*A Theorem on Integral Transforms with an Application to Prediction Theory**

J. CHOVER

Communicated by W. FELLER

1. Introduction. The main theorem of this paper characterizes those continuous functions $f(t)$ on $a \leq t \leq b$ which can be represented as

$$(1) \quad f(t) = \int_a^b r(t-s) dg(s),$$

where $r(t)$ is a given kernel function of a certain class R (see Definition 1) and $g(s)$ is of bounded variation on $a \leq s \leq b$. Some discussion (Sec. 3) is also given of approximation to a solution by solutions of finite algebraic systems.

The type of integral kernel to be considered here was suggested by (and Theorem I has application to) the following problem in prediction theory. Let x_t , $-\infty < t < +\infty$, be a (say) real-valued, second-order-stationary stochastic process with $\mathbb{E}x_t \equiv 0$ and continuous covariance function $r(t) = \mathbb{E}x_0x_t$. Let $-\infty < a < b < c < +\infty$. There exists a well determined random variable y which is the optimal linear least-squares prediction of the variable x_c , given the x_s for $a \leq s \leq b$. The following question arises: When does there exist a function $g(s) = g(s; a, b, c)$ of bounded variation on $a \leq s \leq b$ such that

$$(2) \quad y = \int_a^b x_s dg(s)$$

with a suitable definition of the integral? Standard methods show that (2) is true if and only if

$$(3) \quad r(t-c) = \int_a^b r(t-s) dg(s), \quad a \leq t \leq b.$$

Now if one knows, for a given r , that the function $f(t) = r(t-c)$ is in the class of functions representable as in (1), the above question concerning prediction is answered in the affirmative.

The above prediction question was raised for the semi-infinite case,

*Research done with the support of the Office of Naval Research under Contract No. N7 ONR 28507 Task Order 7, and, in part, of the United States Army under Contract No. DA-11-022-ORD-2059.

$-\infty < s \leq b$, by WIENER [5]; and, under an additional restriction, BEURLING [1] found a necessary and some sufficient conditions for the analogue of (3) to hold in the semi-infinite case. Both WIENER and BEURLING made use of Wiener-Hopf techniques, which do not seem to apply readily in the finite interval case. Some techniques exist for approaching this case provided r has rational spectral density,* but apparently not for more general r .

For the finite interval case, we found [2] that a necessary condition on $r(t)$ for (3) to hold, for any choice of b between fixed a and c , is that r have a cusp at $t = 0$. This, plus other considerations arising from the methods we use, leads us to deal with the following classes of functions.

Let $-\infty < a < b < \infty$.

Definition 1. Let the class R consist of all non-negative continuous functions $r(t)$ on $a - b \leq t \leq b - a$ which are non-decreasing for $a - b \leq t \leq 0$ and non-increasing for $0 \leq t \leq b - a$; which have finite non-decreasing right- and left-handed derivatives (r^*) in the intervals $a - b < t < 0$ and $0 < t < b - a$; and which have the following property: If m is a non-zero Radon measure on $a \leq t \leq b$ then the continuous function $\int_a^b r(t - s)m(ds)$ cannot be identically zero on $a \leq t \leq b$. (The last property is guaranteed, for example, if r is strictly positive-definite.)

Definition 2. Let the class D consist of all complex-valued continuous functions f on $a \leq t \leq b$ for which the left-handed derivative (f^-) is a function of bounded variation in $a < t < b$.

Theorem I states that if r is in R , then the class of functions f representable as in (1) coincides exactly with D .

It seems surprising that the class of representable functions does not depend directly on the individual r .

The key notion in the proof of Theorem I is to use the discontinuity of slope at the origin, of an r in R , to convert equation (1) into a Fredholm equation of Type II in the setting of a linear operation in a space of measures.

2. The main theorem. We shall let C be the Banach space of all continuous complex-valued functions on $[a, b]$ with the sup norm and B be its dual, interpreted as the space of Radon measures on $[a, b]$ with the variational norm ($\| \cdot \|$). B_0 will be the closed subspace of B consisting of all measures which assign zero mass to the points a and b . We note that for any f in D , the function f^- induces a measure m_f in B_0 . For any r in R , the function r^- induces a Radon measure v_0 on $[a - b, b - a]$ which imparts negative mass $-\lambda$ to the point 0 and is of total mass $\leq \lambda$ on the remainder of the interval.

Theorem I. If r is in R , then the class $\mathcal{F}$ of functions f representable as in (1) coincides exactly with D .

* Cf. for example, D. SLEPIAN [4] and C. L. DOLPH & M. A. WOODBURY [3].

Proof. Part I. That $\mathfrak{F} \subset D$ is easy to see. For let m be in B and let $f(t) = \int_a^b r(t-s)m(ds)$; fix t in (a, b) . The Lebesgue dominated convergence theorem may be invoked, giving

$$(4) \quad f^-(t) = \int_a^b r^-(t-s)m(ds).$$

The variation $\text{var}_{(a,b)}(f^-) \leq \{\text{var}_{[a-b, b-a]}(r^-)\} \|m\| < \infty$. Thus f is in D .

Part II. $D \subset \mathfrak{F}$: We convert the problem to one of Fredholm Type II by writing $v_0 = v - \lambda\delta_0$ (δ_0 = measure with mass 1 concentrated at $t = 0$) and examining the corresponding transformation of the subspace B_0 .

We note that the measure $v = v_0 + \lambda\delta_0$ defined on $[a-b, b-a]$ is purely non-negative, imparts zero mass to the origin, and is of total mass $\leq \lambda$.

Let J denote the interval (a, b) , and define the transformation $Q: B_0 \rightarrow B_0$ as follows: For m in B_0

$$(5) \quad (Qm)(E) = (v * m)(E \cap J),$$

where E is any Borel subset of $[a, b]$, $\cap$ denotes set intersection, and $v * m$ denotes convolution of the measures v and m , both considered, for the moment, as defined on subsets of $-\infty < t < \infty$. It is obvious that Qm is always in B_0 and that Q is a bounded linear transformation with operator norm $\|Q\| \leq \text{var}_{[a-b, b-a]}(v) \leq \lambda$.

If, in Part I, we choose m in B_0 , then the fact that (4) holds for all t in (a, b) may be written in terms of measures as $m_{f^-}(E) = (v_0 * m)(E \cap J)$, or as

$$(4') \quad m_{f^-} = (Q - \lambda I)m$$

where I denotes the identity transformation of $B_0 \rightarrow B_0$ and, of course, m_{f^-} determines f^- only up to a constant function.

(The restriction of Q to B_0 rather than to B enables us to use the identity I in (4'). The full space B will be recovered later.)

We wish to invert the operator $Q - \lambda I$, and so inspect $\|Q\|$ and $\|Q\|_{s.p.} = \lim_{n \rightarrow \infty} \|Q^n\|^{1/n}$ more closely. Now if $\lambda > \|Q\|$ then λ certainly is not in the spectrum of Q . We need then consider only the case $\lambda = \|Q\| = \text{var}_{[a-b, b-a]}(v)$.

Suppose that for an integer $n \geq 2$ it is true that $\|Q^n\| = \|Q\|^n$, i.e., $\sup \|Q^n m\| = \|Q\|^n$, where the supremum is extended over all m in B_0 with $\|m\| = 1$. In view of the non-negativity of v , the supremum may be restricted to non-negative m in B_0 with $m(J) = 1$. Now by omitting the process of cutting down to J , which appears in the definition (equation (5)) of Q , one can only increase the amount of mass present ($m \geq 0$), and so we can write

$$\|Q^n m\| = (Q^n m)(J) \leq (v^n * m)(J) \leq \|Q\|^n,$$

where $v^n = v * v * \dots * v$, $n - 1$ $*$ -operations. Thus we must conclude from the above supposition that

$$(6) \quad \sup_m (v^n * m)(J) = \|Q\|^n.$$

Since $(v^n * m)(J) = \int v^n(J_t) m(dt) \leq \sup_{(a,b)} v^n(J_t) m(J) \leq \|Q\|^n$, where $J_t = \{s - t, s \text{ in } J\}$, we can replace (6) by

$$(6') \quad \sup_{(a,b)} v^n(J_t) = \|Q\|^n = \{\text{var}_{[a-b, b-a]}(v)\}^n.$$

But (6') can hold only if the (topological) support of the measure v^n is contained in the interval $[a - b, b - a]$.

However, since, for any measure $w \neq \delta_0$ and a given $c > 0$, there is an integer $k \geq 1$ such that the support of w^k is not wholly contained in the interval $[-c, c]$, we can conclude that there must be some integer $n \geq 1$ such that $\|Q^n\| < \|Q\|^n = \lambda^n$. This immediately implies that λ is not in the spectrum of Q .

Thus for any r in R , the associated transformation $Q - \lambda I: B_0 \rightarrow B_0$ is always invertible.

Let now f be in D . To it is associated m_f in B_0 . Let $w_f = (Q - \lambda I)^{-1} m_f$, which is in B_0 . Let

$$g(t) = \int_a^b r(t-s) w_f(ds),$$

for t in $[a, b]$. By Part I we know that the measure $m_g = (Q - \lambda I) w_f = m_f$. Thus for the function $f - g$ in D we have a left-handed derivative $f^- - g^-$ which generates the measure $m_f - m_g = 0$, i.e., $f^- - g^- \equiv \text{const.}$ Hence $f(t) - g(t) \equiv L(t)$ is a linear function on $[a, b]$.

To complete the proof of Part II we need only show that any linear function $L(t)$ has a representation as in (1). Let

$$f_1(t) = r(t-a) = \int_a^b r(t-s) \delta_1(ds),$$

where δ_1 , in B , gives measure 1 to the point a and is zero elsewhere. Let

$$f_2(t) = r(t-b) = \int_a^b r(t-s) \delta_2(ds),$$

where δ_2 , in B , gives measure 1 to the point b and is zero elsewhere. Now, for $i = 1, 2$, the f_i are in D , and we can define $w_i = (Q - \lambda I)^{-1} m_{f_i}$. Let

$$g_i(t) = \int_a^b r(t-s) w_i(ds).$$

Then, as above, we find that

$$L_i(t) = f_i(t) - g_i(t) = \int_a^b r(t-s) \tilde{w}_i(ds)$$

are linear functions, where $\tilde{w}_i = \delta_i - w_i$ are in B .

The L_i are linearly independent, for if there were constants c_1, c_2 , not both zero, such that

$$c_1 L_1 + c_2 L_2 = \int_a^b r(t-s) (c_1 \tilde{w}_1 + c_2 \tilde{w}_2)(ds) = 0$$

for t in $[a, b]$, then this would contradict the properties of a function r in R , since the measure $c_1\bar{w}_1 + c_2\bar{w}_2$ has at least mass c_1 at a or c_2 at b and is thus non-zero. Now any linear function $L(t)$, being a linear combination of L_1 and L_2 , can be represented as in (1); and Part II is proved.

We note that it follows from Definition 1 that if r is in R , then $r(t - c)$ is in D and therefore has a representation as in (3), and with real-valued g .

It is interesting to note how the method of proof of Theorem I breaks down in the case $-\infty < t \leq b$. The natural setting would seem to be to take for C the set of continuous functions vanishing at $-\infty$, for B the bounded complex measures on $-\infty < t \leq b$, and for B_0 those measures of B assigning zero mass to the point b . Functions in R should be required to vanish at infinity.

The breakdown occurs in the fact that, for any r in R , $\lambda = \|Q\|_{s.p.}$. This is because (6') is now replaced by

$$(6'') \quad \sup_{-\infty < t < b} v^n((-\infty, b - t)) = \|Q\|^n;$$

and no matter how widely the operation of convolving v $n - 1$ times may spread the resulting mass distribution, that mass will be recaptured in $(-\infty, b - t)$ as $t \downarrow -\infty$, so that (6'') must indeed hold for all $n \geq 1$. Hence $\|Q\|_{s.p.} = \|Q\| = \lambda$.

3. An approximation theorem. We give an illustration here of how the operator method of Sec. 2 may be used to estimate the function g of equation (1), once its existence is guaranteed by Theorem I. We shall make rather strong assumptions on r and f so as to facilitate the presentation of the approach.

Let $-\infty < a < b < \infty$. Let r be in R and be even, strictly positive-definite, and with no corners except at $t = 0$. Let f be in D , also with no corners. Let k be an integer and $\gamma = k^{-1}(b - a)$, and consider the algebraic system

$$(7) \quad f(a + i\gamma) = \sum_{j=0}^k r((i - j)\gamma)x_j, \quad i = 0, 1, \dots, k.$$

This will have a solution $x_0, \dots, x_k$, the matrix $(r((i - j)\gamma))$ being non-singular since r is strictly positive-definite. Theorem I guarantees the existence of a function g of bounded variation such that (1) holds. Let $g_k(s)$ be zero at $s = a$, have a jump of magnitude x_i at $s = a + j\gamma$, and be constant elsewhere in $[a, b]$. How close is g_k to g , and does the difference diminish as $k \uparrow \infty$? Clearly, the variation of $g - g_k$ is too strict an index of discrepancy; a uniform-pointwise-convergence criterion is indicated.

It will be convenient, for this purpose, to consider the objects of B as functions g of bounded variation normalized to be zero at a . For g in B we define a norm

$$N(g) = \sup |g(t)| + |g(b)|,$$

where the supremum is extended over all points t in $[a, b]$ which are continuity points of g . In terms of N we can state the following result.

Theorem II. *If r , f , g , and g_k are as given above and if, furthermore, $|r^-(b-a)| > \frac{1}{2} |r^-(0)|$, then $N(g - g_k) \rightarrow 0$ as $k \rightarrow \infty$. (The rather *ad hoc* condition on r^- in the hypotheses is introduced to avoid detailed consideration of spectral norms in the proof and could be improved.)*

Proof of Theorem II. Let r^- , λ , and v be as in Sec. 2. The function $f^-(t)$ may be extended to be uniformly continuous on $[a, b]$. Also the function $q(t) = v([a, t])$ ($= r^-(t)$ for t in $(a-b, 0]$ and $= r^-(t) + \lambda$ for t in $(0, b-a]$) may be so extended. Given $\epsilon > 0$, choose the integer $k_\epsilon > 0$ so that for all $k \geq k_\epsilon$ and $\gamma = k^{-1}(b-a)$ it will be true that both f^- and q differ in absolute value by less than ϵ over any interval of length less than γ .

Let now $k \geq k_\epsilon$. One can investigate the approximation error $N(g - g_k)$ by means of two auxiliary approximations, as follows.

(1) Let $F(t) = f(t)$ for $t = a + i\gamma$ ($i = 0, 1, 2, \dots, k$) and be linear in between. Let $S(t) = r(t)$ for $t = i\gamma$ ($i = -k, \dots, 0, \dots, k$) and be linear in between. Then F is in D , and $S(t)$ is strictly positive-definite (the sum of triangles) and in R . F has a representation:

$$(8) \quad F(t) = \int_a^b r(t-s) dG(s), \quad t \text{ in } [a, b], \quad G \text{ in } B.$$

One can show (A) that $N(g - G) < \text{const.} \cdot \epsilon$.

(2) F also has a representation:

$$(9) \quad F(t) = \int_a^b S(t-s) dH(s), \quad t \text{ in } [a, b], \quad H \text{ in } B.$$

But H must coincide with g_k , under proper normalization; for the function

$$F_0(t) = \int_a^b S(t-s) dg_k(s) = f(t) = F(t)$$

at $t = a + i\gamma$ ($i = 0, 1, \dots, k$) and is linear in between, and therefore is $\equiv F(t)$. Since S is strictly positive-definite, (9) determines H uniquely, and $H = g_k$.

One can show (B) that since both (8) and (9) hold, $N(G - H) = N(G - g_k) < \text{const.} \cdot \epsilon$. Hence $N(g - g_k) < C\epsilon$, where the constant $C = C(f, r)$, the desired inequality.

Sketch of (A): Let $\bar{B}$ denote the completion of B under N . Let $\bar{B}_0$ denote the completion of B_0 in $\bar{B}$, a closed subspace of $\bar{B}$. Then, for g in B_0 , we may transcribe the definition (Equation (5)) of Q , in terms of functions of bounded variation, as

$$(5') \quad (Qg)(t) = \int_a^b [q(t-s) - q(t-a)] dg(s), \quad t \text{ in } [a, b].$$

An easy calculation shows that

$$N(Qg) \leq 2 \{ \text{var}_{[a-b, b-a]}(v) \} N(g),$$

i.e., the N -operator norm of Q , $N(Q) \leq 2 \text{ var } (v)$. (Examples show that this

increase by a factor of 2 over $\|Q\|$ is the minimum possible in general.) Q may be considered extended to $\bar{B}_0$ with the same N -norm.

Let φ^- and Φ^- result from normalizing the functions f^- and F^- , respectively, at a . As in Sec. 2, we have $\varphi^- = (Q - \lambda I)g$ and $\Phi^- = (Q - \lambda I)G$.

For the above choice of k , we conclude, by the mean value theorem, that $|F^-(t) - f^-(t)| < \epsilon$ for all t ; hence $N(\Phi^- - \varphi^-) < \text{const.} \cdot \epsilon$.

Now the condition $|r^-(b - a)| > \frac{1}{2} |r^-(0)|$ merely insures that $\lambda > N(Q)$; hence $N(g - G) \leq N((Q - \lambda I)^{-1}) \cdot \text{const.} \cdot \epsilon$.

Sketch of (B): Let λ_1 , q_1 , and Q_1 correspond to the kernel $S(t)$, and put $\lambda = \lambda_1 + \Delta\lambda$ and $Q = Q_1 + \Delta Q$. Then we have $(Q - \lambda I)G = (Q_1 - \lambda_1 I)g_k$. Now, by the choice of $S(t)$, $|S^-(b - a)| - \frac{1}{2} |S^-(0)| \geq |r^-(b - a)| - \frac{1}{2} |r^-(0)|$. Hence $Q_1 - \lambda_1 I$ is invertible also, and we get

$$g_k - G = (Q_1 - \lambda_1 I)^{-1}(\Delta Q - \Delta\lambda I)(Q - \lambda I)^{-1}\Phi^-.$$

Again, for the above choice of k and by the mean value theorem, one sees that the N -norm of $\Delta Q - \Delta\lambda I$ is small; detailed calculation yields $N(g_k - G) \leq \text{const.} \cdot \epsilon$, the constant depending only on f and r .

REFERENCES

- [1] A. BEURLING: Unpublished material.
- [2] J. CHOVER: *Conditions on the realization of prediction by measures*, Duke Math. J., **25** (2), pp. 305-310.
- [3] C. L. DOLPH & M. A. WOODBURY: *On the relation between Green's functions and covariances of certain stochastic processes and its application to unbiased linear prediction*, Trans. Amer. Math. Soc., **72** (3) (1952), pp. 519-550.
- [4] D. SLEPIAN: *Estimation of signal parameters in the presence of noise* (Appendices), Trans. I.R.E. PGIT-3(1954), pp. 82-87.
- [5] N. WIENER: *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Cambridge, 1949.

University of Wisconsin
Madison, Wisconsin