herm}_N(\mathbb{C})$ is given in three equivalent ways: (i) as the square root of the sum of the squares of the absolute values of all coordinates, (ii) as the function $H \mapsto \sqrt{\mathrm{tr}(HH^*)}$, (iii) as the square root of the sum of the squares of the absolute values of the eigenvalues.
- $\mathcal{U}(N)$ is a non-Abelian (real) Lie subgroup[2] of $\mathrm{GL}_N(\mathbb{C})$ of dimension N^2.
- $\mathbb{T}^N$ is an Abelian (closed) Lie subgroup of $\mathcal{U}(N)$ of dimension N.
- D_N is an $\mathbb{R}$-vector subspace of $\mathrm{Herm}_N(\mathbb{C})$ of dimension N. The map $\mathrm{diag} : \mathbb{R}^N \xrightarrow{\sim} D_N$ is the isomorphism we will use.
- S_N is a finite subgroup of $\mathrm{GL}_N(\mathbb{C})$.

If $U \in \mathcal{U}(N)$ and $\Lambda \in \mathrm{Mat}_{N\times N}(\mathbb{R})$ is a *diagonal* matrix, the matrix $H := U\Lambda U^*$ is Hermitian.[3] This gives a map

$$\psi : D_N \times \mathcal{U}(N) \to \mathrm{Herm}_N(\mathbb{C}) \qquad (\Lambda, U) \mapsto U\Lambda U^*$$

The content of the unitary diagonalization theorem for Hermitian matrices[4] is precisely that the map ψ is *surjective*. For a fixed matrix $H \in \mathrm{Herm}_N(\mathbb{C})$, the corresponding matrix $\Lambda \in D_N$ is $\Lambda = \mathrm{diag}(\rho_1, \ldots, \rho_N)$ where the ρ_i are the eigenvalues of H, and the corresponding matrix $U \in \mathcal{U}(N)$ is given by *some* orthonormal basis of eigenvectors for H.[5] The matrix Λ is unique *up to permuting its diagonal entries*. The matrix U *fails* to be unique in three different ways:

[1] Warning: $\mathrm{Herm}_N(\mathbb{C})$ is *not* a $\mathbb{C}$-vector space.

[2] A Lie group is a smooth manifold with a compatible group structure. $\mathrm{GL}_n(\mathbb{R})$ and $\mathrm{GL}_n(\mathbb{C})$ are canonical examples of Lie groups.

[3] It will be necessary for simplicity to use this form of diagonalization (as opposed to $U^*\Lambda U$). This will have no effect on the end result.

[4] C.f. Lang's *Algebra*, Ch. XV, §6, Corollary 6.5. See also `https://math.umd.edu/~hking/Normal.pdf` for a self-contained discussion.

[5] Specifically, it is the *columns* of U that give the orthonormal eigenbasis.

- If we permute the entries of Λ, we should also permute the entries of U accordingly. Precisely, this means that if $\psi(\Lambda, U) = H$ then for any permutation matrix $P \in S_N$, we have $\psi(P^{-1}\Lambda P, UP) = H$ as well.

- If $v_1, \ldots, v_N$ are an orthonormal basis of eigenvectors for $H \in \mathrm{Herm}_N(\mathbb{C})$, then so is $e^{i\theta_1}v_1, \ldots, e^{i\theta_N}v_N$ for any $\theta_1, \ldots, \theta_n \in \mathbb{R}$. It follows that if $\phi(\Lambda, U) = H$ then also $\phi(\Lambda, UT) = H$ for any $T \in \mathbb{T}^N$.

- If two eigenvalues of H coincide, i.e. if $\rho_i = \rho_j$ for $i < j$, then there is a lot more freedom in the choice of U. This is because in the eigenspace corresponding to ρ_i, we may choose an arbitrary orthonormal basis, of which there are uncountably many if the dimension of this eigenspace is at least 2.

Nevertheless, the following is a very natural question. Suppose we have an integrable function $F : \mathrm{Herm}_N(\mathbb{C}) \to \mathbb{C}$ which is *invariant under conjugation*, i.e. $F(UHU^*) = F(H)$ for all $H \in \mathrm{Herm}_N(\mathbb{C})$, $U \in \mathcal{U}(N)$.[6] This essentially means that the value of F depends only on the eigenvalues of its argument.[7] Then it is natural to wish to rewrite the integral of F over $\mathrm{Herm}_N(\mathbb{C})$ as an integral over $D_N \cong \mathbb{R}^N$, which is (almost) the space of possible eigenvalues. Doing this for arbitrary functions $F : \mathrm{Herm}_N(\mathbb{C}) \to \mathbb{C}$ is (almost) the problem of computing the pushforward measure of the Lebesgue measure under the map $\mathrm{Herm}_N(\mathbb{C}) \to \mathbb{R}^N$ sending a matrix to its eigenvalues.

The error in the above explanation is that the "eigenvalues" map $\mathrm{Herm}_N(\mathbb{C}) \to \mathbb{R}^N$ is *not* well-defined since the matrix only determines an *unordered (multi)set* of eigenvalues, not an ordered tuple of eigenvalues. This problem can be resolved by restricting to the full-measure subset of $\mathrm{Herm}_N(\mathbb{C})$ consisting of Hermitian matrices with eigenvalues having *pairwise distinct absolute values*, and sending a Hermitian matrix H to the ordered tuple $(\rho_1, \ldots, \rho_N)$ where $\rho_1, \ldots, \rho_N$ are the eigenvalues of H and $|\rho_1| < |\rho_2| < \cdots < |\rho_N|$. The image of this function will be the region

$$C := \{(x_1, \ldots, x_N) \in \mathbb{R}^N \mid x_1 < \cdots < x_N\}$$

There are $N!$ such regions—obtained by permuting indices in all possible ways—which cover $\mathbb{R}^N$ (up to a set of measure 0). These regions are obtained from one another through reflections in $\mathbb{R}^N$. It follows that we can express any integral over C as an integral over $\mathbb{R}^N$ multiplied by a factor $\frac{1}{N!}$. This discussion justifies the existence of a solution to the following problem.

Problem. *Let* $f : \mathbb{R}^N \to \mathbb{C}$ *be an integrable* symmetric *function. That is,* $f(x_1, \ldots, x_N) = f(x_{\pi(1)}, \ldots, x_{\pi(N)})$ *for any permutation* $\pi : \{1, \ldots, N\} \to \{1, \ldots, N\}$*. We obtain a well-defined function* F *on* $\mathrm{Herm}_N(\mathbb{C})$ *by setting* $F(H) := f(\rho_1, \ldots, \rho_N)$ *where* $\rho_1, \ldots, \rho_N$ *are the eigenvalues of* $H \in \mathrm{Herm}_N(\mathbb{C})$ *(in some order). Find*

[6] Natural examples include functions that depend only on the trace and determinant of the argument matrix.

[7] This is true on the dense open subset of $\mathrm{Herm}_N(\mathbb{C})$ consisting of matrices with distinct eigenvalues. For matrices with repeated eigenvalues, there are more conjugacy classes. However, since the set of these has measure 0, we can safely ignore this technicality.

a measurable function $\delta : \mathbb{R}^N \to \mathbb{C}$—independent of f—such that

$$\int_{H \in \mathrm{Herm}_N(\mathbb{C})} F(H)\, d\lambda(H) = \int_{\Lambda \in D_N} f(\Lambda)\delta(\Lambda)\, d\lambda(\Lambda)$$

Here we get the Lebesgue measure on $\mathrm{Herm}_N(\mathbb{C})$ *and* D_N *via identifications* $\mathrm{Herm}_N(\mathbb{C}) \cong \mathbb{R}^{N^2}$ *and* $D_N \cong \mathbb{R}^N$.[8]

Solution. Consider the map $\psi : D_N \times \mathcal{U}(N) \to \mathrm{Herm}_N(\mathbb{C})$ defined above. As we observed, the value of ψ remains invariant under the action of multiplication by $\mathbb{T}^N$ on the *right*. We therefore get an induced map

$$\overline{\psi} : D_N \times \mathcal{U}(N) / \mathbb{T}^N \to \mathrm{Herm}_N(\mathbb{C})$$

where $\mathcal{U}(N)/\mathbb{T}^N$ is the *left-coset* space of $\mathbb{T}^N$ in $\mathcal{U}(N)$.[9] Over the generic point in $\mathrm{Herm}_N(\mathbb{C})$, the map $\overline{\psi}$ has a fibre (i.e. pre-image) of size $N!$.

We wish to compute the Jacobian of the map $\overline{\psi}$. We start by doing this at a point $(\Lambda, \overline{I_N})$ where $\Lambda \in D_N$ and $\overline{I_N}$ is the image of the identity matrix in $\mathcal{U}(N)/\mathbb{T}^N$. To this end, first note that the tangent space at I_N of the manifold $\mathcal{U}(N)$[10] (lying inside $\mathrm{Mat}_{N\times N}(\mathbb{C})$) is precisely $i\,\mathrm{Herm}_N(\mathbb{C})$.[11] The tangent space at I_N of the subgroup $\mathbb{T}^N$ is given by the subspace iD_N of $i\,\mathrm{Herm}_N(\mathbb{C})$. It follows that the tangent space to $\mathcal{U}(N)/\mathbb{T}^N$ at $\overline{I_N}$ is $i\,\mathrm{Herm}_N(\mathbb{C})/iD_N$.

For $1 \le j, k \le N$, let e_{jk} be the matrix in $\mathrm{Mat}_{N\times N}(\mathbb{C})$ with a 1 in the jk-th entry and 0s everywhere else. Define also

$$r_{jk} := \begin{cases} e_{jj} & \text{if } j = k \\ \frac{1}{\sqrt{2}}(e_{jk} + e_{kj}) & \text{if } j \ne k \end{cases} \qquad s_{jk} := \frac{i}{\sqrt{2}}(e_{jk} - e_{kj})$$

Then the set

$$B_0 := \{r_{jk} \mid 1 \le j \le k \le N\} \sqcup \{s_{jk} \mid 1 \le j < k \le N\}$$

forms an orthonormal $\mathbb{R}$-basis of $\mathrm{Herm}_N(\mathbb{C})$. Note that the matrices $\{r_{jj} \mid 1 \le j \le N\}$ are an orthonormal basis for D_N. It follows that (the images of) the matrices in

$$B_1 := \{r_{jk} \mid 1 \le j < k \le N\} \sqcup \{s_{jk} \mid 1 \le j < k \le N\}$$

[8] The identifications will be made such that the natural Euclidean norms on either side are also identified. This condition alone is enough to ensure that the function δ is uniquely determined.

[9] See Theorem 5 of Chapter 4 and its corollary on p. 138 in Godement's *Introduction to the Theory of Lie Groups* for a proof that this coset-space is a manifold. In fact, $\mathcal{U}(N)$ inherits the structure of a Riemannian manifold from its embedding into $\mathrm{Mat}_{N\times N}(\mathbb{C}) \cong \mathbb{R}^{2N^2}$. Since the (translation) action of $\mathbb{T}^N$ on $\mathcal{U}(N)$ is by isometries with respect to this metric, the quotient $\mathcal{U}(N)/\mathbb{T}^N$ inherits a canonical Riemannian metric. It is with respect to this metric that we are taking Jacobians.

[10] aka the *Lie* algebra of $\mathcal{U}(N)$

[11] The meaning of this is that the inclusion of $\mathcal{U}(N)$ in $\mathrm{Mat}_{N\times N}(\mathbb{C})$ induces an inclusion of tangent spaces at I_N. The image of this inclusion is precisely the space of skew-Hermitian matrices. To see this, we *differentiate* the defining equation of $\mathcal{U}(N)$, namely $UU^* = I_N$. The derivative of $U \mapsto UU^*$ is $A \mapsto AU^* + UA^*$. Setting $U = I_N$ and returning to the original equation gives $A + A^* = 0$, i.e. $A^* = -A$. This is equivalent to $(-iA)^* = -iA$. That is, A is in $i\,\mathrm{Herm}_N(\mathbb{C})$. Justifying these steps is an exercise in differential topology.

form an orthonormal basis of $\mathrm{Herm}_N(\mathbb{C})/D_N$.

By the product rule, the derivative of ψ at a point $(\Lambda, U) \in D_N \times \mathcal{U}(N)$ is given by

$$\mathscr{D}\psi(\Lambda, U) : (\eta, \upsilon) \mapsto \upsilon\Lambda U^* + U\eta U^* + U\Lambda\upsilon^*, \qquad (\eta, \upsilon) \in D_N \times i\,\mathrm{Herm}_N(\mathbb{C})$$

Setting $U = I_N$ and $\upsilon = iH$ we get

$$\mathscr{D}\psi(\Lambda, I_N) : (\eta, iH) \mapsto \eta + iH\Lambda - i\Lambda H, , \qquad (\eta, H) \in D_N \times \mathrm{Herm}_N(\mathbb{C})$$

An orthonormal basis for $D_N \times i\,\mathrm{Herm}_N(\mathbb{C})$ is given by the union of $\{(r_{jj}, 0) \mid 1 \le j \le N\}$ and $\{(0, ib) \mid b \in B_0\}$. Evaluating the derivative on this basis and writing $\Lambda = \mathrm{diag}(\rho_1, \dots, \rho_N)$ we get

$$(r_{jj}, 0) \mapsto r_{jj}, \qquad (0, r_{jj}) \mapsto 0, \qquad (0, ir_{jk}) \mapsto (\rho_k - \rho_j)s_{jk}, \qquad (0, is_{jk}) \mapsto (\rho_j - \rho_k)r_{jk}$$

An orthonormal basis for $D_N \times i\,\mathrm{Herm}_N(\mathbb{C})/iD_N$ is given by the union of $\{(r_{jj}, 0) \mid 1 \le j \le N\}$ and $\{(0, ib) \mid b \in B_1\}$. The same formulae (restricted to $j < k$) give the derivative of $\overline{\psi}$ with respect to this basis. It follows that the absolute value of the Jacobian determinant of $\overline{\psi}$ at the point $(\Lambda, \overline{I_N})$ is

$$\det \mathscr{D}\overline{\psi} = \prod_{1 \le j < k \le N} (\rho_j - \rho_k)^2 \qquad \text{where } \Lambda = \mathrm{diag}(\rho_1, \dots, \rho_N)$$

In fact the same is true at all points $(\Lambda, \overline{U}) \in D_N \times \mathcal{U}(N)/\,\mathbb{T}^N$. To see this, fix $\overline{U_0} \in \mathcal{U}(N)/\,\mathbb{T}^N$ and a lift $U_0 \in \mathcal{U}(N)/\,\mathbb{T}^N$ (i.e. U_0 mapsto to $\overline{U_0}$ under the quotient map $\mathcal{U}(N) \twoheadrightarrow \mathcal{U}(N)/\,\mathbb{T}^N$). We can write $\overline{\psi}$ as the composition

$$(\Lambda, U) \longmapsto (\Lambda, U_0^* U) \xmapsto{\ \overline{\psi}\ } U_0^* U \Lambda U^* U_0 \xmapsto{H \mapsto U_0 H U_0^*} U\Lambda U^*$$

The leftmost and rightmost maps are isometries (hence have no effect on the Jacobian). Taking the derivative and evaluating at $U = U_0$, we see that the Jacobian of $\overline{\psi}$ at U_0 is equal to the Jacobian at $U_0^* U_0 = I_n$.

Let μ be the measure on $\mathcal{U}(N)/\,\mathbb{T}^N$ induced on this manifold by its Riemannian metric. It follows from the change-of-variables formula for manifolds—and the fact that ψ is (generically) a $N!$-to-one function—that, writing $\Lambda = \mathrm{diag}(\rho_1, \dots, \rho_N)$ gives

$$\begin{aligned} N! \int_{H \in \mathrm{Herm}_N(\mathbb{C})} F(H)\, d\lambda(H) &= \int_{(\Lambda, U) \in D_N \times \mathcal{U}(N)/\,\mathbb{T}^N} F(\psi(\Lambda, U)) \prod_{1 \le j < k \le N} (\rho_j - \rho_k)^2 \; d(\lambda \times \mu)(\Lambda, U) \\ &= \left(\int_{\Lambda \in D_N} f(\Lambda) \prod_{1 \le j < k \le N} (\rho_j - \rho_k)^2 \, d\lambda(\Lambda) \right) \left(\int_{U \in \mathcal{U}(N)/\,\mathbb{T}^N} 1 \, d\mu(U) \right) \end{aligned}$$

where in the second step we have used Tonelli's theorem. Thus for some positive constant C, we have

$$\int_{H \in \mathrm{Herm}_N(\mathbb{C})} F(H)\, d\lambda(H) = C \int_{\substack{\Lambda \in D_N \\ \Lambda = \mathrm{diag}(\rho_1, \dots, \rho_N)}} f(\rho_1, \dots, \rho_N) \prod_{1 \le j < k \le N} (\rho_j - \rho_k)^2 \, d\lambda(\Lambda)$$

To figure out the value of this constant, we plug in the function

$$f(x_1, \dots, x_N) := e^{-\frac{1}{2}(x_1^2 + \cdots + x_N^2)}$$

On the matrix side, this is given by

$$F(H) = e^{-\operatorname{tr}(HH^*)/2}$$

Noting that $H \mapsto \operatorname{tr}(HH^*)$ is just the usual Euclidean norm on $\operatorname{Herm}_N(\mathbb{C}) \cong \mathbb{R}^{N^2}$, the left-hand-side of the above equation is simply a Gaussian integral in $\mathbb{R}^{N \times N}$. Its value is given by

$$\int_{H \in \operatorname{Herm}_N(\mathbb{C})} F(H)\, d\lambda(H) = \int_{\mathbb{R}^{N^2}} e^{-\frac{1}{2}(y_1^2 + \cdots + y_{N^2}^2)}\, dy_1 \dots dy_{N^2} = (2\pi)^{N^2/2}$$

On the other side, the integral is given by

$$\int_{\mathbb{R}^N} e^{-\sum_{j=1}^N \rho_j^2/2} \prod_{1 \le j < k \le N} (\rho_j - \rho_k)^2\, d\rho_1 \dots d\rho_N$$

This is a special case of Mehta's integral (with parameter $\gamma = 1$).[12] Its value is given by

$$\int_{\mathbb{R}^N} e^{-\sum_{j=1}^N \rho_j^2/2} \prod_{1 \le j < k \le N} (\rho_j - \rho_k)^2\, d\rho_1 \dots d\rho_N = (2\pi)^{N/2} \prod_{j=1}^N j!$$

Thus

$$C = (2\pi)^{\frac{N^2 - N}{2}} \prod_{j=1}^N (j!)^{-1}$$

and so the solution to the problem is given by

$$\delta(\rho_1, \dots, \rho_N) = (2\pi)^{\frac{N^2 - N}{2}} \prod_{j=1}^N (j!)^{-1} \prod_{1 \le j < k \le N} (\rho_j - \rho_k)^2$$

□

Problem 6.

Solution. As an example we consider $\Omega = \{1, 2, 3, 4\}$, $\mathfrak{M} = 2^\Omega$, and $\mathbb{P}(\{n\}) = 1/4$ for all $n \in \Omega$. We define the events $A = \{1, 2\}$, $B = \{2, 3\}$, $C = \{1, 3\}$, each having probability $1/2$. Moreover, they are pairwise independent, as we can check:

$$\mathbb{P}(A \cap B) = \mathbb{P}\big(\{1\}\big) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B),$$

and similarly for the other pairs. However, the sequence A, B, C is not fully independent since:

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

□

[12] See the paper "Some Macdonald-Mehta Integrals by Brute Force" by Frank G. Garvan for a discussion. See also `https://en.wikipedia.org/wiki/Selberg_integral`.

Problem 7.

Solution. The quadratic equation $Xt^2 + Yt + Z = 0$ has real roots iff $Y^2 \geq 4XZ$. This event has probability:

$$\mathbb{P}\big(Y(\omega)^2 \geq 4X(\omega)Z(\omega)\big) = \int_0^\infty \int_0^\infty \int_0^\infty \mathbf{1}_{t^2 \geq 4us} d\mu(u) d\mu(t) d\mu(s)$$
$$= \int_0^\infty \int_0^\infty \mu\big((0, t^2/4s]\big) d\mu(t) d\mu(s) = \int_0^\infty \int_0^\infty F\bigg(\frac{t^2}{4s}\bigg) d\mu(t) d\mu(s)$$

□

Problem 8.

Solution. We denote by $f(x) = \frac{d\mathbb{P}_X}{d\lambda}(x)$ the Radon-Nikodym derivative. Thus, $f(-x) = f(x)$ and:

$$\mathbb{P}_X(A) = \mathbb{P}\big(X^{-1}(A)\big) = \int_A f(x) d\lambda(x)$$

For any set $A \subset \mathbb{R}$ we define:

$$\sqrt{A_+} := \big\{x \in [0, \infty) : \ x^2 \in A \cap [0, \infty)\big\}$$

Then, we can write:

$$(X^2)^{-1}(A) = \big\{X \in \sqrt{A_+}\big\} \cup \big\{X \in -\sqrt{A_+}\big\}$$

and note the intersection of the two sets is $X^{-1}\big(\{0\}\big)$. We get:

$$\mathbb{P}_{X^2}(A) = \mathbb{P}_X(\sqrt{A_+}) + \mathbb{P}_X(-\sqrt{A_+}) - \mathbb{P}_X(\{0\}) = \int_{\sqrt{A_+}} f(x) d\lambda(x) + \int_{-\sqrt{A_+}} f(x) d\lambda(x) - \int_{\{0\}} f(x) d\lambda(x)$$

The last term on the RHS is zero. Using $f(-x) = f(x)$ and a change of variables, we get that:

$$\mathbb{P}_{X^2}(A) = 2 \int_{\sqrt{A_+}} f(x) d\lambda(x) = \int_{A \cap [0,\infty)} t^{-1/2} f\big(t^{1/2}\big) d\lambda(t)$$

We conclude that:

$$\frac{d\mathbb{P}_{X^2}}{d\lambda}(x) = x^{-1/2} \frac{d\mathbb{P}_X}{d\lambda}\big(x^{1/2}\big) \cdot \mathbf{1}_{[0,\infty)}$$

□

Problem 9.

Solution. See Chapter III of Widder: The Laplace transform. One could read Section 1 as an introduction, then Section 2 (up to the proof of Theorem 2b), and Section 4 which completes the proof of the Hausdorff moment problem. □

Problem 10.

Solution. See Theorem 1 in Chapter 14 of Cheney and Light: A Course in Approximation Theory for a detailed exposition of this result. □

CHAPTER III

THE MOMENT PROBLEM

1. Statement of the Problem

The *moment problem of Hausdorff*, sometimes called the *little moment problem* is the following. Given a sequence of numbers

$$\{\mu_n\}_0^\infty : \mu_0, \mu_1, \mu_2, \cdots ; \tag{1}$$

we may ask under what conditions it is possible to determine a function $\alpha(t)$ of bounded variation in the interval (0, 1) such that

$$\mu_n = \int_0^1 t^n \, d\alpha(t) \qquad (n = 0, 1, 2, \cdots) \tag{2}$$

Any such sequence will be called a *moment sequence*. It is evident that not every sequence (1) has the form (2) since (2) implies that

$$|\mu_n| \leqq V[\alpha(t)]_0^1$$

the quantity on the right being the variation of $\alpha(t)$ on the interval (0, 1). That is, every moment sequence is bounded. It was F. Hausdorff [1921*a*] who first obtained necessary and sufficient conditions that a sequence should be a moment sequence.

In section 6.1 of Chapter II we showed that a sequence can have at most one representation (2) if $\alpha(t)$ is a normalized function of bounded variation. That is,

$$\alpha(0) = 0, \qquad \alpha(t) = \frac{\alpha(t+) + \alpha(t-)}{2} \qquad (0 < t < 1).$$

Since normalization of the function $\alpha(t)$ does not change the value of the integral (2) we may assume without loss of generality that $\alpha(t)$ is normalized. This we do throughout the present chapter without further repetition of the fact.

Equations (2) may be regarded as a transformation of the function $\alpha(t)$ into the sequence $\{\mu_n\}$. This transformation is closely related to the Laplace transform, is in fact the discrete analogue of the latter. For, if we replace the integer n by the variable s in (2) and then make the change of variable $t = e^{-u}$, we obtain

$$\mu_s = \int_0^\infty e^{-su} \, d[-\alpha(e^{-u})].$$

It is thus clear that much light will be thrown on the Laplace transform by a solution of the moment problem.

2. Moment Sequence

We introduce several definitions:

DEFINITION 2*a*.

$$\Delta^k \mu_n = \sum_{m=0}^{k} (-1)^m \binom{k}{m} \mu_{n+k-m} \qquad (k = 0, 1, 2, \cdots).$$

DEFINITION 2*b*.

$$\lambda_{k,m}(x) = \binom{k}{m} x^m (1-x)^{k-m} \qquad (k, m = 0, 1, 2, \cdots).$$

DEFINITION 2*c*.

$$\lambda_{k,m} = \binom{k}{m} (-1)^{k-m} \Delta^{k-m} \mu_m \qquad (k, m = 0, 1, 2, \cdots).$$

DEFINITION 2*d*. *A Bernstein polynomial* $B_k[f(x)]$ *for a function* $f(x)$, *defined on the interval* (0, 1), *is*

$$B_k[f(x)] = \sum_{m=0}^{k} f\left(\frac{m}{k}\right) \lambda_{k,m}(x).$$

The degree of the polynomial is k unless

$$\sum_{m=0}^{k} (-1)^{k-m} \binom{k}{m} f\left(\frac{m}{k}\right) = 0,$$

when it is of lower degree or identically zero. For example,

$$B_k[1] = \sum_{m=0}^{k} \lambda_{k,m}(x) = 1.$$

DEFINITION 2*e*. *The sequence* $\{\mu_n\}_0^\infty$ *satisfies Condition A if a constant* L *exists such that*

$$\sum_{m=0}^{k} |\lambda_{k,m}| < L \qquad (k = 0, 1, 2, \cdots).$$

For example, if

(1) $$\mu_n = \int_0^1 t^n \, d\alpha(t) \qquad (n = 0, 1, 2, \cdots)$$

with $\alpha(t)$ of bounded variation in (0, 1), then

$$\sum_{m=0}^{k} \lambda_{k,m} = \sum_{m=0}^{k} \binom{k}{m} \int_0^1 t^m (1-t)^{k-m} \, d\alpha(t) = \int_0^1 d\alpha(t)$$

$$\sum_{m=0}^{k} |\lambda_{k,m}| \leqq \int_0^1 |d\alpha(t)| = V[\alpha(t)]_0^1.$$

That is, Condition A is necessary that the sequence $\{\mu_n\}_0^\infty$ should have the form (1). In particular the sequences

$$\left\{\frac{1}{n+1}\right\}_0^\infty, \qquad \{c^n\}_0^\infty \qquad\qquad (0 < c \leqq 1)$$

satisfy Condition A.

DEFINITION 2*f*. *If* $P_n(x)$ *is the polynomial*

$$P_n(x) = \sum_{m=0}^{n} a_m x^m,$$

an operator $M[P_n(x)]$, *called the moment of* $P_n(x)$ *with respect to the sequence* $\{\mu_n\}$, *is defined as*

$$M[P_n(x)] = \sum_{m=0}^{n} a_m \mu_m.$$

For example,

$$M[x^n] = \mu_n, \qquad \sum_{m=0}^{k} M[\lambda_{k,m}(x)] = \sum_{m=0}^{k} \lambda_{k,m} = \mu_0.$$

If μ_n has the representation (1) then

$$M[P_n(x)] = \int_0^1 P_n(t)\, d\alpha(t).$$

Note that the operator is applicable only to polynomials.

We shall now prove that Condition A is also sufficient that the sequence $\{\mu_n\}_0^\infty$ should have the representation (1). We need a preliminary result.

LEMMA 2. *If n is a positive integer, then*

$$\lim_{k\to\infty} \prod_{i=0}^{n-1} \frac{kx - i}{k - i} = x^n$$

uniformly for $0 \leqq x \leqq 1$.

This is clear since each factor of the product approaches x uniformly in the interval $0 \leqq x \leqq 1$.

We now prove:

THEOREM 2*a*. *If the sequence* $\{\mu_n\}_0^\infty$ *satisfies Condition* A, *then*

$$\mu_n = \lim_{k\to\infty} M[B_k[x^n]] \qquad (n = 0, 1, 2, \cdots).$$

For, by the binomial theorem we have for $k > n > 0$

$$x^n = x^n[(1 - x) + x]^{k-n} = \sum_{m=0}^{k-n} \binom{k-n}{m} x^{m+n}(1-x)^{k-m-n}$$

$$= \sum_{m=n}^{k} \frac{m(m-1)\cdots(m-n+1)}{k(k-1)\cdots(k-n+1)} \lambda_{k,m}(x).$$

$$\mu_n = M[x^n] = \sum_{m=n}^{k} \frac{m(m-1)\cdots(m-n+1)}{k(k-1)\cdots(k-n+1)} \lambda_{k,m}.$$

$$\mu_n - M[B_k[x^n]] = \sum_{m=n}^{k} \left\{ \frac{m(m-1)\cdots(m-n+1)}{k(k-1)\cdots(k-n+1)} - \left(\frac{m}{k}\right)^n \right\} \lambda_{k,m}$$

$$- \sum_{m=0}^{n-1} \left(\frac{m}{k}\right)^n \lambda_{k,m}$$

$$= \sum_{m=n}^{k} \left\{ \frac{ky(ky-1)\cdots(ky-n+1)}{k(k-1)\cdots(k-n+1)} - y^n \right\} \lambda_{k,m}$$

$$- \sum_{m=0}^{n-1} \left(\frac{m}{k}\right)^n \lambda_{k,m},$$

where $y = m/k$. Let ϵ be an arbitrary positive number. By Lemma 2 we see that we can determine k_0 such that for $k > k_0$

$$\left| \frac{ky(ky-1)\cdots(ky-n+1)}{k(k-1)\cdots(k-n+1)} - y^n \right| < \epsilon$$

$$\left(y = \frac{m}{k}, m = n, n+1, \cdots k\right),$$

and such that

$$\left| \sum_{m=0}^{n-1} \left(\frac{m}{k}\right)^n \lambda_{k,m} \right| < \left(\frac{n}{k}\right)^n L < \epsilon \qquad (k > k_0).$$

Hence

$$| \mu_n - M[B_k[x^n]] | < \epsilon L + \epsilon \qquad (k > k_0).$$

This gives us the desired result if $n = 1, 2, \cdots$. If $n = 0$

$$\mu_0 = M[B_k[1]].$$

By use of this result we can prove easily the main result of this section

THEOREM 2b. *A necessary and sufficient condition that* $\{\mu_n\}_0^\infty$ *should be a moment sequence is that it should satisfy Condition A.*

We have already seen that the condition is necessary. To prove it sufficient define a step-function $\alpha_k(t)$ which is normalized and has jumps $\lambda_{k,m}$ at points m/k,

$$\alpha_k\left(\frac{m}{k}+\right) - \alpha_k\left(\frac{m}{k}-\right) = \lambda_{k,m} \qquad (m = 1, 2, \cdots, k-1),$$

$$\alpha_k(0+) = \lambda_{k,0}, \qquad \alpha_k(0) = 0,$$

$$\alpha_k(1-) = \sum_{m=0}^{k-1} \lambda_{k,m};$$

$$\alpha_k(1) = \sum_{m=0}^{k} \lambda_{k,m} = \mu_0.$$

Then

$$M[B_k[x^n]] = \int_0^1 t^n \, d\alpha_k(t),$$

and by Theorem 2a,

$$\mu_n = \lim_{k\to\infty} \int_0^1 t^n \, d\alpha_k(t).$$

The total variation of $\alpha_k(t)$ is clearly

$$\sum_{m=0}^{k} |\lambda_{k,m}|,$$

which has an upper bound L, independent of k. Hence by Helly's theorem, i.e., Theorem 16.3 of Chapter I, there exists a subsequence $\{\alpha_{k_j}(t)\}_{j=0}^{\infty}$ of the sequence $\{\alpha_k(t)\}_0^{\infty}$ which approaches a limit $\alpha^*(t)$, of bounded variation in $0 \leqq t \leqq 1$. But

$$\mu_n = \lim_{j\to\infty} \int_0^1 t^n \, d\alpha_{k_j}(t)$$

$$= \lim_{j\to\infty} n \int_0^1 t^{n-1}[\alpha_{k_j}(1) - \alpha_{k_j}(t)]\, dt \qquad (n = 1, 2, \cdots)$$

$$\mu_0 = \lim_{j\to\infty} \alpha_{k_j}(1) = \alpha^*(1) = \int_0^1 d\alpha^*(t).$$

Since

$$|\alpha_{k_j}(1) - \alpha_{k_j}(t)| < 2L,$$

we may employ the Lebesgue limit theorem and obtain*

$$\mu_n = n \int_0^1 t^{n-1}[\alpha^*(1) - \alpha^*(t)]\, dt \qquad (n = 1, 2, \cdots)$$

$$= \int_0^1 t^n \, d\alpha^*(t) \qquad (n = 0, 1, 2, \cdots),$$

which is what we were to prove.

* One could avoid the integration by parts if Theorem 16.4 of Chapter I were used. The Lebesgue limit theorem is perhaps more familiar.

If $\alpha^*(t)$ is not normalized, we normalize it and denote the resulting function by $\alpha(t)$. Then by the uniqueness theorem, i.e., Theorem 6.1 of Chapter II, $\alpha^*(t) = \alpha(t)$ in the set E of points of continuity of $\alpha(t)$ Hence at these points

$$\lim_{j\to\infty} \alpha_{k_j}(t) = \alpha(t)$$

Since every subsequence of $\{\alpha_k(t)\}_0^\infty$ has in it a subsequence which approaches $\alpha(t)$ at points of E we have

$$\lim_{k\to\infty} \alpha_k(t) = \alpha(t) \qquad (t \in E) \tag{2}$$

It can be shown in fact that (2) holds throughout the interval (0, 1)

We may use Theorem 2*b* to prove an important result of F. Riesz [1909] concerning linear functionals.

DEFINITION 2*g*. *To each function $f(x)$ continuous on $0 \leqq x \leqq 1$ let there correspond a number $L[f(x)]$. This correspondence is said to define a linear functional if*

$$L[c_1f_1(x) + c_2f_2(x)] = c_1L[f_1(x)] + c_2L[f_2(x)] \tag{a}$$

for every pair of constants c_1, c_2 and every pair of continuous functions $f_1(x), f_2(x)$;

$$| L[f(x)] | \leqq M \, \| f(x) \|, \tag{b}$$

where M is some positive constant and $\| f(x) \|$ is the maximum value of $| f(x) |$ on $0 \leqq x \leqq 1$.

For example, if

$$L[f(x)] = f(\tfrac{1}{2}),$$

r if

$$L[f(x)] = \int_0^1 f(x)\,dx,$$

we see easily that $L[f(x)]$ is a linear functional. In fact by reference to Chapter I we see that if

$$L[f(x)] = \int_0^1 f(x)\,d\alpha(x) \tag{3}$$

with $\alpha(x)$ of bounded variation on $0 \leqq x \leqq 1$, then conditions (*a*) and (*b*) are satisfied with M equal to the total variation of $\alpha(x)$ on $0 \leqq x \leqq 1$. Riesz's result is that (3) defines the most general linear functional defined on the set of continuous functions. We give a proof due to T. H. Hildebrandt and I. J. Schoenberg [1933].

THEOREM 2*c*. *Every linear function l $L[f(x)]$ defined on the set of functions continuous in $0 \leqq x \leqq 1$ has the form (3) with $\alpha(x)$ of bounded variation on $0 \leqq x \leqq 1$.*

To prove this set

$$L[x^n] = \mu_n \qquad (n = 0, 1, 2, \cdots)$$

We show first that the sequence $\{\mu_n\}_0^\infty$ satisfies Condition A. We mus determine a constant N such that

$$\sum_{m=0}^{k} |\lambda_{k,m}| = \sum_{m=0}^{k} \binom{k}{m} |\Delta^{k-m}\mu_m| < N \qquad (k = 0, 1, 2, \cdots) \tag{4}$$

But by choosing $\epsilon_m = \pm 1$ suitably we have

$$\sum_{m=0}^{k} |\lambda_{k,m}| = \sum_{m=0}^{k} \epsilon_m \binom{k}{m} (-1)^{k-m} \Delta^{k-m} \mu_m$$

$$= L\left[\sum_{m=0}^{k} \epsilon_m \binom{k}{m} x^m (1 - x)^{k-m}\right].$$

Here we have used property (a) of the functional L and observed that

$$L[x^m(1 - x)^{k-m}] = (-1)^{k-m}\Delta^{k-m}\mu_m .$$

Since

$$\left\| \sum_{m=0}^{k} \epsilon_m \binom{k}{m} x^m (1 - x)^{k-m} \right\| \leqq \left\| \sum_{m=0}^{k} \binom{k}{m} x^m (1 - x)^{k-m} \right\| = 1,$$

we see by use of (b) that

$$\sum_{m=0}^{k} |\lambda_{k,m}| \leqq M,$$

so that (4) holds with $M = N$. Hence by Theorem 2b

$$L[x^n] = \int_0^1 x^n \, d\alpha(x) \qquad (n = 0, 1, 2, \cdots) \tag{5}$$

for some function $\alpha(x)$ of bounded variation on $0 \leqq x \leqq 1$.

Now let $f(x)$ be any function continuous on $0 \leqq x \leqq 1$ and let ϵ be an arbitrary positive number. By Weierstrass's theorem we can determine a polynomial $P(x)$ such that

$$|f(x) - P(x)| \leqq \epsilon \qquad (0 \leqq x \leqq 1).$$

By (5), (a) and (b) it is clear that

$$L[f(x)] = L[f(x) - P(x)] + L[P(x)] = L[f(x) - P(x)] + \int_0^1 P(x)\, d\alpha(x)$$

$$\left| L[f(x)] - \int_0^1 f(x)\, d\alpha(x) \right| \leqq M \, \| f(x) - P(x) \| + \left| \int_0^1 [P(x) - f(x)]\, d\alpha(x) \right|$$

$$\leqq M\epsilon + \epsilon \int_0^1 |d\alpha(x)|.$$

Hence (3) follows, and our theorem is proved.

It can also be shown that Theorem 2b follows from Theorem 2c. Hence the problem of determining the general linear functional on the set of continuous functions is equivalent to that of determining the set of all moment sequences.

3. An Inversion Operator

Let us now introduce a new operator on the sequence $\{\mu_n\}_0^\infty$ by the following definition.

DEFINITION 3. *An operator $L_{k,t}\{\mu_n\}$ is defined by the relation*

$$L_{k,t}\{\mu\} = L_{k,t}\{\mu_n\} = (k+1)\lambda_{k,[kt]} \qquad (k = 1, 2, \cdots, 0 \leqq t \leqq 1).$$

The notation $[kt]$ means the largest integer contained in kt. By means of this operator we can prove:

THEOREM 3. *If $\{\mu_n\}_0^\infty$ satisfies Condition A then*

$$\mu_n - \mu_\infty = \lim_{k\to\infty} \int_0^1 t^n L_{k,t}\{\mu\}\, dt \qquad (n = 0, 1, 2, \cdots).$$

For, by the law of the mean

$$\int_0^1 t^n L_{k,t}\{\mu\}\, dt = \frac{(k+1)}{k} \sum_{m=0}^{k-1} \lambda_{k,m} \left(\frac{m+\theta_m}{k}\right)^n,$$

where

$$0 < \theta_m < 1 \qquad (m = 0, 1, \cdots, k-1).$$

But we saw in section 2 that

$$\mu_n = \lim_{k\to\infty} \sum_{m=0}^{k} \left(\frac{m}{k}\right)^n \lambda_{k,m} \qquad (n = 0, 1, \cdots)$$

$$= \lim_{k\to\infty} \left[\lambda_{k,k} + \sum_{m=0}^{k-1} \left(\frac{m}{k}\right)^n \lambda_{k,m}\right].$$

To evaluate the first term we have

$$\lambda_{k,k} = \mu_k = \int_0^1 t^k\, d\alpha(t)$$

$$\mu_k = \{\alpha(1) - \alpha(1-)\} + k \int_0^1 t^{k-1}\{\alpha(1-) - \alpha(t)\}\, dt \quad (k = 1, 2, \cdots)$$

$$\overline{\lim_{k\to\infty}}\, |\mu_k - \alpha(1) + \alpha(1-)| \leqq \overline{\lim_{t\to 1-}}\, |\alpha(1-) - \alpha(t)| = 0$$

$$\mu_\infty = \alpha(1) - \alpha(1-).$$

Thus μ_∞ must exist under Conditions A, and it remains only to show that

$$\lim_{k\to\infty} \sum_{m=0}^{k-1} \left\{ \left(\frac{m+\theta_m}{k}\right)^n - \left(\frac{m}{k}\right)^n \right\} \lambda_{k,m} = 0$$

By the law of the mean we have

$$\left(\frac{m+\theta_m}{k}\right)^n - \left(\frac{m}{k}\right)^n = \theta_m \frac{n}{k} \left(\frac{m+\theta_m'}{k}\right)^{n-1}$$

where

$$0 < \theta_m' < \theta_m < 1 \qquad (m = 0, 1, \cdots, k-1).$$

Hence

$$\left| \sum_{m=0}^{k-1} \left\{ \left(\frac{m+\theta_m}{k}\right)^n - \left(\frac{m}{k}\right)^n \right\} \lambda_{k,m} \right| \leqq \frac{n}{k} \sum_{m=0}^{k-1} | \lambda_{k,m} | < \frac{nL}{k},$$

so that the theorem is established.

4. Completely Monotonic Sequences

We now introduce the notion of a completely monotonic sequence.

DEFINITION 4. *The sequence* $\{\mu_n\}_0^\infty$ *is completely monotonic if its elements are non-negative and its successive differences are alternately non-positive and non-negative*

$$(-1)^k \Delta^k \mu_n \geqq 0 \qquad (n, k = 0, 1, 2, \cdots)$$

An equivalent form for the definition is

$$\lambda_{k,m} \geqq 0 \qquad (m, k = 0, 1, 2, \cdots).$$

For example, the sequences

(1) $$\left\{\frac{1}{n+1}\right\}_0^\infty \qquad \{c^n\}_0^\infty \qquad (0 < c \leqq 1)$$

are completely monotonic. Note that this class is included in the class of sequences which satisfy Condition A. For

$$\sum_{m=0}^{k} | \lambda_{k,m} | = \sum_{m=0}^{k} \lambda_{k,m} = \mu_0 = L.$$

We can now prove:

THEOREM 4*a*. *A necessary and sufficient condition that the sequence* $\{\mu_n\}_0^\infty$ *should have the expression*

(2) $$\mu_n = \int_0^1 t^n \, d\alpha(t) \qquad (n = 0, 1, 2, \cdots),$$

where $\alpha(t)$ *is non-decreasing and bounded for* $0 \leqq t \leqq 1$, *is that it should be completely monotonic.*

For the necessity of the condition we have

$$(-1)^k \Delta^k \mu_n = \int_0^1 t^n (1-t)^k \, d\alpha(t) \geqq 0 \qquad (n, k = 0, 1, 2, \cdots).$$

For the sufficiency we see at once that the given sequence must have the form (2) with $\alpha(t)$ of bounded variation on (0, 1) by Theorem 2*b*. But we showed in section 2 that if $\alpha(t)$ is normalized

$$\lim_{k \to \infty} \alpha_k(t) = \alpha(t)$$

at all points of continuity of $\alpha(t)$. But $\alpha_k(t)$ is non-decreasing since its jumps, $\lambda_{k,m}$, are non-negative. It follows that $\alpha(t)$ is non-decreasing if properly defined at its points of discontinuity. This completes the proof of the theorem.

THEOREM 4*b*. *A necessary and sufficient condition that the sequence* $\{\mu_n\}$ *should satisfy Condition A is that it should be the difference of two completely monotonic sequences.*

This is obvious since $\alpha(t)$ is of bounded variation if and only if it is the difference of two bounded non-decreasing functions.

It is easily seen directly that the sequences (1) have the form (2). In the first case $\alpha(t)$ is the non-decreasing function t; in the second it is a step-function with jump unity at $t = c$.

5. Function of L^p

In this section we discuss sequences §2 (1) where $\alpha(t)$ is the integral of a function of class L^p $(p > 1)$. That is,

$$\mu_n = \int_0^1 t^n \varphi(t) \, dt \qquad (n = 0, 1, 2, \cdots)$$

$$\int_0^1 | \varphi(t) |^p \, dt < \infty.$$

We introduce a condition which will guarantee that a sequence will have this form.

DEFINITION 5. *The sequence* $\{\mu_n\}_0^\infty$ *satisfies Condition B for a given number* $p > 1$ *if there exists a constant L such that*

$$(k+1)^{p-1} \sum_{m=0}^{k} | \lambda_{k,m} |^p < L \qquad (k = 0, 1, 2, \cdots).$$

For example, the sequence

$$\left\{ \frac{1}{n + 1 - \theta} \right\}_0^\infty \qquad (0 < \theta < 1)$$

14

Completely Monotone Functions

Some of the theory of completely monotone functions is necessary for our later work on radial basis functions in Chapter 15. We begin with a definition.

Definition. A function f is said to be **completely monotone** on $[0, \infty)$ if

1. $f \in C[0, \infty)$
2. $f \in C^{\infty}(0, \infty)$
3. $(-1)^k f^{(k)}(t) \geq 0$ for $t > 0$ and $k = 0, 1, 2, \ldots$

Such functions exist in great abundance. Here are some examples that can be quickly verified directly from the definition:

1. $f(t) = a \qquad (a \geq 0)$
2. $f(t) = (t + a)^b \qquad (a > 0 \geq b)$
3. $f(t) = e^{-at} \qquad (a \geq 0)$

A famous theorem of Bernstein and Widder gives a complete characterization of this function class. This theorem states, in effect, that a function is completely monotone if and only if it is the Laplace transform of a nonnegative bounded Borel measure.

The theorem has an equivalent formulation in terms of the Riemann-Stieltjes integral. If γ is a nondecreasing function, we define

$$\int_a^b f(t)\, d\gamma(t) = \lim \sum_{i=1}^{n} f(\xi_i)[\gamma(t_i) - \gamma(t_{i-1})]$$

The limit here is similar to the one used in the Riemann integral. Thus the limit is equal to a number L if for each $\varepsilon > 0$ there corresponds a $\delta > 0$ such that for any partition $a = t_0 < t_1 < \cdots < t_n = b$ satisfying $\max_i |t_i - t_{i-1}| < \delta$ and for any points ξ_i satisfying $t_{i-1} \leq \xi_i \leq t_i$ we have

$$\left| L - \sum_{i=1}^{n} f(\xi_i)[\gamma(t_i) - \gamma(t_{i-1})] \right| < \varepsilon$$

Of course, as in the Riemann integral, the defining limit may not exist in some cases. (It does exist if f is continuous.) Notice that when $\gamma(x) = x$, we recover the familiar Riemann integral. Introductions to the Riemann-Stieltjes integral can be found in Widder's book [W4] and in Hewitt-Stromberg [HewS].

THEOREM 1. (Bernstein–Widder) *A function* $f : [0, \infty) \to [0, \infty)$ *is completely monotone if and only if there is a nondecreasing bounded function* γ *such that* $f(t) = \int_0^\infty e^{-st}\, d\gamma(s)$.

Proof. The easier half of the proof is to show that if γ is as stated then the integral defines a completely monotone function. The derivatives of f are obtained by differentiating under the integral. (The validity of this procedure is addressed in Theorem 5.) We obtain

$$f^{(k)}(t) = \int_0^\infty (-s)^k e^{-st}\, d\gamma(s)$$

The sign of $f^{(k)}(t)$ is clearly $(-1)^k$. To test the continuity of f at 0, note first that

$$f(0) = \int_0^\infty d\gamma(s) = \gamma(\infty) - \gamma(0)$$

On the other hand, by the Monotone Convergence Theorem,

$$\lim_{t\downarrow 0} f(t) = \lim_{t\downarrow 0} \int_0^\infty e^{-st}\, d\gamma(s) = \int_0^\infty \lim_{t\downarrow 0} e^{-st}\, d\gamma(s) = \int_0^\infty d\gamma(s) = \gamma(\infty) - \gamma(0)$$

For the other half of the proof, suppose that f is completely monotone on $[0, \infty)$. As explained later, the sequence $f(n/m)$, $(n = 0, 1, 2, \dots)$ is completely monotone for any $m \in \mathbb{N}$. By the Hausdorff Moment Theorem (Theorem 2 below), there is a nondecreasing bounded function β_m such that

$$f\left(\frac{n}{m}\right) = \int_0^1 s^n\, d\beta_m(s) \qquad (n = 0, 1, 2, \dots)$$

Also, from Theorem 2 we can assume that $\beta_m(0) = 0$ and that for every s, $\beta_m(s) = \frac{1}{2}[\beta_m(s+0) + \beta_m(s-0)]$. Replacing n by nm in an equation above, we have

$$f(n) = \int_0^1 s^{nm}\, d\beta_m(s) = \int_0^1 s^n\, d\beta_m(s^{1/m})$$

By the uniqueness part of Theorem 2, $\beta_m(s^{1/m}) = \beta_1(s)$. Putting $\gamma(s) = -\beta_1(e^{-s})$ and $s = e^{-\sigma}$, we have

$$\begin{aligned} f\left(\frac{n}{m}\right) &= \int_0^1 s^{n/m}\, d\beta_1(s) = \int_{0+}^1 s^{n/m}\, d\beta_1(s) \\ &= \int_\infty^0 e^{-\sigma n/m}\, d\beta_1(e^{-\sigma}) = \int_0^\infty e^{-ns/m}\, d\gamma(s) \end{aligned}$$

By continuity, this leads to

$$f(t) = \int_0^\infty e^{-st}\, d\gamma(s)$$ ■

The measure-theoretic version of the Bernstein-Widder Theorem states that f is completely monotone on $[0, \infty)$ if and only if it is the Laplace transform of a nonnegative, finite-valued, regular, Borel measure on $[0, \infty)$. The term **regular** when applied to a nonnegative measure ν means that

$$\nu(A) = \sup \nu(K) = \inf \nu(O)$$

where K ranges over the compact sets contained in A, and O ranges over the open sets containing A.

Some of the results needed in the preceding proof are given here. For a sequence $\mu = [\mu_0, \mu_1, \mu_2, \dots]$, the forward difference operation is defined by

$$\Delta\mu = [\mu_1 - \mu_0, \mu_2 - \mu_1, \dots]$$

This defines Δ as a linear operator, and its powers are defined in the usual way. If $(-1)^k \Delta^k \mu \geq 0$ for $k = 0, 1, 2, \cdots$, then μ is said to be a **completely monotone sequence.** Examples are $\mu_n = (n+1)^{-1}$ and $\mu_n = \lambda^n$, if $0 < \lambda \leq 1$. The famous Moment Theorem of Hausdorff is as follows:

THEOREM 2. (Hausdorff) *In order that a sequence* $[\mu_0, \mu_1, \dots]$ *be completely monotone, it is necessary and sufficient that it be the moment sequence of a nondecreasing bounded function* β *on* $[0, 1]$; *that is,*

$$\mu_n = \int_0^1 t^n \, d\beta(t)$$

If we insist that $\beta(0) = 0$ *and* $\beta(t) = \frac{1}{2}[\beta(t+0) + \beta(t-0)]$ *for every* t, *then* β *is uniquely determined by the sequence* μ.

The forward difference operator is also defined for functions:

$$(\Delta f)(x) = f(x+1) - f(x)$$

Its powers obey the equation

$$(\Delta^k f)(x) = \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} f(x+j) = \sum_{j=0}^{k} \binom{k}{j} (-1)^j f(x+k-j)$$

See [AS], page 882 or [Stef], page 10. Since these operations are special cases of divided differences, we can write (using standard notation for divided differences)

$$(\Delta f)(x) = f[x, x+1]$$
$$(\Delta^2 f)(x) = f[x, x+1, x+2], \text{ etc.}$$

If $f^{(k)}$ exists and is continuous, then for a suitable ξ between x and $x+k$,

$$(\Delta^k f)(x) = f[x, x+1, \dots, x+k] = \frac{1}{k!} f^{(k)}(\xi)$$

See, for example, [KinC], page 357. These remarks allow us to conclude that if f is completely monotone, then (for each m) the sequence $f(n/m)$ is completely monotone. Indeed, we can let $F(x) = f(x/m)$ and $\mu_n = f(n/m)$ so that

$$(\Delta^k \mu)_n = (\Delta^k F)(n) = \frac{1}{k!} F^{(k)}(\xi) = \frac{1}{k! m^k} f^{(k)}(\xi/m)$$

The sign of this last term is $(-1)^k$, by the complete monotonicity of f.

Examples. By taking various functions γ, we can generate interesting completely monotone functions:

1. If $\gamma(s) = 1 - e^{-s}$ then $f(t) = (t + 1)^{-1}$.
2. If $\gamma(s) = s$ for $0 \leq s \leq 1$ and $\gamma(s) = 1$ for $s > 1$, then $f(t) = (1 - e^{-t})/t$

One can obtain further examples by using the following theorems.

THEOREM 3. *The family of all completely monotone functions on* $[0, \infty)$ *is algebraically closed under the formation of linear combinations with positive coefficients.*

THEOREM 4. *If f and g are completely monotone on* $[0, \infty)$, *then so is fg.*

Proof. By Leibniz's rule,

$$D^k(fg) = \sum_{j=0}^{k} \binom{k}{j} (D^j f)(D^{k-j} g)$$

The sign of $(D^j f)(D^{k-j} g)$ is $(-1)^j(-1)^{k-j} = (-1)^k$. ■

Now we address the question of whether "differentiation under the integral" is a valid procedure. Thus, we seek suitable hypotheses to make the following equation correct:

$$\frac{d}{dt}\int_X f(x, t)\, d\mu(x) = \int_X \frac{\partial f}{\partial t}(x, t)\, d\mu(x) \tag{1}$$

The setting is as follows. A measure space $(X, \mathcal{A}, \mu)$ is prescribed. Thus X is a set, $\mathcal{A}$ is a σ-algebra of subsets of X, and $\mu : \mathcal{A} \to [0, \infty]$ is a measure. An open interval (a, b) is also prescribed. The function f is defined on $X \times (a, b)$ and takes values in $\mathbb{R}$. Select a point t_0 in (a, b) where $(\partial f/\partial t)(x, t_0)$ exists for all x. What further assumptions are needed in order that Equation (1) shall be true for $t = t_0$? Let us assume that

(2) For each t in (a, b), the function $x \mapsto f(x, t)$ belongs to $L^1(X, \mathcal{A}, \mu)$.

(3) There exists a function $g \in L^1(X, \mathcal{A}, \mu)$ such that

$$\left| \frac{f(x, t) - f(x, t_0)}{t - t_0} \right| \leq g(x) \qquad (x \in X,\ a < t < b,\ t \neq t_0)$$

THEOREM 5. *Under the hypotheses given above, Equation* (1) *is true for the point* $t = t_0$.

Proof. By Hypothesis (2) we are allowed to define

$$\phi(t) = \int_X f(x, t)\, d\mu(x)$$

Now the derivative $\phi'(t_0)$ exists if and only if for each sequence t_n converging to t_0 we have

$$\phi'(t_0) = \lim_{n\to\infty} \frac{\phi(t_n) - \phi(t_0)}{t_n - t_0} = \lim_{n\to\infty} \int_X \frac{f(x, t_n) - f(x, t_0)}{t_n - t_0}\, d\mu(x)$$

By Hypothesis (3), the integrands in the preceding equation are bounded in magnitude by the single L^1-function g. The Lebesgue Dominated Convergence Theorem allows an interchange of limit and integral. Hence

$$\phi'(t_0) = \int_X \lim_{n\to\infty} \frac{f(x, t_n) - f(x, t_0)}{t_n - t_0}\, d\mu(x) = \int_X \frac{\partial f}{\partial t}(x, t_0)\, d\mu(x) \qquad \blacksquare$$

This proof is given by Bartle [Bart1]. Related theorems can be found in advanced calculus books, such as [W4], page 352, and in McShane's book [McS]. A useful corollary of Theorem 5 is as follows.

THEOREM 6. *Let $(X, \mathcal{A}, \mu)$ be a measure space such that $\mu(X) < \infty$. Let $f : X \times (a, b) \to \mathbb{R}$. Assume that for each n, $(\partial^n f/\partial t^n)(x, t)$ exists, is measurable, and is bounded on $X \times (a, b)$. Then*

$$\text{(4)} \qquad \frac{d^n}{dt^n} \int_X f(x, t)\, d\mu(x) = \int_X \frac{\partial^n f}{\partial t^n}(x, t)\, d\mu(x) \qquad (n = 1, 2, \ldots)$$

Proof. Since $\mu(X) < \infty$, any bounded measurable function on X is integrable. To see that Hypothesis (3) of the preceding theorem is true, use the mean value theorem:

$$\left| \frac{f(x, t) - f(x, t_0)}{t - t_0} \right| = \left| \frac{\partial f}{\partial t}(x, \xi) \right| \le M$$

where M is a bound for $|\partial f/\partial t|$ on $X \times (a, b)$. By the preceding theorem, Equation (4) is valid for $n = 1$. The same argument can be repeated to give an inductive proof for all n. ■

Further references on the Bernstein-Widder Theorem are [Be]; [Cho], vol. 2, page 239; [Phe1], page 11; [Phe2], page 155; [SG]; [W1]; [W3], page 157; and [W5], page 162. For the moment problem, see [W5], [ST], [Lan], and [Ak1]. For the Riemann-Stieltjes integral, consult [W4].

Problems

1. Prove that for any nondecreasing bounded function γ, the function $f(t) = \int_1^\infty (t + s)^{-1}\, d\gamma(s)$ is completely monotone on $[0, \infty)$. Generalize this result.

2. Prove that if $0 < a < b$ then the function

$$f(t) = \log \frac{t + b}{t + a}$$

is completely monotone on $[0, \infty)$.

3. Determine whether the function $f(t) = \pi/2 - \tan^{-1} t$ is completely monotone on $[0, \infty)$.

4. Prove that if g is a nonnegative member of $L^1[0, \infty)$, then its Laplace transform is completely monotone on $[0, \infty)$.

5. Prove that a polynomial of degree one or more cannot be completely monotone on $[0, \infty)$.

6. A function f is defined to be completely monotone on $(0, \infty)$ if $f \in C^\infty(0, \infty)$ and $(-1)^k f^{(k)}(t) \geq 0$ for all k and t. Verify that the function $f(t) = t^{-1}$ is completely monotone on $(0, \infty)$ but not on $[0, \infty)$.

7. (Continuation) Prove that if γ is nondecreasing on $[0, \infty)$ and if the integral $\int_0^\infty e^{-st}\, d\gamma(s)$ exists for all t, then the resulting function of t is completely monotone on $(0, \infty)$.

8. What must be assumed of the function g if the function $t \mapsto g(-t)$ is to be completely monotone on $[0, \infty)$?

9. What must be assumed of the function g if $f \circ g$ is to be completely monotone on $[0, \infty)$ whenever f is completely monotone on $[0, \infty)$?

10. If f is completely monotone on $[0, \infty)$, does it follow that the function $t \mapsto f(\sqrt{t})$ is completely monotone on $[0, \infty)$?

11. Let f be completely monotone on $[0, \infty)$. Prove that $\lim_{t\to\infty} t^k f^{(k)}(t) = 0$ for $k = 1, 2, 3, \ldots$. (Stronger results are known. For example, the function $t \mapsto t^k f^{(k)}(t)$ is integrable. See a paper by R. E. Williamson in *Duke J. Math. 23* (1956), 189–207, or the book [SG]. What is the value of the limit when $k = 0$?

12. By considering the definition of the Riemann-Stieltjes integral, show that if γ has a jump discontinuity of magnitude c at t_0, then the integral $\int_a^b f(t)\, d\gamma(t)$ will contain a term $cf(t_0)$.

13. (Continuation) Show that at a point of discontinuity of γ, say t_0, the integral $\int f(t)\, d\gamma(t)$ is not changed by redefining γ at t_0 by the equation $\gamma(t_0) = \frac{1}{2}[\gamma(t_0+) + \gamma(t_0-)]$.

14. Prove the formula for $\Delta^k f$ given in the text.

15. A function f is said to be **absolutely monotone** on an interval (a, b) if $f^{(k)}(t) \geq 0$ for $k = 0, 1, 2, \ldots$ and for all t in (a, b). Prove that a function g is completely monotone on $(0, \infty)$ if and only if the function $t \mapsto g(-t)$ is absolutely monotone on $(-\infty, 0)$.

16. Prove Theorem 3.

17. Carry out the inductive proof that is needed to establish Theorem 6.

18. Prove that the family CM of all completely monotone functions on $[0, \infty)$ is closed under translation by a positive number. Thus if $f \in CM$ and $c > 0$, then $t \mapsto f(t + c)$ is also in CM.

19. Prove that a function f is completely monotone on $(0, \infty)$ if and only if there exists a Borel measure μ on $(0, 1]$ such that $f(t) = \int_0^1 x^t\, d\mu(x)$. (The change of variable $x = e^{-y}$ is useful.)

References

[AS] Abramowitz, M., and I. A. Stegun. *Handbook of Mathematical Functions.* National Bureau of Standards, Washington, 1964. Reprint, Dover, New York.

[Ak1] Akhiezer, N. I. *The Classical Moment Problem.* New York, 1955.

[Bart1] Bartle, R. G. *Elements of Integration Theory.* Wiley, New York, 1966. Reprinted, enlarged, and retitled: *Elements of Integration and Lebesgue Measure,* 1995.

[Be] Bernstein, S. N. "Sur les fonctions absolument monotones." *Acta Math. 52* (1929), 1–66.

[Cho] Choquet, G. *Lectures on Analysis* (3 vols). W. A. Benjamin, New York, 1969.

[HewS] Hewitt, E., and K. Stromberg. *Real and Abstract Analysis.* Springer-Verlag, New York, 1965.

[KinC] Kincaid, D., and W. Cheney. *Numerical Analysis.* 2nd ed., Brooks/Cole, Pacific Grove, CA, 1996.

[Lan] Landau, H. J. (ed.). *Moments in Mathematics.* Amer. Math. Soc., Providence, RI, 1987.

[McS] McShane, E. J. *Integration.* Princeton University Press, Princeton, NJ, 1944.

[Phe1] Phelps, R. R., "Lectures on Choquet's Theorem." Van Nostrand, New York, 1966. Rev. ed., *Ergebnisse der Math.* Springer-Verlag, Berlin, 1984.

[Phe2] Phelps, R. R. "Integral representation for elements of convex sets." In *Studies in Functional Analysis,* ed. by R. G. Bartle. Math. Assoc. of America, 1980, 115–157.

[SG] Shilov, G. E., and B. L. Gurevich. *Integral, Measure and Derivative: A Unified Approach.* Prentice-Hall, Englewood Cliffs, NJ, 1966.

[ST] Shohat, J. A., and J. D. Tamarkin. *The Problem of Moments.* Amer. Math. Soc., Providence, RI, 1943.

[Stef] Steffensen, J. F. *Interpolation.* Chelsea, New York, 1950.

[W1] Widder, D. V. "Necessary and sufficient conditions for the representation of a function as a Laplace integral." *Trans. Amer. Math. Soc. 33* (1932), 851–892.

[W3] Widder, D. V. *An Introduction to Transform Theory.* Pure and Applied Mathematics series, vol. 42. ed. by P. A. Smith and S. Eilenberg. Academic Press, New York, 1971.

[W4] Widder, D. V. *Advanced Calculus.* 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1961. Reprint, Dover, New York.

[W5] Widder, D. V. *The Laplace Transform.* Princeton University Press, Princeton, NJ, 1946.