JOURNAL OF FUNCTIONAL ANALYSIS 38, 255–272 (1980)

The Classical Moment Problem: Hilbertian Proofs

H. J. LANDAU

Bell Laboratories, Murray Hill, New Jersey 07974

Communicated by the Editors

Received March 1979

In view of its connection with a host of important questions, the classical moment problem deserves a central place in analysis. It is usually treated by methods striking in their virtuosity, but difficult to motivate. Here we describe an elementary approach which establishes the structural results and exposes the Hilbert space origins of the arguments.

INTRODUCTION

In view of its connection with a host of important topics—spectral representation for operators, positive definiteness, study of harmonic functions in a half-plane, partial fractions, T-spaces, prediction theory, inverse problems—the classical moment problem deserves a central place in analysis. It is usually treated by methods striking in their virtuosity, but difficult to motivate. Our object here is to elucidate them by exposing their roots in the basic notions of Hilbert space. As a by-product, we obtain a new and simple proof of the fundamental structural result.

THE MOMENT PROBLEM AND ITS SOLUTION

The classical moment problem consists of asking whether a prescribed real sequence $1 = \sigma_0, \sigma_1, \sigma_2, \ldots$, can be represented as the sequence of successive moments of some positive measure, i.e., in the form

$$\sigma_k = \int_{-\infty}^{\infty} x^k \, d\mu(x), \qquad k = 0, 1, \ldots, \tag{1}$$

with $d\mu \geqslant 0$ and, if so, whether the measure in question, called a *representing measure*, is uniquely determined. If (1) holds, then, for any choice of complex $\{a_i\}$, $\sum_{j,k=0}^{N} a_j \bar{a}_k \sigma_{j+k} = \int_{-\infty}^{\infty} | \sum_{j=0}^{N} a_j x^j |^2 \, d\mu(x)$, and if $\mu(x)$ has more than a

0022-1236/80/110255–18$02.00/0

finite number of points of increase, which is generally assumed, this quadratic form is positive definite. Thus the condition

$$\sum_{j,k=0}^{N} a_j \bar{a}_k \sigma_{j+k} \geqslant 0, \tag{2}$$

with equality only if $a_j = 0, j = 0,..., N$, is evidently necessary for the existence of the representation (1); it is proved also to be sufficient:

THEOREM A. *If condition* (2) *is satisfied, the moment problem has a solution.*

THEOREM B. *Either the representing measure is unique or, given any real point* α, *there exists a representing measure for which* α *is a point of positive mass.*

We view the dichotomy expressed in Theorem B—that the representing measures form either a very small or a very large collection—as the basic structural result.

We outline the usual proof; details can be found in the excellent exposition [1]. Several approaches to Theorem A are available. One can begin with the truncated sequence $\sigma_0, ..., \sigma_{2n}$ and show the existence of representing measures $d\mu_n(x)$ for it. Two distinct lines of reasoning can be followed here. One, described in the beautiful memoir of M.G.Krein [2], is based on convexity, and leads to the far-reaching generalizations embodied in the idea of Tchebycheff spaces [3]. The other strongly suggests Hilbert space. It uses $\sigma_0, ..., \sigma_{2n}$ to define a linear functional L in the space of polynomials of degree $2n$ by the rule $L(x^k) = \sigma_k$. In view of (2), one can introduce the sequence $1 = P_0, P_1, ..., P_n$ of polynomials orthonormal with respect to L (in the sense that $L(P_i \bar{P}_j) = \delta_{ij}$); here P_k has degree k and a positive leading coefficient. The existence of representing measures follows from Lagrange interpolation at the zeros of certain particular ("quasi-orthogonal") polynomials constructed from the $P_k(x)$. Finally, Helly's theorem is invoked to conclude that a subsequence of the $\{d\mu_n\}$ converges to a representing measure for the full sequence $\{\sigma_j\}$. Alternatively to both of these arguments, one can apply to L a general extension principle of M. Riesz, which asserts that a linear functional originally given on a submanifold, and positive on the intersection of that submanifold with a convex cone, can be extended to the entire space with positivity on the cone preserved. Each of these methods establishes Theorem A. The proof of Theorem B is more intricate.

(a) One shows that if $P_n(\alpha) \neq 0$ there exists a representing measure $d\mu_n(x)$ for the truncated problem which concentrates a mass of $[\sum_{k=0}^{n} | P_k(\alpha)|^2]^{-1}$ at the point $x = \alpha$.

(b) The polynomials $\{P_k(x)\}$ satisfy a 3-term recursion (analogous to a Sturm–Liouville differential equation) of the form $xP_k(x) = b_{k-1}P_{k-1}(x) + a_k P_k(x) + b_k P_{k+1}(x)$, in which $\{a_k\}$ are real and $\{b_k\}$ are positive. In addition to $\{P_k(x)\}$, this equation has a second polynomial solution, $\{Q_k(x)\}$. One proves that if $\sum_{k=0}^{\infty} | P_k(\lambda)|^2 < \infty$ and $\sum_{k=0}^{\infty} | Q_k(\lambda)|^2 < \infty$ for one point λ in the complex plane, the same is true for every λ.

(c) One next takes the problem into the complex domain by introducing the Stieltjes transform $w_\lambda = \int_{-\infty}^{\infty} d\mu_n(x)/(x - \lambda)$, with $\text{Im}\ \lambda > 0$. Using conformal mapping properties of linear fractional transformations, one shows that as $d\mu_n$ varies over the set of representing measures for the truncated problem, the point w_λ sweepts out a closed disk $\Delta_n{}^\lambda$ in the upper half-plane. As n increases the corresponding disks are nested, and hence converge either to a disk or to a point. One then connects this phenomenon with the behavior of the 3-term recursion by showing that, when $\text{Im}\ \lambda > 0$, both solutions are square-summable at λ and only if the limit of $\Delta_n{}^\lambda$ is a disk.

(d) It follows from (c) and (b) that if $\Delta_n{}^\lambda$ converges to a point at a single value of λ in $\text{Im}\ \lambda > 0$, it does so at each such λ. Thereupon, the Stieltjes transform $\int d\mu(x)/(x - \lambda)$ is uniquely prescribed in $\text{Im}\ \lambda > 0$, whatever the choice of the representing measure, and since that transform can be inverted, the representing measure must likewise be unique. In the opposite case, if $\Delta_n{}^\lambda$ approaches a disk for a single value of λ, the representing measure is manifestly not unique. Moreover, by (c) and (b), $\sum | P_k(\alpha)|^2 < \infty$ for each α, and so, from (a), corresponding to each real point α a representing measure can be found which concentrates a non-zero mass there.

An Elementary Hilbert Space Demonstration

The arguments just described rely considerably on methods from the theory of complex variables and differential equations. We will now recast them in a Hilbert space setting and show that they represent answers to very simple and natural questions. In this form, their analytic intricacy can be substantially reduced.

We begin by observing that the given truncated moment sequence $1 = \sigma_0, \ldots, \sigma_{2n}$ can, in view of (2), be used to define a scalar product in the space Π_n *of polynomials of degree n* by the rule

$$[x^j, x^k] = \sigma_{j+k}, \qquad 0 \leqslant j, k \leqslant n.$$

Indeed, this is merely a reinterpretation of the functional L described earlier. This scalar product has the property that, for each T_{n-1} and $U_{n-1} \in \Pi_{n-1}$,

$$[xT_{n-1}, U_{n-1}] = [T_{n-1}, xU_{n-1}]. \tag{3}$$

Conversely, any such scalar product on Π_n generates a sequence $\sigma_{j+k} = [x^j, x^k]$, $0 \leqslant j, k \leqslant n$, which satisfies (2). Let us note that, while knowledge of

σ_j for $0 \leqslant j \leqslant 2n$ does not define a scalar product in Π_{n+1}, it is nevertheless sufficient to determine the orthogonal projection of Π_{n+1} onto Π_{n-1}, since to calculate this, only the quantities $[x^j, x^k]$, $j \leqslant n+1$, $k \leqslant n-1$, are required. This observation will prove important in the subsequent argument.

Our goal is to represent the scalar product as L^2 with respect to a positive measure, i.e., in the form

$$[T_n, U_n] = \int_{-\infty}^{\infty} T_n(x)\,\overline{U_n(x)}\,d\mu(x), \qquad d\mu(x) \geqslant 0. \tag{4}$$

The Evaluation Polynomials

Let us approach the problem by considering the linear functional which assigns to each polynomial $S_n \in \Pi_n$ its value at λ. Since this functional is bounded, it can be represented as the scalar product of S_n with a polynomial which we term the *evaluation polynomial* or, as it is sometimes called, the reproducing kernel.

Definition 1. The evaluation polynomial $E_n{}^\lambda \in \Pi_n$ is that which satisfies

$$[S_n, E_n{}^\lambda] = S_n(\lambda) \tag{5}$$

for each $S_n \in \Pi_n$.

Our interest in such polynomials stems from the fact that any set of $n+1$ mutually orthogonal evaluation polynomials generates a measure satisfying (4).

Definition 2. A measure satisfying (4) which consists of $n+1$ point masses is termed an elementary atomic representing measure.

Proposition 1. *There is a* 1–1 *correspondence between elementary atomic representing measures and sets of $n+1$ mutually orthogonal evaluation polynomials.*

Proof. Suppose $d\mu_n(x)$ is an elementary atomic representing measure, with mass $m_i > 0$ at $x = \alpha_i$, $i = 0,..., n$. Then by (4), $[S_n, T_n] = \sum S_n(\alpha_i)\,\overline{T_n(\alpha_i)}\,m_i$, so that the polynomials $E_n^{\alpha_i} = \Pi_{j\neq i}(x-\alpha_j)/m_i\Pi_{j\neq i}(\alpha_i - \alpha_j)$, with values $E_n^{\alpha_i}(\alpha_j) = m_i^{-1}\delta_{ij}$, evidently form $n+1$ mutually orthogonal evaluation polynomials. Conversely, suppose $\{E_n^{\alpha_i}\}$, $i = 0,..., n$, are $n+1$ mutually orthogonal evaluation polynomials, so that $\{E_n^{\alpha_i}/\| E_n^{\alpha_i}\|\}$ form an orthonormal basis in Π_n. Then, expanding S_n in this basis yields, by (5),

$$S_n = \sum_{i=0}^{n} \left[S_n, \frac{E_n^{\alpha_i}}{\| E_n^{\alpha_i}\|}\right] \frac{E_n{}^i}{\| E_n^{\alpha_i}\|} = \sum_{i=0}^{n} \frac{S_n(\alpha_i)}{\| E_n^{\alpha_i}\|} \frac{E_n^{\alpha_i}}{\| E_n^{\alpha_i}\|},$$

whence

$$[S_n, T_n] = \sum_{i=0}^{n} S_n(\alpha_i) \overline{T_n(\alpha_i)} \frac{1}{\| E_n^{\alpha_i} \|^2}.$$

The right-hand side can now be written as $\int S_n(x) \overline{T_n(x)}\, d\mu(x)$, with $d\mu$ the elementary atomic measure having mass $\| E_n^{\alpha_i} \|^{-2}$ at $x = \alpha_i$. ∎

In consequence of Proposition 1, let us seek representing measures by looking for mutually orthogonal sets of evaluation polynomials. Evidently, from its definition, we can construct E_n^λ explicitly as

$$E_n^\lambda = \sum \phi_k(x) \overline{\phi_k(\lambda)} \tag{6}$$

with $\{\phi_k\}$ any orthonormal basis in Π_n. A convenient such basis consists of the polynomials $\{P_k(x)\}_{k=0}^n$, $P_0(x) = 1$, found by orthogonalizing the successive powers of x by the Gram–Schmidt procedure. Let us begin by collecting a few facts concerning these. Throughout, the only property of polynomials we will use is that, for $n > 0$, S_n has at least one zero, and if $S_n(\lambda) = 0$ then $S_n(x) = (x - \lambda) S_{n-1}(x)$.

PROPOSITION 2. *$P_k(x)$ has real coefficients, and k real and distinct zeros; P_k and P_{k+1} have no zeros in common.*

Proof. $P_k(x)$ has real coefficients by construction, and degree no smaller than k, or else x^k lies in the linear span of $1, \ldots, x^{k-1}$, contradicting the definiteness of the scalar product; let the leading coefficient be positive. If P_k has a non-real zero λ, then since its coefficients are real, $\bar\lambda$ is also a zero, hence $P_k(x)/(x - \lambda)(x - \bar\lambda)$ is a polynomial of degree $k - 2$, so that, by definition of P_k, $[P_k(x)/(x - \lambda)(x - \bar\lambda), P_k] = 0$. But writing the right-hand P_k as $(x - \lambda) P_k/(x - \lambda)$, and using (3), the scalar product becomes $\| P_k/(x - \lambda)\|^2 > 0$, a contradiction. By the same argument, there cannot be any multiple zeros. If P_k and P_{k+1} have a (necessarily real) common zero α, $P_k(x)/(x - \alpha)$ is a polynomial of degree $k - 1$, so $[P_k/(x - \alpha), P_{k+1}] = 0$ by definition of P_{k+1}. But by (3), $[P_k/(x - \alpha), P_{k+1}] = [P_k, P_{k+1}/(x - \alpha)]$ so $P_{k+1}/(x - \alpha)$, of degree k, being orthogonal to P_k, must be of degree $k - 1$, a contradiction. ∎

Inserting the orthonormal set $\{P_k\}$ into (6) we obtain

$$E_n^\lambda = \sum_{k=0}^{n} P_k(x) \overline{P_k(\lambda)}. \tag{7}$$

By definition, E_n^λ is orthogonal to all the polynomials of Π_n which vanish at λ, i.e., to all those of the form $(x - \lambda) S_{n-1}$. This immediately provides the following useful characterization of E_n^λ.

PROPOSITION 3. *Let $\mathscr{P}_{n-1}$ denote the orthogonal projection onto the subspace Π_{n-1}: explicitly, $\mathscr{P}_{n-1}T = \sum_{k=0}^{n-1} [T, P_k] P_k$, so that $\mathscr{P}_{n-1}$ is, by our earlier remark, defined unambiguously for $T \in \Pi_{n+1}$. Then E_n^λ is the solution to the equations*

$$\mathscr{P}_{n-1}(x - \bar{\lambda}) T_n = 0, \tag{8}$$

$$[1, T_n] = 1. \tag{9}$$

If $T_n \neq 0$ satisfies only (8), *then* $[1, T_n] \neq 0$ *and* $T_n/\overline{[1, T_n]} = E_n^\lambda$.

Proof. Condition (8) states that $(x - \bar{\lambda}) T_n$ is orthogonal to all polynomials of Π_{n-1}. Now if $S_{n-1} \in \Pi_{n-1}$, we see by the definition of E_n^λ and by (3)

$$0 = [(x - \lambda) S_{n-1}, E_n^\lambda] = [S_{n-1}, (x - \bar{\lambda}) E_n^\lambda],$$

so that $\mathscr{P}_{n-1}(x - \bar{\lambda}) E_n^\lambda = 0$, while (9) holds by definition. Conversely, if T_n satisfies (8) and (9), then for each $S_n \in \Pi_n$

$$\begin{aligned}[S_n, T_n] &= [S_n(\lambda) + (x - \lambda) S_{n-1}, T_n] \\ &= S_n(\lambda)[1, T_n] + [S_{n-1}, (x - \bar{\lambda}) T_n] = S_n(\lambda),\end{aligned}$$

so that T_n coincides with E_n^λ. Finally, if $T_n \neq 0$ satisfies (8), then

$$\begin{aligned}0 < \| T_n \|^2 &= [T_n, T_n] = [T_n(\lambda) + (x - \lambda) T_{n-1}, T_n], \qquad \text{with} \quad T_{n-1} \in \Pi_{n-1}, \\ &= T_n(\lambda)[1, T_n] + [T_{n-1}, (x - \bar{\lambda}) T_n] = T_n(\lambda)[1, T_n].\end{aligned}$$

Thus $[1, T_n] \neq 0$ and $T_n/\overline{[1, T_n]}$ satisfies both (8) and (9). ∎

Now to continue our search for mutually orthogonal evaluation polynomials let us note that, by definition, $[E_n^\lambda, E_n^\nu] = E_n^\lambda(\nu)$, hence E_n^ν is orthogonal to E_n^λ whenever

$$E_n^\lambda(\nu) = 0. \tag{10}$$

Information concerning the zeros of E_n^λ is easy to find from first principles.

PROPOSITION 4. *The degree of E_n^λ is at least $n - 1$, and equals $n - 1$ if and only if λ is one of the zeros of P_n. If α is real and not a zero of P_n, E_n^α has n distinct real zeros $\beta_1, \ldots, \beta_n$, and E_n^α together with $\{E_n^{\beta_i}\}$ forms a set of $n + 1$ mutually orthogonal evaluation polynomials. If α is one of the zeros of P_n, the evaluation polynomials at the n zeros of P_n all have degree $n - 1$, are mutually orthogonal, and are orthogonal to P_n.*

Proof. By definition, $E_n^\lambda(\lambda) = [E_n^\lambda, E_n^\lambda] > 0$, so that λ is never a zero of E_n^λ. By (7) and Proposition 2, the degree of E_n^λ is at least $n - 1$, since $P_n(\lambda)$ and $P_{n-1}(\lambda)$ cannot both vanish; it is $n - 1$ if and only if λ is one of the zeros $\alpha_1, \ldots, \alpha_n$ of P_n. In this case, by Proposition 3, $E_n^{\alpha_i}$ is a scalar multiple of $P_n(x)/(x - \alpha_i)$.

These are mutually orthogonal by (10) and, having degree $n - 1$, are each orthogonal to P_n. If α is real and not a zero of P_n, then E_n^α, of degree n, is real on the reals, hence its complex zeros occur in conjugate pairs. If ν is a complex zero, then $E_n^\alpha(x)(x - \alpha)^2/(x - \nu)(x - \bar{\nu})$ has degree n and vanishes at α. Thereupon, by definition of E_n^α, $[E_n^\alpha(x - \alpha)^2/(x - \nu)(x - \bar{\nu}), E_n^\alpha] = 0$, but by (3) the scalar product is $\| E_n^\alpha(x)(x - \alpha)/(x - \nu)\|^2 \neq 0$, a contradiction; the same argument eliminates the possibility of multiple real zeros. We conclude that E_n^α has n distinct real zeros $\beta_1, ..., \beta_n$, different from α. Now by Proposition 3, $E_n^\alpha(x)(x - \alpha)/(x - \beta_i)$ coincides with a scalar multiple of $E_n^{\beta_i}$, and these polynomials, together with E_n^α, are mutually orthogonal by (10). ▮

In the light of Proposition 4 we see that, corresponding to each α which is not a zero of P_n, we can find a set of $n + 1$ mutually orthogonal evaluation polynomials, which in turn, by Proposition 1, generate an elementary atomic representing measure for the truncated moment problem. This measure has mass $\| E_n^\alpha \|^{-2}$ at $x = \alpha$. We now give a simple expression in closed form for E_n^λ, which shows how these measures behave as α varies.

PROPOSITION 5.

$$E_n^\lambda(x) = c_n \frac{P_{n+1}(x)\, P_n(\bar{\lambda}) - P_n(x)\, P_{n+1}(\bar{\lambda})}{x - \bar{\lambda}},$$

with $c_n = [xP_n, P_{n+1}] > 0$.

Proof. By Proposition 3, $(x - \bar{\lambda})\, E_n^\lambda(x)$, a polynomial of degree $n + 1$, is orthogonal to Π_{n-1}, so its general form is a linear combination of $P_n(x)$ and $P_{n+1}(x)$. More specifically, the determination of $P_{n+1}(x)$ requires knowledge of σ_{2n+1} and σ_{2n+2} which we do not necessarily yet have. Nevertheless, whatever these quantities may be, so long as they satisfy (2), the subspace of linear combinations of P_n and P_{n+1} remains unchanged. This follows, of course, from the fact that $\mathscr{P}_{n-1}$ is well defined on Π_{n+1}, but we can also see it explicitly, since P_{n+1} is a scalar multiple of $x^{n+1} - \mathscr{P}_n x^{n+1} = x^{n-1} - \sum_{k=0}^{n} [x^{n+1}, P_k]\, P_k(x)$. In this formula, a change in the value of σ_{2n+1} will affect only the coefficient of P_n, while a change in σ_{2n+2} simply rescales P_{n+1}, so that the set of all linear combinations of P_{n+1} and P_n is not affected. Thus

$$(x - \bar{\lambda})\, E_n^\lambda(x) = a(\lambda)\, P_{n+1}(x) + b(\lambda)\, P_n(x)$$

and evaluation at $\bar{\lambda}$ gives

$$0 = a(\lambda)\, P_{n+1}(\bar{\lambda}) + b(\lambda)\, P_n(\bar{\lambda}),$$

whence

$$(x - \bar{\lambda})\, E_n^\lambda(x) = c(\lambda)[P_{n+1}(x)\, P_n(\bar{\lambda}) - P_n(x)\, P_{n+1}(\bar{\lambda})].$$

Replacing E_n^λ on the left by (7), and forming the scalar product with P_{n+1}, we conclude that $c(\lambda) = c_n = [xP_n, P_{n+1}]$ is independent of λ. Since xP_n has leading term $\gamma_n x^{n+1}$, $\gamma_n > 0$, while that of P_{n+1} is $\gamma_{n+1}x^{n+1}$, we see that xP_n differs from $(\gamma_n/\gamma_{n+1}) P_{n+1}$ by a polynomial of degree n, to which P_{n+1} is orthogonal. Thus $[xP_n, P_{n+1}] = (\gamma_n/\gamma_{n+1}) \| P_{n+1} \|^2 > 0$. ∎

The formula of Proposition 5 yields a simple description of the zeros of E_n^α, on which our interest has centered. To see what these are, let us observe that, from it,

$$0 < \| E_n^\alpha \|^2 = E_n^\alpha(\alpha) = c_n\{P'_{n+1}(\alpha) P_n(\alpha) - P'_n(\alpha) P_{n+1}(\alpha)\}$$
$$= -c_n P_{n+1}^2(\alpha) \frac{d}{dx}\left\{\frac{P_n(x)}{P_{n+1}(x)}\right\}\Bigg|_{x=\alpha}.$$

This shows that $P_n(x)/P_{n+1}(x)$ decreases between each two zeros of P_{n+1}, at which points it approaches $\pm\infty$, so that the zeros of P_n must interlace those of P_{n+1}. The zeros $\{\beta_i\}$ of E_n^α together with α, are the solutions to $P_n(x)/P_{n+1}(x) = P_n(\alpha)/P_{n+1}(\alpha)$. As a sketch of the function $P_n(x)/P_{n+1}(x)$ shows, they evidently lie one in each interval between successive zeros of P_{n+1} and increase as α increases in its interval; all $n + 1$ of the points $\{\alpha, \beta_1, ..., \beta_n\}$ are determined by fixing a single one. As described in Proposition 4, there are $n + 1$ of them except when α is a zero of P_n, whereupon there are n. We have thus shown that each truncated moment problem has a one-parameter family of elementary atomic representing measures; the parameter can conveniently be taken to be the prescribed value of $P_n(x)/P_{n+1}(x)$.

We conclude our discussion of E_n^λ by showing that it can also be characterized as the solution of a natural extremal problem.

Proposition 6. *$E_n^\lambda(x)/E_n^\lambda(\lambda)$ is the polynomial in Π_n of least norm having value 1 at λ; equivalently, $\| E_n^\lambda \|^{-1} = \inf_{S_{n-1}\in\Pi_{n-1}} \| 1 - (x - \lambda) S_{n-1} \|$. The largest mass that can be concentrated at α by a representing measure for the truncated problem is $\| E_n^\alpha \|^{-2}$.*

Proof. If $S_n(\lambda) = 1$, then $[S_n, E_n^\lambda] = 1$; hence by Schwarz's inequality $\| S_n \| \geq \| E_n^\lambda \|^{-1}$, with equality if and only if $S_n = E_n^\lambda(x)/E_n^\lambda(\lambda)$. If $d\mu(x)$ is a representing measure with mass m_α at $x = \alpha$, then

$$\| E_n^\alpha \|^2 = \int | E_n^\alpha(x)|^2 \, d\mu(x) \geq m_\alpha \, | E_n^\alpha(\alpha)|^2 = m_\alpha \| E_n^\alpha \|^4. \quad ∎$$

Proof of Theorem A

To return to the full moment problem, given an infinite positive definite sequence $1 = \sigma_0, \sigma_1, ...,$ we can, by Propositions 4 and 1, construct elementary atomic representing measures μ_n for each truncated sequence $\sigma_0, \ldots, \sigma_{2n}$ and by Helly's theorem find a limiting measure which then has the prescribed moments. Moreover, if we fix a real point α and for each n select that measure μ_n which has mass $\| E_n^\alpha \|^{-2}$ at α, the limiting measure will assign to α the mass $\lim_{n\to\infty} \| E_n^\alpha \|^{-2} = \{\sum_{k=0}^{\infty} | P_k(\alpha)|^2\}^{-1}$. In view of Proposition 6, this is the maximum mass that a representative measure can concentrate at α.

Connection with Related Problems

The representation of Proposition 5 is the Christoffel–Darboux formula, here derived very simply. Similarly, the linear combinations of P_n and P_{n+1} with real coefficients are the "quasiorthogonal polynomials," whose significance is clear from our point of view, since by Proposition 5 their zeros determine the mutually orthogonal evaluation polynomials, and these in turn lead immediately to representing measures.

The operator $\mathscr{P}_{n-1}xT_n$ corresponds to the 3-term recursion satisfied by the orthogonal polynomials $\{P_k(x)\}$. This is the discrete analogue, with k corresponding to t, of a second-order Sturm–Liouville differential operator D defined on $0 \leqslant t < \infty$; more precisely, in this identification the value of a function at t corresponds to the component of an element of the Hilbert space along the vector P_k. The spectral theory for such operators [4–6] asserts that if $\phi(t, \lambda)$ is the solution to the problem $D\phi(t, \lambda) = \lambda\phi(t, \lambda)$, viewed as an initial value problem, the mapping $F(\lambda) = \int_0^\infty f(t)\phi(t, \lambda)\, dt$ generates a unitary transformation of $f \in L^2(0, \infty)$ onto $L^2(d\mu)$ for an appropriate $d\mu(\lambda)$, termed a spectral measure. In the context of differential equations, it is natural to prove this by imposing a self-adjoint boundary condition at a right-hand endpoint $t = T$, showing that those $\phi(t, \lambda_j)$ which satisfy it form a mutually orthogonal system in $L^2(0, T)$, expanding $f \in L^2(0, T)$ in this system, and finally letting $T \to \infty$. Here the role of $\phi(t, \lambda)$ in $0 \leqslant t \leqslant T$ is played by E_n^λ. The self-adjoint boundary condition at $t = T$ corresponds to

$$aP_{n+1}(\lambda) - bP_n(\lambda) = 0, \tag{11}$$

so that we see at once from Proposition 5 that the values of λ_j for which (11) holds yield mutually orthogonal $\{E_n^{\lambda_j}\}$. The spectral measure now corresponds to a limit of the representing measures $d\mu_n$ of Proposition 1, generated here as $n \to \infty$, with a/b fixed; it is therefore nothing other than a representing measure for the moment problem to which the 3-term recursion corresponds.

Similarly, the present analogy clarifies some of the issues concerning the so-called inverse problem—that of determining the differential equation from its eigenvalues—which is of particular interest in speech synthesis. It is known [7–11] that in general two sets of eigenvalues, corresponding to different self-adjoint boundary conditions, are required. We can see this at once in our current formulation, for the eigenvalues corresponding to a particular boundary condition of the form (11) are the zeros of $aP_{n+1}(\lambda) - bP_n(\lambda)$. If these are given, then $aP_{n+1}(x) - bP_n(x)$ is determined (up to a constant factor), and if they are given for a second, distinct, condition, then $cP_{n+1}(x) - dP_n(x)$ is available as well. This, however, is sufficient to find $P_{n+1}(x)$ and $P_n(x)$ separately, up to a single constant factor, and from there, by means of Proposition 5, the evaluation polynomials $E^\alpha{}_n$ for any α. These in turn uniquely determine the scalar product, hence everything about the operator $\mathscr{P}_{n-1}xT_n$. Specifically, starting with $C(x) = \gamma(aP_{n+1}(x) - bP_n(x))$ and $D(x) = \delta(cP_{n+1}(x) - dP_n(x))$, the expression $C(x)D(\bar{\lambda}) - D(x)C(\bar{\lambda}) = \gamma\delta(bc - ad)(P_{n+1}(x)\, P_n(\bar{\lambda}) - P_n(x)\, P_{n+1}(\bar{\lambda}))$. Thus by Proposition 5, $(C(x)D(\bar{\lambda}) - D(x)C(\bar{\lambda}))/(x - \bar{\lambda}) = kE_n{}^\lambda(x)$, with k a fixed constant, whence, for real α, $C'(\alpha)D(\alpha) - D'(\alpha)C(\alpha) = kE_n{}^\alpha(\alpha) = k\| E_n{}^\alpha \|^2$. The constant k can now be found from the fact that, by Propositions 1 and 5, $\sum \| E_n^{\alpha_j} \|^{-2} = 1$, where $\{\alpha_j\}$ are the $n + 1$ known zeros of $aP_{n+1}(x) - bP_n(x)$, $a \neq 0$. In turn, by Proposition 1, the quantities $\| E_n^{\alpha_j} \|^{-2}$ here determined specify an elementary atomic representing measure, hence the entire scalar product.

Limiting Behavior

To examine the behavior of representing measures as n increases, we focus briefly on the equation which generates $E_n{}^\lambda$.

Proposition 7. *Given a Hilbert space element A and scalar a, the equation*

$$\mathscr{P}_{n-1}(x - \bar{\lambda})\, T_n = \mathscr{P}_{n-1}A, \qquad [1, T_n] = a, \tag{12}$$

has a unique solution $T_n \in \Pi_n$. As n increases, successive solutions T_n change by orthogonal increments, hence converge if and only if $\| T_n \|$ remains bounded.

Proof. A solution is unique since, by Proposition 3, if $\mathscr{P}_{n-1}(x - \bar{\lambda})\, T_n = 0$ with $[1, T_n] = 0$, then necessarily $\| T_n \| = 0$. It follows that the map $\mathscr{P}_{n-1}(x-\bar{\lambda})S_n$ takes the $(n - 1)$-dimensional subspace of Π_n on which $[1, S_n] = 0$ onto Π_{n-1}. Letting V_n be the element mapped on $\mathscr{P}_{n-1}A$, the solution of (12) is $T_n = V_n + \bar{a}E_n{}^\lambda$. Next let us note that $\mathscr{P}_{n-1}(x - \bar{\lambda})T$ depends only on the component $\mathscr{P}_nT$. For decomposing T into components in, and orthogonal to, Π_n, we obtain $T = \mathscr{P}_nT + \tau$, with $[S_n, \tau] = 0$. It follows that $0 = [(x - \lambda)\, S_{n-1}, \tau] = [S_{n-1}, (x - \bar{\lambda})\tau]$, or equivalently that $\mathscr{P}_{n-1}(x - \bar{\lambda})\tau = 0$. We conclude that

$$\mathscr{P}_{n-1}(x - \bar{\lambda})T = \mathscr{P}_{n-1}(x - \bar{\lambda})\, \mathscr{P}_nT. \tag{13}$$

Now if

$$\mathscr{P}_n(x - \bar{\lambda})\, T_{n+1} = \mathscr{P}_nA, \qquad [1, T_{n+1}] = a,$$

then applying $\mathscr{P}_{n-1}$ yields $\mathscr{P}_{n-1}(x - \bar{\lambda})\, T_{n+1} = \mathscr{P}_{n-1}A$, and so by (13), $\mathscr{P}_{n-1}(x - \bar{\lambda})\, \mathscr{P}_n T_{n+1} = \mathscr{P}_{n-1}A$; moreover, since $[1, T_{n+1}]$ depends only on the component of T_{n+1} in the space Π_0, evidently $a = [1, T_{n+1}] = [1, \mathscr{P}_n T_{n+1}]$. Thus $\mathscr{P}_n T_{n+1}$ and T_n each satisfy (12) and by uniqueness they coincide. This implies that T_{n+1} differs from T_n by an increment orthogonal to T_n. We can conveniently express this in the orthonormal basis of $\{P_k\}$ by writing $T_n = \sum_{j=0}^{n} a_j P_j$, whence $\{T_n\}$ converges if and only if $\sum_{j=0}^{n} |a_j|^2 = \| T_n \|^2$ remains bounded as $n \to \infty$. ∎

For reasons which will become clear in the sequel, let us introduce $F_n^\lambda(x)$, the polynomial of Π_n which solves

$$\mathscr{P}_{n-1}(x - \bar{\lambda})\, F_n^\lambda = 1, \qquad [1, F_n^\lambda] = 0. \tag{14}$$

Paralleling our description of E_n^λ, it is natural to ask for the effect of F_n^λ in the scalar product.

Proposition 8. $[S_n, F_n^\lambda] = [(S_n(x) - S_n(\lambda))/(x - \lambda), 1]$.

Proof.

$$\begin{aligned}
[S_n, F_n^\lambda] &= \left[S_n(\lambda) + (x - \lambda)\,\frac{S_n(x) - S_n(\lambda)}{x - \lambda}, F_n^\lambda\right] \\
&= S_n(\lambda)[1, F_n^\lambda] + \left[\frac{S_n(x) - S_n(\lambda)}{x - \lambda}, (x - \bar{\lambda}) F_n^\lambda\right] \\
&= S_n(\lambda)[1, F_n^\lambda] + \left[\frac{S_n(x) - S_n(\lambda)}{x - \lambda}, \mathscr{P}_{n-1}(x - \bar{\lambda}) F_n^\lambda\right],
\end{aligned}$$

the last equality stemming from the fact that $(S_n(x) - S_n(\lambda))/(x - \lambda) \in \Pi_{n-1}$. The defining properties (14) of F_n^λ complete the proof. ∎

In view of Proposition 8, let us introduce the operator $\mathscr{A}_\lambda$ defined for polynomials by

$$\mathscr{A}_\lambda S = \frac{S(x) - S(\lambda)}{x - \lambda}. \tag{15}$$

$\mathscr{A}_\lambda$ takes each space Π_n into itself, and we see that, for each n,

$$\mathscr{A}_{\bar{\lambda}} P_n = \frac{P_n(x) - P_n(\bar{\lambda})}{x - \bar{\lambda}} = [F_n^\lambda, P_n]\,[E_n^\lambda - E_n^\lambda, P_n]\, F_n^\lambda, \tag{16}$$

since both expressions satisfy $\mathscr{P}_{n-1}(x - \bar{\lambda})\, S_n = -P_n(\bar{\lambda})$ with $[S_n, 1] = [F_n^\lambda, P_n]$, and such a solution is unique by Proposition 7. It follows from (16) that if E_n^λ and F_n^λ both converge as $n \to \infty$, the operator $\mathscr{A}_\lambda$ is well behaved.

In this connection, let us note from the outset that Proposition 7 applies to E_n^λ and F_n^λ, since, by Proposition 3 and (14), each is defined by an equation of the form (12).

PROPOSITION 9. *If E_n^λ and F_n^λ converge as $n \to \infty$, the operator A_λ is completely continuous: for T orthogonal to Π_n, $\| \mathscr{A}_\lambda T \| \leqslant \epsilon_n \| T \|$, where $\epsilon_n \to 0$ as $n \to \infty$.*

Proof. If E_n^λ and F_n^λ have limits E and F as $n \to \infty$, (16) shows that $\mathscr{A}_{\bar\lambda} P_n$ approaches a linear combination of the latter with coefficients which, by Proposition 7, are the components of E and F in the orthonormal basis $\{P_k\}$, and hence are square-summable over n. Indeed, suppose T is orthogonal to Π_n, so that $T = \sum_{n+1}^N \tau_k P_k$, $\| T \|^2 = \sum | \tau_k |^2$. Then by (16)

$$\mathscr{A}_{\bar\lambda} T = \sum_{k=n+1}^{N} \tau_k \{[F, P_k] E_k^\lambda - [E, P_k] F_k^\lambda\}$$

and by the Minkowski and Schwarz inequalities,

$$\| \mathscr{A}_{\bar\lambda} T \| / \| T \| \leqslant \| E \| \left\{ \sum_{n+1}^{N} |[F, P_k]|^2 \right\}^{1/2} + \| F \| \left\{ \sum_{n+1}^{N} |[E, P_k]|^2 \right\}^{1/2};$$

the right-hand side evidently approaches 0 as $n \to \infty$. Since, from their definition, $F_n^{\bar\lambda}(x) = \overline{F_n^\lambda(x)}$ and $E_n^{\bar\lambda}(x) = \overline{E_n^\lambda(x)}$, replacement of λ by $\bar\lambda$ establishes the proposition also for $\mathscr{A}_\lambda$. ▮

By exploiting complete continuity in the usual way, we are led directly to the following conclusion.

PROPOSITION 10. *If E_n^λ and F_n^λ both converge as $n \to \infty$ for a single point λ, real or complex, they likewise converge for every point ν.*

Proof. As we have seen in Proposition 6, $\| E_n^\lambda \|$ measures the largest value which a polynomial of unit norm can attain at $x = \lambda$. To connect the behavior of a polynomial at ν and at λ, let us write simply

$$S_n(\nu) = S_n(\lambda) + (\nu - \lambda) \frac{S_n(\nu) - S_n(\lambda)}{\nu - \lambda}$$

or, equivalently,

$$[S_n , E_n^\nu] = [S_n , E_n^\lambda] + (\nu - \lambda)[\mathscr{A}_\lambda S_n , E_n^\nu], \tag{17}$$

whence

$$[(I - (\nu - \lambda) \mathscr{A}_\lambda) S_n , E_n^\nu] = [S_n , E_n^\lambda], \tag{18}$$

for each $S_n \in \Pi_n$.

Let us show next that for each γ, the operator $I - \gamma\mathscr{A}_\lambda$, defined for polynomials, has a bounded inverse. For in the contrary case, there exists a sequence of polynomials $\{v_n\}$ with $\| v_n \| = 1$ such that

$$\|(I - \gamma\mathscr{A}_\lambda) v_n \| \to 0. \tag{19}$$

Since, by Proposition 9, $\mathscr{A}_\lambda$ is completely continuous, there exists a subsequence $\{v_j\}$ such that $\{\mathscr{A}_\lambda v_j\}$ converges, so by (19), $\{v_j\}$ likewise converges, to an element v (not necessarily a polynomial), with $\| v \| = 1$, and

$$(I - \gamma\mathscr{A}_\lambda)v = 0. \tag{20}$$

Now let us decompose v into components $S_n \in \Pi_n$ and T_n orthogonal to Π_n, do the same with $\mathscr{A}_\lambda v$, and rewrite (20) as

$$\{S_n - \gamma\mathscr{A}_\lambda S_n - \gamma\mathscr{P}_n\mathscr{A}_\lambda T_n\} + \{T_n - \gamma(I - \mathscr{P}_n)\mathscr{A}_\lambda T_n\} = 0.$$

Since by its definition, $\mathscr{A}_\lambda$ takes Π_n into itself, the bracketed terms represent components in Π_n and orthogonal to Π_n, respectively, hence their sum vanishes only if each does separately. Thus

$$S_n - \gamma\mathscr{A}_\lambda S_n = \gamma\mathscr{P}_n\mathscr{A}_\lambda T_n \tag{21}$$

and

$$\| T_n \| = \| \gamma(I - \mathscr{P}_n)\mathscr{A}_\lambda T_n \| \leqslant | \gamma | \| \mathscr{A}_\lambda T_n \|, \tag{22}$$

the last inequality stemming from the fact that $I - \mathscr{P}_n$ is a projection (onto the orthogonal complement of Π_n), and hence diminishes norm. Now in view of Proposition 9, let us choose n sufficiently large that $| \gamma | \cdot \epsilon_n \leqslant \rho < 1$, whereupon $| \gamma | \| \mathscr{A}_\lambda T_n \| \leqslant \rho \| T_n \|$. Then by (22), $T_n = 0$, so that from (21), $\gamma\mathscr{A}_\lambda S_n = S_n$. But since $\mathscr{A}_\lambda$ reduces the degree of a polynomial, this equation has only the trivial solution. Thus $v = S_n + T_n = 0$, contradicting the requirement that $\| v \| = 1$, and thereby establishing the boundedness of $(I - \gamma\mathscr{A}_\lambda)^{-1}$.

Again, since $(I - \gamma\mathscr{A}_\lambda) S_n = 0$ has only the trivial solution, $(I - \gamma\mathscr{A}_\lambda)$ maps the finite-dimensional space Π_n onto itself. Accordingly, returning to (18), let S_n satisfy

$$(I - (\nu - \lambda)\mathscr{A}_\lambda) S_n = E_n^\nu.$$

By the boundedness of $(I - \gamma\mathscr{A}_\lambda)^{-1}$, there exists a constant C, independent of n, such that $\| S_n \| \leqslant C \| E_n^\nu \|$. But now we find from (18)

$$\| E_n^\nu \|^2 = |[S_n , E_n^\lambda]| \leqslant \| S_n \| \| E_n^\lambda \| \leqslant C \| E_n^\nu \| \| E_n^\lambda \|.$$

Consequently, $\| E_n^\nu \| \leqslant C \| E_n^\lambda \|$, and convergence of E_n^ν follows from Proposition 7. By Proposition 8, F_n^ν and F_n^λ likewise satisfy (17), and so the identical argument applies to F_n^ν. ∎

Theorem B: A Simple Proof

The approach we have taken leads to an elementary, direct proof which obviates the need for complex methods and for the Stieltjes transform.

Proposition 11. *Suppose $d\tau(x)$ is a representing measure for the moment problem, with τ normalized so that $\tau(-\infty) = 0$, $\tau(\infty) = 1$. If, for real α, $\| E_n^\alpha \| \to \infty$ as $n \to \infty$, then $\tau(\alpha)$ is uniquely determined.*

Proof. Our argument here follows that of [3, p. 85]. Suppose that $d\mu_n(x)$ is an elementary atomic representing measure for the truncated problem, with $\mu_n(-\infty) = 0$, $\mu_n(\infty) = 1$. Then

$$\int_{-\infty}^{\infty} x^k d(\mu_n - \tau) = 0, \qquad k = 0, \dots, 2n,$$

and an integration by parts shows that $\mu_n(x) - \tau(x)$ is orthogonal in $L^2(dx)$ to all polynomials of degree $2n - 1$. Consequently, $\mu_n(x) - \tau(x)$ must have at least $2n$ sign changes, or else one could match them by those of a polynomial of degree $2n - 1$, contradicting the orthogonality. Since μ_n is a monotone step function which rises at $n + 1$ points, we see that τ must intersect it at each of the rises, with the possible exception of the first or last, in which case τ must be 0 or 1 there, respectively. It follows that, if $d\mu_n$ has a mass point at $x = \alpha$, the value of τ at α lies between $\mu_n(\alpha-)$ and $\mu_n(\alpha+)$. Thus if the mass of $d\mu_n$ at $x = \alpha$ approaches 0 as $n \to \infty$, or equivalently, by Proposition 1, if $\| E_n^\alpha \| \to \infty$, the value of each $\tau(x)$ for which $d\tau$ is a representing measure is determined at $x = \alpha$. ▌

In consequence of Proposition 11, if the measure is not unique, the set of points α at which $\| E_n^\alpha \| \to \infty$ cannot be dense; hence there exists an interval I and a constant C such that

$$\| E_n^\alpha \| \leq C, \qquad \text{for} \quad \alpha \in I. \tag{23}$$

We can conclude the argument by means of the next observation.

Proposition 12. *If E_n^α satisfies* (23), *then E_n^ν converges as $n \to \infty$ for each ν.*

Proof. Since, by Proposition 1, in constructing the elementary atomic measures, the mass at a point $\alpha \in I$ is $\| E_n^\alpha \|^{-2} > C^{-2}$, and since the total mass is 1, there cannot be more than C^2 points of mass contained in I. We can therefore select a subsequence $\{d\mu_j\}$, $j \to \infty$, of these measures so that, for some point

$\beta \in I$, the distance of β from each mass point of $d\mu_j$ exceeds a certain $\delta > 0$, independently of j. Thereupon, for any polynomial $S_j \in \Pi_j$,

$$\left\| \frac{S_j(x) - S_j(\beta)}{x - \beta} \right\| = \left\| \frac{S_j(x) - S_j(\beta)}{x - \beta} \right\|_{d\mu_j} \leqslant \frac{1}{\delta} \| S_j(x) - S_j(\beta) \|$$
$$\leqslant \frac{1}{\delta} (\| S_j \| + |[S_j, E_j^\beta]|)$$
$$\leqslant \| S_j \| \frac{1}{\delta} (1 + C) = C_2 \| S_j \|.$$

Consequently, using Proposition 8,

$$|[S_j, F_j^\beta]| \leqslant \left\| \frac{S_j(x) - S_j(\beta)}{x - \beta} \right\| \leqslant C_2 \| S_j \|,$$

whence, letting S_j coincide with F_j^β, we see that $\| F_j^\beta \| \leqslant C_2$. Thus F_n^β and E_n^β both converge as $n \to \infty$, and by Proposition 10, so does E_n^ν for each ν. ∎

Evidently, Propositions 11 and 12 constitute a proof of Theorem B, for they show that if the representing measure is not unique, $\| E_n^\alpha \|^2 \to l_\alpha < \infty$ for each α, so that, by considering limits of sequences of elementary atomic measures $d\mu_n(x)$ with masses $\| E_n^\alpha \|^{-2}$ at $x = \alpha$, we see that there exists a representing measure carrying the positive mass l_α^{-1} at $x = \alpha$. Theorem B is thus established.

This line of argument also shows that the basic dichotomy of Theorem B is reflected in the fact that, as $n \to \infty$, the zeros of $P_n(x)$ become either everywhere dense or nowhere dense on the real line.

Theorem B: The Classical Argument

From our present vantage point we can also give a simple interpretation of the classical approach to uniqueness. As we have outlined, it consists of introducing the Stieltjes transform $\int d\mu(x)/(x - \lambda)$, with λ in the upper half-plane and $d\mu$ a representing measure, describing when this is uniquely determined, and passing to μ by an inversion formula. The above quantity can be thought of as the scalar product of 1 with $1/(x - \bar{\lambda})$ in $L^2(d\mu)$ or equivalently, since $1 \in \Pi_n$, as the scalar product of 1 with the projection of $1/(x - \bar{\lambda})$ onto Π_n in $L^2(d\mu)$. (It can also be viewed as $[R_\lambda 1, 1]$, with R_λ the resolvent corresponding to multiplication by x in $L^2(d\mu)$; we choose the former notion as being more elementary.) Accordingly, let $d\mu$ be a representing measure for the truncated problem, and let $V_{\mu,n}^\lambda(x)$ denote the projection of $1/(x - \bar{\lambda})$ onto Π_n in $L^2(d\mu)$ (different representing measures will in general produce different polynomials $V_{\mu,n}^\lambda$). By definition

$$\frac{1}{x - \bar{\lambda}} = V_{\mu,n}^\lambda(x) + U(x), \tag{24}$$

with $U(x)$ (not necessarily a polynomial) orthogonal in $L^2(d\mu)$ to all polynomials of Π_n. To see how $V^\lambda_{\mu,n}$ in fact depends on μ, let

$$w = \int \frac{d\mu(x)}{x-\lambda} = \int \overline{V^\lambda_{\mu,n}(x)}\, d\mu(x) = [1, V^\lambda_{\mu,n}], \tag{25}$$

the last equality coming from the fact that $d\mu$ represents the scalar product on Π_n. From (24),

$$1 - (x-\bar\lambda)\, V^\lambda_{\mu,n}(x) = (x-\bar\lambda)\, U(x),$$

which shows that $(x-\bar\lambda)U(x)$ is a polynomial of degree $n+1$. Consequently $[S_{n-1}, (x-\bar\lambda)U(x)]$ is defined by the moments $\sigma_0, ..., \sigma_{2n}$ alone, and we find

$$\begin{aligned}[S_{n-1}, (x-\bar\lambda)\, U(x)] &= \int S_{n-1}(x)\overline{(x-\bar\lambda)\, U(x)}\, d\mu(x)\\ &= \int (x-\lambda)\, S_{n-1}(x)\, \overline{U(x)}\, d\mu(x) = 0,\end{aligned}$$

the first equality stemming from the representing property of $d\mu$, and the last from the definition of U. Thus $(x-\bar\lambda)U(x)$ is orthogonal to all polynomials of Π_{n-1}, i.e.,

$$0 = \mathscr{P}_{n-1}(x-\bar\lambda)U = \mathscr{P}_{n-1}(1-(x-\bar\lambda)\, V^\lambda_{\mu,n}(x))$$

so that

$$\mathscr{P}_{n-1}(x-\bar\lambda)\, V^\lambda_{\mu,n}(x) = 1. \tag{26}$$

By virtue of (25), (26), (8), (9), (14), and Proposition 7 we now see that

$$V^\lambda_{\mu,n}(x) = \bar w E_n{}^\lambda(x) + F_n{}^\lambda(x),$$

and since $F_n{}^\lambda$ is independent of μ, $V^\lambda_{\mu,n}$ depends on μ only through the quantity w. We observe simply that, since $V^\lambda_{\mu,n}$ is a projection of $1/(x-\bar\lambda)$ in $L^2(d\mu)$, its norm cannot exceed that of $1/(x-\bar\lambda)$, so that

$$\begin{aligned}\|\bar w E_n{}^\lambda + F_n{}^\lambda\|^2 \leqslant \int \frac{d\mu(x)}{|x-\lambda|^2} &= \frac{1}{\lambda-\bar\lambda}\left(\int \frac{d\mu(x)}{x-\lambda} - \int \frac{d\mu(x)}{x-\bar\lambda}\right)\\ &= \frac{w-\bar w}{\lambda-\bar\lambda}.\end{aligned} \tag{27}$$

Now, with $\operatorname{Im}\lambda > 0$, let Δ^λ represent the set of points w generated by (25), as $d\mu$ varies over the set of representing measures for the full moment problem; by (27) and Proposition 7, for each $w \in \Delta^\lambda$, $\bar w E_n{}^\lambda + F_n{}^\lambda$ converges as $n \to \infty$.

Now if Δ^λ consists of a single point for each λ in $\operatorname{Im} \lambda > 0$, $\int d\mu(x)/(x-\lambda)$ is uniquely determined in the upper half-plane, and the Stieltjes inversion formula shows μ to be unique. On the other hand, if for a single λ, Δ^λ contains more than one point, the representing measure evidently cannot be unique; moreover, the convergence of $\bar{w}_1 E_n{}^\lambda + F_n{}^\lambda$ and $\bar{w}_2 E_n{}^\lambda + F_n{}^\lambda$ implies that of all linear combinations, hence also of $E_n{}^\lambda$ and $F_n{}^\lambda$. By Proposition 10, $\| E_n{}^\alpha \|^2$ then approaches a finite limit l_α, for each α, and so there exists a representing measure carrying the mass l_α^{-1} at $x = \alpha$. This establishes Theorem B.

We can complete our account of the classical circle of ideas by determining exactly $\Delta_n{}^\lambda$, the set of points w generated by (25), as $d\mu$ varies over the measures which represent the scalar product on Π_n. Each such point satisfies (27), which, by (25), is equivalent to

$$\left\| \bar{w} E_n{}^\lambda + F_n{}^\lambda - \frac{1}{\lambda - \bar{\lambda}} \right\|^2 \leqslant \frac{1}{| \lambda - \bar{\lambda} |^2}.$$

Expanding the left-hand side and completing the square shows that this inequality describes a closed disk in the upper half of the w-plane. The boundary of this disk corresponds to equality in (27) which in turn is equivalent to the fact that $1/(x-\bar{\lambda})$ is already in Π_n in $L^2(d\mu)$. This certainly happens if $d\mu$ is one of the elementary atomic representing measures, for then all of $L^2(d\mu)$ coincides with Π_n ; but also conversely, since if $1/(x-\bar{\lambda}) = V_{\mu,n}^\lambda(x)$ in $L^2(d\mu)$, then $d\mu$ has its support only on the $n+1$ zeros of $1/(x-\bar{\lambda}) - V_{\mu,n}^\lambda(x)$. The situation therefore is that $\Delta_n{}^\lambda$ coincides with this disk, and for w on its rim $1-(x-\bar{\lambda})\, V_{\mu,n}^\lambda(x)$ vanishes at a set of points $\{\alpha_i\}_{i=0}^n$, where, according to Propositions 1 and 5, $P_n(x) - \tau P_{n+1}(x) = 0$ with some real $\tau \neq 0$. Since both functions belong to Π_{n+1}, coincidence of their zeros implies that

$$1-(x-\bar{\lambda})\{\bar{w} E_n{}^\lambda(x) + F_n{}^\lambda(x)\} = c\{P_n(x) - \tau P_{n+1}(x)\}, \tag{28}$$

for some constant c; we determine c by evaluating (28) at $x = \bar{\lambda}$. Now setting $x = \lambda$ in (28), we conclude that the rim of $\Delta_n{}^\lambda$ is given by

$$\frac{1}{\lambda-\bar{\lambda}} - \bar{w} E_n{}^\lambda(\lambda) - F_n{}^\lambda(\lambda) = \frac{1}{\lambda-\bar{\lambda}} \, \frac{P_n(\lambda) - \tau P_{n+1}(\lambda)}{\overline{P_n(\lambda) - \tau P_{n+1}(\lambda)}},$$

as τ varies over the reals. Since $E_n{}^\lambda(\lambda) = \| E_n{}^\lambda \|^2$, and the right-hand side has constant absolute value, we see explicitly that the center of $\Delta_n{}^\lambda$ is at $\| E_n{}^\lambda \|^{-2}((\bar{\lambda}-\lambda)^{-1} - \overline{F_n{}^\lambda(\lambda)})$ and its radius is $| \lambda - \bar{\lambda} |^{-1} \| E_n{}^\lambda \|^{-2}$. Thus Δ^λ is a non-degenerate disk if and only if $E_n{}^\lambda$ converges.

Since by (27), for a non-real point λ some linear combination of $E_n{}^\lambda$ and $F_n{}^\lambda$ always converges, the convergence of $E_n{}^\lambda$ alone suffices to ensure the convergence of $E_n{}^\lambda$ and $F_n{}^\lambda$ separately, hence by Propositition 10 also the convergence of

E_n^ν for each ν. This completes the classical argument for the geometric version of the dichotomy of Theorem B: Δ^λ is either a point or a non-degenerate disk, independently of the choice of non-real λ.

References

1. N. I. Akhiezer, "The Classical Moment Problem," Hafner, New York, 1965.
2. M. G. Krein, The ideas of P. L. Čebyšev and A. A. Markov in the theory of limiting values of integrals and their further development, *Uspehi Mat. Nauk. (NS)* **6** (1951), 3–120; *Amer. Math. Soc. Transl. Ser. 2* **12** (1959), 1–122.
3. S. J. Karlin and W. J. Studden, "Tchebycheff Systems: With Applications in Analysis and Statistics," Interscience, New York, 1966.
4. E. A. Coddington and N. Levinson, "Theory of Ordinary Differential Equations," McGraw–Hill, New York, 1955.
5. N. I. Akhiezer and I. M. Glazman, "Theory of Linear Operators in Hilbert Space," Ungar, New York, 1966.
6. B. M. Levitan and I. S. Sargsyan, "Introduction to Spectral Theory: Self Adjoint Ordinary Differential Operators," Amer. Math. Soc., Providence, R.I., 1975.
7. G. Borg, Eine Umkehrung der Sturm–Liouvilleschen Eigenwertaufgabe, *Acta Math.* **78**, No. 2 (1945), 1–96.
8. I. M. Gelfand and B. M. Levitan, On the determination of a differential equation from its spectral function, *Izv. Akad. Nauk SSSR* **15** (1957), 309–360; *Amer. Math. Soc. Transl. Ser. 2* **1** (1955), 253–304.
9. M. G. Krein, Solution of the Inverse Sturm–Liouville problem, *Dokl. Akad. Nauk SSSR* **76**, No. 1 (1951), 21–24.
10. M. M. Sondhi and B. Gopinath, Determination of vocal tract shape from impulse Response at the lips, *J. Acoust. Soc. Amer.* **49**, No. 6 (Part 2) (1971), 1867–1873.
11. L. Gårding, The inverse of vowel articulation, *Ark. Mat.* **15** (1977), 63–86.