1. Introduction. Let \((X, S, \mu)\) be a probability measure space. Here \(X\) is a set of points \(x\), \(S\) is a \(\sigma\)-algebra of subsets of \(X\), and \(\mu\) is a \(\sigma\)-additive measure on \(S\) with \(\mu(X) = 1\). Let \(V = L^2(X, S, \mu)\) be the real Hilbert space of \(S\)-measurable functions \(f(x)\) with \(\int_X f^2 d\mu < \infty\).

For \(f, g \in V\), we write \((f, g) = \int_X f \cdot g d\mu\) where \((f \cdot g)(x) = f(x) \cdot g(x)\), and \(||f|| = (f, f)^{1/2}\). Convergence in \(V\) is defined as usual in the norm topology, that is to say, \(\lim_{n \to \infty} f_n = f\) means \(\lim_{n \to \infty} ||f_n - f|| = 0\). If \(f\) and \(g\) are functions in \(V\) such that \(||f - g|| = 0\), they are regarded as identical and we write \(f = g\). \(f \leq g\) means that \(\mu\{x : f(x) < g(x)\} = 0\). For any real \(\alpha\), the function which is equal to \(\alpha\) for every \(x\) is also denoted by \(\alpha\). A function \(f\) in \(V\) is said to be bounded if there exist \(\alpha\) and \(\beta\) such that \(\alpha \leq f \leq \beta\).

A function on \(X\) which takes only the values 0 and 1 is called a characteristic function. For any set \(A \subset X\), \(\chi_A\) denotes the characteristic function which equals 1 on \(A\) and vanishes on \(X - A\). If \(S_1\) and \(S_2\) are subclasses of \(S\), we write \(S_1 \subset S_2\) if corresponding to each set \(A \in S_1\) there exists a \(B \in S_2\) such that \(\chi_B = \chi_A\); we write \(S_1 = S_2\) if \(S_1 \subset S_2\) and \(S_2 \subset S_1\).

Let \(W\) be a subspace \((\equiv\) closed linear manifold) in \(V\). \(W\) is algebraic if it contains the function 1 (and therefore every constant function) and if, for any two bounded functions \(f\) and \(g\) in \(W\), the function \(f \cdot g\) is also in \(W\). \(W\) is bounded (so to speak) if the set of bounded functions in \(W\) is dense in \(W\). \(W\) is measurable if there exists a (necessarily unique) \(\sigma\)-subalgebra of \(S\), \(S^*\) say, such that \(W = L^2(X, S^*, \mu)\), that is to say, \(W\) is the set of all \(S^*\)-measurable functions in \(V\).

Let \(T\) be the (orthogonal) projection to the subspace \(W\). \(T\) is constant-preserving if \(T\alpha = \alpha\) for every constant function \(\alpha\). \(T\) is positive if \(f \geq 0\) implies \(Tf \geq 0\).

Let \(S^0\) be the smallest \(\sigma\)-algebra of sets of \(X\) such that each \(f\) in \(W\) is an \(S^0\)-measurable function. Let \(S_0\) be the class of all sets \(A \subset X\) such that \(\chi_A\) is in \(W\). While \(S_0\) is not necessarily a \(\sigma\)-algebra, it is a nonempty class of sets, with \(S_0 \subset S^0 \subset S\).

The main result of this note can then be stated as follows.

**Theorem.** The following propositions are equivalent:

Received by the editors November 18, 1954.

\(^1\) This work was supported in part by the Office of Naval Research under contract N6onr-271, T.O. XI.
(i) \( T \) is constant-preserving and positive.
(ii) \( W \) is algebraic and bounded.
(iii) \( S_0 = S^0 \).
(iv) \( W = \mathcal{L}_2(X, S^0, \mu) \).

The proof of the theorem is given in the next section.
By the equivalence of (ii) and (iv) we have

**Corollary 1.** A subspace is algebraic and bounded if and only if it is measurable.

This result is essentially about the set \( \mathcal{L}_\infty \) of bounded functions in \( V \), and it may be worthwhile to state it entirely in terms of \( \mathcal{L}_\infty \) as follows. Regard \( \mathcal{L}_\infty \) as a linear algebra under pointwise multiplication of functions, and for the moment suppose closure in \( \mathcal{L}_\infty \) to mean closure in the usual \( \mathcal{L}_\infty \) topology. Then Corollary 1 can be shown to be equivalent to the proposition that a closed linear subalgebra of \( \mathcal{L}_\infty \) contains 1 if and only if it is a measurable algebra, that is to say, it is the closure of the linear algebra generated by a set of characteristic functions including 1.

It is easy to show (cf. the last paragraph of the following section) that if \( S^* \) is a \( \sigma \)-subalgebra of \( S \) then the projection to \( \mathcal{L}_2(X, S^*, \mu) \) is exactly the conditional expectation operator relative to \( S^* \). In other words, a transformation \( T \) is a conditional expectation if and only if \( T \) is the projection to a measurable subspace.\(^2\) In consequence, the equivalence of (i) and (iv) yields the following characterization of conditional expectation.

**Corollary 2.** A transformation \( T \) on \( V \) into itself is a conditional expectation if and only if \( T \) is linear, idempotent, self-adjoint, constant-preserving, and positive.

This result is closely related to a characterization of conditional expectation operators on \( \mathcal{L}_1 \) which was obtained by Moy [1, Theorem 2.2]. An essential part of the proof which follows is based on an argument of Moy.

2. **Proof of the theorem.** Since \( T \) is a projection, we have

\[
T(\alpha f + \beta g) = \alpha T(f) + \beta T(g),
\]

\[
T^2 f = Tf,
\]

and

\[
(Tf, g) = (f, Tg)
\]

\(^2\) This fact, which motivated the work presented here, was pointed out to the writer by L. J. Savage.
for all $\alpha, \beta, f$ and $g$; moreover, since $T$ is the projection to $W$,

\[(4) \quad W = \{ Tf : f \in V \},\]

and also

\[(5) \quad W = \{ f : Tf = f, f \in V \} .\]

(Cf., e.g., [2].)

We shall prove the theorem by showing that (i)$\rightarrow$(ii)$\rightarrow$(iii)$\rightarrow$(iv)$\rightarrow$(i).

Suppose then that (i) holds, that is,

\[(6) \quad T\alpha = \alpha \quad \text{for every } \alpha \quad \text{and} \quad \text{(7)} \quad Tf \geq 0 \quad \text{whenever } f \geq 0.\]

It follows from (1) that (6) and (7) are equivalent to

\[(8) \quad \alpha \leq Tf \leq \beta \quad \text{whenever } \alpha \leq f \leq \beta \quad \text{and} \quad \text{(9)} \quad Tf \geq Tg \quad \text{whenever } f \geq g.\]

Consider a fixed $g$ in $W$. Define $f_n(x) = g(x)$ if $|g(x)| \leq n$ and $=0$ otherwise, for $n = 1, 2, \ldots \ldots$. Clearly, $\{f_n\}$ is a sequence in $V$ such that $\lim_{n \to \infty} f_n = g$. Set $g_n = Tf_n$. Then, for each $n$, $g_n \in W$ by (4), and $-n \leq g_n \leq n$ by (8). Since $\|Tf\| \leq \|f\|$ for all $f$ (by (3)), and since $Tg = g$ by (5), we have $\|g_n - g\| = \|Tf_n - Tg\| = \|T(f_n - g)\| \leq \|f_n - g\|$, so that $\lim_{n \to \infty} g_n = g$. Since $g$ is arbitrary, $W$ is thus shown to be bounded.

According to (5) and (6), $W$ does contain every constant function. It remains therefore to show that $f \in W$, $g \in W$ implies $f \cdot g \in W$, provided that $f$ and $g$ are bounded. It will suffice to show that

\[(10) \quad f \in W \text{ implies } f^2 \in W \text{ provided that } f \text{ is bounded.}\]

For, if $f$ and $g$ are bounded functions in $W$ then $f + g$ is such a function also, and it will follow from (10) that $f^2$, $g^2$, and $(f + g)^2$ are in $W$; consequently $f \cdot g = (f + g)^2/2 - f^2/2 - g^2/2$ is in $W$.

We proceed to establish (10). Choose and fix a bounded $f \in W$, and define $g = Tf^2 - f^2$. In view of (5), we have to show that $g = 0$.

Consider the parabola $v = u^2$ in the $uv$-plane, and for any real $r$ let $v = a, u + b$, be the tangent to the parabola at the point $u = r$, $v = r^2$ ($a_r = 2r$, $b_r = -r^2$). Let $R$ be a countable everywhere dense set (e.g. the rational points) of the real line. Then, for each $u$, $\infty < u < \infty$, $u^2 \geq a, u + b$, for every $r$, and $u^2 = \sup_{r \in R} \{ a_r u + b_r \}$. 

We have $f^2 \geq a rf + b_r$ for each fixed $r$. Hence, for each $r$, $Tf^2 \geq T(a rf + b_r) = a rf + b_r$, by (9), (1), (5), and (6). It follows that

$$Tf^2 \geq \sup_{r \in R} \{a rf + b_r\} = f^2.$$ 

Thus $g \geq 0$. We observe next that, for any $h \in V$,

$$\int_X hd\mu = (h, 1)$$

(11)

$$= (h, T 1) \text{ by (6)}$$

$$= (Th, 1) \text{ by (3)}$$

$$= \int_X Thd\mu.$$ 

Since $Tg = T^2f^2 - Tf^2 = Tf^2 - Tf^2 = 0$ by (1) and (2), it follows from (11) that $\int_X g d\mu = 0$, and hence $g = 0$. This completes the proof that (i) $\rightarrow$ (ii).

Suppose now that (ii) holds. We shall show first that $S_0$ is a $\sigma$-algebra. Since $W$ contains the function 1, $\chi_A \in W$ implies $\chi_{X-A} = 1 - \chi_A \in W$; since $W$ is closed under pointwise multiplication of functions, $\chi_A \in W, \chi_B \in W$ implies $\chi_{A \cap B} = \chi_A \cdot \chi_B \in W$; and since $W$ is a closed subset of $V$, if $f_1, f_2, \ldots$ is a sequence of characteristic functions in $W$ such that $f_r \cdot f_s = 0$ for $r \neq s$, $g = \sum f_r$ is a characteristic function in $W$. By referring to the definition of $S_0$, we see that the class $S_0$ is closed under complementation, under intersection, and under countable union of disjoint sets; $S_0$ is therefore a $\sigma$-algebra. We proceed to show that $S_0 \subset S_0$; this will establish (iii), since $S_0 \subset S_0$ in any case.

Let $f$ be a bounded function in $W$, and suppose that $\alpha \leq f \leq \beta$. Let $I$ denote the interval $[\alpha, \beta]$, and let $M$ be the class of Borel measurable sets of $I$. Define $\nu(C) = \mu(f^{-1}(C))$ for $C \in M$. Then $\nu$ is a probability measure on $M$. It is well known (cf., e.g., [3]) that the set of polynomial functions on $I$ is dense in $X(f^* M, \nu)$. Let us consider a fixed $C \in M$. There exists a sequence $\{p_n\}$ of polynomials on $I$ such that $\lim_{n \to \infty} p_n = \chi_C$. Writing $A = f^{-1}(C)$, we have $\|p_n - \chi_A\| = \|p_n(f) - \chi_A\|$ for every $n$, so that $\lim_{n \to \infty} p_n(f) = \chi_A$. Since $W$ is algebraic, and $f$ is a bounded function in $W$, we have $p_n(f) \in W$ for each $n$; hence, since $W$ is closed, $\chi_A \in W$, that is to say, $A = f^{-1}(C) \in S_0$. Since $C$ is arbitrary, $f$ is $S_0$-measurable.

We have therefore shown that every bounded $f \in W$ is an $S_0$-measurable function. Since $W$ is bounded, it follows that every $f \in W$ is $S_0$-measurable. Consequently $S_0 \subset S_0$, by the definition of $S_0$. This completes the proof that (ii) $\rightarrow$ (iii).
Suppose next that $S_0$ is a $\sigma$-algebra, which is less than supposing that (iii) holds. Consider an $f \in L_2(X, S_0, \mu)$. There exists a sequence $\{f_n\}$ such that each $f_n$ is of the form $\sum_{i=1}^{k} \alpha_i g_i$ where $g_1, g_2, \ldots, g_k$ are $S_0$-measurable characteristic functions, and also such that $\lim_{n \to \infty} f_n = f$. It follows from the definition of $S_0$ that $\{f_n\}$ is a sequence in $W$; hence $f \in W$, since $W$ is closed. Since $f$ is arbitrary, we have $L_2(X, S_0, \mu) \subseteq W$. On the other hand, $f \in W$ implies (by the definition of $S^0$) that $f$ is $S^0$-measurable and therefore in $L_2(X, S^0, \mu)$, so that $W \subseteq L_2(X, S^0, \mu)$. Thus $L_2(X, S_0, \mu) \subseteq W \subseteq L_2(X, S^0, \mu)$, provided only that $S_0$ is a $\sigma$-algebra; if in fact $S_0 = S^0$ the three subspaces must be identical. This proves, in particular, that (iii) $\rightarrow$ (iv).

Suppose, finally, that (iv) holds. Let $U$ be the conditional expectation operator relative to $S^0$, that is to say, for each $f$ in $L_2(X, S, \mu)$ let $Uf$ be the unique function in $L_2(X, S^0, \mu)$ such that $\int_A f d\mu = \int_A Uf d\mu$ for all $A \in S^0$. The existence of such a function follows easily from the Radon-Nikodym theorem (cf., e.g., [4]). Consider an $f \in V$. Since $Tf \in W$ by (4), the hypothesis (iv) implies that $Tf \in L_2(X, S^0, \mu)$.

Also, for any $A \in S^0$ we have

$$\int_A f d\mu = (f, \chi_A)$$

$$= (Tf, \chi_A) \text{ by (iv) and (5)}$$

$$= (Tf, \chi_A) \text{ by (3)}$$

$$= \int_A Tf d\mu.$$  

It follows hence by the uniqueness of conditional expectation that $Tf$ differs from $Uf$ on a set of $\mu$-measure zero. Since $f$ is arbitrary, we conclude that $T = U$ on $V$. (It follows incidentally that $f \in V$ implies $Uf \in V$, $\|Uf\| \leq \|f\|$). As is well known, $U$ is constant-preserving and positive, so that (i) holds. This completes the proof of the theorem.

3. Applications to a finite-dimensional function space. By way of an example, let $X = \{1, 2, \ldots, n\}$, $S$ = the class of all sets of $X$, and $\mu$ = the uniform distribution on $X$, that is to say, $\mu(\{x\}) = 1/n$ for $x = 1, 2, \ldots, n$. In this case $V$ is the real vector space of all real-valued functions on $X$, every $f \in V$ is bounded, and every linear manifold in $V$ is closed.

Let $W$ be a linear manifold in $V$. By definition, $W$ is algebraic if it contains every constant function, and if $W$ is closed under pointwise
multiplication of functions. It follows from Corollary 1 that a \( k \)-dimensional linear manifold \( W \) is algebraic if and only if there exists a partition of \( X \) into \( k \) mutually exclusive nonempty sets, \( X = \bigcup_{r=1}^{k} X_r \), say, such that \( W \) is the set of all functions of the form \( f(x) = \alpha_r \) for \( x \in X_r \) \((r = 1, 2, \ldots, k)\). The verification is omitted.

For each \( i = 1, 2, \ldots, n \) let \( f_i(x) = 1 \) for \( x = i \) and \( f_i(x) = 0 \) for \( x \neq i \). Then \( \{f_1, f_2, \ldots, f_n\} \) is a basis in \( V \). Let \( \{t_{ij}\} \) be an \( n \times n \) symmetric probability matrix (i.e. \( t_{ij} = t_{ji}, t_{ij} \geq 0, \sum_{j=1}^{n} t_{ij} = 1 \)), and for \( f = \sum_{i=1}^{n} \alpha_i f_i \) let \( Tf = \sum_{i=1}^{n} \alpha_i T f_i \), where \( T f_i = \sum_{j=1}^{n} t_{ij} f_j \). Then the transformation \( T \) is linear, self-adjoint, constant-preserving, and positive.

Clearly, \( T \) is idempotent if and only if the matrix \( \{t_{ij}\} \) is idempotent (i.e. \( \sum_{i=1}^{n} t_{ij} t_{ij} = t_{ij} \)). Moreover, it follows easily from the definition of a conditional probability that \( T \) is a conditional expectation, with a \( k \)-dimensional subspace as its range, if and only if there exists a partition of \( n \) into \( k \) positive integers, \( n = n_1 + n_2 + \cdots + n_k \) say, such that (except for a permutation of rows and columns) \( \{t_{ij}\} \) is the matrix

\[
M_k(n_1, n_2, \ldots, n_k) = \begin{bmatrix}
N_1 & 0 & \cdots & 0 \\
0 & N_2 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & N_k
\end{bmatrix},
\]

where \( N_r \) is the \( n_r \times n_r \) matrix with each element equal to \( 1/n_r \).

It now follows from Corollary 2 that a symmetric \( n \times n \) probability matrix of rank \( k \) is idempotent if and only if there exists a partition of \( n \) into \( k \) positive integers \( n_1, n_2, \ldots, n_k \) such that the matrix is a permutation of the rows and columns of \( M_k(n_1, n_2, \ldots, n_k) \).

References