From: OPTIMAL ESTIMATION
IN APPROXIMATION THEORY
Edited by Charles A. Micchelli
and Theodore J. Rivlin
(Plenum Publishing Corporation, 1977)

n-WIDTHS AND OPTIMAL INTERPOLATION OF TIME- AND BAND-LIMITED FUNCTIONS

Avraham A. Melkman

IBM Thomas J. Watson Research Center

Yorktown Heights, New York 10598

1. INTRODUCTION

Let $\mathscr{L}^2(R)$ denote the space of complex valued square integrable functions on the real line. Define

$$\mathscr{G} = \{f \in \mathscr{L}^2(R) \Big| \int_{|t|>T} |f(t)|^2\, dt \le \epsilon_T^2, \int_{|\omega|>\sigma} |\hat{f}(\omega)|^2 d\omega \le \eta_\sigma^2\} \quad (1)$$

with $\hat{f}$ the Fourier transform of f. Following Slepian [7] $\mathscr{G}$ may be regarded as the class of functions time-limited to $(-T,T)$, at level ϵ_T, and bandlimited to $(-\sigma,\sigma)$, at level η_σ (only $f \equiv 0$ is strictly time- and band-limited). For this class we consider the n-widths and the optimal recovery of a function from its sampled values as well as the interrelation between them.

Section 2 is devoted to n-widths. It is based on the classical papers of Landau, Pollak and Slepian [3,4,8] and Slepian [7]. We have attempted to make the proofs more geometrical in nature and to provide a unified framework. The main result for the n-widths is

$$d_n^2(\mathscr{G};\mathscr{L}^2(R)) = \frac{1}{2}\,\frac{(\epsilon_T+\eta_\sigma)^2}{1-\sqrt{\lambda_n}} + \frac{1}{2}\,\frac{(\epsilon_T-\eta_\sigma)^2}{1+\sqrt{\lambda_n}} \quad (2)$$

where λ_n are decreasing numbers dependent only on σT. This result can be put to two uses of particular interest. By the definition of the zero-width as the radius of $\mathscr{G}$ it follows that for all $f \in \mathscr{G}$, $||f||^2 \le (\epsilon_T^2+\eta_\sigma^2 + 2\sqrt{\lambda_0}\,\epsilon_T\,\eta_\sigma)/(1-\lambda_0)$. This inequality may then be turned around to provide, for any f, a relationship between the fraction of its norm outside $(-T,T)$ and outside $(-\sigma,\sigma)$, the uncertainty relationship of Landau and Pollak [3]. Another use of formula (2) rests upon an examination of λ_n for σT large. This shows that for any $\epsilon^2 > \epsilon_T^2 + \eta_\sigma^2$ and $\delta > 0$. σT can be chosen large enough so that for $n = (1+\delta)4\sigma T$ $d_n^2 < \epsilon^2$ while for $n = (1-\delta)4\sigma T$ $d_n^2 > \epsilon^2$. Slepian [7], in rephrasing Landau and Pollak's [4] dimension theorem, interpreted this fact to mean that the approximate dimension of the set of time- and band-limited signals is asymptotically $4\sigma T$ as σ or T becomes large.

In section 3 we turn to the optimal recovery of a function from its values at a fixed sampling set $\{s_i\}_1^n$. To make the problem meaningful $\eta_\sigma=0$ is required which turns $\mathcal{G}$ into an ellipsoidal set. Consequently it follows from general arguments, e.g., Micchelli and Rivlin in these Proceedings, that the optimal procedure is to interpolate with the set of functions $\{K(t,s_i)\}_1^n$, $K(t,t')$ a suitable kernel (if the ϵ_T^2 bound is replaced by $\int_{-\infty}^{\infty}|f(t)|^2\,dt \le 1$ the resulting kernel is simply $\sin 2\pi\sigma(t-t')/\pi(t-t')$). Since the procedure is a particular kind of linear approximation the worst error that can be encountered in recovery of the class $\mathcal{G}$ cannot be less than the n-width d_n. A natural question is therefore whether this bound can be achieved. This is indeed the case, the corresponding points being the zeros of the "worst" function for the n-width problem. Thus as a fringe benefit this approach identifies an (additional) optimal approximating subspace, which moreover achieves the n-width by the simple device of interpolation. Similar results have been described by Melkman and Micchelli [6] for totally positive kernels. The present result shows that the condition of total positivity is not necessary.

2. n-WIDTHS

We briefly review some basics of n-widths; for more details consult Lorentz [5]. The n-width of a subset $\mathcal{A}$ of a normed linear space X is obtained by considering its distance from an arbitrary n-dimensional subspace $X_n \subset X$, and then looking for the optimal approximating subspace X_n^* which achieves the least distance, the n-width with respect to X,

$$d_n(\mathcal{A};X) = \inf_{X_n \subset X} \sup_{f\in\mathcal{A}} \inf_{g\in X_n} ||f-g|| \tag{3}$$

When X is a Hilbert space $\mathcal{A}$ is typically an ellipsoidal set for which n-widths results are known in great generality. Although the present set $\mathcal{G}$ is not ellipsoidal it is illuminating to consider a simple example of this kind:

$$X = R^3 \text{ and } \mathcal{A} = \{(x_0,x_1,x_2) \mid \sum_{i=o}^{2} x_i^2/\lambda_i \le 1\}$$

with $\lambda_0 \ge \lambda_1 \ge \lambda_2 > 0$. Then the zero width is the radius of the smallest ball containing $\mathcal{A}$, $d_0(\mathcal{A};R^3) = \sqrt{\lambda_0}$. For the 1-width one has to look for the optimal line through the origin. The major axis is clearly such a line, $d_1(\mathcal{A};R^3) = \sqrt{\lambda_1}$, but any other line in the x_0-x_2 plane with slope not exceeding $\sqrt{(\lambda_1-\lambda_2)/(\lambda_0-\lambda_1)}$ is optimal too.

The same line of reasoning shows that non-uniqueness of the optimal subspace prevails for general ellipsoidal sets, e.g., Karlovitz [1], and even in the case at hand which is not quite ellipsoidal. But more on that in the next section.

The above example intimates that a convenient representation for $\mathcal{G}$, defined in (1), will greatly facilitate the calculation of its n-widths. To that end let us introduce the following notation.

Denote by $\mathcal{D}$ the space of functions $f \in \mathcal{L}^2(R)$ vanishing outside $(-T,T)$ (strictly time-limited), by $\mathcal{B}$ those with Fourier transform vanishing outside $(-\sigma,\sigma)$ (strictly band-limited). Let D and B be the projections on these spaces, i.e.,

$$Df(t) = \begin{cases} f(t) & |t| \le T \\ 0 & |t| > T \end{cases} \tag{4}$$

$$Bf(t) = \int_{-\infty}^{\infty} f(t') \frac{\sin 2\pi\sigma(t-t')}{\pi(t-t')}\,dt' \tag{5}$$

corresponding to time-limiting and band-limiting. Landau and Pollak [3] made the crucial observation that the minimum angle formed between $\mathcal{D}$ and $\mathcal{B}$ is non-zero, meaning

$$\sup_{\substack{d\in\mathcal{D}\\ b\in\mathcal{B}}} \frac{\mathrm{Re}(d,b)}{||d||\,||b||} < 1$$

with the usual inner product

$$(d,b) = \int_{-\infty}^{\infty} d(t)\, \bar{b}(t)dt, \quad (f,f) = ||f||^2.$$

Using this property they prove

Lemma 1.

The space $\mathcal{D} + \mathcal{B}$ is closed. Consequently any $f \in \mathcal{L}^2(R)$ may be decomposed as

$$f = d+b+g \text{ with } d \in \mathcal{D},\ b \in \mathcal{B},\ Dg = Bg = 0.$$

There is no convenient representation for the multitude of functions g with $Dg = Bg = 0$, but fortunately they play no essential role here. The spaces $\mathcal{D}$ and $\mathcal{B}$ on the other hand possess bases with remarkable properties as shown by Slepian and Pollak [8], Landau [2].

Let ψ_i and λ_i, $i = 0,1,\ldots$ be the eigenfunctions and eigenvalues of the integral equation

$$\int_{-T}^{T} \frac{\sin 2\pi\sigma(t-t')}{\pi(t-t')} \psi(t') = \lambda\psi(t) \tag{6}$$

Noting that it has a completely continuous, positive definite, symmetric kernel the λ_i may be assumed nonnegative decreasing to zero and the ψ_i real and normalized $(\psi_i,\psi_j) = \delta_{ij}$. Denoting

$$\phi_i(t) = \frac{1}{\sqrt{\lambda_i}} D\psi_i(t) \tag{7}$$

the integral equation implies

$$B\phi_i = \sqrt{\lambda_i}\, \psi_i, \quad (\phi_i,\phi_j) = \delta_{ij}.$$

Clearly $\psi_i \in \mathcal{B}$, $\phi_i \in \mathcal{D}$. Out of the many other properties of these functions to be found in the cited references we will find the following particularly useful.

1. The set $\{\psi_i\}_0^\infty$ is complete in $\mathcal{B}$, the set $\{\phi_i\}_o^\infty$ in $\mathcal{D}$.
2. ψ_i has exactly i simple zeros in $(-T,T)$.
3. $1>\lambda_0>\lambda_1> \ldots$, $\lim_{j\to\infty} \lambda_j = 0$.
4. The eigenvalues depend only on the product σT,

$$\lambda_{[4\sigma T]+1}(\sigma T) \le .5, \quad \lambda_{[4\sigma T]-1}(\sigma T) \ge .5,$$

$$\lim_{\sigma T\to\infty} \lambda_n(\sigma T) = \begin{cases} 0 & n = [(1+\eta)4\sigma T] \\ 1 & n = [(1-\eta)4\sigma T] \end{cases}$$

5. Of all functions $f \in \mathcal{B}$ orthogonal to $\psi_0,\ldots,\psi_n$ the one most concentrated in $(-T,T)$ is ψ_{n+1}, i.e., this function achieves

$$\sup_{\substack{f \in \mathcal{B} \\ (f,\psi_i)=0\ i=o,\ldots,n}} \frac{||Df||^2}{||f||^2} = \lambda_{n+1}$$

The last property provides the rationale behind the ψ_i: it leads directly to the integral equation $BDf = \lambda f$, i.e., (6). It may be worth mentioning that the ψ_i are also (regular) solutions of a singular Sturm-Liouville type differential equation, the prolate spheroidal wave equation, so that much is known about them.

Property 1 and Lemma 1 immediately provide the sought after representation for $\mathcal{G}$, which is then used to calculate its n-widths.

Lemma 2.

Any function $f \in \mathcal{L}^2(R)$ has the representation

$$f(t) = \sum_{i=o}^{\infty} (d_i\phi_i(t) + b_i\psi_i(t)) + g(t) \tag{8}$$

where $Dg = Bg = 0$. *If in addition* $f \in \mathcal{G}$, *i.e.*, $||(1-D)f||^2 \le \epsilon_T^2$, $||(1-B)f||^2 \le \eta_\sigma^2$ *then*

$$\sum_{i=o}^{\infty} |b_i|^2(1-\lambda_i) + ||g||^2 \le \epsilon_T^2 \tag{9}$$

$$\sum_{i=o}^{\infty} |d_i|^2(1-\lambda_i) + ||g||^2 \le \eta_\sigma^2 \tag{10}$$

Theorem 1.

$$d_n(\mathcal{G};\mathcal{L}^2(R)) = \frac{1}{2}\frac{(\epsilon_T+\eta_\sigma)^2}{1-\sqrt{\lambda_n}} + \frac{1}{2}\frac{(\epsilon_T-\eta_\sigma)^2}{1+\sqrt{\lambda_n}} \tag{11}$$

and any set of the form $\{\delta_i\phi_i(t) + \beta_i\psi_i(t)\}_o^{n-1}$, $|\delta_i|^2+|\beta_i|^2 > 0$, $\mathrm{Re}\ \delta_i\beta_i \ge 0$, *spans an optimal subspace.*

Proof.

We show first that the right hand side is an upper bound by calculating the distance between $\mathcal{G}$ and the particular space spanned by

$$g_i(t) = \delta_i \phi_i(t) + \beta_i \psi_i(t).$$

Let $P_n f$ denote the projection of f given by (8) on $\{g_i\}_o^{n-1}$.

A short calculation using $(\phi_i,\phi_j) = (\psi_i,\psi_j) = \delta_{ij}, (\phi_i,\psi_j) = \sqrt{\lambda_i}\,\delta_{ij}$ shows

$$||d_i\phi_i + b_i\psi_i - \frac{(f,g_i)}{(g_i,g_i)} g_i||^2$$

$$= [|b_i|^2+|d_i|^2 - \frac{|\bar{\delta}_i d_i+\bar{\beta}_i b_i|^2 + 2\sqrt{\lambda_i}\,(|b_i|^2+|d_i|^2)\mathrm{Re}\,\delta_i\bar{\beta}_i}{|\delta_i|^2+|\beta_i|^2 + 2\sqrt{\lambda_i}\mathrm{Re}\;\delta_i\bar{\beta}_i}](1-\lambda_i)$$

$$\le (|b_i|^2+|d_i|^2)(1-\lambda_i)$$

since $\mathrm{Re}\;\delta_i\bar{\beta}_i \ge 0$. Hence

$$||f-P_n f||^2$$

$$\le \sum_{i=o}^{n-1} (|b_i|^2+|d_i|^2)(1-\lambda_i) + \sum_{i=n}^{\infty} [|b_i|^2+|d_i|^2 + 2\sqrt{\lambda_i}\;\mathrm{Re}\; d_i\,\bar{b}_i]+||g||^2$$

$$\le \frac{\sum_{i=o}^{\infty}|b_i|^2(1-\lambda_i)+||g||^2+\sum_{i=o}^{\infty}|d_i|^2(1-\lambda_i)+||g||^2+2\sqrt{\lambda_n}\sum_{i=o}^{\infty}|b_i d_i|(1-\lambda_i)}{1-\lambda_n}$$

$$\le \frac{\epsilon_T^2 + \eta_\sigma^2 + 2\sqrt{\lambda_n}\;\epsilon_T\eta_\sigma}{1-\lambda_n}$$

making use of $(\phi_i,g) = (\psi_i,g) = 0$, $0 < \lambda_i < \lambda_n < 1$ $i \ge n$ and Cauchy-Schwarz. Examination of the inequalities reveals that equality is attained only for $h_n = (\eta_\sigma\phi_n + \epsilon_T\,\psi_n)/\sqrt{1-\lambda_n}$.

In order to show that no n-dimensional subspace can improve this bound we use the standard technique of finding an n+1-dimensional ball of this radius contained in $\mathcal{G}$, i.e., a set $\{h_i\}_o^n$ such that

$$||\sum_{i=o}^{n} a_i h_i||^2 \le (\epsilon_T^2 + \eta_\sigma^2 + 2\sqrt{\lambda_n}\;\epsilon_T\eta_\sigma)/(1-\lambda_n) \quad \text{implies}$$

$\sum_{i=o}^{n} a_i h_i \in \mathcal{G}$. Indeed, the distance of this ball from an arbitrary n-dimensional subspace equals its radius, for there always is a function on the ball which is orthogonal to the subspace and hence its best approximation is zero. Taking

$$h_i = (\eta_\sigma\phi_i + \epsilon_T\,\psi_i)/\sqrt{1-\lambda_i}$$

one has $||(1-D)h_i|| = \epsilon_T$, $||(1-B)h_i|| = \eta_\sigma$ and hence $\sum_{i=o}^{n} a_i h_i \in \mathcal{G}$ if $\sum_{i=o}^{n}|a_i|^2 \le 1$. That the ball is indeed contained in $\mathcal{G}$ follows therefore from

$$||\sum_{i=o}^{n} a_i g_i||^2 = \sum_{i=o}^{n}|a_i|^2\,\frac{\epsilon_T^2+\eta_\sigma^2+2\sqrt{\lambda_i}\;\epsilon_T\eta_\sigma}{1-\lambda_i} \ge \sum_{i=o}^{n}|a_i|^2\,\frac{\epsilon_T^2+\eta_\sigma^2+2\sqrt{\lambda_n}\epsilon_T\eta_\sigma}{1-\lambda_n}.$$

Remark. From the proof it follows that $h_n = (\eta_\sigma\phi_n + \epsilon_T\,\psi_n)/\sqrt{1-\lambda_n}$ is the unique function with the largest norm of all those in $\mathcal{G}$ orthogonal to $\{h_i\}_o^{n-1}$. In this sense the system $\{h_i\}$ is the most natural one to approximate with, though from the n-width point of view one may choose the approximating subspace independently from ϵ_T and η_σ.

In this theorem $\mathcal{G}$ is approximated on the whole real line. If one is interested in obtaining an approximation only on $(-T,T)$ then $\mathcal{G}$ should be considered as a subset of $\mathcal{L}^2(-T,T)$ instead of $\mathcal{L}^2(R)$. This is done by Slepian [7] who, using variational arguments derived the bounds

$$\lambda_n\epsilon_T^2/(1-\lambda_n) \le d_n(\mathcal{G};\mathcal{L}^2(-T,T)) \le (\sqrt{\lambda_n}\;\epsilon_T+\eta_\sigma)^2/(1-\lambda_n).$$

The next theorem shows that the upper bound is sharp.

Theorem 2.

$$d_n(\mathcal{G};\mathcal{L}^2(-T,T)) = \frac{(\sqrt{\lambda_n}\;\epsilon_T+\eta_\sigma)^2}{1-\lambda_n} \qquad (12)$$

and $\{\phi_i\}_o^{n-1}$ spans an optimal subspace.

Proof.

The proof of the previous theorem showed that, with P_n the projection on $\{\phi_i\}_0^{n-1}$,

$$||f-P_nf||^2 \le \frac{||(1-D)f||^2+||(1-B)f||^2+2\sqrt{\lambda_n}\,||(1-D)f||\cdot||(1-B)f||}{1-\lambda_n}$$

Since $D\phi_i = \phi_i$, $f-P_nf = (1-D)f+D(f-P_nf)$ and so

$$||D(f-P_nf)||^2 \le \lambda_n\frac{||(1-D)f||^2+||(1-B)f||^2+2\sqrt{\lambda_n}\,||(1-D)f||\cdot||(1-B)f||}{1-\lambda_n}$$

establishing the right hand of (12) as an upper bound. The ball argument again shows that equality holds.

Two consequences of these results have particularly interesting interpretations. The first one addresses the question to what extent a signal can be concentrated in the time interval $(-T,T)$ and simultaneously in the frequency interval $(-\sigma,\sigma)$. This problem was solved by Landau and Pollak [3]. We give here an equivalent version which instead shows how small the fractional concentrations outside the time and band intervals, $||(1-D)f||/||f||$ and $||(1-B)f||/||f||$, can be.

Corollary 1. For $f \in \mathscr{L}^2(R)$ let $||f|| = 1$. Then the possible values of $\epsilon_T = ||(1-D)f||$ and $\eta_\sigma = ||(1-B)f||$ fill up the unit square $[0,1]\times[0,1]$ except for the points $(0,1)$, $(1,0)$ and the region inside the ellipse

$$\frac{1}{2}\,\frac{(\epsilon_T+\eta_\sigma)^2}{1-\sqrt{\lambda_0}} + \frac{1}{2}\,\frac{(\epsilon_T-\eta_\sigma)^2}{1+\sqrt{\lambda_0}} = 1. \qquad (13)$$

Proof.

The zero-width of theorem 1 establishes the ellipse as a boundary since, by definition, $||f|| \le d_0(\mathscr{S};\mathscr{L}^2(R))$ for any $f \in \mathscr{S}$. The boundary is attained by the functions $h_0 = (\eta_\sigma\phi_0+\epsilon_T\psi_0)/\sqrt{1-\lambda_0}$. The point $(1,0)$ is not permitted since a strictly band-limited function cannot vanish identically in an interval. However the points $(1,\lambda_i)$ come arbitrarily close and are attained by $(1-D)\psi_i/\sqrt{1-\lambda_i}$. The point $(1,1)$ corresponds to all functions in $\mathscr{L}^2(R)-(\mathscr{D}+\mathscr{B})$. Finally all other points may be reached by the device of shifting frequency via a factor $e^{2\pi i\delta t}$. This leaves ϵ_T intact and increases η_σ to 1 as δ increases.

The other application we want to bring concerns the notion that there are $4\sigma T$ signals of duration $2T$ and bandwidth σ. It is based upon an analysis of the eigenvalues as functions of σT, as summarized in property 5. The idea is to show that $4\sigma T$ functions suffice to approximate $\mathscr{S}$ on the real line to within the order of magnitude of $\epsilon_T^2 + \eta_\sigma^2$.

Corollary 2. Let $N(\epsilon_T,\eta_\sigma;\epsilon)$ be the least integer such that $d_N(\mathscr{S};\mathscr{L}^2(R)) \le \epsilon$. Then

a. Landau and Pollak [4]: for $\epsilon^2=2(\epsilon_T^2+\eta_\sigma^2+\sqrt{2}\epsilon_T\eta_\sigma)$ and all σT

$$[4\sigma T]-1 \le N(\epsilon_T,\eta_\sigma;\epsilon) \le [4\sigma T]+1$$

b. Slepian [7]: for any $\epsilon^2 > \epsilon_T^2 + \eta_\sigma^2$ and $\delta > 0$ σT can be chosen large enough so that

$$(1-\delta)[4\sigma T] \le N(\epsilon_T,\eta_\sigma;\epsilon) \le (1+\delta)[4\sigma T]$$

3. OPTIMAL INTERPOLATION

Consider the problem of recovery of $f \in \mathscr{S}$, on the real line or $(-T,T)$, from the knowledge of its values at a fixed set of sampling points $\{s_i\}_1^n$. Since point evaluations have no meaning in the context of $\mathscr{S}$ we confine attention to the subset $\mathscr{S}_0$ for which $\eta_\sigma = 0$

$$\mathscr{S}_0 = \{f \in \mathscr{L}^2(R) \mid ||(1-D)f|| \le \epsilon_T,\ Bf = f\} \qquad (14)$$

Thus $\mathscr{S}_0$ is a subset of $\mathscr{B}$ the Paley Wiener class of entire functions of exponential type $\le 2\pi\sigma$. For comparison we shall also consider the recovery problem for $f \in \hat{\mathscr{B}} = \{f \in \mathscr{B} \mid ||f||^2 \le 1\}$.

Taking the latter problem first, recall that $\mathscr{B}$ has the reproducing kernel

$$K(t,t') = \frac{\sin 2\pi\sigma(t-t')}{\pi(t-t')} \qquad (15)$$

i.e., $f(t) = (K(t,\cdot),f)$ for all $f \in \mathscr{B}$ (this may also be deduced from (5) and $Bf = f$). For this situation an optimal scheme for recovery on $(-T,T)$, or even pointwise, is well known, see, e.g., the survey of Micchelli and Rivlin, and consists of projecting f on $\{K(t,s_i)\}_1^n$. Indeed, projection obviously decreases the norm and moreover it uses only the available information because the orthogonality conditions imply it is equivalent to interpolation

$$0 = (K(\cdot,s_k),\ f - \sum_{i=1}^{n} a_i K(\cdot,s_i)) = f(s_k) - \sum_{i=1}^{n} a_i K(s_k,s_i).$$

Hence the worst error that can be encountered as f ranges over $\mathscr{B}$ occurs for zero data,

$$E(\hat{\mathscr{B}};S) = \max_{\substack{f \in \hat{\mathscr{B}} \\ f(s_i)=0 \quad i=1,\dots,n}} ||Df||$$

The procedure of projection-interpolation with the functions $\sin 2\pi\sigma(t-s_i)/\pi(t-s_i)$ is of course widely used in practice. However this procedure is not optimal for $\mathscr{G}_o$, for in that case we should use the kernel $K_o(t,t')$ which reproduces $\mathscr{B}$ with respect to the inner product

$$\langle f,g\rangle = (f,(1-D)g) \tag{16}$$

This kernel can be defined implicitly by the integral equation

$$K_T(t,t') = K(t,t') + \int_{-T}^{T} K(t,s)\, K_T(s,t')ds \tag{17}$$

or explicitly by its expansion

$$K_T(t,t') = \sum_{i=o}^{\infty} \frac{\psi_i(t)\psi_i(t')}{1-\lambda_i}. \tag{18}$$

Proceeding as before one finds that an optimal recovery procedure is projection with respect to $\langle\cdot,\cdot\rangle$ on $\{K_T(t,s_i)\}_1^n$ which again is equivalent to interpolation by this set. The error in recovery of $f \in \mathscr{G}_o$ on $(-T,T)$ is therefore

$$E(\mathscr{G}_o;S) = \max_{\substack{f \in \mathscr{G}_o \\ f(s_i)=0,i=1,\dots,n}} ||Df|| = \epsilon_T E(\hat{\mathscr{B}};S)/(1-E(\hat{\mathscr{B}};S)) \tag{19}$$

Since the optimal recovery scheme involves linear approximation from a particular n-dimensional subspace, the inequality

$$E(\mathscr{G}_o;S) \ge d_n\ (\mathscr{G}_o;\mathscr{L}^2(-T,T))$$

is immediate. We want to show next that with a propitious choice of the points equality can be achieved. A similar result will hold for $\hat{\mathscr{B}}$. The following fact will be needed.

Lemma 3.

Let $f^* \in \mathscr{B}$ achieve $E(\hat{\mathscr{B}};S)$ or $E(\mathscr{G}_o;S)$. Then the only zeros of f^* in $[-T,T]$ are those s_i lying in the interval.

Proof.

Suppose f^* does have an additional zero, s_o, in $[-T,T]$. Consider the function

$$f_1(t) = f^*(t)\,\frac{T^2-ts_o}{T(t-s_o)}.$$

f_1 is an entire function of exponential type $2\pi\sigma$ and $f_1 \in \mathscr{L}^2(R)$ because the same is true for f^* and $f^*(s_o) = 0$. Hence by the Paley-Wiener theorem $f_1 \in \mathscr{B}$. However $|f_1(t)| \ge |f^*(t)|$ for $|t| \le T$, $|f_1(t)| \le |f^*(t)|$ for $|t| \ge T$ and therefore

$$\frac{||Df_1||}{||(1-D)f_1||} > \frac{||Df^*||}{||(1-D)f^*||}$$

Thus f^* cannot possibly attain $E(\mathscr{G}_o;S)$ nor, from (19), $E(\hat{\mathscr{B}};S)$.

Theorem 3.

Let the (simple) zeros of $\psi_n(t)$ in $(-T,T)$ be $\Xi = \{\xi_i\}_1^n$. Then the recovery error of the optimal scheme based on these points equals the n-width

$$E(\hat{\mathscr{B}};\Xi) = \sqrt{\lambda_n} \tag{20}$$

$$E(\hat{\mathscr{G}}_o;\Xi) = \epsilon_T\sqrt{\lambda_n}/(1-\lambda_n) \tag{21}$$

Thus Ξ is an optimal sampling set.

Proof.

We prove (20) since (21) follows from it via (19). What is needed is an explicit evaluation of max $||Df||^2/||f||^2$, the maximum being taken over the space $f \in \mathscr{B}$, $f(\xi_i) = 0$ $i=1,\dots,n$. This is most easily done by considering the reproducing kernel for this space.

$$K^n(t,t') = \frac{K\begin{pmatrix} t, \xi_1, \ldots, \xi_n \\ t', \xi_1, \ldots, \xi_n \end{pmatrix}}{K\begin{pmatrix} \xi_1, \ldots, \xi_n \\ \xi_1, \ldots, \xi_n \end{pmatrix}}$$

where the numerator and denominator are appropriate Fredholm determinants of $K(t,t')$. With this notation $E(\mathcal{B};\Xi)$ is the largest eigenvalue of the integral operator $H = K^n D K^n$ with kernel

$$H(s,t) = \int_{-T}^{T} K^n(s,t')K^n(t',t)dt'.$$

The eigenfunction associated with the top eigenvalue is the "worst" function for recovery. Clearly $\psi_n(t)$ is an eigenfunction of this operator, with eigenvalue λ_n, since $(K^n(t,\cdot),\psi_n) = \psi_n(t)$, $(K^n(t,\cdot),D\psi_n) = \lambda_n\psi_n(t)$, by the choice of Ξ. The question is whether it is the largest eigenvalue. Any eigenfunction associated with a different eigenvalue is orthogonal to $\psi_n(t)$ on $(-T,T)$. Moreover, all eigenfunctions must vanish at the ξ_i. Since these are the only zeros of ψ_n in $(-T,T)$ the orthogonality implies that all other eigenfunctions (which may be taken real) must vanish in addition somewhere else in $(-T,T)$. Lemma 3 now finishes the proof of (20).

That $d_n(\hat{\mathcal{B}};\mathcal{L}^2(-T,T)) = \sqrt{\lambda_n}$ follows by a standard argument, cf. Lorentz [5], once it is observed that $\hat{\mathcal{B}}$ is an ellipsoidal set with respect to $\mathcal{L}^2(-T,T)$. Indeed, while any $f \in \mathcal{L}^2(-T,T)$ may be expanded as

$$f(t) = \sum_{i=o}^{\infty} c_i\phi_i(t)$$

the condition $f \in \hat{\mathcal{B}}$ is equivalent to $\sum_{i=o}^{\infty} |c_i|^2/\lambda_i \le 1$.

<u>Corollary 3.</u> <u>The subspace spanned by</u> $\{DK_o(t,\xi_i)\}_1^n$ <u>is an optimal n-dimensional approximating subspace for</u> $\mathcal{G}_o$, <u>with respect to both</u> $\mathcal{L}^2(-T,T)$ <u>and</u> $\mathcal{L}^2(-\infty,\infty)$, <u>in addition to the subspaces given in theorems 1, 2.</u>

<u>Proof.</u>

Optimality for $\mathcal{L}^2(-T,T)$ is inherent in theorem 3. From it the optimality on the whole real line follows since, denoting by $P_n f$ the interpolant of f by $\{DK_o(t,\xi_i)\}_1^n$

$$||f-P_nf||^2 = ||(1-D)f+D(f-P_nf)||^2 \le \epsilon_T^2 + \epsilon_T^2 \frac{\lambda_n}{1-\lambda_n} = \frac{\epsilon_T^2}{1-\lambda_n}$$

which equals $d_n(\mathcal{G}_o;\mathcal{L}^2(-\infty,\infty))$, by theorem 1.

<u>Remark.</u> We have been informed that Logan (unpublished) obtained a similar theorem on the basis of a formula, valid for all $f \in \mathcal{B}$, expressing $||Df||^2 - \lambda_n||f||^2$ as a weighted sum of $\{|f(x_k)|^2\}$ where $\{x_k\}$ are the zeros of $\psi_n(t)$ in $(-\infty,\infty)$ and the weights are positive for $|x_k| < T$ and negative otherwise.

The above proof has been patterned after Melkman and Micchelli [6]. There we considered sets of the form $f = Kh$, $||h|| \le 1$ with K a totally positive kernel, and used the total positivity to arrive at similar results. Though total positivity does not hold here, e.g., $\sin 2\pi\sigma(t-t')/\pi(t-t')$ is not TP for all t,t', an important consequence of it, the zero properties of the eigenfunctions, continues to hold. This, together with lemma 3, was the crux of the proof. An open problem is what the most general conditions are under which a conclusion such as theorem will be valid. The following example demonstrates that the zero properties in themselves are neither sufficient nor necessary.

Let $\{f_i\}_0^2$ be real orthonormal functions, $\lambda_0 \ge \lambda_1 \ge \lambda_2 > 0$, and consider the set of functions

$$f(t) = \sum_{i=o}^{2} a_i f_i(t) \qquad \sum_{i=o}^{2} a_i^2/\lambda_i \le 1 .$$

An optimal recovery scheme when given $f(\xi)$ consists of interpolation by

$$K(t,\xi) = \sum_{i=o}^{2} \lambda_i f_i(t) f_i(\xi)$$

and the largest error E encountered occurs for zero data. Thus, taking ξ to be a zero of f_1, one has to find the maximum of $\Sigma\, a_i^2$ under the conditions $\Sigma\, a_i^2/\lambda_i = 1$, $a_o f_o(\xi) + a_2 f_2(\xi) = 0$.

$$E^2 = \max_{a_2} \left[\left(\frac{f_2(\xi)}{f_o(\xi)} \right)^2 (\lambda_0 - \lambda_1) - (\lambda_1 - \lambda_2) \right] a_2^2 + \lambda_1$$

Hence $E^2 = \lambda_1$, the 1-width, if and only if

$$\left| \frac{f_2(\xi)}{f_0(\xi)} \right| \le \sqrt{\frac{\lambda_1 - \lambda_2}{\lambda_0 - \lambda_1}}$$

This condition can be made to hold, by choosing λ_i, when ξ is not the only zero of f_1, or made to fail when it is.

References

1. L. A. Karlovitz, Remarks on variational characterization of eigenvalues and n-width problems, J. Math. Anal. Appl. 53 (1976), 99-110.

2. H. J. Landau, The eigenvalue behavior of certain convolution equations, Trans. Am. Math. Soc. 115 (1965), 242-256.

3. H. J. Landau and H. O. Pollak, Prolate spherical wave functions, Fourier analysis and uncertainty. II, Bell System Tech. J. 40 (1961), 65-84.

4. ______________, Prolate spheroidal wave functions. Fourier analysis and uncertainty. III, Bell System Tech. J. 41 (1962), 1295-1336.

5. G. G. Lorentz, Approximation of functions. Holt, Rinehart and Winston, 1966.

6. A. A. Melkman and C. A. Micchelli, On non-uniqueness of optimal subspaces for L^2 n-width, IBM Research Report, RC 6113 (1976).

7. D. Slepian, On bandwidth, Proc. IEEE, 64 (1976), 292-300.

8. D. Slepian and H. O. Pollak, Prolate spheroidal wavefunctions, Fourier analysis and uncertainty, I, Bell System Tech. J. 40 (1961), 43-64.