A DEVICE FOR STUDYING HAUSDORFF MOMENTS

BY

CHANDLER DAVIS

1. Introduction. The Hausdorff moment problem [9, pp. 1, 8-9] asks for necessary and sufficient conditions on the numbers \( \mu_n \) in order that there exist a distribution function \( \Phi \) on \([0, 1] \) such that

\[
\text{for all } n \in I, \quad \mu_n = \int_0^1 t^n d\Phi(t).
\]

Here \( I = \{0, 1, 2, \cdots \} \). The reduced Hausdorff moment problem [9, p. 77] asks the same question where \( I \) is a proper subset of \( \{0, 1, 2, \cdots \} \), usually a finite subset. If \( I \) is allowed to be any subset, this includes the first problem.

It is known (cf. [11, Theorem 10.30]) and easy to prove that one condition is the existence of a matrix \( A = (A_{ij}) \) such that \( 0 \leq A \leq 1 \) and

\[
\text{for all } n \in I, \quad \mu_n = (A^n)_{00}.
\]

The matrix \( A \) may be chosen to be a Jacobi matrix, that is, \( A_{ij} = 0 \) for \( |i-j| > 1 \); and \( A_{n,n+1} \geq 0 \) may be required. If \( I \) is \( \{0, 1, 2, \cdots \} \), \( A \) is then determined uniquely, assuming the convention that any invariant subspace of \( A \) orthogonal to the 0th coordinate subspace will be ignored. (Assuming, that is, that if \( A_{k,k+1} = A_{k+1,k} = 0 \) then in \( A = (A_{ij}) \) indices will be let run up to \( k \) only.) For finite \( I \), \( A \) is not in general determined uniquely.

This paper gives a convenient canonical form for \( A \). There is little trouble in including in this result the generalized sort of moment problem introduced by Nagy [6], where \( \mu_n \) above are in \( \mathbb{R} \), the set of bounded self-adjoint operators on a Hilbert space \( \mathcal{H} \). One virtue of the generalized problem is its application to the classical problem; see Proposition 4 below. Accordingly some of the facts outlined in the preceding paragraph may as well be proved for the generalized problem; this is done in §2, which in fact is essentially a recitation in the wider setting of the proof for the classical case. The main result is in §3. The following sections mostly examine the classical problem in light of it. I have found the canonical form handy for getting numerical bounds on moments, but will not discuss this use further.

Throughout the paper I use the notation \( \bar{\alpha} = 1 - \alpha \), where \( \alpha \) may be a number or an operator. If it is an operator this requires some understanding as to the space on which it operates, but I think I have avoided ambiguity.

Presented to the Society February 25, 1956; received by the editors September 10, 1956.
2. Jacobi matrices.

**Theorem 1.** Let $\mu_n \in \mathcal{B}$ for $n \in I$, where $I \subseteq \{0, 1, 2, \ldots \}$. The following are equivalent:

(i) There exists a function $\Phi$ on $[0, 1]$ to $\mathcal{B}$ such that $\Phi(0^-) = 0 \leq \Phi(t_i)$ $\leq \Phi(t_i) \leq \Phi(1) = 1$ for $0 \leq t_i \leq t_j \leq 1$, and such that (1) holds.

(ii) There exists a matrix $A = (A_{ij})^*_{i,j=0}$, with $A_{ij} \in \mathcal{B}$, $A_{ij} \geq 0$, and $A_{ij} = 0$ for $|i-j| > 1$; such that $0 \leq A \leq 1$ and (2) holds.

**Proof.** Take $I = \{0, 1, 2, \ldots \}$, because from this the result for any subset of indices will follow.

Neumark's theorem [7](2) says that $\Phi(t)$ with the properties described in (i) can be expressed as $\Phi(t) = PE(t)P$, where $E(t)$ is a resolution of the identity in a Hilbert space $\mathcal{K}$ containing $\mathcal{K}$, and $P$ is the projection on $\mathcal{K}$ onto $\mathcal{K}$. That is, (i) is equivalent to the existence of such $\mathcal{K}$ and an operator $A$ on $\mathcal{K}$, $0 \leq A \leq 1$, such that for $n \in I$, $\mu_n = PAP$. Now it is straightforward to define orthogonal projections $P_1$, $P_2$, $\ldots$ such that relative to the subspaces $P_i\mathcal{K}$ the matrix of $A$ has the desired form; as follows.

The $P_i$ and $A_{ij}$ will be defined inductively together with operators $E_i$ such that $E_i$ maps $P_i\mathcal{K}$ isometrically onto $\mathcal{K}$, and $E_iP_j = 0$ for $i \neq j$, $P_i = E_i^*E_i$. Let $E_0 = P_0 = P$ above, that is, $\mathcal{K}$ is identified by the identity mapping with $P_0\mathcal{K}$; $A_{00} = P_0AP_0$. Suppose $A_{ij} \in \mathcal{B}$ and $E_i$, satisfying all requirements, have been defined whenever $i, j \leq k$. Let $A_{k+1,k} \in \mathcal{B}$ and $E_{k+1}$ satisfy the following:

\[
(P_kAP_k = E_{k+1}^*A_{k+1,k}E_k; \\
A_{k+1,k} \geq 0; E_{k+1}^*E_k \text{ an isometry on the range of } E_k^*A_{k+1,k}E_k \geq 0.
\]

(This is a polar resolution, for which see e.g. [8, §110].) As noted above, $E_{k+1}^*$ is zero on $\mathcal{K}^\perp$. But also (3) may fail to define $E_{k+1}^*$ because $E_k^*A_{k+1,k}E_k$ as an operator on $P_k\mathcal{K}$ may have a nullspace. If so, let $E_{k+1}^*E_k$ map it isometrically onto a subspace of the nullspace of $A$ orthogonal to $\sum_k P_k\mathcal{K}$; this is always possible because $\mathcal{K}$ is a space being constructed and can be augmented if desired by a new orthogonal subspace on which $A$ is defined to be 0. Last, of course, define $P_{k+1} = E_{k+1}^*E_{k+1}$ and $P_{k+1}AP_{k+1} = E_{k+1}^*A_{k+1,k+1}E_{k+1}$.

It is now clear that (i) implies (ii). The converse deduction is the same only easier.

I left the facts on uniqueness out of the statement of the theorem. The operator $A$ given by Neumark's theorem is unique up to isomorphism if we require (as we clearly may) that there is no subspace of $\mathcal{K}$ invariant under $A$ and orthogonal to $\mathcal{K}$ [7]. But the construction above introduced such in-

(1) With the usual understanding that $\Phi(0^-)$ is defined. We may as well assume $\Phi(t) = \Phi(t^+)$. 

(2) This is an appropriate occasion to acknowledge in print my mistake in announcing this theorem as new [1]. My method, it happened, was different from Neumark's, being an extended version of a device of E. A. Michael [4, Thm. 2].
essential subspaces of $\mathcal{K}$. Before uniqueness can be asserted $\mathcal{K}$ must be pared down to that subspace which actually plays a part. This is done (at the cost of some clumsiness of statement) in Theorem 3 below.

3. The canonical form.

**Theorem 2.** Let $^{(3)} \mathbf{A} = ((A_{ij}))_{i,j \geq 0}, A_{ij}$ bounded operators on $\mathcal{K}; A = A^*; \text{and } A_{ij} = 0 \text{ for } |i-j| > 1$. If $A$ has any invariant subspace orthogonal to the 0th coordinate Hilbert space, assume $A$ annihilates it. In order that $0 \leq A \leq 1$, it is necessary and sufficient that there exist $\eta_i \in \mathcal{B} (i = 1, 2, \cdots)$ with $0 \leq \eta_i \leq 1$, and partial isometries $\xi_i$ defined on $\mathcal{K}$ into $\mathcal{K}$, such that

\begin{align}
A_{i,i} &= \xi_i^* (\eta_{2i})^{1/2} \eta_{2i-1} (\eta_{2i})^{1/2} + (\eta_{2i})^{1/2} \eta_{2i+1} (\eta_{2i})^{1/2} \xi_i, \\
A_{i-1,i} &= A_{i-1,i}^* = \xi_i^* (\eta_{2i-2})^{1/2} (\eta_{2i-1})^{1/2} (\eta_{2i})^{1/2} \xi_i.
\end{align}

Here the convention is $\eta_0 = 0, \xi_0 = 1$.

Some general facts about matrices will be given as lemmas. In the lemmas, let each $\mathcal{K}_i$ be a Hilbert space; let each $a_{ij}$ be a bounded operator on $\mathcal{K}_j$ to $\mathcal{K}_i$; let $a_{ji} = a_{ij}^*$ and $a_{ii} \geq 0$.

**Lemma 1.** In order that

\begin{pmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{pmatrix} \equiv 0,

it is necessary and sufficient that $a_{21}a_{11}^{-1}$ be bounded on the range of $a_{11}$ and that $b = a_{22} - a_{21}a_{11}^{-1}a_{12} \geq 0$.

**Proof.** Case I. $a_{21}a_{11}^{-1}$ is densely defined but unbounded.

Then there is a sequence $\{x_\nu\}$ of unit elements of $\mathcal{K}_1$ such that $\|a_{11}x_\nu\|$ approaches zero but $\|a_{21}x_\nu\|$ does not. Selecting a subsequence if necessary gives that for some $\delta > 0$ and for all $\nu, \|a_{21}x_\nu\| > \delta$. As a temporary convenience, identify $\mathcal{K}_1$ with $\mathcal{K}_2$, in such a way that $a_{21} \geq 0$. (This may always be done, by enlarging one or the other $\mathcal{K}_i$ if needed.) Then, by expanding \[ \|a_{21}x_\nu\|^2 \leq \|a_{21}\|^2, \]

one easily computes that $(a_{21}x_\nu, x_\nu) > \delta^2/\|a_{21}\| > 0$. Now consider

\begin{align}
P(x, y) &= (a_{11}x, x) + (a_{12}y, x) + (a_{21}x, y) + (a_{22}y, y).
\end{align}

I must show that, for some choice of $x, y \in \mathcal{K}_1$, $P(x, y) < 0$. This is accomplished by letting $\nu$ be so large that $\|a_{11}x_\nu\| < \delta^2/\|a_{22}\|$; whence

\begin{align}
P(\|a_{22}\| x_\nu, -\delta^2 x_\nu) < 0
\end{align}

may be readily verified.

\(^{(3)}\) The assumption $A_{ij} \geq 0$ has been dropped. Keeping it would not obviate the nuisance of mentioning "phases" $\xi_i$ anyhow.
Case II. \( a_{11} \) has a nullspace. Then \( a_{21} \) may be assumed to be zero there, and \( a_{32}a_{11}^{-1} \) may be defined to be zero there. (The proof is like the preceding one, but simpler.) Granted this, Case III applies.

Case III. \( a_{21}a_{11}^{-1} \) exists as a bounded operator on \( X_1 \) to \( X_2 \). From this follow the existence and boundedness of \( a_{11}^{-1}a_{12}, a_{21}a_{11}^{-1/2} \), and \( a_{11}^{-1/2}a_{12} \).

Now (6) may be rewritten, by making the substitution \( a_{22} = a_{21}a_{11}^{-1}a_{12} + b \) and simplifying, as

\[
P(x, y) = \| a_{11} x + a_{11}^{-1/2} a_{12} y \|^2 + (b, y).
\]

Clearly \( b \geq 0 \) implies \( P(x, y) \geq 0 \); half of the lemma is proved. In case \( b \leq 0 \), choose \( y \in X_2 \) so \( (b, y) < 0 \), and let \( x = -a_{11}^{-1}a_{12}y \). For this choice, \( P(x, y) < 0 \).

**Lemma 2.** In order that

\[
\begin{bmatrix}
a_{11} & a_{12} & 0 \\
a_{21} & a_{22} & a_{23} \\
0 & a_{32} & a_{33}
\end{bmatrix} \succeq 0,
\]

it is necessary and sufficient that

\[
a_{22} = b + c, \quad \text{with} \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & b \end{pmatrix} \succeq 0, \quad \begin{pmatrix} c & a_{23} \\ a_{32} & a_{33} \end{pmatrix} \succeq 0.
\]

The definite matrix is being expressed as a sum of simpler definite matrices.

**Proof.** (8) is equivalent to

\[
a_{22} \succeq a_{21}a_{11}^{-1}a_{12} + a_{23}a_{33}^{-1}a_{32} \succeq 0.
\]

This results by applying Lemma 1 to \( X_1 \) and \( X_2 \), then to \( X_3 \) and \( X_2 \).

But Lemma 1 may also be applied to the pair of spaces \( X_1 \oplus X_3 \) and \( X_2 \). This says (7) is equivalent to

\[
a_{22} \succeq (a_{21} a_{23}) \begin{pmatrix} a_{11} & 0 \\ 0 & a_{33} \end{pmatrix}^{-1} \begin{pmatrix} a_{12} \\ a_{32} \end{pmatrix} \succeq 0,
\]

which is the same as (9).

**Lemma 3.** In order that there exist \( a_{22} \) such that

\[
0 \leq \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \leq 1,
\]

it is necessary and sufficient that \( 0 \leq a_{11} \leq 1 \) and \( a_{21}a_{11}^{-1}a_{12}a_{12}^{-1}a_{12} \leq 1 \).

**Proof.** Although I have chosen to state Lemma 3 in this simple form, a slight generalization is needed below. What I will prove is the generalization. Impose as upper bound on the matrix
not 1, but instead
\[
\begin{pmatrix}
\hat{b} & 0 \\
0 & 1
\end{pmatrix},
\] with \(0 \leq \hat{b}\).

This requires \(0 \leq a_{11} \leq \hat{b}\), or, what is equivalent, the existence of \(c\) such that \(0 \leq c \leq 1\) and \(a_{11} = b^{1/2}c^{1/2}\). The problem is to find a condition which, together with \(0 \leq c \leq 1\), is necessary and sufficient for the matrix to satisfy the inequalities.

Use Lemma 1 twice: for the matrix to be \(\geq 0\), \(a_{22} \geq a_{21}a_{11}^{-1}a_{12}\); and for the matrix to be \(\leq 0\), \(\leq \begin{pmatrix}
\hat{b} & 0 \\
0 & 1
\end{pmatrix}\), \(\hat{a}_{22} \geq (a_{21})(b - a_{11})^{-1}(-a_{12})\).

Such \(a_{22}\) can exist if and only if
\[
1 = a_{22} + \hat{a}_{22} \geq a_{21}(a_{11}^{-1} + (b - a_{11})^{-1})a_{12}
= a_{21}b^{-1/2}(c^{-1} + \hat{c}^{-1})b^{-1/2}a_{12}
= a_{21}b^{-1/2}c^{-1}b^{-1/2}a_{12} = a_{21}b^{-1/2}c^{-1}b^{-1/2}a_{12}.
\]

(As in Lemma 1, the formula makes sense even though some inverses may fail to exist everywhere.) When \(b = 1\), \(c = a_{11}\), giving Lemma 3 as stated.

Proof of Theorem 2. To say that \(\geq 0\) is to say that \(A \geq 0\) and \(\bar{A} \leq 0\). Both \(A\) and \(\bar{A}\) are given as Jacobi matrices with operator elements; \((\bar{A})_{ii} = (A_{ii})^{-1}\), \((\bar{A})_{i-1,i} = -A_{i-1,i}\).

According to (4), \(A_{00} = \eta_1\); so all the theorem says about \(A_{00}\) is \(0 \leq A_{00} \leq 1\). This is evidently necessary and sufficient for the existence of \(A_{01}, A_{11}, \ldots\) such that \(0 \leq A \leq 1\).

Proceed by induction on \(k\). Inductive hypothesis: That \(A_{00}, A_{01}, \ldots, A_{k-2,k-1}, A_{k-1,k-1}\) satisfy (4) and (5), with \(\eta_i\) and \(\xi_i\) as described, is necessary and sufficient for the existence of \(A_{k-1,k}, A_{kk}, \ldots\), such that the \(A\) in the statement of the theorem will satisfy \(0 \leq A \leq 1\).

The theorem will have been proved once the following is deduced from the inductive hypothesis: Given \(A_{00}, A_{01}, \ldots, A_{k-2,k-1}, A_{k-1,k-1}\) expressed in the prescribed form. That \(A_{k-1,k}\) (and of course its adjoint \(A_{k,k-1}\)) and \(A_{kk}\) be expressible in the prescribed form, consistently with the preceding \(A_{ij}\), is necessary and sufficient for the existence of \(A_{k,k+1}, A_{k+1,k+1}, \ldots\) such that \(0 \leq A \leq 1\).

Furthermore we may simplify the discussion by assuming in the proof that \(A_{k,k+1} = A_{k+1,k+1} = \cdots = 0\), or that the matrix expression of \(A\) is
$(k+1) \times (k+1)$. To see this, use $P_i$, the projections on the coordinate Hilbert spaces, as in the proof of Theorem 1. In general $0 \leq A \leq 1$ implies $0 \leq (P_0 + \cdots + P_k)A(P_0 + \cdots + P_k) \leq 1$; while this relation, though it does not imply $0 \leq A \leq 1$, does imply the existence of some choice of $A_{k,k+1}$, $A_{k+1,k+1}$, etc. such that $0 \leq A \leq 1$—namely, the choice that all be zero. So take $P_{k+1} = P_{k+2} = \cdots = 0$.

As another simplification, we may assume $\zeta_0 = \cdots = \zeta_{k-1} = 1$ hereafter. (The most that could possibly be involved here is replacing one replica of $P_0 \mathcal{K}$ by another. If $\zeta_i$ annihilates some of the range of an operator it premultiplies, either that operator or $\zeta_i$ may be redefined.)

Consider $\mathcal{K}$, the Hilbert space where $A$ is defined, as the sum of the following three spaces: $\mathcal{K}_1 = (P_0 + \cdots + P_{k-2}) \mathcal{K}$, $\mathcal{K}_2 = P_{k-1} \mathcal{K}$, $\mathcal{K}_3 = P_k \mathcal{K}$. Decompose $A$ in this way: $a_{22} = A_{k-1,k-1}$, etc.

Now to derive conditions on $A_{k-1,k}$ and $A_{k,k-1} = (A_{k-1,k})^*$, apply Lemma 2.

By the inductive hypothesis, the g.l.b. (in the poset of bounded Hermitian operators) of allowable values for $A_{k-1,k-1}$ is $\eta_{2k-2}^{-1/2} \eta_{2k-3}^{-1/2} \eta_{2k-2}^{-1/2}$. Hence, for $A \geq 0$ the requirement on $a_{23}$ is

\[
\begin{pmatrix}
d & a_{23}
da_{32} & a_{33}
\end{pmatrix} \leq 0,
\]

with $d = A_{k-1,k-1} - \eta_{2k-2}^{-1/2} \eta_{2k-3}^{-1/2} \eta_{2k-2}^{-1/2}$. The same reasoning derives a requirement from the condition $A \geq 0$. The g.l.b. of allowable values for $A_{k-1,k-1}$ is $\eta_{2k-2}^{-1/2} \eta_{2k-3}^{-1/2} \eta_{2k-2}^{-1/2}$, so the condition on $a_{23}$ is

\[
\begin{pmatrix}
d' & -a_{23}
-a_{32} & a_{33}
\end{pmatrix} \leq 0,
\]

with $d' = A_{k-1,k-1} - \eta_{2k-2}^{-1/2} \eta_{2k-3}^{-1/2} \eta_{2k-2}^{-1/2} = -\eta_{2k-2}^{-1/2} \eta_{2k-1}^{-1/2} \eta_{2k-2}^{-1/2}$. Combining the two conditions on $a_{23}$,

\[
0 \leq \begin{pmatrix} d & a_{23} \\ a_{32} & a_{33} \end{pmatrix} \leq \begin{pmatrix} \eta_{2k-2} & 0 \\ 0 & 1 \end{pmatrix},
\]

because $d + d' = \eta_{2k-2}$. By the generalization of Lemma 3 (or, when $k = 1$, Lemma 3 itself), this is equivalent to

\[
(10) \quad 1 \geq A_{k-1,k} \geq \begin{pmatrix} -1/2 & -1/2 \\ -1/2 & -1/2 \end{pmatrix} \begin{pmatrix} \eta_{2k-2} & -1 \\ -1 & \eta_{2k-2} \end{pmatrix} \begin{pmatrix} -1/2 & -1/2 \\ -1/2 & -1/2 \end{pmatrix} A_{k-1,k}.
\]

(Remember that $a_{23} = A_{k-1,k}$.)

$A_{k-1,k}$ may be written in the form (5) for some bounded $\eta_{2k} \geq 0$ and partial isometry $\xi_k$, just by virtue of the fact, contained in (10), that

\[
\begin{pmatrix} -1/2 & -1/2 \\ -1/2 & -1/2 \end{pmatrix} \begin{pmatrix} \eta_{2k-2} & -1 \\ -1 & \eta_{2k-2} \end{pmatrix} \begin{pmatrix} -1/2 & -1/2 \\ -1/2 & -1/2 \end{pmatrix} A_{k-1,k}
\]

is bounded. The closure of the range of $\eta_{2k}$ may be assumed contained in that of $\xi_k$. But when (5) is substituted in, (10) reduces readily to $\xi_{k}^* \xi_k \geq \xi_{k}^* \eta_{2k} \xi_{k}^* \xi_k$, which is equivalent to $\eta_{2k} \leq 1$. 

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use
It remains to consider $A_{kk} = a_{33}$.

By Lemmas 2 and 1, the condition equivalent to $A \geq 0$ is this:

$$a_{22} = b + c, \quad \text{with} \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & b \end{pmatrix} \geq 0, \quad \text{and} \quad a_{33} \geq a_{32}c^{-1}a_{23}.$$

The existence of some choice for $c$ which will allow a solution, is ensured by the argument just concluded about $a_{23}$. The g.l.b. of acceptable values for $a_{32}c^{-1}a_{23}$, hence for $a_{33}$, is attained when $c$ attains its l.u.b., hence when $b$ attains its g.l.b.; this is not hard to prove. Also any $a_{33}$ satisfying the resulting inequality will be consistent with $A \geq 0$. Now the g.l.b. of $b$ under the restriction

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & b \end{pmatrix} \geq 0$$

is, by the inductive hypothesis, $\eta_{2k-2}^{-1/2} \eta_{2k-3}^{-1/2}$; the corresponding value for $c$ is $\eta_{2k-1}^{-1/2} \eta_{2k-2}^{-1/2}$. Using this, and using (5) for $a_{23}$ and $a_{32}$, the condition on $a_{33}$ equivalent to $A \geq 0$ becomes

$$(11) \quad a_{33} \geq \xi_k^* (\eta_{2k})^{1/2} \eta_{2k-1} (\eta_{2k})^{1/2} \xi_k.$$

By similar reasoning, the condition on $a_{33}$ equivalent to $\bar{A} \geq 0$ becomes

$$(12) \quad \bar{a}_{33} \geq \xi_k^* (\eta_{2k})^{1/2} \eta_{2k-1} (\eta_{2k})^{1/2} \xi_k.$$

Assuming (5) as in analogous situations above that $a_{33} = \xi_k^* a_{33}$, we see that the only effect of $\xi_k \neq 1$ is to oblige us to deal with $\xi_k^* a_{33} \xi_k$ in the rest of this paragraph, so assume $\xi_k = 1$. Now $a_{33} - \eta_{2k}^{-1/2} \eta_{2k-1} \eta_{2k}^{-1/2}$ is an operator which whenever (11) holds is $\geq 0$ and which whenever (12) holds is $\leq \eta_{2k}$. Hence it may be written in the form $\eta_{2k}^{-1/2} \eta_{2k+1} \eta_{2k}^{-1/2}$ for some $\eta_{2k+1}$ with $0 \leq \eta_{2k+1} \leq 1$. This expresses $a_{33} = A_{kk}$ in the required form (4).

The proof is complete.

Next I combine Theorems 1 and 2, in a somewhat superior formulation.

**Definition.** Let $\mathcal{K} = \mathcal{K}_0 \supseteq \mathcal{K}_1 \supseteq \mathcal{K}_2 \supseteq \cdots$ be Hilbert spaces; let $A_{ij}$ be a bounded operator on $\mathcal{K}_j$ to $\mathcal{K}_i$, for $i,j = 0, 1, 2, \cdots$. Then $A = ((A_{ij}))$ is an operator with domain in $\mathcal{K} = \sum_i \mathcal{K}_i$, where $\mathcal{K}_i$ is a replica of $\mathcal{K}_i$, defined in the natural way: denoting $x = \sum \oplus x_i$ and $Ax = \sum (Ax)_i$, then $(Ax)_i = \sum_i A_{ij}x_j$. Evidently $A$ need not thereby be everywhere defined. This is called an expression of $A$ as a pruned matrix with respect to the given descending sequence of subspaces. The usual rule for matrix multiplication of course applies to pruned matrices.

**Theorem 3.** Let $\mu_n \in \mathbb{R}$ for $n \in I$, where $I \subseteq \{0, 1, 2, \cdots \}$. The following are equivalent:

---

(5) See footnote 4.
(i) There exists a function $\Phi$ on $[0, 1]$ to $\mathbb{R}$ such that $\Phi(0^-) = 0 \leq \Phi(t)$, $\Phi(t) = 1$ for $0 \leq t_1 \leq t_2 \leq 1$, and such that

$$
\mu_n = \int_0^1 t^n \Phi(t) \, dt.
$$

(1)

(ii) For $i = 1, 2, \ldots$, there exist subspaces $\mathfrak{A}_i$, $\mathfrak{A}_{i+1} \subseteq \mathfrak{A}_i \subseteq \mathfrak{A}_0 = \mathfrak{A}$; and there exist $\eta_i \in \mathbb{R}$, with $0 \leq \eta_i \leq 1$, with the closure of the range of $\eta_{2i}$ being $\mathfrak{A}_i$, the range of $\eta_{2i+1}$ being $\mathfrak{A}_i$, and the range of $\eta_{2i-1}$ being $\mathfrak{A}_i$; such that the pruned matrix $A = (A_{ij})$ defined by

$$
A_{2i} = \frac{1}{2} (\eta_{2i} - \eta_{2i-1}) (\eta_{2i+1} - \eta_{2i})^{1/2},
$$

$$
A_{2i+1} = \frac{1}{2} (\eta_{2i+1} - \eta_{2i-1}) (\eta_{2i+1} - \eta_{2i})^{1/2},
$$

$$
A_{ij} = 0 \text{ for } |i-j| > 1,
$$

satisfies

(2)

$$
\mu_n = (A^n)_{00}.
$$

Here the convention is $\eta_0 = 0$. Necessarily $0 \leq A \leq 1$.

If $I = \{0, 1, 2, \ldots\}$, then the $\eta_i$ and the $\mu_n$ uniquely determine each other. In fact, $\eta_i$ is determined by those $\mu_p$ with $p \leq n$, and conversely, $\mu_n$ is determined by those $\eta_p$ with $p \leq n$.

No detailed proof need be given. The unsightly conditions put on the $\mathfrak{A}_i$ and the $\eta_i$ bar the introduction of inessential subspaces to $\mathfrak{A}$, and the uniqueness assertion of Neumark's theorem can be invoked\(^{(6)}\). The strong uniqueness statement of the last sentence of Theorem 3 now involves no ideas not already enlisted in the proof of Theorems 1 and 2. A full proof may be supplied by the reader immune to tedium.

Note. Since $\eta_{2i}$ is always zero on $\mathfrak{A}_i$ and $i = i$, it might be suggested that it and $\tilde{\eta}_{2i}$ be defined only on $\mathfrak{A}_i$. However, it turns out to be more natural to define $\eta_{2i} = 0$ and $\tilde{\eta}_{2i} = 1$ on $\mathfrak{A}_i$. Odd subscripts are different: $\eta_{2i+1}$ and $\tilde{\eta}_{2i+1}$ enter just as symmetrically as $A$ and $\tilde{A}$ do. We may reasonably stick to $\eta_{2i+1} = \tilde{\eta}_{2i+1} = 0$ on $\mathfrak{A}_i$.

4. The classical case. Here the Hilbert space $\mathfrak{A}$ is 1-dimensional, so members of $\mathfrak{A}$ are real numbers. In particular, $\mu_n \in [0, 1]$. In (4') and (5'), everything commutes, and $A_{i-1,i} \geq 0$. It may be that $\mathfrak{A}_i = \mathfrak{A}$ for all $i$, in which case $A$ has rows and columns indexed $0, 1, 2, \ldots$; or, from $i = k$ on, $\mathfrak{A}_i$ may be zero, in which case $A$ is $k \times k$.

Theorem 1 is more closely related than it might appear to the standard solution [9, Theorem 1.5] of the Hausdorff moment problem. One considers the mapping $M$ on polynomials defined by

$$
M(a_n t^n + \cdots + a_0) = a_n \mu_n + \cdots + a_0.
$$

The condition put on the $\mu_n$ is that $k$th order differences all be non-negative.

\(^{(6)}\) See the last paragraph of §2.
for all $k$. This is proved equivalent to the positivity of $M$ as a mapping on polynomials on $[0, 1]$. Neumark’s theorem [7] may be regarded as asserting that such $M$ is obtainable as a homomorphism to another $C^*$-algebra followed by a projection [10, §1]. But any $C^*$-algebra which is a homomorphic image of the $C^*$-algebra of continuous functions on $[0, 1]$, is generated by a single operator $A$, the image of the polynomial $t$. There must be an operator $A$ and a projection $P$ such that $\mu P = M(t^n)P = PA^nP$. As above, if $E(t)$ is the spectral resolution of $A$ and $\Phi(t) = PE(t)P$, then $\mu = \int_0^1 t^nd\Phi(t)$. Tools used in [6] extend this to the case where the $\mu_n$ are operators.

The foregoing remarks are not advanced as an improved or even an alternate proof of the Hausdorff moment theorem. They avoid no difficulty of the standard proof, and they entail new ones. The aim is merely to make the relation explicit.

But now, the operator homomorphism having been introduced, the canonical form of Theorem 3 above is made available. I will apply this to the classical case.

5. Relation between the parameters and the distribution. Throughout the rest of the paper, except where otherwise stated, $\eta_n$ are numbers, $A$ is the associated matrix in the sense of Theorem 3, with numerical parameters $\eta_n$, and $\Phi(t)$ is the associated distribution.

Proposition 1. The set of points of nonconstancy of $\Phi(t)$ is exactly the spectrum of $A$.

This is obvious from Theorem 1.

Definition (cf. [12; 5, §9]). A distribution on $[0, 1]$ is of degree $m$, with $m = \frac{1}{2}, \frac{3}{2}, \ldots$, provided it is concentrated in $m + \frac{1}{2}$ distinct points $\alpha_0 < \alpha_1 < \ldots < \alpha_{m-\frac{1}{2}}$, and either $\alpha_0 = 0$ or $\alpha_{m-\frac{1}{2}} = 1$ (but not both). In the former case the degree may be written $m_*$, in the latter case $m^*$.

A distribution on $[0, 1]$ is of degree $m$, with $m = 1, 2, \ldots$, provided either it is concentrated in $m$ distinct points $\alpha_i$, $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_{m-1} < 1$, in which case the degree may be written $m^*$; or it is concentrated in $m + 1$ distinct points $0 = \alpha_0 < \alpha_1 < \ldots < \alpha_m = 1$, in which case the degree may be written $m_*$.

Proposition 2. Let the sequence $\{\eta_n\}$ terminate at $\eta_{2m}$, where $m = \frac{1}{2}, 1, 3/2, 2, \ldots$. That is, let $\eta_{\frac{1}{2}} > 0, \ldots, \eta_{2m-1} \eta_{2m-1} > 0, \eta_{2m} \eta_{2m} = 0$. If $\eta_{2m} = 0$, $\Phi$ is of degree $m_*$; if $\eta_{2m} = 1$, $\Phi$ is of degree $m^*$. Conversely, all $\Phi$ of finite degree arise in this way.

Proof. Let $k = 1, 2, \ldots$. If $\eta_{2k} = 0$, $A$ is $k \times k$ and can have at most $k$ distinct eigenvalues. Then by Proposition 1, $\Phi$ has at most $k$ points of nonconstancy. Similarly if $\eta_{2k-1} = 0$, $\eta_{2k-1} = 1$, or $\eta_{2k-2} = 1$.

Conversely, suppose the distribution $\Phi$ concentrated in exactly $k$ distinct
points. The construction of Neumark's theorem in this case\(^{(7)}\) yields an \(A\) which is \(k \times k\). By the last paragraph of Theorem 3, no different \(A\) will do. Therefore \(\{\eta_n\}\) terminates at \(\eta_{2k} = 0, \eta_{2k-1} = 0, \eta_{2k-1} = 1, \) or \(\eta_{2k-2} = 1\).

If \(\eta_{2k-1} = 0\), \(A_{k-1,k-1}\) is \(\eta_{2k-3}\eta_{2k-2}\), and, by Theorem 3, diminishing \(A_{k-1,k-1}\) while leaving other matrix elements unchanged would make \(A \geq 0\) cease to hold. But if there was a positive lower bound to the spectrum of \(A\), a sufficiently small positive multiple of any positive semidefinite matrix could be subtracted from \(A\) keeping the result \(\geq 0\). Therefore \(0\) is an eigenvalue of \(A\). Similarly, if \(\eta_{2k-1} = 1\), \(0\) and \(1\) are eigenvalues of \(A\).

Conversely, let \(\Phi\) as above have \(0\) as a point of nonconstancy. Now it may seem that the corresponding eigenspace of \(A\) could be orthogonal to the \((k-1)\)th coordinate space; suppose this, and let the \(j\)th coordinate space be the highest to which it is not orthogonal. Then reduction of \(A_{jj}\) would be inconsistent with \(A \geq 0\). It follows that either \(\eta_{2j+1} = 0\) or \(\eta_{2j+1} = 1\) (otherwise diminishing of \(\eta_{2j+1}\) would be an available way of diminishing \(A_{jj}\)). If \(j < k-1\), this is a contradiction. For \(j = k-1\), it gives the desired conclusion that either \(\eta_{2k-1} = 0\) or \(\eta_{2k-2} = 1\). Similarly, if \(\Phi\) has \(1\) as a point of nonconstancy, either \(\eta_{2k-1} = 1\) or \(\eta_{2k-2} = 1\).

The statement of Proposition 2 is exactly the expression of the facts just proved in terms of the preceding definition.

In this proof I bypassed the question of whether in general an eigenvector of \(A\) can be orthogonal to the last coordinate space. The question is worth settling, though. To begin with, no eigenvector of \(A\) can be orthogonal to the 0th coordinate space, as already remarked. But also any admissible finite-dimensional \(A\) remains admissible when the order of rows and columns is reversed. (The proof of this fact is omitted; it involves retracing some of Theorem 2, in the classical case. Of course, reversing the order of rows and columns does not merely reverse the sequence of parameters \(\eta_n\), but replaces them by an entirely new sequence.) Therefore no eigenvector of \(A\) can be orthogonal to the last coordinate space. This is used in Proposition 4 below.

**Proposition 3.** The other \(\eta_p\) being held constant, \(\mu_n\) is a strictly increasing linear function of \(\eta_n\).

**Proof.** I mean to imply by granting the existence of \(\eta_n\) that the earlier parameters do not have extreme values. The later parameters, on the other hand, are without effect on \(\mu_n\), so \(\eta_{n+1} = 0\) may be assumed.

Let \(B\) differ from \(A\) only in having \(\eta_n = 0\).

**Case I.** \(n = 2i+1\). Then \(A\) is \((i+1) \times (i+1)\) with last entry \(A_{ii} = \eta_{2i+1}\eta_{2i+2} + \eta_{2i+2}\eta_{2i+1}\). The only entry in which \(B\) differs is \(B_{ii} = \eta_{2i-1}\eta_{2i}\). Therefore \(\eta_{2i-1}\eta_{2i}\).

\(^{(7)}\) Alternative constructions exist in this simple case and even somewhat more generally, e.g., [3].
\[ \mu_{2i+1} = (A^{2i+1})_{00} = ((B + (A - B))^{2i+1})_{00} \]
\[ = (B^{2i+1})_{00} + B_{01}B_{12} \cdots B_{i-1,i}(A - B)_{i,i}B_{i,i-1} \cdots B_{21}B_{10} \]
\[ = (B^{2i+1})_{00} + \eta_{i-1}^2 \eta_{2i} \cdots \eta_{2i-1}^2 \eta_{2i} \eta_{2i+1}, \]
and only the last term involves \( \eta_{2i+1} \).

Case II. \( n = 2i \). Since \( A \) is \((i+1) \times (i+1)\) and \( B \) is \( i \times i \), adjoin to \( B \) a zero last row and column.

\[ A - B = \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & (\eta_{2i-2} \eta_{2i-1} \eta_{2i-1})_{1/2} \\
0 & (\eta_{2i-2} \eta_{2i-1} \eta_{2i})_{1/2} & \eta_{2i-1} \eta_{2i}
\end{pmatrix}, \]

where the zero initial row stands for \( i - 1 \) such, and likewise for columns.

\[ (A^{2i-1})_{01} = ((B + (A - B))^{2i-1})_{01} \]
\[ = (B^{2i-1})_{01} + B_{01}B_{12} \cdots B_{i-2,i-1}(A - B)_{i-1,i-1}(A - B)_{i,i-1}B_{i-1,i-2} \cdots B_{21}B_{10} \]
\[ = (B^{2i-1})_{01} + \eta_{i-1} \eta_{2i} \eta_{3i} \cdots \eta_{2i-3} \eta_{2i-2} \eta_{2i-1} \eta_{2i-2} \eta_{2i}, \]

and by a similar argument \( (A^{2i-1})_{00} = (B^{2i-1})_{00} \). Therefore

\[ \mu_{2i} = (A_{2i})_{00} = (B_{2i})_{00} + \eta_{i-1} \eta_{2i} \eta_{3i} \cdots \eta_{2i-3} \eta_{2i-2} \eta_{2i-1} \eta_{2i}, \]

and only the last term involves \( \eta_{2i} \).

The following proposition, suggested naturally by the last two, is much harder to prove.

Proposition 4. Let the sequence \( \{ \eta_m \} \) terminate at \( \eta_{2m} \), where \( m = 1/2, 1, 3/2, 2, \cdots \). Let \( \alpha_0 < \alpha_1 < \cdots \) be the eigenvalues of \( A \), and let \( i \) be such that \( 0 < \alpha_i < 1 \). If \( \eta_{2m} = 0 \), \( \alpha_i \) is a monotone strictly increasing function of \( \eta_{2m-1} \); if \( \eta_{2m} = 1 \), \( \alpha_i \) is a monotone strictly decreasing function of \( \eta_{2m-1} \).

Proof. Let \( k = 1, 2, \cdots \).

Case I. \( \eta_{2k+2} = 0 \). Then \( A \) is \((k+1) \times (k+1)\), and \( \eta_{2k+1} \) appears only in \( A_{kk} = \eta_{2k-1} \eta_{2k} + \eta_{2k} \eta_{2k+1} \). Therefore to increase \( \eta_{2k+1} \) is to add to \( A \) a positive semidefinite matrix having a strictly positive eigenvalue in a subspace not orthogonal to any eigenspace of \( A \) (see the remark following Proposition 2).

The conclusion follows in this case, by virtue of the following general property of matrices [8, p. 236].

Lemma 4. Let \( \alpha_0 < \alpha_1 < \cdots < \alpha_n \) be the eigenvalues of the \((n+1) \times (n+1)\) Hermitian matrix \( A \); let \( \alpha_0' < \alpha_1' < \cdots < \alpha_n' \) be the eigenvalues of \( A + B \), where \( B \geq 0 \). Then \( \alpha_i \leq \alpha_i' \), with equality if and only if the eigenvector corresponding to \( \alpha_i \) is annihilated by \( B \).

Now the remaining cases of Proposition 4 will be proved by reducing them similarly to Lemma 4.
Case II. \( \eta_{2k+1} = 0 \). Again \( A \) is \((k+1) \times (k+1)\). For any value of \( \eta_k \) it has 0 as a simple eigenvalue, so it is natural to change basis and restrict attention to the complementary \( k \)-dimensional invariant subspace.

Now in order to do this I will in a formal sense reduce the case of arbitrary \( k \) to the first case, \( \eta_k = 0 \). Keep the given \( A \), but regard it as a \( 2 \times 2 \) matrix, whose top left entry \( A_{00} \) is itself a \( k \times k \) matrix (in contrast to the former \( A_{00} \)), and whose bottom right entry \( A_{11} \) is the former \( A_{kk} \). In the terminology of Theorem 3, \( \mathcal{C}_0 \) is \( k \)-dimensional, \( \mathcal{C}_1 \) 1-dimensional. Also by \( \eta_1 \) we shall now mean the matrix \( A_{00} \); and \( A_{01} = A_{10}^* = (\eta_1 \eta_1)^{1/2} \eta_2^{1/2} \), \( A_{11} = \eta_2^{1/2} - \eta_1 \eta_2^{1/2} \). The new \( \eta_2 \) is the former \( \eta_{2k} \)—on the subspace \( \mathcal{C}_1 \), not on the whole space; \( \eta_2 \) is zero on \( \mathcal{C}_1 \), so \( \eta_2 \) does not commute with \( \eta_1 \).

So regarded, \( A \) has as its domain 2-vectors with components \( x_0 \in \mathcal{C}_0 \) and \( x_1 \in \mathcal{C}_1 \). The range consists of those of the form

\[
y = \begin{pmatrix} (\eta_1)^{1/2} x_0 \\ (\eta_2)^{1/2} \eta_1^{1/2} x_0 \end{pmatrix}, \quad x_0 \in \mathcal{C}_0.
\]

I have to consider the eigenvalues of \( A \) acting on such \( y \), that is, to consider minimax values of \( \langle Ay, y \rangle/\|y\|^2 \). This can be replaced by a minimax problem on \( \mathcal{C}_0 \). Set \( G = \eta_1 + \eta_1^{1/2} \eta_2 \eta_1^{1/2} \). Then \( \|y\|^2 \) reduces to \( \langle Gx_0, x_0 \rangle \); or \( \|y\| = \|w_0\| \) if \( w_0 = G^{1/2} x_0 \). A further computation gives \( \langle Ay, y \rangle = \langle G^2 x_0, x_0 \rangle = \langle Gw_0, w_0 \rangle \). The problem is therefore to find the dependence on \( \eta_2 \) of each eigenvalue of \( G \).

To increase \( \eta_2 \) to a new value \( \eta'_2 \) is to add to \( \eta_1 + \eta_1^{1/2} \eta_2 \eta_1^{1/2} \) a positive semi-definite matrix, hence no eigenvalue can be decreased, by Lemma 4. But can the \( i \)-th eigenvalue be unchanged? Only if the \( i \)-th eigenvector \( w_{0i} \) is annihilated by \( \eta_1^{1/2} (\eta'_2 - \eta_2) \eta_1^{1/2} \). This operator is a numerical multiple of \( \eta_1^{1/2} \eta_2 \eta_1^{1/2} \); hence \( w_{0i} \) would have to be an eigenvector of \( \eta_1 \), hence of \( \eta_1^{1/2} \); hence \( w_{0i} \) would have to be annihilated by \( \eta_1 \). Suppose this is the case. Adjoining to \( w_{0i} \) a zero last component gives an eigenvector of \( A \) which is orthogonal to the last component subspace—a contradiction (see again the remark following Proposition 3). Hence all eigenvalues are strictly increasing.

Case III. \( \eta_{2k+1} = 1 \). Apply Case II to \( \tilde{A} \).

Case IV. \( \eta_{2k+2} = 1 \). This time \( A \) is \((k+2) \times (k+2)\) and has 0 and 1 as simple eigenvalues. The proof copies Case II.

Rewrite \( A \), letting \( \eta_1 \) now be \( k \times k \) (\( \mathcal{C}_0 \) \( k \)-dimensional), letting \( \eta_2 \) be the former \( \eta_{2k} \) (on the 1-dimensional \( \mathcal{C}_1 \) only), and letting \( \eta_3 \) be the former \( \eta_{2k+1} \) (on \( \mathcal{C}_1 \)). Since \( \eta_k \rightarrow \eta_3 \),

\[
A = \begin{pmatrix}
\eta_1 & (\eta_1 \eta_1)^{1/2} (\eta_2)^{1/2} \\
(\eta_2)^{1/2} (\eta_1 \eta_1)^{1/2} & (\eta_2)^{1/2} (\eta_1 \eta_1)^{1/2} + \eta_2 \eta_3 + (\eta_2 \eta_3 \eta_3)^{1/2} \\
0 & (\eta_2 \eta_3 \eta_3)^{1/2} \\
\end{pmatrix}.
\]

The range of \( A \tilde{A} \) consists of those vectors of the form

---

(8) This expression is justified only by using the convention in the last paragraph of §3: \( \tilde{\eta}_2 \) acts on all \( \mathcal{C} \). It is nonsingular there, for \( \eta_1 < 1 \), and \( \tilde{\eta}_2 \) is 1 on \( \mathcal{C}_1 \).
y = \begin{pmatrix} (\eta_1 \widetilde{\eta}_1)^{1/2} (\eta_2)^{1/2} x_0 \\ (\eta_2)^{1/2} (\eta_1 - \eta_0) (\eta_2)^{1/2} x_0 \\ - (\eta_2 \eta_3 \widetilde{\eta}_3)^{1/2} x_0 \end{pmatrix}, \quad x_0 \in \mathbb{C}_0.

I have to consider \((Ay, y)/\|y\|^2\) for such \(y\). Set \(H = \eta_2^2 \eta_1 \widetilde{\eta}_2^2 + \eta_2 \widetilde{\eta}_3\). Then\(^7\)
\[\|y\|^2 = (H \widetilde{H} x_0, x_0) = \|w_0\|^2 \text{ if } w_0 = (H \widetilde{H})^{1/2} x_0.\] A longer computation yields\(^8\)
\[(Ay, y) = (H^2 \widetilde{H} x_0, x_0) = (H w_0, w_0).\] The problem is therefore to find the dependence on \(\eta_3\) of each eigenvalue of \(H\).

To increase \(\eta_3\) to a new value \(\eta_3'\) is to subtract from \(\eta_2^2 \eta_1 \widetilde{\eta}_2^2 + \eta_2 \widetilde{\eta}_3\) a positive semidefinite matrix which is 0 on \(\mathbb{C}_0^2\). The \(i\)th eigenvalue must be strictly decreased; justifying the word "strictly" is even easier here than at the end of Case II.

From Theorem 3 and Propositions 1–4, involving the \(\eta_n\), some familiar relationships between the \(\mu_n\) and \(\Phi\) are immediate. (In the following statements, \(m = 1/2, 1, 3/2, 2, \ldots\).)

From Theorem 3 and Propositions 1 and 2—If \(\mu_1, \ldots, \mu_{2m-1}\) are the first \(2m-1\) moments of any distribution, they are the first \(2m-1\) moments of a distribution of degree at most \(m\).

From Theorem 3 and Propositions 2 and 3—Consider the set of all distributions having \(\mu_1, \ldots, \mu_{2m-1}\) as the first \(2m-1\) moments. Assume the set contains no distribution of degree less than \(m\). Then the maximum value of the \(2m\)th moment within the set is attained for just one distribution: the unique one with degree \(m^*\). Similarly the minimum is attained just for the unique distribution with degree \(m^*_n\).

From the foregoing with Proposition 4—If \(\Phi_1\) and \(\Phi_2\) are different distributions, both of degree \(m^*\) (or less), or else both of degree \(m^*_n\) (or less), and having the same first \(2m-2\) moments, then their points of nonconstancy are interlocking sets of real numbers; except that they may have 0 and-or 1 in common.

6. Relation between the parameters and associated determinants. If \(a_{ij} = \mu_{i+j}\) for \(i, j = 0, 1, \ldots, k\), then define \(\det (a_{ij}) = \Delta_{2k}\). If \(a_{ij} = \mu_{i+j+1}\) for \(i, j = 0, 1, \ldots, k\), then define \(\det (a_{ij}) = \Delta_{2k+1}\). If \(a_{ij} = \mu_{i+j+1} - \mu_{i+j+2}\) for \(i, j = 0, 1, \ldots, k-1\), then define \(\det (a_{ij}) = \Delta_{2k}\). If \(a_{ij} = \mu_{i+j} - \mu_{i+j+1}\) for \(i, j = 0, 1, \ldots, k\), then define \(\det (a_{ij}) = \Delta_{2k+1}\).

It is known (see e.g. [5, Chap. 4]) that a finite sequence of numbers \(\mu_n, \mu_0 = 1\), can be the beginning of a nonextreme Hausdorff moment sequence if and only if all the \(\Delta_n\) which can be formed from them are positive. Also it is clear from the definitions that each \(\Delta_n\) depends only on \(\mu_1, \ldots, \mu_n\); and that \(\Delta_{2n}\) is linear strictly increasing in \(\mu_n\) but \(\Delta_{2n+1}\) is linear strictly decreasing in \(\mu_n\). The facts concerning the \(\eta_n\) proved in Theorem 3 and Proposition 3 show a strong resemblance. Indeed it is a matter of elementary algebra (which I omit) to deduce from the cited facts that
\[\eta_1 \eta_2 \ldots \eta_{n-1} \eta_n = \Delta_{2n}/\Delta_{2n-2}, \quad \eta_1 \eta_2 \ldots \eta_n = \Delta_{2n}/\Delta_{2n-2},\]
which may alternatively be expressed
This is simple enough, yet I do not know how to prove it any more directly.

7. Subsidiary remarks.

Remark 1. The representation of operators in Theorem 2 is in a sense in close analogy to their representation by the spectral theorem. Consider only the classical case (i.e., the $P_i$ of Theorem 1 are 1-dimensional). The spectral theorem puts operator $A$, if it has only point spectrum, in the form of a sum of multiples of orthogonal projections. Theorem 2 puts $A$ in the form of a sum of multiples of 1-dimensional projections $Q_i$, $i = 1, 2, 3, \cdots$, where the $Q_iQ_j$ are required to be zero, not for $i \neq j$, but only for $|i - j| > 1$. I specify the $Q_i$: $Q_iP_j = 0$ unless $j = i - 1$ or $j = i$, while in the coordinate system adapted to $P_{i-1}$ and $P_i$,

\[
(\tilde{\eta}_{2i-2}\tilde{\eta}_{2i-1} + \tilde{\eta}_{2i-1}\eta_{2i})Q_i = \begin{pmatrix}
\tilde{\eta}_{2i-2}\eta_{2i-1} & (\tilde{\eta}_{2i-2}\eta_{2i-1} - \tilde{\eta}_{2i-1}\eta_{2i})^{1/2} \\
(\tilde{\eta}_{2i-2}\eta_{2i-1} - \tilde{\eta}_{2i-1}\eta_{2i})^{1/2} & \tilde{\eta}_{2i-1}\eta_{2i}
\end{pmatrix}.
\]

Then $A = \sum_i (\tilde{\eta}_{2i-2}\tilde{\eta}_{2i-1} + \tilde{\eta}_{2i-1}\eta_{2i})Q_i$.

From this point of view Theorem 2 is not in its most general form, to be sure, because of its second sentence. Removing this limitation is no problem.

Remark 2. Suppose the questions taken up in this paper are asked for distributions on some other set than $[0, 1]$. In order for the same considerations to apply immediately, that set must be a finite interval $[a, b]$. The parametrization of a Jacobi matrix $B$, $a \leq B \leq b$, deserves to be given explicitly. It is in terms of $\xi_1, \eta_2, \xi_3, \eta_4, \cdots$, with $a \leq \xi_1 \leq b$, $0 \leq \eta_i \leq 1$. These parameters may be operators in $B$, that is, this is not restricted to the classical case. Define $\xi = a + b - \xi$. Instead of (4'), (5'), we have

\begin{align*}
(4'') & \quad B_{ii} = (\tilde{\eta}_{2i})^{1/2}\tilde{\xi}_{2i-1}(\eta_{2i})^{1/2} + (\tilde{\eta}_{2i})^{1/2}\xi_{2i+1}(\tilde{\eta}_{2i})^{1/2}, \\
(5'') & \quad B_{i-1,i} = B_{i,i-1} = (\tilde{\eta}_{2i-2})^{1/2}(\xi_{2i-1}\tilde{\xi}_{2i-1} - ab)^{1/2}(\eta_{2i})^{1/2},
\end{align*}

$i, j = 0, 1, 2, \cdots$; and $B_{ij} = 0$ for $|i - j| > 1$. (Again, $\eta_0 = 0$ is the convention.) For $a = 0, b = 1$, this reduces to (4'), (5'), with the obvious notational equivalences.

Remark 3. The canonical form of Theorem 3 suggests as a byproduct still another parametrization of all Hausdorff moment sequences. Define $x_1 = \eta_1 - \tilde{\eta}_1$, and for $n > 1$,

\[
x_n = 2^{n-1}\{\eta_1\tilde{\eta}_1 \cdots \eta_{n-1}\tilde{\eta}_{n-1}\}^{1/2}(\eta_n - \tilde{\eta}_n).
\]

Alternatively, this could be written in terms of angles $\theta_n \in [0, \pi]$ defined by $\cos \theta_n = \eta_n - \tilde{\eta}_n$; the advantage would be the more "geometric" formulas $x_1 = \cos \theta_1$, and for $n > 1$, $x_n = \sin \theta_1 \cdots \sin \theta_{n-1} \cos \theta_n$.

\[\text{(9) In the unconventional terminology of Guttman [2], the $Q_i$ (or rather vectors in their respective ranges) must form a "perfect simplex."}\]
In any case, the facts are easy to prove: By Theorem 3, Hausdorff moment sequences correspond 1-1 to sequences of numbers \( \eta_n \in [0, 1] \), \( n = 1, 2, \ldots \), provided any sequence is regarded as terminating at the first \( \eta_n \) which is equal to 0 or 1. But (13) gives for each such sequence a unique sequence of real numbers \( x_n, n = 1, 2, \ldots \). It is easy to verify from the identity \( (\eta - \bar{\eta})^2 + 4\eta \bar{\eta} = 1 \), that \( \sum x_n^2 \leq 1 \). Conversely, \( x_n \) real and \( \sum x_n^2 \leq 1 \) imply that (13) may be solved successively for \( \eta_n \) until some \( \eta_n \) is 0 or 1. The unit ball (unit sphere and its interior) in real Hilbert space of countably infinite dimensionality has been put in 1-1 correspondence with Hausdorff moment sequences, hence (see e.g. [13, Theorem 6.1]) with distribution functions on \([0, 1]\).

\( \eta_n = 0 \) or 1 if and only if \( x_{n+1} = x_{n+2} = \ldots = 0 \). Such points lie on, but are far from exhausting, the unit sphere in the space. The corresponding distributions are exactly those concentrated in a finite number of points.

The distribution \( \Phi \) corresponding to the center of the sphere may be shown (using [5, §25]) to be given by

\[
d\Phi(t) = \frac{dt}{\pi(t(1-t))^{1/2}}.
\]

**Bibliography**


COLUMBIA UNIVERSITY,
NEW YORK, N. Y.