ILLINOIS JOURNAL OF MATHEMATICS
Volume 22, Number 4, December 1978

SPLINE SPACES ARE OPTIMAL FOR L^2 n-WIDTH

BY

AVRAHAM A. MELKMAN[1] AND CHARLES A. MICCHELLI

1. Introduction

Let $X = (X, \|\cdot\|)$ be a normed linear space, $\mathscr{K}$ a subset of X and X_n an n-dimensional linear subspace of X. The Kolmogorov n-width of $\mathscr{K}$ relative to X is defined by

$$d_n(\mathscr{K}; X) = d_n(\mathscr{K}) = \inf_{X_n} \sup_{x \in \mathscr{K}} \inf_{y \in X_n} \|x - y\|.$$

X_n is called an optimal subspace for $\mathscr{K}$ provided that

$$d_n(\mathscr{K}) = \sup_{x \in \mathscr{K}} \inf_{y \in X_n} \|x - y\|.$$

This concept of n-width was introduced by Kolmogorov in [8] and in his paper he finds the exact value of the n-width for

$$W^{2,r}[0, 1] = \{f: f^{(r-1)} \text{ abs. cont. on } (0, 1), \|f^{(r)}\| \leq 1\}$$
$$(\|\cdot\| = L^2 \text{ norm on } [0, 1]).$$

Roughly speaking Kolmogorov showed that the n-width corresponds to the nth eigenvalue of a boundary value problem and an optimal subspace is spanned by the first n eigenfunctions. Kolmogorov claimed that $W^{2,r}[0, 1]$ has a unique optimal subspace and as late as Tihomirov [13] this error was overlooked. It was first observed to be false by Karlovitz in [4] while in Ioffe and Tihomirov [2] it is conjectured that $W^{2,r}[0, 1]$ has an optimal spline subspace.

Subsequently, Karlovitz [5] explored the question of nonuniqueness of optimal subspaces in a general setting. The related question for min max and max min characterization of eigenvalues has been treated in Weinstein and Stenger's book [17].

A main goal of this paper is to prove that $W^{2,r}$ admits, for all r, optimal spline subspaces. There are in fact two; one of degree $r - 1$ and another of degree $2r - 1$.

Before stating exactly our result for $W^{2,r}$ we wish to point out that an effort

Received August 18, 1976.

[1] Part of the work was conducted while the first author was at IBM. T. J. Watson Research Center, Yorktown Heights, New York.

© 1978 by the Board of Trustees of the University of Illinois
Manufactured in the United States of America

has been made to present this result in as general a setting as we are aware that it applies. A large portion of the paper deals with this general point of view via the notion of oscillation kernel. The importance of this concept to integral equations and Sturm-Liouville differential equations is the subject of the book [1]. Our results for $W^{2,r}$ follow below.

The eigenvalue problem

$$(1.1)\quad (-1)^r y^{(2r)}(x) = \mu y(x), \qquad y^{(i)}(0) = y^{(i)}(1) = 0,\ i = 0, 1, \ldots, r-1,$$

has positive simple eigenvalues $0 < \mu_{1,r} < \mu_{2,r} < \cdots < \mu_{n+1,r} < \cdots$ and a corresponding set of complete orthonormal eigenfunctions, $y_{1,r}(x)$, $y_{2,r}(x), \ldots$, $y_{n+1,r}(x), \ldots$. The function $y_{n+1,r}(x)$ has exactly n simple zeros in $(0, 1)$, given by

$$0 < \xi_{1,r} < \xi_{2,r} < \cdots < \xi_{n,r} < 1$$

and its rth derivative $y_{n+1,r}^{(r)}$ has exactly $N = n + r$ zeros in $(0, 1)$,

$$0 < \eta_{1,r} < \eta_{2,r} < \cdots < \eta_{n+r,r} < 1.$$

The n-width of $W^{2,r}[0, 1]$ is given by

$$d_n(W^{2,r}[0, 1]; L^2[0, 1]) = \begin{cases} \infty, & n < r \\ \mu_{n-r+1,r}^{-1/2}, & n \geq r, \end{cases}$$

and the space of spline functions of degree $r - 1$ with knots at $\xi_{1,r}, \ldots, \xi_{n,r}$,

$$X_{n+r}^1 = [1, x, \ldots, x^{r-1}, (x - \xi_{1,r})_+^{r-1}, \ldots, (x - \xi_{n,r})_+^{r-1}],$$

is an optimal subspace for the $(n + r)$-width of $W^{2,r}[0, 1]$ ($x_+^{r-1} = x^{r-1}$, $x \geq 0$, zero otherwise, and $[f_1, \ldots, f_m]$ = linear space spanned by $f_1, \ldots, f_m$). Furthermore, interpolation of $f \in W^{2,r}[0, 1]$ at $\eta_{1,r}, \ldots, \eta_{n+r,r}$ by an element in X_{n+r}^1 is an optimal method of approximating $W^{2,r}[0, 1]$.

In addition, the space of natural splines,

$$X_N^2 = \{S \in [1, x, \ldots, x^{2r-1}, (x - \eta_{1,r})_+^{2r-1}, \ldots, (x - \eta_{n+r,r})_+^{2r-1}]: \\ S^{(i)}(0) = S^{(i)}(1) = 0,\ i = r, \ldots, 2r - 1\},$$

is an optimal subspace for the $(n + r)$-width of $W^{2,r}[0, 1]$ and interpolation at $\eta_{1,r}, \ldots, \eta_{n+r,r}$ is an optimal method of approximating $W^{2,r}[0, 1]$.

We also include in Section 4 a matrix formulation of our results on totally positive integral operators, as well as in Section 3 the computation of n-widths under restricted approximation.

This latter problem allows us to answer the following question of optimal estimation.

Given f in a certain set and sampled function values $f(x_1), \ldots, f(x_s)$, where is the best place to sample f at n additional places to obtain the most information about it?

Some of the results presented here were announced in [10].

2. Statement of problem

Let H denote the Hilbert space of real-valued, square-summable functions on [0, 1]. We denote the norm and inner product on H by

$$\|f\|^2 = (f, f), \qquad (f, g) = \int_0^1 f(t)g(t)\,dt,$$

respectively.

Let $K(x, y)$ be a continuous function on $[0, 1] \times [0, 1]$ and define

$$\mathscr{K} = \left\{\int_0^1 K(x, y)h(y)\,dy \colon \|h\| \le 1\right\}.$$

$\mathscr{K}$ is the image of the unit ball in H under the completely continuous operator

$$(Kh)(x) = \int_0^1 K(x, y)h(y)\,dy.$$

Whenever T is an integral operator on H we will use the notation $T(x, y)$ for the kernel of T. Thus the adjoint of K which we denote by K^* has a kernel given by $K^*(x, y) = K(y, x)$. We will be frequently concerned with the combinations K^*K and KK^*. These operators have as kernels

$$(K^*K)(x, y) = (K(\cdot, x), K(\cdot, y)) \quad \text{and} \quad (KK^*)(x, y) = (K(x, \cdot), K(y, \cdot)).$$

K^*K is a completely continuous positive semi-definite symmetric operator with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge \cdots \ge 0$, and corresponding orthonormal eigenfunctions,

$$K^*K\phi_n = \lambda_n \phi_n, \quad (\phi_n, \phi_m) = \delta_{n,m}, \; m, n = 1, 2, \ldots. \tag{2.1}$$

In addition, we define $\psi_n = K\phi_n$ and observe that

$$KK^*\psi_n = \lambda_n \psi_n, \quad (\psi_n, \psi_m) = \lambda_n \delta_{nm}, \; n, m = 1, 2, \ldots. \tag{2.2}$$

The following theorem is a familiar result for the Kolmogorov n-width of $\mathscr{K}$; see Shapiro [16, p. 188].

THEOREM 2.1. *$d_n(\mathscr{K}; H) = \lambda_{n+1}^{1/2}$ and $X_n^0 = [\psi_1, \ldots, \psi_n]$, the linear subspace spanned by $\psi_1, \ldots, \psi_n$, is an optimal subspace for $\mathscr{K}$.*

To comment further on this interesting result we require the following idea. A continuous kernel K is said to be totally positive provided that the Fredholm determinant

$$K\begin{pmatrix} x_1, \ldots, x_n \\ y_1, \ldots, y_n \end{pmatrix} = \det_{i,j=1,\ldots,n} |K(x_i, y_j)|$$

is nonnegative for all $0 \le x_1 < \cdots < x_n \le 1, 0 \le y_1 < \cdots < y_n \le 1, n = 1, 2, \ldots.$

The following theorem of Kellogg [6], [7] gives useful information about the spectrum of a totally positive symmetric kernel (see Gantmacher and Krein [1] for various important extensions and applications of Kellogg's theorem).

THEOREM 2.2. *Let $K(x, y)$ be a totally positive symmetric kernel such that*

$$K\begin{pmatrix} x_1, \dots, x_n \\ x_1, \dots, x_n \end{pmatrix} > 0, \quad 0 < x_1 < \cdots < x_n < 1.$$

Then all the eigenvalues of K are positive and simple, $\lambda_1 > \lambda_2 > \cdots > \lambda_n > \cdots > 0$, and the corresponding orthonormal eigenfunctions $Ku_n = \lambda_n u_n$, $n = 1, 2, \dots$, form a Markov system on $(0, 1)$, that is,

$$U\begin{pmatrix} 1, \dots, n \\ x_1, \dots, x_n \end{pmatrix} = \det |u_i(x_j)| > 0, \quad 0 < x_1 < \cdots < x_n < 1, n = 1, 2, \dots.$$

Consequently, the $(n + 1)$st eigenfunction u_{n+1} has exactly n simple zeros in $(0, 1)$.

We will call a totally positive kernel *nondegenerate* if

$$\dim [K(x_1, \cdot), \dots, K(x_n, \cdot)] = \dim [K(\cdot, x_1), \dots, K(\cdot, x_n)] = n$$

for all $0 < x_1 < \cdots < x_n < 1$, $n = 1, 2, \dots$.

LEMMA 2.1. *If K is a nondegenerate totally positive kernel then the functions ϕ_{n+1} and ψ_{n+1}, defined above have exactly n simple zeros in $(0, 1)$,*

$$\phi_{n+1}(\xi_j) = \psi_{n+1}(\eta_j) = 0, \quad j = 1, 2, \dots, n,$$
$$0 < \xi_1 < \cdots < \xi_n < 1, \quad 0 < \eta_1 < \cdots < \eta_n < 1.$$

Proof. This lemma is an immediate consequence of Theorem 2.2. We argue as follows: The basic composition formula [3, p. 17] applied to K^*K gives

$$(K^*K)\begin{pmatrix} x_1, \dots, x_n \\ y_1, \dots, y_n \end{pmatrix} = \underset{0<\sigma_1<\cdots<\sigma_n<1}{\int_0^1 \cdots \int_0^1} K\begin{pmatrix} \sigma_1, \dots, \sigma_n \\ x_1, \dots, x_n \end{pmatrix} K\begin{pmatrix} \sigma_1, \dots, \sigma_n \\ y_1, \dots, y_n \end{pmatrix} d\sigma_1 \cdots d\sigma_n.$$

Thus K^*K satisfies the hypothesis of Theorem 2.2 and consequently ϕ_{n+1} has exactly n simple zeros in $(0, 1)$. Similarly, KK^* satisfies the hypothesis of Theorem 2.2 and thus ψ_{n+1} also has n simple zeros. We may now prove:

THEOREM 2.3. *Let K be a nondegenerate totally positive kernel. Then*

$$X_n^0 = [\psi_1, \dots, \psi_n], \qquad X_n^1 = [K(\cdot, \xi_1), \dots, K(\cdot, \xi_n)],$$

and

$$X_n^2 = [(KK^*)(\cdot, \eta_1), \ldots, (KK^*)(\cdot, \eta_n)]$$

are optimal subspaces for $\mathscr{K}$.

Proof. We begin by showing that X_n^1 is an optimal subspace for $\mathscr{K}$. Let P be the orthogonal projection of H onto X_n^1. Then

$$\delta(\mathscr{K}; X_n^1) = \|K - PK\| = \sup_{\|h\| \le 1} \|Kh - PKh\|$$

and hence $\delta(\mathscr{K}; X_n^1)^2$ is the largest eigenvalue of the operator $T = K^*(I - P)K$. An easy calculation shows that the kernel of T is given by

$$T(x, y) = \frac{(K^*K)\begin{pmatrix} x, \xi_1, \ldots, \xi_n \\ y, \xi_1, \ldots, \xi_n \end{pmatrix}}{(K^*K)\begin{pmatrix} \xi_1, \ldots, \xi_n \\ \xi_1, \ldots, \xi_n \end{pmatrix}}.$$

We begin the proof by demonstrating that λ_{n+1} is an eigenvalue of T. To this end, observe that for any $h \in H$ such that $(K^*h)(\xi_i) = 0$, $i = 1, \ldots, n$, then necessarily $Ph \equiv 0$. Therefore, because $K^*(K\phi_{n+1})(\xi_i) = \lambda_{n+1}\phi_{n+1}(\xi_i) = 0$, $i = 1, \ldots, n$,

$$T\phi_{n+1} = K^*(K\phi_{n+1} - PK\phi_{n+1}) = K^*K\phi_{n+1} = \lambda_{n+1}\phi_{n+1}.$$

Now, to show that λ_{n+1} is the largest eigenvalue of T we define $T_0(x, y) = |T(x, y)|$. Then $T_0(x, y) = \operatorname{sgn} \phi_{n+1}(x) \operatorname{sgn} \phi_{n+1}(y) T(x, y)$ is a symmetric nonnegative kernel with a nonnegative eigenfunction $|\phi_{n+1}(x)|$ and corresponding eigenvalue λ_{n+1}. Since T_0 has a nonnegative eigenfunction it is a familiar result that the corresponding eigenvalue must be the largest eigenvalue of T_0. Let us give the easy proof of this fact.

Suppose λ is the largest eigenvalue for T_0 and $f(x)$ the corresponding eigenfunction. Then $\lambda |f(x)| \le T_0(|f|)(x)$, $x \in [0, 1]$, and therefore

$$\begin{aligned} &\lambda(|f|, |\phi_{n+1}|) \\ &\quad \le (T_0(|f|), |\phi_{n+1}|) = (|f|, T_0(|\phi_{n+1}|)) = \lambda_{n+1}(|f|, |\phi_{n+1}|). \end{aligned}$$

Since $(|f|, |\phi_{n+1}|) > 0$ we conclude that $\lambda \le \lambda_{n+1}$. But we also have by the definition of λ that $\lambda_{n+1} \le \lambda$. Thus $\lambda = \lambda_{n+1}$. We conclude that $\|K - PK\| = \lambda_{n+1}^{1/2}$ and X_n^1 is an optimal subspace for $\mathscr{K}$.

Let us now prove the optimality of X_n^2. Suppose now, Q represents the orthogonal projection of H onto the subspace $[K(\eta_1, \cdot), \ldots, K(\eta_n, \cdot)]$. Thus the operator KQ takes H onto X_n^2 and $\delta(\mathscr{K}; X_n^2) \le \|K - KQ\| = \|K^* - QK^*\|$. Now, using our previous arguments (replace K by K^*) we conclude that $\|K^* - QK^*\| = \lambda_{n+1}^{1/2}$. Thus X_n^2 is also an optimal subspace for $\mathscr{K}$ and the proof is complete.

Below we give some examples of Theorem 2.3. The n-width in $L^\infty[0, 1]$ for our first example was computed in [11].

Example 2.1. Given any nonnegative real numbers $t_1, \ldots, t_m$ we define a polynomial of degree $r = 2m$ by $q_r(x) = \prod_{j=1}^{m} (x^2 - t_j^2)$. Let $D = d/dx$ and consider the set

$$\mathscr{D}_2 = \{f : f^{(r-1)} \text{ abs. cont. on } (0, 1), f^{(r)} \in L^2[0, 1], f^{(2k)}(0) = f^{(2k)}(1) = 0,$$
$$k = 0, 1, \ldots, m-1, \|q_r(D)f\| \le 1\}.$$

Then $\mathscr{D}_2 = \mathscr{K}$ where $K(x, y)$ is the Green's function for the differential operator

$$\prod_{j=1}^{m} \left(\frac{d^2}{dx^2} - t_j^2\right) f = 0, \qquad f^{(2k)}(0) = f^{(2k)}(1) = 0, k = 0, 1, \ldots, m-1.$$

It may be verified that

$$K(x, y) = 2 \sum_{k=1}^{\infty} \frac{\sin k\pi x \sin k\pi y}{q_r(ik\pi)}$$

and $(-1)^m K(x, y)$ is a nondegenerate totally positive kernel [11].

The kernel K is symmetric with eigenvalues $q_r^{-1}(ik\pi)$, $k = 1, 2, \ldots$, and corresponding eigenfunctions $\sin k\pi x$, $k = 1, 2, \ldots$. Hence according to Theorem 2.3,

$$X_n^0 = \left\{ \sum_{j=1}^{n} a_j \sin j\pi x : (a_1, \ldots, a_n) \in R^n \right\},$$

$$X_n^1 = \left\{ \sum_{j=1}^{n} a_j K\left(x, \frac{j}{n+1}\right) : (a_1, \ldots, a_n) \in R^n \right\}$$

and

$$X_n^2 = \left\{ \sum_{j=1}^{n} a_j G\left(x, \frac{j}{n+1}\right) : (a_1, \ldots, a_n) \in R^n \right\}$$

where $G(x, y) = (K(x, \cdot), K(\cdot, y)) = (K^2)(x, y)$ are optimal subspaces for $\mathscr{D}_2$. Note that X_n^1 is a subspace of (generalized) periodic spline functions of order r while X_n^2 is a subspace, of (generalized) spline functions of order $2r$. In [11], it was shown that X_n^1 is also an optimal subspace for

$$\mathscr{D}_\infty = \left\{ \int_0^1 K(x, y) h(y)\, dy : h \in L^\infty[0, 1], \operatorname*{ess\,sup}_{0 \le x \le 1} |h(x)| \le 1 \right\}$$

in $L^\infty[0, 1]$. We conjecture that this fact persists for all $L^p[0, 1]$, $1 \le p \le \infty$.

Example 2.2 (Nondegenerate cyclic Pòlya frequency functions). This example is not quite covered by Theorem 2.3, however, the method of proof is nevertheless applicable.

Here we deal with the convolution operator

$$(K_\phi h)(x) = \int_0^1 \phi(x-y)h(y)\,dy$$

where ϕ, h are in H and are both periodic with period one. We denote this class of functions by H_1.

We cannot expect that the even Fredholm minors of $K_\phi(x, y) = \phi(x-y)$ are always nonnegative since

$$K_\phi\begin{pmatrix} x_1, \ldots, x_{2n} \\ y_1, \ldots, y_{2n} \end{pmatrix} = -K_\phi\begin{pmatrix} x_2, \ldots, x_{2n}, 1+x_1 \\ y_1, \ldots, y_{2n-1}, y_{2n} \end{pmatrix}.$$

However, for the class of cyclic totally positive functions ϕ, that is, those functions for which the odd order Fredholm minors of K_ϕ are nonnegative, it is possible to produce an analog of Theorem 2.3. For a discussion of cyclic totally positive functions see [3, Chapter 9]. When speaking about the cyclic totally positive function $\phi(x)$ we also require, as before, that the kernel $K_\phi(x, y) = \phi(x-y)$ satisfies the nondegeneracy hypothesis,

$$\dim\,[K_\phi(x_1, \cdot), \ldots, K_\phi(x_n, \cdot)] = \dim\,[K_\phi(\cdot, x_1), \ldots, K_\phi(\cdot, x_n)] = n$$

for all $0 \le x_1 < \cdots < x_n < 1$, $n = 1, 2, \ldots$.

Returning to the proof of Theorem 2.3 we can readily see that to identify optimal subspaces for the n-width we only need to know that the $(n+1)$st Fredholm minors of K are nonnegative and, of course, that the $(n+1)$st eigenfunction of K^*K has exactly n simple zeros. Keeping these facts in mind we have the following results for the convolution operator K_ϕ.

First, observe that all the eigenvalues of $K_\phi^* K_\phi$ have double multiplicity, except for the largest. In fact, if $\hat{\phi}(n) = \int_0^1 \phi(y)e^{2\pi i n y}\,dy$ then

$$\lambda_1 = |\hat{\phi}(0)|^2, \qquad \lambda_{2n} = \lambda_{2n+1} = |\hat{\phi}(n)|^2, \; n = 1, 2, \ldots,$$

and the corresponding eigenfunctions are

$$\phi_1(x) = 1, \qquad \{\phi_{2n}(x), \phi_{2n+1}(x)\} = \left\{\frac{1}{\sqrt{2}}\sin 2\pi n x, \frac{1}{\sqrt{2}}\cos 2\pi n x\right\}, \; n = 1, \ldots,$$

(we may choose $\phi_{2n}(x)$ to be either $1/\sqrt{2}\sin 2\pi n x$ or $1/\sqrt{2}\cos 2\pi n x$). Hence for $n \ge 1$,

$$d_{2n-1}(\mathscr{K}_\phi; H_1) = d_{2n}(K_\phi; H_1) = |\hat{\phi}(n)|,$$

and

$$T_{2n-1} = [1, \sin 2\pi x, \cos 2\pi x, \ldots, \sin 2\pi(n-1)x, \cos 2\pi(n-1)x]$$

is an optimal subspace for the $2n-1$ width of $\mathscr{K}_\phi$, while

$$X_{2n}^0 = [1, \sin 2\pi x, \cos 2\pi x, \ldots, \sin 2\pi(n-1)x, \cos 2\pi(n-1)x, \cos 2\pi n x]$$

$$\text{or } = [1, \sin 2\pi x, \cos 2\pi x, \ldots, \sin 2\pi(n-1)x, \cos 2\pi(n-1)x, \sin 2\pi n x]$$

are optimal $2n$-dimensional subspaces.

Since the $2n+1$-eigenfunction $\sin 2\pi n(x-\alpha)$ has $2n$ simple zeros at $\alpha+j/2n$, $j=0,1,\ldots,2n-1$, for $0\le\alpha<1/2n$ we conclude by the methods employed in Theorem 2.3 that

$$X_{2n}^1=\left[\phi(\cdot-\alpha),\ldots,\phi\left(\cdot-\alpha-\frac{2n-1}{2n}\right)\right]$$

$$\text{and}\quad X_{2n}^2=\left[F(\cdot-\alpha),\ldots,F\left(\cdot-\alpha-\frac{2n-1}{2n}\right)\right],$$

$$F(x)=\int_0^1\phi(t)\phi(x+t)\,dt,$$

and optimal subspaces for all α, $0\le\alpha<1/2n$.

Let us again return to the proof of Theorem 2.3 and observe that the method of approximation for the subspace X_n^2 is interpolation of Kh at $\eta_1,\ldots,\eta_n$ by elements of X_n^2. This fact is a consequence of the orthogonality conditions which determine Qh, the orthogonal projection of h onto $[K(\eta_1,\cdot),\ldots,K(\eta_n,\cdot)]$:

$$0=(h-Qh,K(\eta_i,\cdot))=(Kh-KQh)(\eta_i),\quad i=1,\ldots,n,\ h\in H.$$

A similar, but perhaps less obvious, fact holds for the subspace X_n^1. Before we explain this further we need:

LEMMA 2.2. *Let K be a nondegenerate totally positive kernel. Then for $\xi_1,\ldots,\xi_n$, and $\eta_1,\ldots,\eta_n$, defined in Lemma 2.1, we have*

$$K\begin{pmatrix}\eta_1,\ldots,\eta_n\\ \xi_1,\ldots,\xi_n\end{pmatrix}>0.$$

Proof. If the above determinant is zero then there exist constants $\alpha_1,\ldots,\alpha_n$, not all zero, such that the function $f(x)=\sum_{j=1}^n\alpha_jK(x,\xi_j)$ vanishes at $\eta_1,\ldots,\eta_n$. Also, since $\dim[K(\cdot,\xi_1),\ldots,K(\cdot,\xi_n)]=n$ there is an $\eta_0\ne\eta_1,\ldots,\eta_n$ such that $f(\eta_0)\ne0$. We choose a constant d such that $\psi_{n+1}(\eta_0)-df(\eta_0)=0$.

Now, since

$$K\begin{pmatrix}\eta_0,\eta_1,\ldots,\eta_n\\ x_0,x_1,\ldots,x_n\end{pmatrix}\ge0$$

for all $0\le x_0<\cdots<x_n\le1$ (and strictly positive for some choice of $x_0,\ldots,x_n$) there exists a function $g(x)=\sum_{j=0}^n\beta_jK(\eta_j,x)$ which weakly changes signs at $\xi_1,\ldots,\xi_n$: $g(x)(-1)^i\ge0$, $\xi_i\le x\le\xi_{i+1}$, $(\xi_0=0,\ \xi_{n+1}=1)$. In particular,

$g(\xi_i) = 0$, $i = 1, \ldots, n$, and

$$\begin{aligned} 0 &= \sum_{j=0}^{n} \beta_j(\psi_{n+1}(\eta_j) - df(\eta_j)) \\ &= \int_0^1 g(x)\phi_{n+1}(x)\,dx - d \sum_{j=1}^{n} \alpha_j g(\xi_j) \\ &= \int_0^1 g(x)\phi_{n+1}(x)\,dx \\ &= \pm \int_0^1 |\phi_{n+1}(x)|\,|g(x)|\,dx \end{aligned}$$

This contradiction proves the lemma.

Since the above proof requires only that the $(n + 1)$st Fredholm minors of K be nonnegative we have the following periodic version of Lemma 2.2.

LEMMA 2.3. *Let ϕ be a nondegenerate cyclic totally positive function. Let $K_\phi(x, y) = \phi(x - y)$. Then*

$$K_\phi\begin{pmatrix} \alpha, \alpha + \frac{1}{2n}, \ldots, \alpha + \frac{2n-1}{2n} \\ \beta, \beta + \frac{1}{2n}, \ldots, \beta + \frac{2n-1}{2n} \end{pmatrix} \neq 0 \quad \text{for } 0 \leq \alpha, \beta < \frac{1}{2n}$$

where $\phi(n)e^{2\pi i n(\alpha - \beta)}$ is real.

Proof. The proof follows the idea used in Lemma 2.2 and uses the equation

$$\int_0^1 K_\phi(x, y) \sin 2\pi n(y - \beta)\,dy = |\hat{\phi}(n)| \sin 2\pi n(x - \alpha).$$

According to Lemma 2.2 there exists an interpolation operator $L\colon C[0, 1] \to X_n^1$ defined by

$$(Lh)(\eta_i) = h(\eta_i), \quad i = 1, \ldots, n,\ h \in C[0, 1].$$

THEOREM 2.4. *Let K be a nondegenerate totally positive kernel. Then*

$$\|K - LK\| = \lambda_{n+1}^{1/2}.$$

This theorem say that for the class $\mathscr{K}$, approximating $f \in \mathscr{K}$ by Lf is as good as approximating f by Pf where P is the orthogonal projection onto

$$X_n^1 = [K(\cdot, \xi_1), \ldots, K(\cdot, \xi_n)].$$

Proof. The proof is quite similar to the proof of Theorem 2.3. First, we compute the kernel of the operator $R = K - LK$,

$$R(x, y) = (K - LK)(x, y) = \frac{K\begin{pmatrix} x, \eta_1, \ldots, \eta_n \\ y, \xi_1, \ldots, \xi_n \end{pmatrix}}{K\begin{pmatrix} \eta_1, \ldots, \eta_n \\ \xi_1, \ldots, \xi_n \end{pmatrix}},$$

and then demonstrate, as in the proof of Theorem 2.3, that $R^*R\phi_{n+1} = \lambda_{n+1}\phi_{n+1}$. Finally,

$$(R^*R)(x, y) \operatorname{sgn} \phi_{n+1}(x) \operatorname{sgn} \phi_{n+1}(y) \geq 0, \quad x, y \in [0, 1],$$

because

$$(R^*R)(x, y) = \int_0^1 \frac{K\begin{pmatrix} \sigma, \eta_1, \ldots, \eta_n \\ x, \xi_1, \ldots, \xi_n \end{pmatrix} K\begin{pmatrix} \sigma, \eta_1, \ldots, \eta_n \\ y, \xi_1, \ldots, \xi_n \end{pmatrix}}{\left(K\begin{pmatrix} \eta_1, \ldots, \eta_n \\ \xi_1, \ldots, \xi_n \end{pmatrix}\right)^2} d\sigma$$

and we conclude that the largest eigenvalue of R^*R is given by $\lambda_{\max}(R^*R) = \lambda_{n+1}$. Thus the proof is complete.

Lemma 2.3 leads also to a result similar to Theorem 2.4 for convolution operators. Thus, in Example 2.2, interpolation at $\alpha, \ldots, \alpha + (2n-1)/2n$ by the subspace

$$X_{2n}^1 = \left[\phi(\cdot - \beta), \ldots, \phi\left(\cdot - \beta - \frac{2n-1}{2n}\right)\right]$$

is an optimal procedure for the class $\mathscr{K}_\phi$.

3. n-widths under restricted approximation

In this section we study n-widths under restricted approximation. We begin by describing our initial motivation for this problem.

Suppose we sample a function $f(x)$ at s points $\mathbf{x} = (x_1, \ldots, x_s)$, $0 < x_1 < \cdots < x_s < 1$. Given only that $f \in \mathscr{K}$ and the data $\mathbf{f} = (f(x_1), \ldots, f(x_s))$ we wish to find an optimal method of estimating $f(x)$. To describe what we mean by this we let T be any mapping (not necessarily linear) from $R^s \to C[0, 1]$. This mapping determines the estimator $Sf = T\mathbf{f}$ for f and the error, given only that $f \in \mathscr{K}$, is

$$E(\mathbf{x}; S) = \sup_{f \in \mathscr{K}} \|f - Sf\|.$$

We will say that S_0 is an optimal estimator for $\mathscr{K}$, provided that

$$E(\mathbf{x}; S_0) = \inf E(\mathbf{x}; S)$$

where the infimum is taken over all mappings from R^s into $C[0, 1]$; see [12] for a related problem.

Let $P_{\mathbf{x}}$ be the orthogonal projection of H onto the subspace

$$X_s = [K(x_1, \cdot), \ldots, K(x_s, \cdot)].$$

For $f = Kh \in \mathscr{K}$ we define $S_{\mathbf{x}} f = KPh$. Since $Ph = 0$ for any h such that $(Kh)(x_i) = 0$, $i = 1, \ldots, n$, $S_{\mathbf{x}}$ is a well-defined estimator which uses only the information

$$\mathbf{f} = (f(x_1), \ldots, f(x_s)).$$

We define

$$C(\mathbf{x}) = \sup \{\|f\| : f(x_i) = 0, i = 1, \ldots, s, f \in \mathscr{K}\}.$$

THEOREM 3.1. *$S_{\mathbf{x}}$ is an optimal linear estimator for $\mathscr{K}$ and $E(\mathbf{x}) = C(\mathbf{x})$.*

Proof. Let $f \in \mathscr{K}$, $f(x_i) = 0$, $i = 1, \ldots, n$. Then for any mapping $T: R^s \to C[0, 1]$,

$$\|f\| \le \tfrac{1}{2}[\|f - T(0)\| + \|f + T(0)\|] \le \sup_{f \in \mathscr{K}} \|f - Sf\|$$

where $Sf = T\mathbf{f}$. Hence $C(\mathbf{x}) \le E(\mathbf{x})$.

The reverse inequality follows from the following reasoning. First observe that

$$\|f - S_{\mathbf{x}} f\| = \|Kh - KP_{\mathbf{x}} h\| = \|K(h - P_{\mathbf{x}} h)\|.$$

Now, $g = K(h - P_{\mathbf{x}} h) \in \mathscr{K}$ because $\|h - P_{\mathbf{x}} h\| \le \|h\| \le 1$ and also, $g(x_i) = 0$, $i = 1, \ldots, s$ by the definition of $P_{\mathbf{x}}$. Hence

$$E(\mathbf{x}; S_{\mathbf{x}}) = \sup_{f \in \mathscr{K}} \|f - S_{\mathbf{x}} f\| \le C(\mathbf{x}),$$

and the theorem is proved.

According to Theorem 2.3, when K is a totally positive nondegenerate kernel,

$$E(\boldsymbol{\eta}; S_{\boldsymbol{\eta}}) = \|K - KP_{\boldsymbol{\eta}}\| = d_s(\mathscr{K}; H) \le E(\mathbf{x}, S_{\mathbf{x}})$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_s)$, $\psi_{s+1}(\eta_i) = 0$, $i = 1, \ldots, s$. Thus, in this case, the best place to sample $f \in \mathscr{K}$ is at $\eta_1, \ldots, \eta_s$.

Since $\psi_{s+1} = K\phi_{s+1} \in \mathscr{K}$ and $\|\psi_{s+1}\| = \lambda_{s+1}^{1/2}$, Theorem 2.3 says that ψ_{s+1} is the "worst" function in the class $\mathscr{K}$ to approximate by s-dimensional subspaces. Our above remarks say that, by sampling an $f \in \mathscr{K}$ at the zeros of ψ_{s+1}, we obtain the most information about the function f.

Now, we ask the following question. If we wish to fix the first s sample locations at $x_1, \ldots, x_s$, where is the best place to sample $f \in \mathscr{K}$ at n additional locations? Thus we wish to determine

$$\min_{\mathbf{y}} E(\mathbf{z}, S_{\mathbf{z}}) \quad \mathbf{y} = (y_1, \ldots, y_n), \mathbf{z} = (x_1, \ldots, x_s, y_1, \ldots, y_n). \tag{3.1}$$

Since $(f - P_{\mathbf{z}} f)(x_i) = 0$, $i = 1, \ldots, s$, $f \in \mathscr{K}$, we are lead to the computation of the following (restricted) n-width:

$$d_n^s(\mathscr{K}; H) = \inf_{X_{n+s}} \sup_{f \in \mathscr{K}} \inf \{\|f - g\| : (f - g)(x_i) = 0, i = 1, \ldots, s, g \in X_{n+s}\}.$$

The value of $d_n^s(\mathscr{K}; H)$ is identified as follows:
Let

$$\begin{aligned} K_{\mathbf{x}} &= K(I - P_{\mathbf{x}}), \\ K_{\mathbf{x}}^* K_{\mathbf{x}} h_l &= \lambda_l h_l, \quad (h_l, h_m) = \delta_{lm}, \; l, m = 1, 2, \ldots, \\ \lambda_l &= \lambda_l(\mathbf{x}), \\ \lambda_1 &\geq \lambda_2 \geq \lambda_3 \geq \cdots. \end{aligned}$$

Then $d_n^s(\mathscr{K}; H) = \lambda_{n+1}^{1/2}$, and $X_n^0 = [K_{\mathbf{x}} h_1, \ldots, K_{\mathbf{x}} h_n]$ is an optimal subspace (we will prove this fact later). This result, of course, does not require $K(x, y)$ to be a nondegenerate totally positive kernel. However, when this hypothesis is in force we can appeal to some results of Gantmacher and Krein, [1, p. 236–242] and make assertions, as before, about the zeros of the eigenfunctions h_{n+1} and $K_{\mathbf{x}} h_{n+1}$.

The kernel of the integral operator $R_{\mathbf{x}} = K_{\mathbf{x}} K_{\mathbf{x}}^*$ is given by

$$R_{\mathbf{x}}(x, y) = \frac{KK^* \begin{pmatrix} x, x_1, \ldots, x_n \\ y, x_1, \ldots, x_n \end{pmatrix}}{KK^* \begin{pmatrix} x_1, \ldots, x_n \\ x_1, \ldots, x_n \end{pmatrix}}. \tag{3.2}$$

Kernels of this type are studied by Gantmacher and Krein in [1, p. 236–242] where it is pointed out that if $(KK^*)(x, y)$ is the influence function for a "continuum" for which "stationary (hinged) supports" are introduced at $x_1, \ldots, x_n$ then $R_{\mathbf{x}}(x, y)$ is the influence function for the resulting constrained continuum.

Let $\varepsilon(x) = (-1)^i$, $x_i < x < x_{i+1}$, $i = 0, 1, \ldots, n$ $(x_0 = 0, x_{n+1} = 1)$ and define $\tilde{R}_{\mathbf{x}}(x, y) = \varepsilon(x)\varepsilon(y) R_{\mathbf{x}}(x, y)$. When K is a nondegenerate totally positive kernel then by Sylvester's determinant identity [3, p. 3],

$$\tilde{R}_{\mathbf{x}} \begin{pmatrix} y_1, \ldots, y_n \\ y_1, \ldots, y_n \end{pmatrix} > 0, \quad 0 < y_1 < \cdots < y_n < 1,$$

$$\tilde{R}_{\mathbf{x}} \begin{pmatrix} z_1, \ldots, z_n \\ y_1, \ldots, y_n \end{pmatrix} > 0, \quad 0 < z_1 < \cdots < z_n < 1, 0 < y_1 < \cdots < y_n < 1,$$

where the z's and y's are chosen distinct from $x_1, \ldots, x_s$. Let $f_l = K_{\mathbf{x}} h_l$, $l = 1, 2, \ldots$. Then $R_{\mathbf{x}} f_l = \lambda_l(\mathbf{x}) f_l$, $l = 1, 2, \ldots$, and by Gantmacher and Krein [1, p. 236–242] $\lambda_1(\mathbf{x}) > \lambda_2(\mathbf{x}) > \cdots$, and f_{n+1} has exactly $n + s$ simple zeros. Exactly s of them are at $x_1, \ldots, x_s$ because $(K_{\mathbf{x}} h)(x_i) = 0$, $i = 1, \ldots, s$, for any $h \in \mathscr{K}$, and there are n more which we denote by $\eta_1(\mathbf{x}), \ldots, \eta_n(\mathbf{x})$, $0 < \eta_1(\mathbf{x}) < \cdots < \eta_n(\mathbf{x}) < 1$. Furthermore, if

$$\tilde{f}_l(x) = \varepsilon(x) f_l(x)$$

then $\{\tilde{f}_1(x), \ldots, \tilde{f}_n(x), \ldots\}$ is a Markov system on $(0, 1) - \{x_1, \ldots, x_s\}$.

We now have enough information to solve our problem on optimal estimation.

THEOREM 3.2. *Let* $\mathbf{z} = (x_1, \ldots, x_s, y_1, \ldots, y_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ *and suppose* K *is a nondegenerate totally positive kernel. Then*

$$\inf_{\mathbf{y}} E(\mathbf{z}; S_{\mathbf{z}}) = E(\mathbf{z}_0; S_{\mathbf{z}_0}) = \lambda_{n+1}^{1/2}(\mathbf{x}) \quad \textit{where } \mathbf{z}_0 = (x_1, \ldots, x_s, \eta_1(\mathbf{x}), \ldots, \eta_n(\mathbf{x})).$$

Proof. Let us begin by showing that $d_n^s(\mathscr{K}; H) = d_n(\mathscr{K}_{\mathbf{x}}; H) = \lambda_{n+1}^{1/2}(\mathbf{x})$ where,

$$\mathscr{K}_{\mathbf{x}} = \{K_{\mathbf{x}} h \colon \|h\| \le 1\}.$$

Note that, since $K_{\mathbf{x}} h = K(h - P_{\mathbf{x}} h)$, $\mathscr{K}_{\mathbf{x}} = \{f \in \mathscr{K} \colon f(x_i) = 0,\ i = 1, 2, \ldots, s\}$. Clearly,

$$d_n^s(\mathscr{K}; H) \ge \inf_{X_n} \sup_{f \in \mathscr{K}_{\mathbf{x}}} \inf_{g \in X_n} \|f - g\| = d_n(\mathscr{K}_{\mathbf{x}}; H) = \lambda_{n+1}^{1/2}(\mathbf{x}).$$

Let

$$Z_{n+s}^0 = [KK^*(\cdot, x_1), \ldots, KK^*(\cdot, x_s), f_1, \ldots, f_n] \quad \text{and} \quad X_n^0 = [f_1, \ldots, f_n].$$

Then $d_n^s(K; H)$ is bounded by

$$\sup_{f \in \mathscr{K}} \inf \{\|f - g\| \colon f(x_i) - S_{\mathbf{x}} f(x_i) - g(x_i) = 0, g \in X_n^0\}$$

$$= \sup_{f \in \mathscr{K}_{\mathbf{x}}} \inf \{\|f - g\| \colon g \in X_n^0\} = \lambda_{n+1}^{1/2}(\mathbf{x}).$$

the last equality follows from the fact that X_n^0 is an optimal subspace for $\mathscr{K}_{\mathbf{x}}$.

Now, let us prove the theorem. We define $P_{\mathbf{y},\mathbf{x}}$ to be the orthogonal projection of H onto $[K_{\mathbf{x}}(y_1, \cdot), \ldots, K_{\mathbf{x}}(y_n, \cdot)]$. Recall that $P_{\mathbf{z}}$, $P_{\mathbf{x}}$ are respectively the orthogonal projections onto

$$[K(z_1, \cdot), \ldots, K(z_{n+s}, \cdot)] \quad \text{and} \quad [K(x_1, \cdot), \ldots, K(x_s, \cdot)]$$

where $\mathbf{z} = (x_1, \ldots, x_s, y_1, \ldots, y_n)$, $\mathbf{x} = (x_1, \ldots, x_s)$, $\mathbf{y} = (y_1, \ldots, y_n)$. Hence

$$K - KP_{\mathbf{z}} = K_{\mathbf{x}} - K_{\mathbf{x}} P_{\mathbf{y},\mathbf{x}} \tag{3.3}$$

because $I - P_{\mathbf{z}} = (I - P_{\mathbf{x}})(I - P_{\mathbf{z},\mathbf{x}})$. We conclude from Theorem 3.1 that

$$E(\mathbf{z}; P_{\mathbf{z}}) = \|K - KP_{\mathbf{z}}\| = \|K_{\mathbf{x}} - K_{\mathbf{x}}P_{\mathbf{y},\mathbf{x}}\| \geq d_n(\mathscr{K}_{\mathbf{x}}; H) = \lambda_{n+1}^{1/2}(\mathbf{x})$$

for all $\mathbf{y}$. To complete the proof, we will demonstrate that

$$\|K_{\mathbf{x}} - K_{\mathbf{x}}P_{\boldsymbol{\eta},x}\| = \lambda_{n+1}^{1/2}(\mathbf{x}) \quad \text{where } \boldsymbol{\eta} = (\eta_1(\mathbf{x}), \ldots, \eta_n(\mathbf{x})). \tag{3.4}$$

Then combining this fact with (3.3) the proof will be complete.

Let $T = K_{\mathbf{x}} - K_{\mathbf{x}}P_{\boldsymbol{\eta},\mathbf{x}}$. Then using the fact that

$$T = K - KP_{\mathbf{z}0}, \quad \mathbf{z}_0 = (x_1, \ldots, x_s, \eta_1(\mathbf{x}), \ldots, \eta_n(\mathbf{x}))$$

we may compute the kernel of T^*T to be

$$(T^*T)(x, y) = (TT^*)(y, x) = \frac{KK^*\begin{pmatrix} y, x_1, \ldots, x_s, \eta_1(\mathbf{x}), \ldots, \eta_n(\mathbf{x}) \\ x, x_1, \ldots, x_s, \eta_1(\mathbf{x}), \ldots, \eta_n(\mathbf{x}) \end{pmatrix}}{KK^*\begin{pmatrix} x_1, \ldots, x_s, \eta_1(\mathbf{x}), \ldots, \eta_n(\mathbf{x}) \\ x_1, \ldots, x_s, \eta_1(\mathbf{x}), \ldots, \eta_n(\mathbf{x}) \end{pmatrix}}$$

Thus, $T^*T(x, y) \operatorname{sgn} f_{n+1}(x) \operatorname{sgn} f_{n+1}(y) \geq 0$. But, because $T = K_{\mathbf{x}} - K_{\mathbf{x}}P_{\boldsymbol{\eta},\mathbf{x}}$, then (replacing K by $K_{\mathbf{x}}$) as in the proof of Theorem 2.3 we conclude that

$$T^*Tf_{n+1} = \lambda_{n+1}(\mathbf{x})f_{n+1} \quad \text{and} \quad \lambda_{\max}(T^*T) = \lambda_{n+1}(\mathbf{x});$$

hence (3.4) is verified.

According to the above theorem, both

$$X_n^0 = [f_1, \ldots, f_n] \quad \text{and} \quad X_n^2 = [(K_{\mathbf{x}}K_{\mathbf{x}}^*)(\cdot, \eta_1(\mathbf{x})), \ldots, (K_{\mathbf{x}}K_{\mathbf{x}}^*)(\cdot, \eta_n(\mathbf{x}))]$$

are optimal subspaces for $d_n(\mathscr{K}_{\mathbf{x}}; H)$. There is, of course, a third optimal subspace based upon the zeros of $h_{n+1}(x)$ (it can be shown that $h_{n+1}(x)$ has $n + s$ zeros) and again interpolation at the zeros of $f_{n+1}(x)$ is an optimal procedure. However, the discussion of these facts will take us too far from our discussion on optimal estimation. Instead, let us note that the results we have been considering in Section 2 and Section 3 extend when the operator K maps

$$H_\alpha = \left\{f: \int_0^1 f^2(t)\, d\alpha(t) < \infty\right\} \quad \text{onto} \quad H_\beta = \left\{f: \int_0^1 f^2(t)\, d\beta(t) < \infty\right\}$$

provided that $d\alpha$, $d\beta$ are finite measures on [0, 1] which have mass throughout the interval (the case when $d\alpha$, $d\beta$ are discrete corresponds to the matrix version of the results of Section 2 and will be dealt with in Section 4). A particularly relevant choice for $d\alpha$, $d\beta$ is

$$d\alpha(t) = d\beta(t) = dt + \tau \sum_{j=1}^{s} \delta(t - x_j)$$

($\delta(t - x_j)$ is the point mass at x_j) where τ is a nonnegative constant. Then the $(n + s)$-width of

$$\mathscr{K}_\alpha = \left\{Kh: \int_0^1 h^2(t)\, d\alpha(t) \le 1\right\}$$

in H_α varies from the $(n + s)$-width of $\mathscr{K}$ in H, $\tau = 0$, to the restricted n-width of $\mathscr{K}$ in H when $\tau \to \infty$. These facts are not difficult to prove.

We now turn our attention to the matrix version of the results in Section 2.

4. Matrix version

In this section we discuss the matrix version of the results in Section 2.

The inner product of $x, y \in R^N$ is denoted by $(x, y) = \sum_{j=1}^N x_j y_j$, and $\|x\|^2 = (x, x)$. Subscripts will be used to denote components of a vector while superscripts are used to distinguish between vectors.

We assume that $A = (a_{ij})$, $i, j = 1, \ldots, N$, is an $N \times N$ strictly totally positive matrix, that is, all the minors of A are positive. Then by the Gantmacher and Krein Theorem [1], A^*A has eigenvalues and eigenvectors such that

$$\begin{gathered}
A^*Ax^n = \lambda_n x^n, \quad (x^n, x^m) = \delta_{nm}, n, m = 1, \ldots, N, \\
\lambda_1 > \lambda_2 > \cdots > \lambda_N > 0, \\
S^+(Ax^{n+1}) = S^-(Ax^{n+1}) = n, \quad n = 0, 1, \ldots, N - 1, \\
S^+(x^{n+1}) = S^-(x^{n+1}) = n, \quad n = 0, 1, \ldots, N - 1.
\end{gathered}$$

Here $S^-(x)$ equals the number of actual sign changes in the vector x where zero components are discarded, while $S^+(x)$ is the maximum number of sign changes obtainable by adding 1 or -1 to the zero components of x.

Hence $x_1^{n+1} x_N^{n+1} \ne 0$ and if $x_j^{n+1} = 0$, $1 < j < N$, then $x_{j-1}^{n+1} x_{j+1}^{n+1} < 0$. Also, there exist by the above equations indices $0 = l_0 < l_1 < \cdots < l_n < N = l_{n+1}$ such that $\pm x_i^{n+1}(-1)^j \ge 0$, $l_j < i \le l_{j+1}$, $j = 0, 1, \ldots, n$.

Adopt convention that l_j is such that either

$$x_{l_j}^{n+1} x_{l_j+1}^{n+1} < 0 \tag{4.1}$$

or

$$x_{l_j}^{n+1} = 0 \quad \text{and} \quad x_{l_j-1}^{n+1} x_{l_j+1}^{n+1} < 0. \tag{4.2}$$

For each j, $j = 1, \ldots, n$ we define an N-dimensional vector e^j as follows: If (4.1) holds, let

$$(e^j)_k = \begin{cases} |x_k^{n+1}|^{-1}, & k = l_j, l_j + 1, \\ 0, & \text{otherwise,} \end{cases}$$

while if (4.2) arises we let $(e^j)_k = \delta_{l_j k}$. Thus

$$(e^j, x^{n+1}) = 0, \quad j = 1, \ldots, n. \tag{4.3}$$

Similarly, for the vector $y^{n+1} = Ax^{n+1}$ we construct $f^1, \dots, f^n \in R^N$, such that

$$(f^j, y^{n+1}) = 0, \quad j = 1, \dots, n. \tag{4.4}$$

THEOREM 3.1. *Let A be an $N \times N$ strictly totally positive matrix. Then the n-width of $A = \{Ax: \|x\| \le 1\}$, $n < N$, is given by*

$$d_n(A; l^2) = \begin{cases} \lambda_{n+1}^{1/2}, & n = 0, 1, \dots, N-1, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$X_n^0 = [y^1, \dots, y^n], \quad X_n^1 = [Ae^1, \dots, Ae^n], \quad X_n^2 = [AA^*f^1, \dots, AA^*f^n]$$

are optimal subspaces. Furthermore, if R is the $N \times N$ matrix defined by requiring that R: $R^N \to X_n^1$ and $(x - Rx, f^i) = 0$, $i = 1, \dots, n$, $x \in R^N$, for all $x \in R^N$ then $\|A - RA\| = \lambda_{n+1}^{1/2}$.

Proof. The proof of Theorem 3.1 follows the pattern of proof given in Theorems 2.3 and 2.4. We will briefly indicate how our last assertion is proved.

First, let us note that, in obvious symbolic notation,

$$Gx = Ax - RAx = \frac{\begin{vmatrix} (f^1, Ae^1) & \cdots & (f^1, Ae^n) & (f^1, Ax) \\ \vdots & & \vdots & \vdots \\ (f^n, Ae^1) & \cdots & (f^n, Ae^n) & (f^n, Ax) \\ Ae^1 & \cdots & Ae^n & Ax \end{vmatrix}}{\begin{vmatrix} (f^1, Ae^1) & \cdots & (f^1, Ae^n) \\ \vdots & & \vdots \\ (f^n, Ae^1) & \cdots & (f^n, Ae^n) \end{vmatrix}} \tag{4.5}$$

The determinant in the denominator is nonzero because A is strictly totally positive and the $n \times N$ matrices $F = (f_j^i)$, $E = (e_j^i)$ are totally positive (all minors are nonnegative) and have rank n. According to the orthogonality conditions (4.3) and (4.4) $G^*Gx^{n+1} = \lambda_{n+1} x^{n+1}$ and thus we need only verify that

$$(G^*G)_{ij} \operatorname{sgn} x_i^{n+1} \operatorname{sgn} x_j^{n+1} \ge 0. \tag{4.6}$$

Appealing to (4.5), we obtain by direct computation,

$$G_{ij} = \frac{\det\limits_{k,l=1,\dots,n+1} |(f^k, Ae^l)|}{\det\limits_{k,l=1,\dots,n} |(f^k, Ae^l)|}$$

where we define $f^{n+1} = u^i$, the ith unit vector, $(u^k)_l = \delta_{kl}$, $k, l = 1, \dots, N$, and $e^{n+1} = u^j$. The matrix whose columns are composed of the vectors $e^1, \dots, e^{n+1}$ has the property that the signs of all its $(n+1)$st minors are $(-1)^{r+n}$, if

$l_r < j \le l_{r+1}$. A similar result holds for $f^1, \ldots, f^{n+1}$. Hence using the total positivity of A, as well, we obtain

$$G_{ij} \operatorname{sgn} y_i^{n+1} \operatorname{sgn} x_j^{n+1} \ge 0.$$

Thus (4.6) is verified and the proof of the theorem is finished.

5. Further extensions

In this section we will indicate how Theorem 2.3 can be extended to sets of the form $\mathcal{K}_r = X_r + \mathcal{K}$ where X_r is some fixed r-dimensional subspace of H.

Let $k_1(x), \ldots, k_r(x)$ be continuous functions on [0, 1] and define

$$\begin{aligned}(5.1)\quad \mathcal{K}_r &= \left\{\sum_{j=1}^{r} a_j k_j(x) + \int_0^1 K(x, y)h(y)\,dy \colon \|h\| \le 1, (a_1, \ldots, a_r) \in R^r\right\}\\ &= X_r + \mathcal{K}.\end{aligned}$$

The main prototype, for us, of this class of examples is the Sobolev class

$$W^{2,r}[0, 1] = \{f \colon f^{(r-1)} \text{ abs. cont. on } (0, 1), f^{(r)} \in L^2[0, 1], \|f^{(r)}\| \le 1\}$$

which may be written in the form (5.1) by using Taylor's theorem with remainder:

$$(5.2)\qquad f(x) = \sum_{j=0}^{r-1} \frac{f^{(j)}(0)}{j!} x^j + \frac{1}{(r-1)!} \int_0^1 (x - y)_+^{r-1} f^{(r)}(y)\,dy$$

($x_+^{r-1} = x^{r-1}$, $x \ge 0$, zero otherwise).

Let Q_r be the orthogonal projection of H onto $X_r = [k_1, \ldots, k_r]$. Then Theorem 2.1 easily extends to $\mathcal{K}_r$ as follows: We define $K_r = (1 - Q_r)K$. Then $K_r^* K_r$ is a completely continuous, symmetric positive semi-definite operator with eigenvalues $\lambda_{1,r} \ge \lambda_{2,r} \ge \cdots \ge 0$ and corresponding orthonormal eigenfunctions,

$$K_r^* K_r \phi_{n,r} = \lambda_{n,r}, \quad (\phi_{n,r}, \phi_{m,r}) = \delta_{nm}, \; n, m = 1, 2, \ldots.$$

Let $\psi_{n,r} = K_r \phi_{n,r}$. Then

$$K_r K_r^* \psi_{n,r} = \lambda_{n,r} \psi_{n,r}, \quad (\psi_{n,r}, \psi_{m,r}) = \lambda_n \delta_{nm}, \; n, m = 1, 2, \ldots,$$

and

$$d_n(\mathcal{K}_r) = \begin{cases} \infty, & n < r, \\ \lambda_{n-r+1,r}^{1/2}, & n \ge r. \end{cases}$$

When $n \ge r$, $X_n^0 = [k_1, \ldots, k_r, \psi_{1,r}, \ldots, \psi_{n-r,r}]$ is an optimal subspace for the n-width of $\mathcal{K}_r$.

For the analog of Theorems 2.3 and 2.4 we require the following assumptions. For any points $0 \le s_1 < \cdots < s_m \le 1$, $0 \le t_1 < \cdots < t_m \le 1$ and any

$m \leq 0$, the determinant

$$K\begin{pmatrix} 1, \ldots, r, s_1, \ldots, s_m \\ t_1, \ldots, t_r, t_{r+1}, \ldots, t_{r+m} \end{pmatrix} = \begin{vmatrix} k_1(t_1) & \cdots & k_1(t_{m+r}) \\ \vdots & & \vdots \\ k_r(t_1) & \cdots & k_r(t_{m+r}) \\ K(t_1, s_1) & \cdots & K(t_{m+r}, s_1) \\ \vdots & & \vdots \\ K(t_1, s_m) & \cdots & K(t_{m+r}, s_m) \end{vmatrix}$$

is nonnegative. The linear spaces

$$[k_1, \ldots, k_r, K(\cdot, s_1), \ldots, K(\cdot, s_m)] \quad \text{and} \quad [k_1, \ldots, k_r, K(s_1, \cdot), \ldots, K(s_m, \cdot)]$$

have dimension $r + m$ for all $0 < s_1 < \cdots < s_m < 1$, $m = 1, 2, \ldots$, and $\{k_1, \ldots, k_r\}$ is a Chebyshev system on $(0, 1)$, that is

$$k\begin{pmatrix} 1, \ldots, r \\ s_1, \ldots, s_r \end{pmatrix} > 0, \quad 0 < s_1 < \cdots < s_r < 1.$$

It is a fundamental result from the theory of spline functions that these conditions hold in the special case (3.2) [15]. A large class of examples satisfying these assumptions may be obtained from [9] by considering a totally disconjugate differential operator subject to sign consistent boundary conditions.

We will now develop an extension of Theorem 2.3 to $\mathscr{K}_r$, in a series of lemmas. We begin with:

LEMMA 5.1. *$K_r^* K_r(x, y)$ is a nondegenerate totally positive kernel.*

Before we prove this lemma let us note that $K_r(x, y)$ itself is *not* totally positive, since the range of K_r is orthogonal to the Chebyshev subspace X_r. Hence $K_r h$ has at least r sign changes on $(0, 1)$ for every nontrivial $h \in L^2$.

Proof. From the definition of Q_r,

$$(h - Q_r h)(x) = \frac{\begin{vmatrix} (k_1, k_1) & \cdots & (k_1, k_r) & (k_1, h) \\ \vdots & & \vdots & \vdots \\ (k_r, k_1) & \cdots & (k_r, k_r) & (k_r, h) \\ k_1(x) & \cdots & k_r(x) & h(x) \end{vmatrix}}{G(k_1, \ldots, k_r)}$$

where $G(k_1, \ldots, k_r) = \det |(k_i, k_j)|$, the Grammian of $k_1, \ldots, k_r$. Thus

$$(K_r^* K_r h)(x) = \frac{\begin{vmatrix} (k_1, k_1) & \cdots & (k_1, k_r) & (K^* k_1, h) \\ \vdots & & \vdots & \vdots \\ (k_r, k_1) & \cdots & (k_r, k_r) & (K^* k_r, h) \\ (K^* k_1)(x) & \cdots & (K^* k_r)(x) & (K^* K h)(x) \end{vmatrix}}{G(k_1, \ldots, k_r)}$$

and

$$(K_r^* K_r)(x, y) = \frac{\begin{vmatrix} (k_1, k_1) & \cdots & (k_1, k_r) & (k_1, K(\cdot, y)) \\ \vdots & & \vdots & \vdots \\ (k_r, k_1) & \cdots & (k_r, k_r) & (k_r, K(\cdot, y)) \\ (K(\cdot, x), k_1) & \cdots & (K(\cdot, x), k_r) & (K(\cdot, x), K(\cdot, y)) \end{vmatrix}}{G(k_1, \ldots, k_r)}$$

Hence by Sylvester's determinant identity [3], we have

$$(K_r^* K_r)\begin{pmatrix} x_1, \ldots, x_n \\ y_1, \ldots, y_n \end{pmatrix} =$$

$$\frac{\begin{vmatrix} (k_1, k_1) & \cdots & (k_1, k_r) & (k_1, K(\cdot, y_1)) & \cdots & (k_1, K(\cdot, y_n)) \\ \vdots & & \vdots & \vdots & & \vdots \\ (k_r, k_1) & \cdots & (k_r, k_r) & (k_r, K(\cdot, y_1)) & \cdots & (k_r, K(\cdot, y_n)) \\ (K(\cdot, x_1), k_1) & \cdots & (K(\cdot, x_1), k_r) & (K(\cdot, x_1), K(\cdot, y_1)) & \cdots & (K(\cdot, x_1), K(\cdot, y_n)) \\ \vdots & & \vdots & \vdots & & \vdots \\ (K(\cdot, x_n), k_1) & \cdots & (K(\cdot, x_n), k_r) & (K(\cdot, x_n), K(\cdot, y_1)) & \cdots & (K(\cdot, x_n), K(\cdot, y_n)) \end{vmatrix}}{G(k_1, \ldots, k_r)}$$

Using the basic composition formula we obtain

$$K_r^* K_r \begin{pmatrix} x_1, \ldots, x_n \\ y_1, \ldots, y_n \end{pmatrix} = \underset{0<\sigma_1<\cdots<\sigma_{n+r}<1}{\int_0^1 \cdots \int_0^1} K\begin{pmatrix} 1, \ldots, r, x_1, \ldots, x_n \\ \sigma_1, \ldots, \sigma_r, \sigma_{r+1}, \ldots, \sigma_{r+n} \end{pmatrix}$$

$$\times K\begin{pmatrix} 1, \ldots, r, y_1, \ldots, y_n \\ \sigma_1, \ldots, \sigma_r, \sigma_{r+1}, \ldots, \sigma_{n+r} \end{pmatrix} d\sigma_1, \ldots, d\sigma_{n+r}.$$

The lemma now follows from our assumptions on $\mathscr{K}_r$.

LEMMA 5.2. *The set of functions* $\{k_1, k_2, \ldots, k_r, K_r\phi_{1,r}, \ldots, K_r\phi_{n,r}, \ldots\}$ *form a Markov system on* $(0, 1)$.

Proof. According to the identity

$$\begin{vmatrix} k_1(x_1) & \cdots & k_1(x_{n+r}) \\ \vdots & & \vdots \\ k_r(x_1) & \cdots & k_r(x_{n+r}) \\ (K_r\phi_{1,r})(x_1) & \cdots & (K_r\phi_{1,r})(x_{n+r}) \\ \vdots & & \vdots \\ (K_r\phi_{n,r})(x_1) & \cdots & (K_r\phi_{n,r})(x_{n+r}) \end{vmatrix}$$

$$= \underset{0<\sigma_1<\cdots<\sigma_n<1}{\int_0^1 \cdots \int_0^1} \phi_r\begin{pmatrix} 1, \ldots, n \\ \sigma_1, \ldots, \sigma_n \end{pmatrix} K\begin{pmatrix} 1, \ldots, r, \sigma_1, \ldots, \sigma_n \\ x_1, \ldots, x_r, x_{r+1}, \ldots, x_{n+r} \end{pmatrix} d\sigma_1, \ldots, d\sigma_n,$$

where

$$\phi_r\begin{pmatrix}1, \ldots, n \\ \sigma_1, \ldots, \sigma_n\end{pmatrix} = \det |\phi_{i,r}(\sigma_j)|,$$

this lemma follows directly from Lemma 5.1 and Theorem 2.1.

LEMMA 5.3. *For* $n \geq 0$, $\psi_{n+1,r}$ *has exactly* $n + r$ *simple zeros in* $(0, 1)$.

Proof. Since $\psi_{n+1,r}$ is in the range of $I - Q_r$ we conclude that $(\psi_{n+1,r}, k_i) = 0$, $i = 1, 2, \ldots, r$. In addition,

$$\begin{aligned}(\psi_{n+1,r}, K_r\phi_{l,r}) &= (K_r\phi_{n+1,r}, K_r\phi_{l,r}) \\ &= (\phi_{n+1}, K_r^* K_r \phi_{l,r}) \\ &= \lambda_{l,r}(\phi_{n+1}, \phi_{l,r}) \\ &= 0, \quad l = 1, \ldots, n.\end{aligned}$$

Hence $\psi_{n+1,r}$ is orthogonal to the Chebyshev subspace

$$Z_{n+r} = [k_1, k_2, \ldots, k_r, K_r\phi_{1,r}, \ldots, K_r\phi_{n,r}].$$

Thus $\psi_{n+1,r}$ has at least $n + r$ sign changes in $(0, 1)$. However, $\psi_{n+1,r} \in Z_{n+r+1}$. Thus $\psi_{n+1,r}$ has at most $n + r$ zeros in $(0, 1)$. Therefore $\psi_{n+1,r}$ has exactly $n + r$ simple zeros in $(0, 1)$ and the lemma is proved.

We let

$$\phi_{n+1,r}(\xi_{i,r}) = 0, \quad 0 < \xi_{1,r} < \cdots < \xi_{n,r} < 1,$$

$$\psi_{n+1,r}(\eta_{i,r}) = 0, \quad 0 < \eta_{1,r} < \cdots < \eta_{n+r,r} < 1.$$

To state our extension of Theorem 2.3 we define an interpolation operator J_r from $C[0, 1]$ onto $[k_1, \ldots, k_r]$ by requiring that $(h - J_r h)(\eta_{i,r}) = 0$, $i = 1, \ldots, r$, $h \in H$. J_r is given explicitly by

$$(h - J_r h)(x) = \frac{\begin{vmatrix} k_1(\eta_{1,r}) & \cdots & k_r(\eta_{1,r}) & h(\eta_{1,r}) \\ \vdots & & \vdots & \vdots \\ k_1(\eta_{r,r}) & \cdots & k_r(\eta_{r,r}) & h(\eta_{r,r}) \\ k_1(x) & \cdots & k_r(x) & h(x) \end{vmatrix}}{K\begin{pmatrix}1, \ldots, r \\ \eta_{1,r}, \ldots, \eta_{r,r}\end{pmatrix}} \tag{5.3}$$

We set $\bar{K}_r = (I - J_r)K$, then we have:

THEOREM 5.1.

$$d_n(\mathscr{K}_r; H) = \begin{cases} \infty & \text{if } n < r, \\ \lambda_{n-r+1,r}^{1/2} & \text{if } n = r, \ldots, \end{cases}$$

and

$$X^0_{n+r} = [k_1, \ldots, k_r, K\phi_{1,r}, \ldots, K\phi_{n,r}],$$
$$X^1_{n+r} = [k_1, \ldots, k_r, K(\cdot, \xi_{1,r}), \ldots, K(\cdot, \xi_{n,r})],$$
$$X^2_{n+r} = [k_1, \ldots, k_r, (\bar{K}_r\bar{K}^*_r)(\cdot, \eta_{r+1,r}), \ldots, \bar{K}_r\bar{K}^*_r(\cdot, \eta_{n+r,r})]$$

are optimal subspaces for the $(n + r)$*-width of* $\mathscr{K}_r$.

Proof. Since $\delta(\mathscr{K}_r; X^1_{n+r}) = \|K_r - PK_r\|^2$ where P is the orthogonal projection onto $[K_r(\cdot, \xi_{1,r}), \ldots, K_r(\cdot, \xi_{n,r})]$ we conclude from Lemma 3.1 and the argument used to prove Theorem 2.3 (note that the proof only requires $K^*_rK_r$ to be nondegenerate totally positive) that $\delta(\mathscr{K}_r; X^1_{n+r}) = \lambda^{1/2}_{n+1,r}$. Thus X^1_{n+r} is an optimal subspace for the $(n + r)$-width of $\mathscr{K}_r$. To prove the optimality of X^2_{n+r} we observe that

$$\delta(\mathscr{K}_r; X^2_{n+r}) \leq \|\bar{K}_r - \bar{K}_rQ\|^2$$

where Q is now the orthogonal projection onto the subspace

$$[\bar{K}_r(\eta_{r+1}, \cdot), \ldots, \bar{K}_r(\eta_{n+r}, \cdot)].$$

A glance at the proof of Theorem 2.3 shows that X^2_{n+r} is an optimal subspace for $\mathscr{K}_r$, if we can demonstrate that

$$\text{(5.4)} \quad \operatorname{sgn} \bar{K}_r\bar{K}^*_r \begin{pmatrix} x, \eta_{r+1,r}, \ldots, \eta_{n+r,r} \\ y, \eta_{r+1,r}, \ldots, \eta_{n+r,r} \end{pmatrix} = \operatorname{sgn} \psi_{n+r,r}(x) \operatorname{sgn} \psi_{n+r,r}(y), \qquad x, y \in (0, 1).$$

(Observe that $\bar{K}_r\bar{K}^*_r(x, y)$ is not a totally positive kernel.) Let us now compute the Fredholm determinant for $K_rK^*_r$. According to (5.3),

$$\text{(5.5)} \qquad \bar{K}_r(x, y) = \frac{\begin{vmatrix} k_1(\eta_{1,r}) & \cdots & k_r(\eta_{1,r}) & K(\eta_{1,r}, y) \\ \vdots & & \vdots & \vdots \\ k_1(\eta_{r,r}) & \cdots & k_r(\eta_{r,r}) & K(\eta_{r,r}, y) \\ k_1(x) & \cdots & k_r(x) & K(x, y) \end{vmatrix}}{K\begin{pmatrix} 1, \ldots, r \\ \eta_{1,r}, \ldots, \eta_{r,r} \end{pmatrix}}.$$

Since $\bar{K}_r\bar{K}^*_r(x, y) = \int_0^1 \bar{K}_r(x, \sigma)\bar{K}_r(y, \sigma)\, d\sigma$ we have, by the basic composition formula,

$$\bar{K}_r\bar{K}^*_r \begin{pmatrix} x_1, \ldots, x_l \\ y_1, \ldots, y_l \end{pmatrix} = \underset{0<\sigma_1<\cdots<\sigma_l<1}{\int_0^1 \cdots \int_0^1} \bar{K}_r \begin{pmatrix} x_1, \ldots, x_l \\ \sigma_1, \ldots, \sigma_l \end{pmatrix} \bar{K}_r \begin{pmatrix} y_1, \ldots, y \\ \sigma_1, \ldots, \sigma_l \end{pmatrix} d\sigma_1, \ldots, d\sigma_l.$$

Now Sylvester's determinant identity applied to (5.5) gives

$$\bar{K}_r\begin{pmatrix} x_1, \ldots, x_l \\ y_1, \ldots, y_l \end{pmatrix} = \frac{K\begin{pmatrix} 1, \ldots, r, \sigma_1, \ldots, \sigma_l \\ \eta_{1,r}, \ldots, \eta_{r,r} x_1, \ldots, x_l \end{pmatrix}}{K\begin{pmatrix} 1, \ldots, r \\ \eta_{1,r}, \ldots, \eta_{r,r} \end{pmatrix}}.$$

Thus we obtain

$$\bar{K}_r\bar{K}_r^*\begin{pmatrix} x, \eta_{r+1,r}, \ldots, \eta_{r+r,r} \\ y, \eta_{r+1,r}, \ldots, \eta_{n+r,r} \end{pmatrix} = \frac{1}{\left(K\begin{pmatrix} 1, \ldots, r \\ \eta_1, \ldots, \eta_{r,r} \end{pmatrix}\right)^2} \underset{0<\sigma_1<\cdots<\sigma_{n+1}<1}{\int_0^1 \cdots \int_0^1}$$

$$\times K\begin{pmatrix} 1, \ldots, r, \sigma_1, \ldots, \sigma_n, \sigma_{n+1} \\ \eta_{1,r}, \ldots, \eta_{n+1,r}, x \end{pmatrix}$$

$$\times K\begin{pmatrix} 1, \ldots, r, \sigma_1, \ldots, \sigma_n, \sigma_{n+1} \\ \eta_{1,r}, \ldots, \eta_{n+r,r}, y \end{pmatrix} d\sigma_1, \ldots, d\sigma_{n+1}.$$

Hence (5.4) is valid and the proof of the theorem is completed.

LEMMA 5.1.

$$K\begin{pmatrix} 1, \ldots, r, \xi_{1,r}, \ldots, \xi_{n,r} \\ \eta_{1,r}, \ldots, \eta_{n+r,r} \end{pmatrix} > 0.$$

The proof of this lemma parallels the proof of Lemma 2.2 and we omit the details.

We can now define an interpolation operator

$$L_r\colon C[0, 1] \to X_n^1, \qquad X_n^1 = [k_1, \ldots, k_r, K(\cdot, \xi_{1,r}), \ldots, K(\cdot, \xi_{n,r})]$$

by requiring that $(L_r h)(\eta_{i,r}) = h(\eta_{i,r})$, $i = 1, \ldots, n + r$, $h \in C[0, 1]$, and, as before, we may verify that

$$R_r(x, y) = (K - L_r K)(x, y) = \frac{K\begin{pmatrix} 1, \ldots, r, \xi_{1,r}, \ldots, \xi_{n,r}, x \\ \eta_{1,r}, \ldots, \eta_{n+r,r}, y \end{pmatrix}}{K\begin{pmatrix} 1, \ldots, r, \xi_{1,r}, \ldots, \xi_{n,r} \\ \eta_{1,r}, \ldots, \eta_{n+r,r} \end{pmatrix}}$$

and $R_r^* R_r \phi_{n+1,r} = \lambda_{n+1,r} \phi_{n+1,r}$. Thus we have

$$(R_r^* R_r)(x, y) \operatorname{sgn} \phi_{n+1,r}(x) \operatorname{sgn} \phi_{n+1,r}(y) \geq 0, \; x, y \in [0, 1]$$

and:

THEOREM 5.2. $\|K - L_r K\| = \lambda_{n+1,r}^{1/2}$.

Hence, again interpolation is an optimal procedure for estimating the class $\mathscr{K}_r$.

Let us now apply Theorem 5.1 to an example discussed by Kolmogorov in [8].

Example 5.1. $k_i(t) = t^{i-1}$, $i = 1, \ldots, r$, $K(x, y) = 1/(r-1)!$, $x, y \in [0, 1]$. In this case

$$\mathscr{K}_r = \{f: f^{(r-1)} \text{ abs. cont.}, f^{(r)} \in L^2[0, 1], \|f^{(r)}\| \leq 1\} = W^{2,r}[0, 1].$$

The eigenvalue equation $K_r^* K_r \phi_{n,r} = \lambda_{n,r} \phi_{n,r}$, $n = 1, 2, \ldots$, is easily seen to be equivalent to

$$(3.6) \qquad y_{n,r}^{(2r)}(x) = \mu_{n,r} y_{n,r}(x), \qquad y_{n,r}^{(i)}(1) = y_{n,r}^{(i)}(0) = 0,\ i = 0, 1, \ldots, r-1,$$

where $y_{n,r} = K_r^* K_r \phi_{n,r}$ and $\mu_{n,r} \lambda_{n,r} = 1$. Thus the results concerning this example mentioned in the introduction follow from Theorems 5.1 and 5.2.

REFERENCES

1. F. R. GANTMACHER, M. G. KREIN, *Oszillations matrizen, Oszillationskerne and kleine Schwingungen mechanischer Systeme*, Akademie Verlag, Berlin, 1960.
2. A. IOFFE, V. M. TIHOMIROV, *Duality of convex functions and extremum problems*, Russian Math. Surveys, vol. 23 (1968), pp. 53–124.
3. S. KARLIN, *Total positivity*, vol. I, Stanford University Press, Stanford, Calif., 1968.
4. L. A. KARLOVITZ, *On a class of Kolmogorov n-width problems*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur., vol. 53 (1972), pp. 241–245.
5. ———, *Remarks on variational characterization of eigenvalues and n-width problems*, J. Math. Anal. Appl., vol. 53 (1976), pp. 99–110.
6. O. D. KELLOGG, *The oscillation of functions of an orthogonal set*, Amer. J. Math., vol. 38 (1916), pp. 1–5.
7. ———, *Orthogonal function sets arising from integral equations*, Amer. J. Math., vol. 40 (1918), pp. 145–154.
8. A. KOLMOGOROFF, *Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse*, Ann. of Math., vol. 37 (1936), pp. 107–110.
9. AVRAHAM A. MELKMAN, *Interpolation by splines satisfying mixed boundary conditions*, Israel J. Math., vol. 19 (1974), pp. 369–381.
10. ———, "n-widths under restricted approximation," in *Approximation Theory* II, G. G. Lorentz, C. K. Chui, L. L. Schumaker, eds., Academic Press, 1976.
11. C. A. MICCHELLI AND A. PINKUS, *On n-widths in L^∞*, Trans. Amer. Math. Soc., vol. 234 (1977), pp. 139–174.
12. C. A. MICCHELLI, T. J. RIVLIN AND S. WINOGRAD, *The optimal recovery of smooth functions*, Numer. Math., vol. 260 (1976), pp. 191–200.
13. V. M. TIHOMIROV, *Best methods of approximation and interpolation of differentiable functions in the space $C_{[-1,1]}$*, Math. USSR Sbornik, vol. 9 (1969), pp. 275–289.
14. V. M. TIKHOMIROV, *Diameters of sets in functional spaces and the theory of best approximations*, Russian Math. Surveys, vol. 15 (1960), pp. 75–111.
15. I. J. SCHOENBERG AND A. WHITNEY, *On Polya frequency functions III. The positivity of translation determinants with an application to the interpolation problem by spline curves*, Trans. Amer. Math., Soc., vol. 74 (1953), pp. 246–259.
16. H. S. SHAPIRO, *Topics in approximation theory*, Lecture Notes in Mathematics, vol. 187, Springer Verlag, Berlin, 1971.

17. A. Weinstein and W. Stenger, *Methods of intermediate problems for eigenvalues*, Academic Press, New York, 1972.

The Hebrew University
Jerusalem, Israel
Ben Gurion University
Beer Sheva, Israel
IBM Thomas J. Watson Research Center
Yorktown Heights, New York