

Inference from Inadequate and Inaccurate Data, III*

George Backus

UNIVERSITY OF CALIFORNIA, LA JOLLA, CALIF. 92037

Communicated June 12, 1970

Abstract. Having measured D numerical properties of a physical object E which requires many more than D parameters for its complete description, we want to estimate P other numerical properties of E . Continuing the discussion in papers I¹ and II,² the present paper gives estimates when we believe it likely that we can guess an upper bound M on the Hilbert norm not of \mathbf{h}_E , the model representing E in some Hilbert space, but of the orthogonal projection of \mathbf{h}_E onto a sufficiently large subspace. In addition, the present paper simplifies the notation of I, and makes explicit the application of Bayesian subjective probability when there are errors in the data and we want to find the joint probability distribution of more than one prediction.

1. Notation. Old German capitals will stand for closed subspaces of real Hilbert spaces, a capital script letter will denote the orthogonal projection operator³ from the whole Hilbert space onto the subspace named by the same letter in old German, and bold-face lower-case italics will denote vectors in Hilbert spaces. The inner product of \mathbf{f} and \mathbf{g} will be written $\langle \mathbf{f}, \mathbf{g} \rangle$ and also $\mathbf{f} \cdot \mathbf{g}$; $\|\mathbf{f}\|$ will denote $\langle \mathbf{f}, \mathbf{f} \rangle^{1/2}$. If $V: \mathfrak{H} \rightarrow \mathfrak{H}$ is a bounded linear operator sending the Hilbert space \mathfrak{H} into itself, $V \cdot \mathbf{f}$ will denote the vector obtained by applying V to \mathbf{f} . If V^* is the adjoint³ of V , then $V^* \cdot \mathbf{f}$ will sometimes be written $\mathbf{f} \cdot V$, so that $(\mathbf{f} \cdot V) \cdot \mathbf{g} = \mathbf{f} \cdot (V \cdot \mathbf{g})$. If $W: \mathfrak{H} \rightarrow \mathfrak{H}$ is another bounded linear operator then $V \cdot W: \mathfrak{H} \rightarrow \mathfrak{H}$ will be defined by requiring $(V \cdot W) \cdot \mathbf{f} = V \cdot (W \cdot \mathbf{f})$ for all \mathbf{f} in \mathfrak{H} . If S is any subset of \mathfrak{H} , $V \cdot S$ will denote the set of all vectors in \mathfrak{H} which have the form $V \cdot \mathbf{s}$ for some \mathbf{s} in S . The set whose members are $\mathbf{f}_1, \dots, \mathbf{f}_q$ will be written $\{\mathbf{f}_1, \dots, \mathbf{f}_q\}$. If $P(\mathbf{h})$ is any statement about the arbitrary vector \mathbf{h} in \mathfrak{H} then $\{\mathbf{h}: P(\mathbf{h})\}$ will denote the set of all \mathbf{h} in \mathfrak{H} for which $P(\mathbf{h})$ is true.

If \mathfrak{R} and \mathfrak{S} are closed subspaces of Hilbert space \mathfrak{H} and $\mathfrak{S} \subseteq \mathfrak{R}$ then $\mathfrak{R} \ominus \mathfrak{S}$ will denote the set of all vectors in \mathfrak{R} which are orthogonal to every vector in \mathfrak{S} . The space $\mathfrak{H} \ominus \mathfrak{R}$ will be written \mathfrak{R}^\perp , and $\mathfrak{R}: \mathfrak{H} \rightarrow \mathfrak{R}$ and $\mathfrak{R}^\perp: \mathfrak{H} \rightarrow \mathfrak{R}^\perp$ will denote the orthogonal projection operators from \mathfrak{H} onto \mathfrak{R} and \mathfrak{R}^\perp . The identity operator on \mathfrak{H} will be written \mathfrak{I} .

If $V: \mathfrak{H} \rightarrow \mathfrak{H}$ is a bounded linear operator and \mathfrak{R} is a closed subspace of \mathfrak{H} , then $V|_{\mathfrak{R}}$ will denote the operator $\tilde{V}: \mathfrak{R} \rightarrow \mathfrak{H}$ defined on \mathfrak{R} by requiring that $\tilde{V} \cdot \mathbf{r} = V \cdot \mathbf{r}$ for every \mathbf{r} in \mathfrak{R} . If $V \cdot \mathfrak{R} \subseteq \mathfrak{R}$, and $V \cdot \mathfrak{R}^\perp = \{\mathbf{o}\}$, and $(V|_{\mathfrak{R}})^{-1}: \mathfrak{R} \rightarrow \mathfrak{R}$ exists and is bounded, we will say that V lives on \mathfrak{R} , and we will define $V^{\text{inv}}: \mathfrak{H} \rightarrow \mathfrak{H}$ as $V^{\text{inv}} = (V|_{\mathfrak{R}})^{-1} \cdot \mathfrak{R}$. Then V^{inv} is the generalized inverse of V defined by Moore⁴ and Penrose,⁵ and is the only linear mapping $U: \mathfrak{H} \rightarrow \mathfrak{H}$ such that $U \cdot V = V \cdot U = \mathfrak{R}$ and $U \cdot \mathfrak{R}^\perp = \mathbf{o}$. We note that V^* and V^{inv} also

live on \mathfrak{H} and $(V^{\text{inv}})^{\text{inv}} = V$, $(V^*)^{\text{inv}} = (V^{\text{inv}})^*$. Also, if V and W both live on \mathfrak{H} then so does $V \cdot W$, and $(V \cdot W)^{\text{inv}} = W^{\text{inv}} \cdot V^{\text{inv}}$.

If \mathfrak{S} and \mathfrak{T} are two subspaces of \mathfrak{H} , $\mathfrak{S} + \mathfrak{T}$ will denote the subspace consisting of all vectors $s + t$ with s in \mathfrak{S} and t in \mathfrak{T} . If \mathfrak{S} and \mathfrak{T} are mutually orthogonal, $\mathfrak{S} + \mathfrak{T}$ will be written $\mathfrak{S} \oplus \mathfrak{T}$. The angle between \mathfrak{S} and \mathfrak{T} , $\angle(\mathfrak{S}, \mathfrak{T})$, will be defined as the infimum of the positive angles between pairs of vectors (s, t) with s in \mathfrak{S} and t in \mathfrak{T} . The angle between o and any vector is taken to be $\pi/2$. Clearly if $\mathfrak{S} \cap \mathfrak{T} \neq \{o\}$ then $\angle(\mathfrak{S}, \mathfrak{T}) = 0$; the converse is false.⁶

2. Oblique Subspaces. If \mathfrak{S} and \mathfrak{T} are subspaces of Hilbert space \mathfrak{H} such that $\mathfrak{S} \cap \mathfrak{T} = \{o\}$, we can define two linear operators, $\mathfrak{S}_\mathfrak{T}: \mathfrak{S} + \mathfrak{T} \rightarrow \mathfrak{S}$ and $\mathfrak{T}_\mathfrak{S}: \mathfrak{S} + \mathfrak{T} \rightarrow \mathfrak{T}$, by requiring that for any s in \mathfrak{S} and any t in \mathfrak{T} we have $\mathfrak{S}_\mathfrak{T}(s + t) = s$ and $\mathfrak{T}_\mathfrak{S}(s + t) = t$. We call $\mathfrak{S}_\mathfrak{T}$ the projection of $\mathfrak{S} + \mathfrak{T}$ onto \mathfrak{S} parallel to \mathfrak{T} . Evidently $\mathfrak{S}_\mathfrak{T}|_{\mathfrak{S}}$ is the identity operator on \mathfrak{S} and $\mathfrak{S}_\mathfrak{T}|_{\mathfrak{T}}$ is the o operator on \mathfrak{T} , and $\mathfrak{S}_\mathfrak{T} + \mathfrak{T}_\mathfrak{S}$ is the identity operator on $\mathfrak{S} + \mathfrak{T}$.

Lorch⁶ has used the closed graph theorem⁷ to show that when \mathfrak{S} and \mathfrak{T} are closed and $\mathfrak{S} \cap \mathfrak{T} = \{o\}$ the following three statements are equivalent: (i) $\mathfrak{S} + \mathfrak{T}$ is closed; (ii) $\angle(\mathfrak{S}, \mathfrak{T}) > 0$; (iii) $\mathfrak{S}_\mathfrak{T}$ is bounded. He has also shown⁶ that these statements are always true if \mathfrak{S} or \mathfrak{T} is finite-dimensional.

If \mathfrak{S} and \mathfrak{T} are closed subspaces of Hilbert space \mathfrak{H} such that $\mathfrak{S} \cap \mathfrak{T} = \{o\}$ and $\mathfrak{H} = \mathfrak{S} + \mathfrak{T}$, we write $\mathfrak{H} = \mathfrak{S} \dot{+} \mathfrak{T}$. Lorch's remark shows that if $\mathfrak{H} = \mathfrak{S} + \mathfrak{T}$ then $\mathfrak{S}_\mathfrak{T}: \mathfrak{H} \rightarrow \mathfrak{S}$ and $\mathfrak{T}_\mathfrak{S}: \mathfrak{H} \rightarrow \mathfrak{T}$ are bounded. Moreover, $\mathfrak{S}_\mathfrak{T} + \mathfrak{T}_\mathfrak{S} = \mathfrak{I}$, $\mathfrak{S}_\mathfrak{T}^2 = \mathfrak{S}_\mathfrak{T}$, and $\mathfrak{S}_\mathfrak{T} \cdot \mathfrak{T}_\mathfrak{S} = o$.

We also have

$$\mathfrak{S}_\mathfrak{T} \cdot \mathfrak{S} = \mathfrak{S}, \quad \mathfrak{S} \cdot \mathfrak{S}_\mathfrak{T} = \mathfrak{S}_\mathfrak{T}, \quad (1a, b)$$

$$\mathfrak{S}_\mathfrak{T} \cdot \mathfrak{T} = o, \quad \mathfrak{T} \cdot \mathfrak{S}_\mathfrak{T} = \mathfrak{T} - \mathfrak{T}_\mathfrak{S}, \quad (2a, b)$$

$$\mathfrak{S}_\mathfrak{T} \cdot \mathfrak{S}^\perp = \mathfrak{S}_\mathfrak{T} - \mathfrak{S}, \quad \mathfrak{S}^\perp \cdot \mathfrak{S}_\mathfrak{T} = o, \quad (3a, b)$$

$$\mathfrak{S}_\mathfrak{T} \cdot \mathfrak{T}^\perp = \mathfrak{S}_\mathfrak{T}, \quad \mathfrak{T}^\perp \cdot \mathfrak{S}_\mathfrak{T} = \mathfrak{T}^\perp. \quad (4a, b)$$

Formulae (1), (2a) and (3b) are obvious and the others follow from them and the identities $\mathfrak{S} + \mathfrak{S}^\perp = \mathfrak{T} + \mathfrak{T}^\perp = \mathfrak{S}_\mathfrak{T} + \mathfrak{T}_\mathfrak{S} = \mathfrak{H}$.

If $\mathfrak{H} = \mathfrak{S} \dot{+} \mathfrak{T}$, a simple calculation which we omit shows that $\mathfrak{H} = \mathfrak{S}^\perp \dot{+} \mathfrak{T}^\perp$ and $\mathfrak{S}_\mathfrak{T}^* = \mathfrak{T}_\mathfrak{S}^\perp$.

Now suppose that $\mathfrak{H} = \mathfrak{S} \dot{+} \mathfrak{T}$, that $A: \mathfrak{H} \rightarrow \mathfrak{S}$ lives on \mathfrak{S} , that $B: \mathfrak{H} \rightarrow \mathfrak{T}$ lives on \mathfrak{T} , and that $Q: \mathfrak{H} \rightarrow \mathfrak{H}$ is linear. Then an easy calculation, omitted here, shows that $U = A + \mathfrak{S} \cdot Q \cdot \mathfrak{T} + B$ has a bounded inverse on \mathfrak{H} and that

$$U^{-1} = \mathfrak{S}_\mathfrak{T}^* \cdot A^{\text{inv}} \cdot \mathfrak{S}_\mathfrak{T} - \mathfrak{S}_\mathfrak{T}^* \cdot A^{\text{inv}} \cdot Q \cdot B^{\text{inv}} \cdot \mathfrak{T}_\mathfrak{S} + \mathfrak{T}_\mathfrak{S}^* \cdot B^{\text{inv}} \cdot \mathfrak{T}_\mathfrak{S}. \quad (5)$$

When $\mathfrak{H} = \mathfrak{S} + \mathfrak{T}$, there is a useful variational characterization of $\mathfrak{S}_\mathfrak{T}$. Let \mathfrak{U} be any closed subspace of \mathfrak{H} such that $\mathfrak{U} = (\mathfrak{S} \cap \mathfrak{U}) \oplus \mathfrak{T}$. Then for any h in \mathfrak{H} , $\mathfrak{S}_\mathfrak{T} \cdot h$ is that vector s in $\mathfrak{S} \cap (h + \mathfrak{U})$ which minimizes $\|s - h\|$. To see this, note first that $\mathfrak{S}_\mathfrak{T} \cdot h = h - \mathfrak{T}_\mathfrak{S} \cdot h$, so $\mathfrak{S}_\mathfrak{T} \cdot h$ is in $\mathfrak{S} \cap (h + \mathfrak{U})$. If s is in $\mathfrak{S} \cap (h + \mathfrak{U})$ then $s - \mathfrak{S}_\mathfrak{T} \cdot h$ is in $\mathfrak{S} \cap \mathfrak{U}$ and hence orthogonal to \mathfrak{T} . Therefore $\|s - h\|^2 = \|s - \mathfrak{S}_\mathfrak{T} \cdot h - \mathfrak{T}_\mathfrak{S} \cdot h\|^2 = \|s - \mathfrak{S}_\mathfrak{T} \cdot h\|^2 + \|\mathfrak{S}_\mathfrak{T} \cdot h - h\|^2$, and $\|s - h\| > \|\mathfrak{S}_\mathfrak{T} \cdot h - h\|$ unless $s = \mathfrak{S}_\mathfrak{T} \cdot h$.

3. Cylinder Measures. If B is a Borel⁹ subset of finite-dimensional closed subspace \mathfrak{T} of Hilbert space \mathfrak{H} , then $\mathfrak{I}^{-1}(B)$, the set of all \mathbf{h} in \mathfrak{H} such that $\mathfrak{I} \cdot \mathbf{h}$ is in B , is called a cylinder set⁸ based on B . The class $c(\mathfrak{T})$ of all cylinder sets based on Borel subsets of \mathfrak{T} is a σ -ring,⁹ and the class of all cylinder sets in \mathfrak{H} is a ring.^{8,10} A nonnegative real-valued set function μ defined on all cylinder sets in \mathfrak{H} is called a cylinder measure^{8,10} on \mathfrak{H} if $\mu(\mathfrak{H}) = 1$ and if μ is countably additive⁹ on $c(\mathfrak{T})$ for each finite-dimensional subspace \mathfrak{T} of \mathfrak{H} . From the definition, μ is finitely additive on all the cylinder sets of \mathfrak{H} , but μ need not be countably additive there.⁸ On $c(\mathfrak{T})$, μ is a probability measure,⁹ which will be written $\mu_{\mathfrak{I}}$ and called the marginal distribution of μ on \mathfrak{T} , by analogy with probability theory on finite-dimensional spaces.¹¹ The value of $\mu(\mathfrak{I}^{-1}(B))$ will be called the probability that $\mathfrak{I} \cdot \mathbf{h}$ is in B if \mathbf{h} is in \mathfrak{H} .

A function $f: \mathfrak{H} \rightarrow F$ mapping \mathfrak{H} into an arbitrary set F is called a cylinder function based on the finite-dimensional subspace \mathfrak{T} of \mathfrak{H} if, for every \mathbf{h} in \mathfrak{H} , $f(\mathbf{h}) = f(\mathfrak{I} \cdot \mathbf{h})$. If f is a real-valued, vector-valued or tensor-valued cylinder function based on \mathfrak{T} then the integral of f over \mathfrak{H} with respect to the cylinder measure μ is defined as

$$\int_{\mathfrak{H}} f(\mathbf{x}) \mu(d\mathbf{x}) = \int_{\mathfrak{T}} f(\mathbf{t}) \mu_{\mathfrak{I}}(d\mathbf{t}), \quad (6)$$

where the integral on the right is the ordinary integral⁹ over \mathfrak{T} with respect to the probability measure $\mu_{\mathfrak{I}}$.

Suppose that μ is a cylinder measure on \mathfrak{H} and that the linear functional $f: \mathfrak{H} \rightarrow R$ (R is the real line) defined by $f(\mathbf{h}) = \int_{\mathfrak{H}} \mathbf{h} \cdot \mathbf{x} \mu(d\mathbf{x})$ exists and is bounded. Then the unique vector³ $\mathbf{m}_{\mathfrak{H}}$ in \mathfrak{H} such that $f(\mathbf{h}) = \mathbf{m}_{\mathfrak{H}} \cdot \mathbf{h}$ for all \mathbf{h} in \mathfrak{H} is called the center of mass, or mean, of μ , and is written heuristically as $\mathbf{m}_{\mathfrak{H}} = \int_{\mathfrak{H}} \mathbf{x} \mu(d\mathbf{x})$. If \mathfrak{T} is any finite-dimensional subspace of \mathfrak{H} such that $\int_{\mathfrak{T}} \mathbf{t} \mu_{\mathfrak{I}}(d\mathbf{t}) = \mathbf{m}_{\mathfrak{I}}$ exists, then $\mathbf{m}_{\mathfrak{I}}$ is the ordinary mean¹¹ of $\mu_{\mathfrak{I}}$ on \mathfrak{T} . If the mean $\mathbf{m}_{\mathfrak{H}}$ of μ on \mathfrak{H} exists, then for any finite-dimensional subspace \mathfrak{T} of \mathfrak{H} , the mean $\mathbf{m}_{\mathfrak{I}}$ exists and equals $\mathfrak{I} \cdot \mathbf{m}_{\mathfrak{H}}$.

Suppose that μ is a cylinder measure on \mathfrak{H} with mean $\mathbf{m}_{\mathfrak{H}}$ and that the bilinear functional $f(\mathbf{h}, \mathbf{k}) = \int_{\mathfrak{H}} [\mathbf{h} \cdot (\mathbf{x} - \mathbf{m}_{\mathfrak{H}})][(\mathbf{x} - \mathbf{m}_{\mathfrak{H}}) \cdot \mathbf{k}] \mu(d\mathbf{x})$ exists and is bounded. It is symmetric and positive semi-definite, so there exists³ a unique self-adjoint, positive semi-definite linear operator $\mathbf{V}_{\mathfrak{H}}: \mathfrak{H} \rightarrow \mathfrak{H}$ such that for every \mathbf{h} and \mathbf{k} in \mathfrak{H} , $f(\mathbf{h}, \mathbf{k}) = \mathbf{h} \cdot \mathbf{V}_{\mathfrak{H}} \cdot \mathbf{k}$. The operator $\mathbf{V}_{\mathfrak{H}}$ is called the variance of μ , and is written heuristically as $\int_{\mathfrak{H}} (\mathbf{x} - \mathbf{m}_{\mathfrak{H}})(\mathbf{x} - \mathbf{m}_{\mathfrak{H}}) \mu(d\mathbf{x})$. If $\mathbf{V}_{\mathfrak{H}}$ exists and \mathfrak{T} is any finite dimensional subspace of \mathfrak{H} , then the variance tensor $\mathbf{V}_{\mathfrak{I}} = \int_{\mathfrak{T}} (\mathbf{x} - \mathbf{m}_{\mathfrak{I}})(\mathbf{x} - \mathbf{m}_{\mathfrak{I}}) \mu_{\mathfrak{I}}(d\mathbf{x})$ of $\mu_{\mathfrak{I}}$ on \mathfrak{T} exists, and $\mathbf{V}_{\mathfrak{I}} = \mathfrak{I} \cdot \mathbf{V}_{\mathfrak{H}} \cdot \mathfrak{I}$.

The characteristic function $\phi_{\mu}: \mathfrak{H} \rightarrow C$ (C is the complex plane) of a cylinder measure μ on \mathfrak{H} is defined for any \mathbf{h} in \mathfrak{H} by the equation $\phi_{\mu}(\mathbf{h}) = \int_{\mathfrak{H}} e^{i\mathbf{h} \cdot \mathbf{x}} \mu(d\mathbf{x})$.

Gettoor¹² has shown that a function $\phi: \mathfrak{H} \rightarrow C$ is the characteristic function of a cylinder measure iff (if and only if) $\phi(o) = 1$ and $\phi|_{\mathfrak{T}}$ is continuous and positive definite for each finite-dimensional subspace \mathfrak{T} of \mathfrak{H} , and that ϕ uniquely determines its cylinder measure.

Now suppose $\mathfrak{H} = \mathfrak{S} \dot{+} \mathfrak{T}$ with \mathfrak{S} and \mathfrak{T} possibly infinite dimensional. Suppose that $\mu_{\mathfrak{S}}$ and $\mu_{\mathfrak{T}}$ are cylinder measures on \mathfrak{S} and \mathfrak{T} . We claim that there is exactly one cylinder measure on \mathfrak{H} , which we call $\mu_{\mathfrak{S}+\mathfrak{T}}$, with the following property: for any cylinder sets $S \subseteq \mathfrak{S}$ and $T \subseteq \mathfrak{T}$, if $U \subseteq \mathfrak{H}$ is the cylinder set $U = (S + \mathfrak{S}^{\perp}) \cap (T + \mathfrak{T}^{\perp})$, then $\mu_{\mathfrak{S}+\mathfrak{T}}(U) = \mu_{\mathfrak{S}}(S)\mu_{\mathfrak{T}}(T)$. Uniqueness is clear, because any such cylinder measure must have the characteristic function $\phi_{\mathfrak{S}+\mathfrak{T}}(\mathbf{h}) = \phi_{\mathfrak{S}}(\mathfrak{S}_{\mathfrak{T}} \cdot \mathbf{h})\phi_{\mathfrak{T}}(\mathfrak{T}_{\mathfrak{S}} \cdot \mathbf{h})$. But if we define a function $\phi_{\mathfrak{S}+\mathfrak{T}}: \mathfrak{H} \rightarrow C$ by this equation, $\phi_{\mathfrak{S}+\mathfrak{T}}$ satisfies the hypothesis of Gettoor's theorem, and hence does come from a cylinder measure $\mu_{\mathfrak{S}+\mathfrak{T}}$ on \mathfrak{H} . It is easy to check that $\mu_{\mathfrak{S}+\mathfrak{T}}$ has the required relation to its marginal distributions $\mu_{\mathfrak{S}}$ and $\mu_{\mathfrak{T}}$. If we think of $\mu_{\mathfrak{S}+\mathfrak{T}}$, $\mu_{\mathfrak{S}}$ and $\mu_{\mathfrak{T}}$ as probability distributions (even though they may not be countably additive), then we can think of \mathbf{h} , $\mathfrak{S} \cdot \mathbf{h}$ and $\mathfrak{T} \cdot \mathbf{h}$ as random variables, and $\mu_{\mathfrak{S}+\mathfrak{T}}$ has been chosen so that $\mathfrak{S} \cdot \mathbf{h}$ and $\mathfrak{T} \cdot \mathbf{h}$ are independently distributed¹¹ on \mathfrak{H} .

Now suppose that $\mu_{\mathfrak{S}}$ and $\mu_{\mathfrak{T}}$ have means $\mathbf{m}_{\mathfrak{S}}$ and $\mathbf{m}_{\mathfrak{T}}$ and variances $\mathbf{V}_{\mathfrak{S}}$ and $\mathbf{V}_{\mathfrak{T}}$. Then for any \mathbf{h} in \mathfrak{H} ,
$$\int_{\mathfrak{H}} \mathbf{h} \cdot \mathbf{x} \mu_{\mathfrak{S}+\mathfrak{T}}(d\mathbf{x}) = \int_{\mathfrak{S}} \mathbf{h} \cdot (\mathfrak{S}_{\mathfrak{T}}^* \cdot \mathfrak{S} \cdot \mathbf{x} + \mathfrak{T}_{\mathfrak{S}}^* \cdot \mathfrak{T} \cdot \mathbf{x}) \mu_{\mathfrak{S}+\mathfrak{T}}(d\mathbf{x}).$$
 The integrand on the right is the sum of a cylinder function based on the subspace of \mathfrak{S} spanned by $\mathfrak{S}_{\mathfrak{T}} \cdot \mathbf{h}$ and a cylinder function based on the subspace of \mathfrak{T} spanned by $\mathfrak{T}_{\mathfrak{S}} \cdot \mathbf{h}$. Therefore
$$\int_{\mathfrak{H}} \mathbf{h} \cdot \mathbf{x} \mu_{\mathfrak{S}+\mathfrak{T}}(d\mathbf{x}) = \int_{\mathfrak{S}} \mathbf{h} \cdot \mathfrak{S}_{\mathfrak{T}}^* \cdot \mathfrak{S} \mu_{\mathfrak{S}}(d\mathbf{s}) + \int_{\mathfrak{T}} \mathbf{h} \cdot \mathfrak{T}_{\mathfrak{S}}^* \cdot \mathfrak{T} \mu_{\mathfrak{T}}(d\mathbf{t}).$$
 The right hand side is $\mathbf{h} \cdot \mathfrak{S}_{\mathfrak{T}}^* \cdot \mathbf{m}_{\mathfrak{S}} + \mathbf{h} \cdot \mathfrak{T}_{\mathfrak{S}}^* \cdot \mathbf{m}_{\mathfrak{T}}$, so the left-hand side is a bounded linear functional in \mathbf{h} , and $\mu_{\mathfrak{S}+\mathfrak{T}}$ has a mean, $\mathbf{m}_{\mathfrak{S}+\mathfrak{T}}$, given by

$$\mathbf{m}_{\mathfrak{S}+\mathfrak{T}} = \mathfrak{S}_{\mathfrak{T}}^* \cdot \mathbf{m}_{\mathfrak{S}} + \mathfrak{T}_{\mathfrak{S}}^* \cdot \mathbf{m}_{\mathfrak{T}}. \quad (7)$$

If for any \mathbf{h} in \mathfrak{H} we define $\mathbf{V}_{\mathfrak{S}} \cdot \mathbf{h} = \mathbf{V}_{\mathfrak{S}} \cdot \mathfrak{S} \cdot \mathbf{h}$ and $\mathbf{V}_{\mathfrak{T}} \cdot \mathbf{h} = \mathbf{V}_{\mathfrak{T}} \cdot \mathfrak{T} \cdot \mathbf{h}$ then a similar argument shows that $\mu_{\mathfrak{S}+\mathfrak{T}}$ has a variance $\mathbf{V}_{\mathfrak{S}+\mathfrak{T}}: \mathfrak{H} \rightarrow \mathfrak{H}$ and that

$$\mathbf{V}_{\mathfrak{S}+\mathfrak{T}} = \mathfrak{S}_{\mathfrak{T}}^* \cdot \mathbf{V}_{\mathfrak{S}} \cdot \mathfrak{S}_{\mathfrak{T}} + \mathfrak{T}_{\mathfrak{S}}^* \cdot \mathbf{V}_{\mathfrak{T}} \cdot \mathfrak{T}_{\mathfrak{S}}. \quad (8)$$

Comparison with (5) then shows that

$$\mathbf{V}_{\mathfrak{S}+\mathfrak{T}}^{\text{inv}} = \mathbf{V}_{\mathfrak{S}}^{\text{inv}} + \mathbf{V}_{\mathfrak{T}}^{\text{inv}}. \quad (9)$$

4. A Single Prediction from Error-free Data. As in I¹, $\mathbf{d}_1, \dots, \mathbf{d}_D$ are known vectors in the Hilbert space \mathfrak{H} of models which contains the unknown correct model $\mathbf{h}_{\mathbf{E}}$, and we suppose that $\gamma_i = \mathbf{d}_i \cdot \mathbf{h}_{\mathbf{E}}$ has been measured exactly for $i = 1, \dots, D$. Let \mathfrak{D} be the closed subspace of \mathfrak{H} spanned by $\mathbf{d}_1, \dots, \mathbf{d}_D$. Then the data are equivalent to knowledge of the vector $\mathfrak{D} \cdot \mathbf{m}_{\mathbf{E}} = \sum_{i=1}^D \gamma_i \mathbf{d}^i$, where $\mathbf{d}^1, \dots, \mathbf{d}^D$ is the basis for \mathfrak{D} dual to $\mathbf{d}_1, \dots, \mathbf{d}_D$. Using the data, we want to estimate $\mathbf{p} \cdot \mathbf{h}_{\mathbf{E}}$, where \mathbf{p} is another known vector in \mathfrak{H} . The vectors \mathbf{a} for which we can calculate $\mathbf{a} \cdot \mathbf{h}_{\mathbf{E}}$ from the data are precisely the members of \mathfrak{D} . Therefore

we might seek that vector \mathbf{a} in \mathfrak{D} which is most nearly equal to \mathbf{p} and estimate $\mathbf{p} \cdot \mathbf{h}_E$ as $\mathbf{a} \cdot \mathbf{h}_E$. The optimal \mathbf{a} then minimizes $\|\mathbf{a} - \mathbf{p}\|$ under the constraint that \mathbf{a} be in \mathfrak{D} ; thus³ $\mathbf{a} = \mathfrak{D} \cdot \mathbf{p}$, and we estimate $\mathbf{p} \cdot \mathbf{h}_E$ as $(\mathbf{p} \cdot \mathfrak{D}) \cdot \mathbf{h}_E = \mathbf{p} \cdot (\mathfrak{D} \cdot \mathbf{h}_E)$. If we think it very likely that $\|\mathbf{h}_E\| \leq M$, then from Schwarz's inequality the maximum error in our estimate is $\|\mathbf{p} - \mathfrak{D} \cdot \mathbf{p}\|M$.

In some investigations¹³ it is expedient to demand not only that \mathbf{a} be in \mathfrak{D} but also that \mathbf{a} treat n given vectors $\mathbf{h}_1, \dots, \mathbf{h}_n$ exactly; that is, $\mathbf{a} \cdot \mathbf{h}_\nu = \mathbf{p} \cdot \mathbf{h}_\nu$, $\nu = 1, \dots, n$. Then we want to minimize $\|\mathbf{a} - \mathbf{p}\|$ subject to the demand that \mathbf{a} be in $\mathfrak{D} \cap (\mathbf{p} + \mathfrak{F})$ where \mathfrak{F} is the closed subspace of \mathfrak{S} consisting of all vectors in \mathfrak{S} orthogonal to $\{\mathbf{h}_1, \dots, \mathbf{h}_n\}$. If the problem is to be soluble, $\mathfrak{D} \cap (\mathbf{p} + \mathfrak{F})$ must be nonempty, so \mathbf{p} must be in $\mathfrak{D} + \mathfrak{F}$. Therefore we lose no generality by assuming $\mathfrak{S} = \mathfrak{D} + \mathfrak{F}$. Let $\mathfrak{B} = \mathfrak{F} \ominus (\mathfrak{D} \cap \mathfrak{F})$. Then $\mathfrak{S} = \mathfrak{B} + \mathfrak{D}$ and, according to the penultimate paragraph of section 2, the optimal \mathbf{a} is $\mathfrak{D}_{\mathfrak{B}} \cdot \mathbf{p}$. Then we estimate $\mathbf{p} \cdot \mathbf{h}_E$ as $(\mathfrak{D}_{\mathfrak{B}} \cdot \mathbf{p}) \cdot \mathbf{h}_E = \mathbf{p} \cdot (\mathfrak{D}_{\mathfrak{B}}^* \cdot \mathbf{h}_E) = \mathbf{p} \cdot (\mathfrak{D}_{\mathfrak{B}}^* \cdot \mathfrak{D} \cdot \mathbf{h}_E)$. The error in the estimate is

$$|\mathbf{p} \cdot \mathbf{h}_E - (\mathfrak{D}_{\mathfrak{B}} \cdot \mathbf{p}) \cdot \mathbf{h}_E| = |\langle \mathbf{p} - \mathfrak{D}_{\mathfrak{B}} \cdot \mathbf{p}, \mathfrak{F} \cdot \mathbf{h}_E \rangle| \leq \|\mathbf{p} - \mathfrak{D}_{\mathfrak{B}} \cdot \mathbf{p}\|M$$

if we think it likely that $\|\mathfrak{F} \cdot \mathbf{h}_E\| \leq M$. The advantage of constraining \mathbf{a} to lie in $\mathfrak{D} \cap (\mathbf{p} + \mathfrak{F})$ is that then we can estimate our error from a bound merely on $\|\mathfrak{F} \cdot \mathbf{h}_E\|$ rather than on $\|\mathbf{h}_E\|$ itself.

5. A Single Prediction from Erroneous Data. In the foregoing section suppose that $\gamma_i = \mathbf{d}_i \cdot \mathbf{h}_E$, $i = 1, \dots, D$, have not been measured exactly but that the errors $\delta\gamma_i$ have a joint probability distribution with mean \mathbf{o} whose variance matrix $V_{ij} = E(\delta\gamma_i \delta\gamma_j)$ is known. (E means expected value.) Then

$$\mathbf{h}' = \sum_{i=1}^D \gamma_i \mathbf{d}^i \quad (10)$$

is only an estimate for $\mathfrak{D} \cdot \mathbf{h}_E$, and is in error by $\sum_{i=1}^D (\delta\gamma_i) \mathbf{d}^i$. The variance tensor \mathbf{V} for the distribution of \mathbf{h}' in \mathfrak{D} is

$$\mathbf{V} = E[(\mathbf{h}' - \mathfrak{D} \cdot \mathbf{h}_E)(\mathbf{h}' - \mathfrak{D} \cdot \mathbf{h}_E)] = \sum_{i,j=1}^D V_{ij} \mathbf{d}^i \mathbf{d}^j. \quad (11)$$

For any \mathbf{a} in $\mathfrak{D} \cap (\mathbf{p} + \mathfrak{F})$, if we estimate $\mathbf{p} \cdot \mathbf{h}_E$ as $\mathbf{a} \cdot \mathbf{h}'$ we commit an error $(\mathbf{a} - \mathbf{p}) \cdot \mathbf{h}_E + \mathbf{a} \cdot (\mathbf{h}' - \mathfrak{D} \cdot \mathbf{h}_E)$. The expected value of the square of this error is $|(\mathbf{a} - \mathbf{p}) \cdot \mathfrak{F} \cdot \mathbf{h}_E|^2 + \mathbf{a} \cdot \mathbf{V} \cdot \mathbf{a}$. If we think it likely that $\|\mathfrak{F} \cdot \mathbf{h}_E\| \leq M$, then the maximum value which this expected square error can have is $\epsilon(\mathbf{a})^2 = M^2 \|\mathbf{a} - \mathbf{p}\|^2 + \mathbf{a} \cdot \mathbf{V} \cdot \mathbf{a}$. In this situation, we want to choose \mathbf{a} in $\mathfrak{D} \cap (\mathbf{p} + \mathfrak{F})$ so as to minimize not $\|\mathbf{a} - \mathbf{p}\|$ but $\epsilon(\mathbf{a})$. At the optimum \mathbf{a} , $\epsilon(\mathbf{a})$ must be stationary under all variations $\delta\mathbf{a}$ in $\mathfrak{D} \cap \mathfrak{F}$. This enables us to show that if an optimal \mathbf{a} exists it must be

$$\mathbf{a}_0 = \mathfrak{D}_{\mathfrak{B}} \cdot \mathbf{p} - [(\mathfrak{D} \wedge \mathfrak{F}) \cdot (M^2 \mathfrak{D} + \mathbf{V}) \cdot (\mathfrak{D} \wedge \mathfrak{F})]^{inv} \cdot \mathbf{V} \cdot \mathfrak{D}_{\mathfrak{B}} \cdot \mathbf{p} \quad (12)$$

where $\mathfrak{D} \wedge \mathfrak{F}$ is the orthogonal projection operator from \mathfrak{S} onto $\mathfrak{D} \cap \mathfrak{F}$. If \mathbf{a} is in $\mathfrak{D} \cap (\mathbf{p} + \mathfrak{F})$, a simple calculation shows that $\epsilon(\mathbf{a})^2 = \epsilon(\mathbf{a}_0)^2 + M^2 \|\mathbf{a} - \mathbf{a}_0\|^2 + (\mathbf{a} - \mathbf{a}_0) \cdot \mathbf{V} \cdot (\mathbf{a} - \mathbf{a}_0)$. Since \mathbf{V} is positive definite, the stationary point \mathbf{a}_0 is indeed a minimum for $\epsilon(\mathbf{a})$ when \mathbf{a} is in $\mathfrak{D} \cap (\mathbf{p} + \mathfrak{F})$.

We estimate $\mathbf{p} \cdot \mathbf{h}_E$ as $\mathbf{a}_0 \cdot \mathbf{h}$, which is $\mathbf{p} \cdot \tilde{\mathbf{h}}_{\mathfrak{D}+\mathfrak{F}}$, where

$$\tilde{\mathbf{h}}_{\mathfrak{D}+\mathfrak{F}} = \mathfrak{D}_{\mathfrak{G}}^* \cdot \mathbf{h}' - \mathfrak{D}_{\mathfrak{G}}^* \cdot \mathbf{V} \cdot [(\mathfrak{D} \wedge \mathfrak{F}) \cdot (M^2 \mathfrak{D} + \mathbf{V}) \cdot (\mathfrak{D} \wedge \mathfrak{F})]^{inv} \cdot \mathbf{h}'. \quad (13)$$

The calculations by Backus and Gilbert¹⁴ of tradeoff curves between absolute error $\mathbf{a} \cdot \mathbf{V} \cdot \mathbf{a}$ and resolving length (measured by $\|\mathbf{a} - \mathbf{p}\|$) are a special case of (13) in which the dimension of \mathfrak{F}^\perp is 1. The Lagrange multiplier λ which appears in those calculations is, in fact, M^2 .

6. Several Predictions from Error-free Data. Suppose we want to estimate $\mathbf{p}_1 \cdot \mathbf{h}_E, \dots, \mathbf{p}_P \cdot \mathbf{h}_E$, given $\mathbf{p}_1, \dots, \mathbf{p}_P, \mathfrak{D} \cdot \mathbf{h}_E$ and the hypothesis that $\|\mathfrak{F} \cdot \mathbf{h}_E\| \leq M$. We could estimate each $\mathbf{p}_i \cdot \mathbf{h}_E$ separately and calculate the errors in our estimates as in section 4. Clearly, however, these errors cannot all reach their maximum size at once. What we would really like to know is the subset of R^P to which the P -tuple $(\mathbf{p}_1 \cdot \mathbf{h}_E, \dots, \mathbf{p}_P \cdot \mathbf{h}_E)$ is confined by the data and the hypothesis $\|\mathfrak{F} \cdot \mathbf{h}_E\| \leq M$. Let \mathfrak{P} be the subspace of \mathfrak{S} spanned by $\mathbf{p}_1, \dots, \mathbf{p}_P$. Then we want to know what limits are placed on $\mathfrak{O} \cdot \mathbf{h}_E$ by a knowledge of $\mathfrak{D} \cdot \mathbf{h}_E$ and the assumption $\|\mathfrak{F} \cdot \mathbf{h}_E\| \leq M$. That is, we would like to know $S(\mathfrak{D}, \mathfrak{F}, M)$, the set of all vectors $\mathfrak{O} \cdot \mathbf{h}$ such that \mathbf{h} is in \mathfrak{S} , $\mathfrak{D} \cdot \mathbf{h} = \mathfrak{D} \cdot \mathbf{h}_E$, and $\|\mathfrak{F} \cdot \mathbf{h}\| \leq M$. If $S(\mathfrak{D}, \mathfrak{F}, M)$ is to be bounded, we must have $\mathfrak{P} \subseteq \mathfrak{D} + \mathfrak{F}$; we will assume this, and therefore lose no generality in assuming $\mathfrak{S} = \mathfrak{D} + \mathfrak{F}$.

It is easy to verify that the solution to this problem in solid geometry is

$$S(\mathfrak{D}, \mathfrak{F}, M) = \mathfrak{O} \cdot \mathfrak{D}_{\mathfrak{G}}^* \cdot \mathbf{h}_E + \{ \mathbf{n} : \mathbf{n} \text{ in } \mathfrak{N} \text{ and } \mathbf{n} \cdot \mathfrak{N} \cdot \mathfrak{G}_{\mathfrak{D}}^* \cdot \mathfrak{G}_{\mathfrak{D}} \cdot \mathfrak{N} \cdot \mathbf{n} \leq M^2 - \|\mathfrak{D} \wedge \mathfrak{F} \mathbf{h}_E\|^2 \}, \quad (14)$$

where $\mathfrak{B} = \mathfrak{F} \ominus (\mathfrak{D} \cap \mathfrak{F})$ and $\mathfrak{N} = \mathfrak{P} \ominus (\mathfrak{D} \cap \mathfrak{P})$. Thus $S(\mathfrak{D}, \mathfrak{F}, M)$ is a non-degenerate ellipsoid in $\mathfrak{O} \cdot \mathfrak{D}_{\mathfrak{G}}^* \cdot \mathbf{h}_E + \mathfrak{N}$, whose principal axes are determined by the eigenvectors and eigenvalues of $\mathfrak{N} \cdot \mathfrak{G}_{\mathfrak{D}}^* \cdot \mathfrak{G}_{\mathfrak{D}} \cdot \mathfrak{N}$.

7. Several Predictions from Erroneous Data. We continue to use the notation of sections 5 and 6. If $P > 1$ and the data γ_i are inaccurate, we do not know how to generalize efficiently and linearly the approach of section 5. Therefore we adopt a Bayesian point of view. Our belief that $\|\mathfrak{F} \cdot \mathbf{h}_E\|$ is probably not much larger than M can be embodied in an *a priori* subjective probability distribution¹⁵ of $\mathfrak{F} \cdot \mathbf{h}_E$ on \mathfrak{F} . This distribution must be a cylinder measure $\mu_{\mathfrak{F}}$ rather than a countably additive probability measure if \mathfrak{F} is infinite-dimensional. A convenient and natural choice of $\mu_{\mathfrak{F}}$ is the cylinder measure whose characteristic function is $\phi_{\mathfrak{F}}(f) = \exp[-M^2 \|f\|^2/2]$. On any finite-dimensional subspace \mathfrak{T} of \mathfrak{F} , the marginal distribution of $\mu_{\mathfrak{F}}$ is normal with mean \mathbf{o} and variance tensor $M^2 \mathfrak{I}$, so $\mu_{\mathfrak{F}}$ has mean \mathbf{o} and variance tensor $M^2 \mathfrak{F}$. The marginal distribution of $\mu_{\mathfrak{F}}$ on $\mathfrak{D} \cap \mathfrak{F}$ has probability density p_{prej} given by $-2 \ln p_{prej}(\mathfrak{D} \wedge \mathfrak{F} \cdot \mathbf{h}) = \text{constant} + M^{-2} \|\mathfrak{D} \wedge \mathfrak{F} \cdot \mathbf{h}\|^2$.

The distribution of errors in the data can be thought of as a probability distribution on \mathfrak{D} with density p_{data} , mean \mathbf{h}' and variance tensor \mathbf{V} given by (10) and (11). The Bayesian approach¹⁵ leads us to adopt an *a posteriori* subjective probability distribution for $\mathfrak{D} \cdot \mathbf{h}$ in \mathfrak{D} , whose probability density is $p_{\mathfrak{D}}$ with $p_{\mathfrak{D}}(\mathfrak{D} \cdot \mathbf{h}) = K p_{data}(\mathfrak{D} \cdot \mathbf{h}) p_{prej}(\mathfrak{D} \wedge \mathfrak{F} \cdot \mathbf{h})$, where K is a normalizing constant. As in I, this $p_{\mathfrak{D}}$ uses our prejudice about $\|\mathfrak{F} \cdot \mathbf{h}_E\|$ to limit the variance of $\mathfrak{D} \cdot \mathbf{h}_E$ in those directions in $\mathfrak{D} \cap \mathfrak{F}$ for which the data provide no useful limit.

If p_{data} is normal then $p_{\mathfrak{D}}$ is normal, with variance tensor $V_{\mathfrak{D}}$ and mean $h_{\mathfrak{D}}$ given by

$$V_{\mathfrak{D}}^{\text{inv}} = V^{\text{inv}} + M^{-2}\mathfrak{D}\wedge\mathfrak{F} \tag{15}$$

$$h_{\mathfrak{D}} = V_{\mathfrak{D}} \cdot V^{\text{inv}} \cdot h'. \tag{16}$$

Since $\mathfrak{B} \cap \mathfrak{D} = \{o\}$, the only limitation on $\mathfrak{B} \cdot h_E$ comes from our subjective probability $\mu_{\mathfrak{F}}$, which gives for the distribution of $\mathfrak{B} \cdot h_E$ on \mathfrak{B} the cylinder measure $\mu_{\mathfrak{B}}$, whose characteristic function is $\phi_{\mathfrak{B}} = \phi_{\mathfrak{F}}|_{\mathfrak{B}}$. The mean and variance of $\mu_{\mathfrak{B}}$ are $h_{\mathfrak{B}} = o$ and $V_{\mathfrak{B}} = M^2\mathfrak{B}$. We suppose that $\mathfrak{B} \cdot h_E$ and $\mathfrak{D} \cdot h_E$ are distributed independently and that $\mu_{\mathfrak{B}}$ and $\mu_{\mathfrak{D}}$ are the marginal distributions of a single cylinder measure μ on $\mathfrak{R} = \mathfrak{B} \dot{+} \mathfrak{D}$. Then, from section 3, $\mu = \mu_{\mathfrak{B}+\mathfrak{D}}$. (This conclusion can also be reached by a Bayesian argument.) The net effect of the data and our prejudice is to produce for us an *a posteriori* subjective probability distribution, $\mu_{\mathfrak{B}+\mathfrak{D}}$, for $\mathfrak{K} \cdot h$ in \mathfrak{R} . The mean $h_{\mathfrak{B}+\mathfrak{D}}$ and variance $V_{\mathfrak{B}+\mathfrak{D}}$ are given by (see (7) and (9))

$$h_{\mathfrak{B}+\mathfrak{D}} = \mathfrak{D}_{\mathfrak{B}}^* \cdot V_{\mathfrak{D}} \cdot V^{\text{inv}} \cdot h' \tag{17}$$

$$V_{\mathfrak{B}+\mathfrak{D}}^{\text{inv}} = V_{\mathfrak{D}}^{\text{inv}} + V_{\mathfrak{B}}^{\text{inv}}.$$

Using (15) and the fact that $V^{\text{inv}} = M^{-2}\mathfrak{B}$, we see that

$$V_{\mathfrak{B}+\mathfrak{D}}^{\text{inv}} = V^{\text{inv}} + M^{-2}\mathfrak{F}. \tag{18}$$

Unless $\mathfrak{B} \subseteq \mathfrak{B} \dot{+} \mathfrak{D}$ there is no marginal distribution of $\mu_{\mathfrak{B}+\mathfrak{D}}$ on \mathfrak{B} and no way to make our predictions. Therefore we assume $\mathfrak{B} \subseteq \mathfrak{B} \dot{+} \mathfrak{D} = \mathfrak{D} + \mathfrak{F}$. Then we lose no generality by assuming that $\mathfrak{S} = \mathfrak{D} + \mathfrak{F}$. With this assumption, $\mu_{\mathfrak{B}+\mathfrak{D}}$ is a cylinder measure on \mathfrak{S} , and the prediction vector $\mathcal{O} \cdot h_E$ has an *a posteriori* subjective probability distribution with mean $h_{\mathcal{O}} = \mathcal{O} \cdot h_{\mathfrak{B}+\mathfrak{D}}$ and variance tensor $V_{\mathcal{O}} = \mathcal{O} \cdot V_{\mathfrak{B}+\mathfrak{D}} \cdot \mathcal{O}$. The P -tuple of predictions $(p_1 \cdot h_E, \dots, p_P \cdot h_E)$ has a probability distribution in R^P whose mean is $(p_1 \cdot h_{\mathcal{O}}, \dots, p_P \cdot h_{\mathcal{O}})$ and whose variance matrix is $E[p_i \cdot (h_E - h_{\mathcal{O}})(h_E - h_{\mathcal{O}}) \cdot p_j] = p_i \cdot V_{\mathcal{O}} \cdot p_j$.

For readers skeptical of Bayes it will be of interest to note that when the foregoing Bayesian calculus is applied to a single prediction, the result is the same as that obtained in the objectivist, non-Bayesian section 5. We want to show from (13) and (17) that $\tilde{h}_{\mathfrak{B}+\mathfrak{D}} = h_{\mathfrak{B}+\mathfrak{D}}$. Let $U = M^{-2}V$ and $W = \mathfrak{D}\wedge\mathfrak{F} + \mathfrak{D}\wedge\mathfrak{F} \cdot U \cdot \mathfrak{D}_{\mathfrak{B}}$. Then W lives on $\mathfrak{D} \cap \mathfrak{F}$ and U lives on \mathfrak{D} . We have $\tilde{h}_{\mathfrak{B}+\mathfrak{D}} = \mathfrak{D}_{\mathfrak{B}}^* \cdot (\mathfrak{D} - U \cdot W^{\text{inv}}) \cdot h'$ and $h_{\mathfrak{B}+\mathfrak{D}} = \mathfrak{D}_{\mathfrak{B}}^* \cdot (\mathfrak{D} + U \cdot \mathfrak{D}\wedge\mathfrak{F})^{\text{inv}} \cdot h'$. But $\mathfrak{D} + U \cdot \mathfrak{D}\wedge\mathfrak{F} = \mathfrak{C} + \mathfrak{C} \cdot U \cdot \mathfrak{D}\wedge\mathfrak{F} + W$, so, from (5), $(\mathfrak{D} + U \cdot \mathfrak{D}\wedge\mathfrak{F})^{\text{inv}} = \mathfrak{C} - \mathfrak{C} \cdot U \cdot W^{\text{inv}} + W^{\text{inv}} = \mathfrak{D} - U \cdot W^{\text{inv}}$. Here $\mathfrak{C} = \mathfrak{D} \ominus (\mathfrak{D} \cap \mathfrak{F})$.

* Supported in part by the Geophysics section of the National Science Foundation.
¹ Backus, G., *Proc. Nat. Acad. Sci. USA*, **65**, 1 (1970).
² Backus, G., *Proc. Nat. Acad. Sci. USA*, **65**, 281 (1970).
³ Halmos, P. R., *Introduction to Hilbert Space* (New York: Chelsea, 1951).
⁴ Moore, E. H., *Bull. Amer. Math. Soc.*, **26**, 393 (1920).
⁵ Penrose, R., *Proc. Cambridge Phil. Soc.*, **51**, 406 (1955).
⁶ Lorch, E. R., *Trans. Amer. Math. Soc.*, **45**, 217 (1939).
⁷ Dunford, N., and J. T. Schwartz, *Linear Operators I* (New York: Interscience, 1958).
⁸ Gross, L., *Analysis in Function Space*, ed. W. T. Martin and I. E. Segal (Cambridge, Mass.: M. I. T. Press, 1964), pp. 51-68.

- ⁹ Halmos, P. R., *Measure Theory* (New York: Van Nostrand, 1950).
- ¹⁰ Segal, I. E., *Trans. Amer. Math. Soc.*, **81**, 106 (1956).
- ¹¹ Parzen, E., *Modern Probability Theory and its Applications* (New York: Wiley, 1960).
- ¹² Gettoor, R. K., *Pacific J. Math.*, **7**, 885 (1957).
- ¹³ Backus, G., and F. Gilbert, *Geophys. J. Roy. Astron. Soc.*, **16**, 169 (1968).
- ¹⁴ Backus, G., and F. Gilbert, *Phil. Trans. Roy. Soc. London Ser. A*, **266**, 123 (1970).
- ¹⁵ Savage, L. J., *The Foundations of Statistical Inference* (London: Methuen, 1962), p. 22.