Constr. Approx. (1985) 1: 15–62

CONSTRUCTIVE APPROXIMATION

n-Widths of Sobolev Spaces in L^p

Allan Pinkus

Abstract. Let $W_p^{(r)} = \{f: f \in C^{r-1}[0, 1], f^{(r-1)} \text{ abs. cont.}, \|f^{(r)}\|_p < \infty\}$, and set $B_p^{(r)} = \{f: f \in W_p^{(r)}, \|f^{(r)}\|_p \le 1\}$. We find the exact Kolmogorov, Gel'fand, linear, and Bernstein n-widths of $B_p^{(r)}$ in L^p for all $p \in (1, \infty)$. For the Kolmogorov n-width we show that for $n \ge r$ there exists an optimal subspace of splines of degree $r-1$ with $n-r$ fixed simple knots depending on p.

1. Introduction

Let X be a normed linear space and A a subset of X. For ease of exposition we assume that A is closed, convex, and centrally symmetric (i.e., $x \in A$ implies $-x \in A$). We introduce four quantities to be studied for specific choices of A and X. These four quantities are the Kolmogorov, linear, Gel'fand, and Bernstein n-widths. They are defined as follows:

1) The *Kolmogorov n-width of A in X* is given by

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

where the X_n range over all subspaces of X of dimension at most n. The Kolmogorov n-width measures the extent to which n-dimensional subspaces approximate A. If there exists an $X_n^\star$ of dimension at most n for which

$$d_n(A;X) = \sup_{x \in A} \inf_{y \in X_n^\star} \|x - y\|$$

then $X_n^\star$ is said to be optimal for $d_n(A;X)$.

2) The *linear n-width of A in X* is defined by

$$\delta_n(A;X) = \inf_{P_n} \sup_{x \in A} \|x - P_n(x)\|$$

where the P_n range over the set of continuous linear operators of X into X of rank at most n, i.e., whose range is of dimension at most n. The linear n-width replaces the

Date received: December 5, 1983. Communicated by Ronald A. DeVore.
AMS classification: 41A46, 41A15.
Key words and phrases: n-Widths, Total positivity, Eigenvalues, Eigenfunctions, Implicit function theorem.

best approximation of the Kolmogorov n-width by a linear approximant. This is often more useful in practice. If $\delta_n(A;X) = \sup\{\|x - P_n^\star(x)\|: x \in A\}$ and rank $P_n^\star \leq n$, then $P_n^\star$ is said to be optimal for $\delta_n(A;X)$.

3) The *Gel'fand n-width of A in X* is given by

$$d^n(A;X) = \inf_{L_n} \sup_{x \in A \cap L_n} \|x\|$$

where the L_n range over all closed subspaces of X of codimension at most n. The Gel'fand n-width is important as a lower bound in problems of optimal recovery (see, e.g., Traub, Wozniakowski [17] and Micchelli, Rivlin [11]). It measures the extent to which one can determine $x \in A$ given the value of n linear functionals on x. (It is also useful because of its duality relationship with the Kolmogorov n-width.) If $d^n(A;X) = \sup\{\|x\|: x \in A \cap L_n^\star\}$ and codim $L_n^\star \leq n$, then $L_n^\star$ is said to be optimal for $d^n(A;X)$.

4) The *Bernstein n-width of A in X* is given by

$$b_n(A;X) = \sup_{X_{n+1}} \sup\{\lambda: \lambda S(X_{n+1}) \subseteq A\}$$

where $S(X_{n+1}) = \{x: x \in X_{n+1}, \|x\| \leq 1\}$ and X_{n+1} ranges over all subspaces of X of dimension at least $n+1$. [From the definition, X_{n+1} must lie in span (A).] The Bernstein n-width, in general, is a tool used in determining lower bounds for the Kolmogorov and Gel'fand n-widths. In the setting of this paper it has an additional meaning, which will be given later. If dim $X_{n+1}^\star \geq n+1$ and $b_n(A;X)S(X_{n+1}^\star) \subseteq A$, then $X_{n+1}^\star$ is said to be optimal for $b_n(A;X)$.

The topic of n-widths in approximation theory is concerned with the study of these quantities (especially the first two), their properties, characterizations of optimal subspaces or operators (if possible), and the determination of their asymptotic behaviour as $n \uparrow \infty$. In this paper we shall, for specific choices of A and X, characterize optimal subspaces and operators.

Before discussing the specific problem to be considered we state here two well-known relationships among the above four quantities. Since these inequalities are important we sketch a proof.

Theorem 1.1. *Let A be a closed, convex, centrally symmetric subset of a normed linear space X. Then*

$$\delta_n(A;X) \geq d_n(A;X), \; d^n(A;X) \geq b_n(A;X).$$

Sketch of proof. The inequality $\delta_n(A;X) \geq d_n(A;X)$ is a consequence of the definition. To prove that $\delta_n(A;X) \geq d^n(A;X)$, note that any linear, continuous rank n operator P_n may be written in the form

$$P_n(\cdot) = \sum_{i=1}^{n} f_i(\cdot)x_i$$

where the $\{x_i\}_{i=1}^n$ are a basis for the range of P_n and $f_i \in X'$, i.e., continuous linear functionals on X. Set $L_n = \{x: f_i(x) = 0, i = 1, \ldots, n\}$. Then L_n is a closed subspace of codimension $\leq n$ and

$$\sup \{\|x - P_n(x)\|: x \in A\} \geq \sup \{\|x\|: x \in A \cap L_n\},$$

from which follows the desired inequality.

The inequality $d^n(A;X) \geq b_n(A;X)$ is a simple consequence of the fact that given X_{n+1}, $\dim X_{n+1} \geq n + 1$, and L_n, codim $L_n \leq n$, there exists an $x \in X_{n+1}$, $x \neq 0$, such that $x \in L_n$.

The more interesting and more difficult inequality to prove is the last, $d_n(A;X) \geq b_n(A;X)$. It rests on the fact that given X_n and X_{n+1} ($\dim X_n \leq n$, $\dim X_{n+1} \geq n + 1$) there exists an $x \in X_{n+1}$, $x \neq 0$, such that $E(x;X_n) = \inf \{\|x - y\|: y \in X_n\} = \|x\|$, i.e., the zero function is a best approximation to x from X_n. This fact was first proved by Krein, Krasnosel'ski, and Milman [6] as a consequence of the Borsuk Antipodality Theorem. ■

The Sobolev space $W_p^{(r)}[0, 1]$ of real-valued functions on [0, 1] is defined by $W_p^{(r)} = W_p^{(r)}[0, 1] = \{f: f^{(r-1)} \text{ abs.cont.}, f^{(r)} \in L^p\}$. Set $B_p^{(r)} = \{f: f \in W_p^{(r)}, \|f^{(r)}\|_p \leq 1\}$. One of the interesting problems in the study of n-widths has been the determination of the n-widths and the characterization of optimal subspaces and operators when $A = B_p^{(r)}$ and $X = L^q$ for $p, q \in [1, \infty]$. (No a priori relationship is assumed to exist between p and q.) This problem was considered by Kolmogorov [5] in 1936 for $p = q = 2$; it was this article in which the concept of n-widths was introduced.

The case $p = q = \infty$ was solved the Tihomirov [15] in 1969 in a paper that motivated much of the subsequent work on n-widths. He proved the existence, for $n \geq r$ (if $n < r$, then all the n-widths are infinite), of two particular sets of points $0 < \xi_1 < \ldots < \xi_{n-r} < 1$ and $0 < \eta_1 < \ldots < \eta_n < 1$. The subspace

$$X_n^\star = \text{span } \{1, x, \ldots, x^{r-1}, (x - \xi_1)_+^{r-1}, \ldots, (x - \xi_{n-r})_+^{r-1}\}$$

(i.e., splines of degree $r - 1$ with simple knots at the ξ_i) is optimal for $d_n(B_\infty^{(r)}; L^\infty)$. The subspace

$$L_n^\star = \{f: f \in B_p^{(r)}, f(\eta_i) = 0, \quad i = 1, \ldots, n\}$$

is optimal for $d^n(B_\infty^{(r)};L^\infty)$. $P_n^\star$, the operator that interpolates to $f \in B_\infty^{(r)}$ at the $\{\eta_i\}_{i=1}^n$ from $X_n^\star$, is optimal for $\delta_n(B_\infty^{(r)};L^\infty)$, and $\delta_n(B_\infty^{(r)};L^\infty) = d_n(B_\infty^{(r)};L^\infty) = d^n(B_\infty^{(r)};L^\infty) = b_n(B_\infty^{(r)};L^\infty)$. [He also identified an optimal subspace for $b_n(B_\infty^{(r)};L^\infty)$.] Micchelli and Pinkus [10] extended this same result to $p = \infty$, every $q \in [1, \infty]$, and $p \in [1\ \infty]$, $q = 1$. That is, there exist points $\{\xi_i\}_{i=1}^{n-r}$ and $\{\eta_i\}_{i=1}^n$ (which do depend on p and q) such that $X_n^\star$ is optimal for d_n, $L_n^\star$ is optimal for d^n, and $P_n^\star$ is optimal for δ_n. In all these cases we also have $d_n = d^n = \delta_n$, although b_n is unknown except when $p = q \in \{1, \infty\}$. Melkman and Micchelli [8] returned to the case $p = q = 2$ that had been considered by Kolmogorov and proved that here too there were optimal subspaces and operators of the above form, a fact that had eluded Kolmogorov. In a slightly different vein, Tihomirov and Babadjanov [16] proved the analogous result for $p = q \in [1, \infty]$ when $r = 1$, while Makovoz [7] extended this to $p \geq q$ and $r = 1$. The case for $r = 1$

is more tractable. One expects and obtains that the points will be evenly spaced. This is no longer true for $r \geq 2$.

On the basis of the above results it has been conjectured that for all $p \geq q$, $p,q \in [1, \infty]$, there exist points $\{\xi_i\}_{i=1}^{n-r}$ and $\{\eta_i\}_{i=1}^{n}$ (dependent on p and q) such that

i) $X_n^\star$ is optimal for $d_n(B_p^{(r)};L^q)$.
ii) $L_n^\star$ is optimal for $d^n(B_p^{(r)};L^q)$.
iii) $P_n^\star$ is optimal for $\delta_n(B_p^{(r)};L^q)$.
iv) $\delta_n(B_p^{(r)};L^q) = d_n(B_p^{(r)};L^q) = d^n(B_p^{(r)};L^q)$.

[Actually, (iii) and (iv) imply (i) and (ii).] In this paper we prove these four conjectures for all $p = q \in [1, \infty]$ and every r, and also determine $b_n(B_p^{(r)};L^p)$, which has the same value as these other n-widths.

The quantity $b_n(B_p^{(r)};L^p)$ is related to Markov-Bernstein-type inequalities. It is not difficult to prove that

$$b_n(B_p^{(r)};L^p) = \sup_{X_{n-r+1}} \inf_{f^{(r)} \in X_{n-r+1}} \frac{\|f\|_p}{\|f^{(r)}\|_p},$$

where X_{n-r+1} ranges over all subspaces of dimension at least $n-r+1$. A Markov-Bernstein-type inequality is an inequality of the form

$$\|f^{(r)}\|_p \leq C\|f\|_p$$

that is valid for all f in a certain subspace (e.g., algebraic polynomials of some fixed degree). The constant C is understood to be the smallest constant for which the inequality holds. If we minimize this constant C when varying over all subspaces of dimension $n-r+1$, then the resulting value is simply $[b_n(B_p^{(r)};L^p)]^{-1}$. From an optimal subspace for $b_n(B_p^{(r)};L^p)$ we obtain a subspace for which this minimum value is obtained.

We prove our result by showing that $\delta_n(B_p^{(r)};L^p) = b_n(B_p^{(r)};L^p)$ and that (iii) holds. From Theorem 1.1 it follows that both $d_n(B_p^{(r)};L^p)$ and $d^n(B_p^{(r)};L^p)$ are equal to this value. From (iii) we see that $X_n^\star$ is optimal for $d_n(B_p^{(r)};L^p)$ (since $X_n^\star$ is the range of the optimal operator $P_n^\star$ of $\delta_n(B_p^{(r)};L^p)$) and $L_n^\star$ is optimal for $d^n(B_p^{(r)};L^p)$ (since $L_n^\star$ is the null space of $P_n^\star$).

Let us outline how this result is obtained. For $n < r$, all n-widths are infinite. We therefore assume that $n \geq r$. Also, since the cases $p = 1$ and $p = \infty$ have been solved, we consider only $p \in (1, \infty)$. The idea of the proof is the following: Assume that there exists a function $f_n \in W_p^{(r)}$ satisfying

$$(1.1) \qquad \int_0^1 x^i |f_n(x)|^{p-1} \operatorname{sgn}(f_n(x))\, dx = 0, \qquad i = 0, 1, \ldots, r-1$$

and

$$(1.2) \qquad \frac{1}{(r-1)!} \int_0^1 (x-y)_+^{r-1} |f_n(x)|^{p-1} \operatorname{sgn}(f_n(x))\, dx = \lambda_n^p |f_n^{(r)}(y)|^{p-1} \operatorname{sgn}(f_n^{(r)}(y))$$

for all $y \in [0, 1]$, where λ_n is some constant and f_n has exactly n sign changes on $[0, 1]$. (The concept of sign changes will be made more precise later.) It then follows that f_n has exactly n simple zeros in $[0, 1]$ (i.e., at its sign changes), which we denote by $\{\eta_i\}_{i=1}^n$, and $f_n^{(r)}$ has exactly $n-r$ simple zeros (sign changes) in (0, 1) at $\{\xi_i\}_{i=1}^{n-r}$. Equations (1.1) and (1.2) come from considering the zeros of the Gateaux derivative of the ratio $\|f\|_p/\|f^{(r)}\|_p$. In Section 2 we prove that from (1.1) and (1.2) one obtains $\delta_n(B_p^{(r)};L^p) \le \lambda_n$ and $\lambda_n \le b_n(B_p^{(r)};L^p)$ so that equality obtains. In addition, the first inequality (the upper bound) is proved using the the interpolation operator $P_n^\star$. The result therefore follows if we can prove, for $n \ge r$, the existence of $f_n \in W_p^{(r)}$ with n sign changes which satisfies (1.1) and (1.2). This, for us, was the more difficult result. It is proven in Sections 5 and 6 by going to a "smoothed" version of the problem and then discretizing to the matrix case.

Every $f \in W_p^{(r)}$ may be written in the form

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

This is simply Taylor's formula with the remainder in integral form. We see that, aside from the free r-dimensional subspace spanned by $1, x, \ldots, x^{r-1}$, $B_p^{(r)}$ is essentially of the form

$$A = \left\{ f\colon f(x) = \int_0^1 K(x, y)h(y)\, dy, \qquad \|h\|_p \le 1 \right\}$$

where for $B_p^{(r)}$, $K(x, y)$ is the kernel $(x-y)_+^{r-1}/(r-1)!$ and $h(y) = f^{(r)}(y)$. In problems of n-widths, this form should always be at the back of one's mind. It was found (see, e.g., Micchelli and Pinkus [9, 10] and Melkman and Micchelli [8]) that the essential property of $B_p^{(r)}$ that allows us to calculate its n-widths (in the known cases) is the fact that the kernel $(x-y)_+^{r-1}$ is totally positive. In other words, the result that was presented previously may and should be considered as a special case of a more general result.

Let $K_p = \{f\colon f(x) = \int_0^1 K(x, y)h(y)\, dy,\ \|h\|_p \le 1\}$. We prove in Section 4 that if K is extended totally positive, then there exist distinct points $\{\xi_i\}_{i=1}^n$ and $\{\eta_i\}_{i=1}^n$ in (0, 1) such that

$$X_n^\star = \text{span } \{K(\cdot, \xi_1), \ldots, K(\cdot, \xi_n)\}$$

is optimal for $d_n(K_p;L^p)$,

$$L_n^\star = \{f\colon f \in K_p,\ f(\eta_i) = 0,\ i = 1, \ldots, n\}$$

is optimal for $d^n(K_p;L^p)$, interpolation to $f \in K_p$ at the $\{\eta_i\}_{i=1}^n$ from $X_n^\star$ is optimal for $\delta_n(K_p;L^p)$, and $\delta_n(K_p;L^p) = d_n(K_p;L^p) = d^n(K_p;L^p) = b_n(K_p;L^p)$.

This is essentially the result we are obtaining for $A = B_p^{(r)}$. Two problems arise. The first is that the kernel $(x-y)_+^{r-1}$ is totally positive in a weak sense. The second is the existence of the free r-dimensional subspace spanned by algebraic polynomials of degree $r-1$. The overcoming of these problems involves a fair degree of technical work. However, we cannot overemphasize the fact that the main ideas are connected with K_p and total positivity (or more precisely with the variation diminishing property that holds for such kernels).

To prove our result for K_p we consider $\|Kh\|_p/\|h\|_p$ ($Kh(x) = \int_0^1 K(x, y)h(y)\, dy$) and the zeros of the Gateaux derivatives of this ratio. The resulting equation is

$$(1.3) \qquad \int_0^1 K(x, y)|Kh(x)|^{p-1} \operatorname{sgn}(Kh(x))\, dx = \lambda^p |h(y)|^{p-1} \operatorname{sgn}(h(y))$$

for all $y \in [0, 1]$. We prove the existence of solutions $(\lambda_n, h_n)_{n=0}^{\infty}$ to (1.3) where $\lambda_0 > \lambda_1 > \lambda_2 > \ldots > 0$ and h_n has exactly n sign changes. This is a direct generalization of the eigenvalue problem

$$(1.4) \qquad \int_0^1 \left(\int_0^1 K(x, y)K(x, z)\, dx \right) h(z)\, dz = \lambda^2 h(y).$$

Equation (1.4) is simply (1.3) with $p = 2$. The Gel'fand and Bernstein n-widths may, in this setting, be considered as generalizations of the minmax and maxmin characterizations, respectively, of the eigenvalues of the self-adjoint, nonnegative, compact operator induced by the kernel $M(y, z) = \int_0^1 K(x, y)K(x, z)\, dx$. In the determination of $X_n^\star$ and $L_n^\star$, the $\{\xi_i\}_{i=1}^n$ are the zeros of h_n, and the $\{\eta_i\}_{i=1}^n$ the zeros of Kh_n. Furthermore, the value of the n-widths is λ_n. For $p = 2$, this was the result obtained by Melkman and Micchelli [8].

This paper is organized as follows. In Section 2 we prove the results on n-widths of $B_p^{(r)}$ in L^p, modulo the existence of f_n satisfying (1.1) and (1.2). Section 3 is a discussion of total positivity and related matters. These results are needed in the subsequent sections and are gathered together in one section to facilitate presentation. Section 4 parallels Section 2 in that we prove results on n-widths of other classes of functions (e.g., K_p in L^p), modulo the important existence theorems. Finally, in Sections 5 and 6 we prove these existence theorems.

For a survey of the subject of n-widths, the interested reader is referred to Pinkus [13].

2. n-Widths of $B_p^{(r)}$ in L^p

We first prove, modulo one important step in the proof, the major result concerning n-widths of $B_p^{(r)}$ in L^p. Before stating this result, we need a precise definition of zeros and sign changes.

Definition 2.1 *Let I be an interval which is either open, closed, or half-open. Let $f \in C(I)$.*

a) *$S_I(f)$ will denote the number of sign changes (in the usual sense) of f on I. In other words, $S_I(f) = m$ if m is the maximum integer such that there exist points $x_1 < \ldots < x_{m+1}$, $x_i \in I$, for which $f(x_i)f(x_{i+1}) < 0$, $i = 1, \ldots m$. ($S_I(f)$ may be zero or infinite, with the obvious meaning.)*

b) *$Z_I(f)$ will denote the number of distinct zeros of f on I.*

c) *Assume $f \in C^k(I)$. Then $Z_I^\star(f)$ will count the number of zeros of f on I, where a zero is counted up to multiplicity at most $k + 1$. In other words, if $x \in I$, $f(x) = f'(x) = \ldots = f^{(i-1)}(x) = 0$, $f^{(i)}(x) \neq 0$ (or $i = k+1$), then x is counted as a zero of multiplicity* $\min\{i, k+1\}$.

If the subscript I is omitted from any of the above quantities, then it is to be understood that $I = [0, 1]$.

Consider the following equations for $f \in W_p^{(r)}$:

$$\int_0^1 x^i |f(x)|^{p-1} \operatorname{sgn}(f(x))\, dx = 0, \qquad i = 0, 1, \ldots, r-1, \tag{2.1}$$

and

$$\frac{1}{(r-1)!}\int_0^1 (x-y)_+^{r-1} |f(x)|^{p-1} \operatorname{sgn}(f(x))\, dx = \lambda^p |f^{(r)}(y)|^{p-1} \operatorname{sgn}(f^{(r)}(y)), \tag{2.2}$$

for all $y \in [0, 1]$, where λ is some constant. Multiplying each side of (2.2) by $f^{(r)}(y)$, using (2.1) and integrating over y, we obtain $\|f\|_p = \lambda \|f^{(r)}\|_p$. Thus, if f satisfies (2.1) and (2.2), and $\|f\|_p \neq 0$, then $\lambda > 0$ and $\|f^{(r)}\|_p \neq 0$.

Assume f satisfies (2.1) and (2.2), and set

$$\begin{aligned} G(y) &= \frac{1}{(r-1)!}\int_0^1 (x-y)_+^{r-1} |f(x)|^{p-1} \operatorname{sgn}(f(x))\, dx \\ &= \lambda^p |f^{(r)}(y)|^{p-1} \operatorname{sgn}(f^{(r)}(y)). \end{aligned} \tag{2.3}$$

From (2.1) and (2.2),

$$G^{(i)}(0) = G^{(i)}(1) = 0, \qquad i = 0, 1, \ldots, r-1. \tag{2.4}$$

Furthermore, $(-1)^r G^{(r)}(x) = |f(x)|^{p-1} \operatorname{sgn}(f(x))$ [from (2.3)]. Let $q \in (1, \infty)$ satisfy $1/p + 1/q = 1$. Then from (2.2) we have

$$(-1)^r \lambda^q \frac{d^r}{dx^r}[|G^{(r)}(x)|^{q-1} \operatorname{sgn}(G^{(r)}(x))] = |G(x)|^{q-1} \operatorname{sgn}(G(x)). \tag{2.5}$$

The differential equation (2.5), together with the boundary conditions (2.4), is equivalent to (2.1) and (2.2). It is this equation that we will study.

For $p = 2$, (2.4) and (2.5) reduce to

$$\begin{cases} (-1)^r \lambda^2 G^{(2r)}(x) = G(x) \\ G^{(i)}(0) = G^{(i)}(1) = 0, \qquad i = 0, 1, \ldots, r-1. \end{cases} \tag{2.6}$$

The eigenvalue problem (2.6) is well-known. For each nonnegative integer m there exists a G_m, unique up to multiplication by a constant, and λ_m positive, for which (2.6) holds and G_m has exactly m sign changes. Furthermore, $\lambda_0 > \lambda_1 > \ldots$. It is this result that we generalize for (2.5) together with (2.4).

We will, for convenience, assume $r \geq 2$. The results may be extended to the case $r = 1$, although there is a slight technical problem. However, since the end result has already been solved for $r = 1$, we allow ourselves this luxury. Note that if $f \in W_p^{(r)}$ satisfies (2.1) and (2.2), then in fact $f \in C^r[0, 1]$. (More is actually true, but this suffices for our analysis.)

The proof of the following theorem is deferred to Sections 5 and 6.

Theorem 2.1. *Fix $p \in (1, \infty)$ and $r \geq 2$. For each positive integer n, $n \geq r$, there exists an $f_n \in W_p^{(r)}$ such that $S(f_n) = n$ and f_n satisfies* (2.1) *and* (2.2).

Before stating our main result we must know a little more about the zeros and sign changes of f_n and $f_n^{(r)}$. We first have:

Proposition 2.2. *Assume $f \in W_p^{(r)}$, $f \neq 0$, satisfies* (2.1) *and* (2.2), *and* $S(f) < \infty$. *Then $f^{(r)}$ cannot vanish identically on any interval.*

Remark. From (2.2) it follows that $f^{(r)}$ vanishes identically on an interval if and only if f vanishes identically on the same interval.

Proof. For the above f, let G be as previously defined. Assume $f^{(r)}$ vanishes identically on some interval $[a, b]$, $0 \leq a < b \leq 1$. For convenience we shall assume that $a > 0$ and that $f^{(r)}$ does not vanish identically on any subinterval of $[0, a]$. (The other cases are proved in the same manner.) Thus $G^{(i)}(0) = 0$, $i = 0, 1, \ldots, r-1$, $G \in C^r[0, 1]$, and $G^{(i)}(a) = 0$, $i = 0, 1, \ldots, r$, since G vanishes identically on $[a, b]$.

We now apply Rolle's Theorem repeatedly. Assume $f^{(r)}$ has k sign changes on $[0, a]$. Then $G = \lambda^p |f^{(r)}|^{p-1} \operatorname{sgn}(f^{(r)})$ has k sign changes on $[0, a]$, since $\lambda > 0$. Since $G^{(i)}(0) = G^{(i)}(a) = 0$, $i = 0, 1, \ldots, r-1$, it follows that $G^{(r)}$ has at least $k + r$ sign changes on $[0, a]$. Moreover, $G^{(r)}(a) = 0$, and $(-1)^r G^{(r)} = |f|^{p-1} \operatorname{sgn}(f)$, so that f has at least $k + r + 1$ distinct zeros on $[0, a]$. Since $f^{(r)}$ does not vanish identically on any subinterval of $[0, a]$, it follows that $f^{(r)}$ has at least $k + 1$ sign changes on $[0, a]$. This contradicts our definition of k. ■

For the f_n of Theorem 2.1 we need and have the following more precise result. Let G_n satisfy (2.3) with respect to f_n. For the purpose of this section, $Z_I^{\star}(f)$ counts zero up to multiplicity r.

Proposition 2.3. *Let f_n be as in the statement of Theorem 2.1. Then*

$$\begin{aligned} n &= S(f_n) = Z^{\star}(f_n) = S(f_n^{(r)}) + r \\ &= Z_{(0,1)}(f_n^{(r)}) + r = Z_{(0,1)}^{\star}(G_n) + r. \end{aligned}$$

Remark. From the definition of G_n, we obviously have $S(G_n) = S(f_n^{(r)})$. The above proposition states that f_n has exactly n simple zeros (sign changes) on $(0, 1)$ (and does not vanish at the end points), while $f_n^{(r)}$ has exactly $n - r$ simple zeros (sign changes) in $(0, 1)$ (and does not vanish at both end points).

Proof. Since neither f_n nor $f_n^{(r)}$ vanishes identically on any subinterval,

$$\begin{aligned} n &= S(f_n) \leq Z^{\star}(f_n) \leq S(f_n^{(r)}) + r \\ &\leq Z_{(0,1)}(f_n^{(r)}) + r = Z_{(0,1)}(G_n) + r \\ &\leq Z_{(0,1)}^{\star}(G_n) + r = Z^{\star}(G_n) - r, \end{aligned}$$

since $G^{(i)}(0) = G^{(i)}(1) = 0$, $i = 0, 1, \ldots, r-1$. Now

$$\begin{aligned} Z^{\star}(G_n) - r &\leq S(G_n^{(r)}) = S(|f_n|^{p-1} \operatorname{sgn}(f_n)) \\ &= S(f_n) = n. \end{aligned}$$

We therefore have equality throughout, proving the proposition. ■

For f_n as above, let

$$0 < \eta_1 < \ldots < \eta_n < 1$$

denote its zeros (sign changes) and let

$$0 < \xi_1 < \ldots < \xi_{n-r} < 1$$

denote the zeros (sign changes) of $f_n^{(r)}$ on (0, 1).

For ease of exposition, we normalize f_n so that $\|f_n^{(r)}\|_p = 1$ and $f_n^{(r)}(x) > 0$ for $x \in (0, \xi_1)$. Let $\{g_i\}_{i=1}^{n-r+1}$ be defined as follows:

$$g_i^{(j)}(0) = 0, \quad j = 0, 1, \ldots, r-1,$$

$$g_i^{(r)}(x) = \begin{cases} |f_n^{(r)}(x)|, & \xi_{i-1} < x < \xi_i, \\ 0 \quad , & \text{otherwise,} \end{cases}$$

for $i = 1, \ldots, n-r+1$, where $\xi_0 = 0$, $\xi_{n-r+1} = 1$. Thus

$$g_i(x) = \frac{1}{(r-1)!} \int_{\xi_{i-1}}^{\xi_i} (x-y)_+^{r-1} |f_n^{(r)}(y)| \, dy.$$

Note that

$$\begin{aligned} f_n(x) &= \sum_{i=0}^{r-1} a_i^\star x^i + \frac{1}{(r-1)!} \int_0^1 (x-y)_+^{r-1} f_n^{(r)}(y) \, dy \\ &= \sum_{i=0}^{r-1} a_i^\star x^i + \sum_{i=1}^{n-r+1} (-1)^{i+1} g_i(x). \end{aligned}$$

Theorem 2.4. *Fix $p \in (1, \infty)$, $r \geq 2$, and $n \geq r$. Let f_n satisfy the conditions of Theorem 2.1. Let $\{\eta_i\}_{i=1}^n$, $\{\xi_i\}_{i=1}^{n-r}$ and $\{g_i\}_{i=1}^{n-r+1}$ be as defined above. Then*

$$\begin{aligned} \delta_n(B_p^{(r)};L^p) &= d_n(B_p^{(r)};L^p) = d^n(B_p^{(r)};L^p) \\ &= b_n(B_p^{(r)};L^p) = \|f_n\|_p / \|f_n^{(r)}\|_p = \lambda_n. \end{aligned}$$

Furthermore,

a) $X_n^\star = \text{span}\,\{1, x, \ldots, x^{r-1}, (x-\xi_1)_+^{r-1}, \ldots, (x-\xi_{n-r})_+^{r-1}\}$ *is optimal for* $d_n(B_p^{(r)};L^p)$.

b) $L_n^\star = \{f: f \in B_p^{(r)}, f(\eta_i) = 0, i = 1, \ldots, n\}$ *is optimal for* $d^n(B_p^{(r)};L^p)$.

c) $P_n^\star$, *the rank n operator determined by interpolation from* $X_n^\star$ *to* $f \in B_p^{(r)}$ *at the* $\{\eta_i\}_{i=1}^n$ *is optimal for* $\delta_n(B_p^{(r)};L^p)$.

d) $Y_{n+1}^\star = \text{span}\,\{1, x, \ldots, x^{r-1}, g_1(x), \ldots, g_{n-r+1}(x)\}$ *is optimal for* $b_n(B_p^{(r)};L^p)$.

Proof. To prove all of the above results, it suffices to prove that

1) $\|f - P_n^\star f\|_p \leq \lambda_n \|f^{(r)}\|_p$ for all $f \in W_p^{(r)}$, and

2) $\lambda_n S(Y_{n+1}^\star) \subseteq B_p^{(r)}$, where $S(Y_{n+1}^\star) = \{g: g \in Y_{n+1}^\star, \|g\|_p \leq 1\}$.

The results of the theorem then follow from Theorem 1.1 and the form of $P_n^\star$.

We divide the proof of the theorem into a series of steps:

Proposition 2.5. $\lambda_n S(Y_{n+1}^*) \subseteq B_p^{(r)}$.

Proof. We must prove that if $g(x) = \sum_{i=0}^{r-1} a_i x^i + \sum_{i=1}^{n-r+1} b_i g_i(x)$, and $\|g\|_p \le \lambda_n$, then $\|g^{(r)}\|_p \le 1$. Let $\|g_i^{(r)}\|_p^p = c_i$, $i = 1, \ldots, n-r+1$. Then $c_i > 0$, all i, and $\sum_{i=1}^{n-r+1} c_i = \|f_n^{(r)}\|_p^p = 1$. Now $g^{(r)}(x) = \sum_{i=1}^{n-r+1} b_i g_i^{(r)}(x)$, and since the $g_i^{(r)}$ have essentially distinct support, $\|g^{(r)}\|_p = (\sum_{i=1}^{n-r+1} |b_i|^p c_i)^{1/p}$. The claim thereby reduces to proving that

$$\inf_{\mathbf{a},\mathbf{b}\neq\mathbf{0}} \frac{\left\| \sum_{i=0}^{r-1} a_i x^i + \sum_{i=1}^{n-r+1} b_i g_i(x) \right\|_p}{\left(\sum_{i=1}^{n-r+1} |b_i|^p c_i \right)^{1/p}} \ge \lambda_n .$$

This infinum is necessarily attained by some $\hat{g}(x) = \sum_{i=0}^{r-1} \hat{a}_i x_i + \sum_{i=1}^{n-r+1} \hat{b}_i g_i(x)$, where $\hat{\mathbf{a}} = (\hat{a}_0, \ldots, \hat{a}_{r-1})$, and $\hat{\mathbf{b}} = (\hat{b}_1, \ldots, \hat{b}_{n-r+1}) \neq \mathbf{0}$. Set

$$\hat{\mu} = \|\hat{g}\|_p / \|\hat{g}^{(r)}\|_p .$$

Because we are at a minimum value (this value cannot be zero), $\hat{\mu}$ and $\hat{g}$ must satisfy (differentiating with respect to the a_i and b_i)

$$\text{(2.7)} \qquad \int_0^1 x^i |\hat{g}(x)|^{p-1} \operatorname{sgn}(\hat{g}(x))\, dx = 0, \qquad i = 0, 1, \ldots, r-1$$

and

$$\text{(2.8)} \int_0^1 g_k(x) |\hat{g}(x)|^{p-1} \operatorname{sgn}(\hat{g}(x))\, dx = \hat{\mu}^p |\hat{b}_k|^{p-1} \operatorname{sgn}(\hat{b}_k) c_k, \; k = 1, \ldots, n-r+1.$$

Recall that

$$f_n(x) = \sum_{i=0}^{r-1} a_i^* x^i + \sum_{i=1}^{n-r+1} (-1)^{i+1} g_i(x),$$

and f_n satisfies (2.1) and (2.2). Let $\mathbf{a}^* = (a_0^*, a_1^*, \ldots, a_{r-1}^*)$ and $\mathbf{b}^* = (1, -1, \ldots, (-1)^{n+r}) \in \mathbf{R}^{n-r+1}$. We claim that (2.7) and (2.8) are also valid where f_n replaces $\hat{g}$, λ_n replaces $\hat{\mu}$, and $\mathbf{a}^*$, $\mathbf{b}^*$ replaces $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, respectively. To see this, note that (2.7) is (2.1) and that (2.8) follows from (2.2) by multiplying each side of (2.2) by $f_n^{(r)}(y)$ and integrating over (ξ_{k-1}, ξ_k), $k = 1, \ldots, n-r+1$. From the definition of $\hat{\mu}$, we have $\lambda_n \ge \hat{\mu}$. We shall prove that $\lambda_n = \hat{\mu}$. Assume this not to be true, i.e., $\lambda_n > \hat{\mu}$. If $f_n = \gamma\hat{g}$ for some $\gamma \in \mathbf{R}$,, then $\lambda_n = \hat{\mu}$, as is easily seen. Thus $f_n \neq \gamma\hat{g}$ for every $\gamma \in \mathbf{R}$.

Normalize $\hat{g}$ so that $|\hat{b}_i| \le 1$, $i = 1, \ldots, n-r+1$, and $\hat{b}_j = (-1)^{j+1}$ for some j. Thus

$$f_n(x) - \hat{g}(x) = \sum_{i=0}^{r-1} c_i x^i + \sum_{\substack{i=1 \\ i\neq j}}^{n-r+1} ((-1)^{i+1} - \hat{b}_i) g_i(x),$$

and a Rolle's Theorem argument implies that $S(f_n - \hat{g}) \le n-1$.

For any $a,b \in \mathbf{R}$,

$$\operatorname{sgn}(a+b) = \operatorname{sgn}(|a|^{p-1}\operatorname{sgn}(a) + |b|^{p-1}\operatorname{sgn}(b)).$$

Therefore $S(|f_n|^{p-1}\operatorname{sgn}(f_n) - |\hat{g}|^{p-1}\operatorname{sgn}(\hat{g})) \le n-1$. From (2.7), (2.8), and the analogous results for f_n

$$\int_0^1 x^i[|f_n(x)|^{p-1}\operatorname{sgn}(f_n(x)) - |\hat{g}(x)|^{p-1}\operatorname{sgn}(\hat{g}(x))]\,dx = 0, \tag{2.9}$$

$$i = 0, 1, \ldots, r-1,$$

and

$$\int_0^1 g_k(x)[|f_n(x)|^{p-1}\operatorname{sgn}(f_n(x)) - |\hat{g}(x)|^{p-1}\operatorname{sgn}(\hat{g}(x))]\,dx \tag{2.10}$$

$$= [\lambda_n^p(-1)^{k+1} - \hat{\mu}^p|\hat{b}_k|^{p-1}\operatorname{sgn}(\hat{b}_k)]c_k, \qquad k = 1, \ldots, n-r+1.$$

If $\lambda_n > \hat{\mu}$, then the sequence $\{[\lambda_n^p(-1)^{k+1} - \hat{\mu}^p|\hat{b}_k|^{p-1}\operatorname{sgn}(\hat{b}_k)]c_k\}_{k=1}^{n-r+1}$ strictly alternates in sign, since $|\hat{b}_k| \le 1$ for all k. Since $\{1, x, \ldots, x^{r-1}\}$ is a Tchebycheff (T-) system and $\{1, x, \ldots, x^{r-1}, g_1(x), \ldots, g_{n-r+1}(x)\}$ is a weak Tchebycheff (WT-) system (see Section 3), this implies that $S(|f_n|^{p-1}\operatorname{sgn}(f_n) - |\hat{g}|^{p-1}\operatorname{sgn}(\hat{g})) \ge n$. This is a contradiction, and therefore $\lambda_n = \hat{\mu}$. ■

Before proving the upper bound we must first show that $P_n^\star$ is well-defined.

Proposition 2.6. *$P_n^\star$ is well-defined.*

Proof. For ease of exposition, set $K(x, y) = (x-y)_+^{r-1}$ for $x, y \in [0, 1]$, and $K(x, i) = x^{i-1}$ for $i = 1, \ldots, r$. We define

$$K\begin{pmatrix} x_1, & \ldots & , x_{r+m} \\ 1, \ldots, r, y_1, & \ldots & , y_m \end{pmatrix} = \begin{vmatrix} 1 & x_1 & \ldots & x_1^{r-1} & (x_1-y_1)_+^{r-1} & \ldots & (x_1-y_m)_+^{r-1} \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ 1 & x_{r+m} & \ldots & x_{r+m}^{r-1} & (x_{r+m}-y_1)_+^{r-1} & \ldots & (x_{r+m}-y_m)_+^{r-1} \end{vmatrix}.$$

It is a well-known fact (see, e.g., Karlin [3], p. 513) that $K\begin{pmatrix} x_1, & \ldots & , x_{r+m} \\ 1, \ldots, r, y_1, & \ldots & , y_m \end{pmatrix} \ge 0$ for all $0 \le x_1 < \ldots < x_{r+m} \le 1$ and $0 \le y_1 < \ldots < y_m \le 1$, and is strictly positive if and only if $x_i < y_i < x_{i+r}$, $i = 1, \ldots, m$.

$P_n^\star$ is well-defined if and only if

$$K\begin{pmatrix} \eta_1, & \ldots & , \eta_n \\ 1, \ldots, r, \xi_1, & \ldots & , \xi_{n-r} \end{pmatrix} > 0.$$

We must therefore prove that $\eta_i < \xi_i < \eta_{i+r}$, $i = 1, \ldots, n-r$. Because f_n has n zeros in [0, 1] and $f_n^{(r)}$ has $n-r$ sign changes (zeros) in (0, 1), it follows that $f_n^{(i)}$ has exactly $n-i$ zeros (sign changes) in (0, 1) at $0 < \zeta_1^i < \ldots < \zeta_{n-i}^i < 1$, $i = 0$,

$1, \ldots, r$, where $\zeta_j^0 = \eta_j$ and $\zeta_j^r = \xi_j$. From Rolle's Theorem we have $\zeta_j^{i-1} < \zeta_j^i < \zeta_{j+1}^{i-1}$ for $i = 1, \ldots, r$ and $j = 1, \ldots, n-i$. Thus $\eta_i < \xi_i < \eta_{i+r}$, $i = 1, \ldots, n-r$. ■

Proposition 2.7. $\|f - P_n^\star(f)\|_p \le \lambda_n \|f\|_p$ *for all* $f \in W_p^{(r)}$.

Proof. $P_n^\star$ is well-defined and

$$f(x) - (P_n^\star f)(x) = \int_0^1 M(x, y) f^{(r)}(y)\, dy,$$

where

$$M(x, y) = \frac{1}{(r-1)!} \frac{K\begin{pmatrix}\eta_1, & \ldots & , & \eta_n, & x\\ 1, \ldots, r, & \xi_1, & \ldots, & \xi_{n-r}, & y\end{pmatrix}}{K\begin{pmatrix}\eta_1, & \ldots & , & \eta_n\\ 1, \ldots, r, & \xi_1, & \ldots, & \xi_{n-r}\end{pmatrix}}.$$

We must show that

$$\left\| \int_0^1 M(\cdot, y) f^{(r)}(y)\, dy \right\|_p \le \lambda_n \|f^{(r)}\|_p$$

for all $f \in W_p^{(r)}$. To prove this we make use of the following facts:

i) $\int_0^1 M(x, y) f_n^{(r)}(y)\, dy = f_n(x)$. This is valid since $(P_n^\star f_n)(x) = 0$.

ii) $\operatorname{sgn}(f_n(x)) M(x, y) \operatorname{sgn}(f_n^{(r)}(y)) \ge 0$ for all $x, y \in [0, 1]$.

iii) $\int_0^1 M(x,y) |f_n(x)|^{p-1} \operatorname{sgn}(f_n(x))\, dx = \lambda^p |f_n^{(r)}(y)|^{p-1} \operatorname{sgn}(f_n^{(r)}(y))$, for all $y \in [0, 1]$. To prove this, note that

$$\begin{aligned}
&\int_0^1 M(x, y) |f_n(x)|^{p-1} \operatorname{sgn}(f_n(x))\, dx \\
&\quad = \frac{1}{(r-1)!} \int_0^1 (x-y)_+^{r-1} |f_n(x)|^{p-1} \operatorname{sgn}(f_n(x))\, dx \\
&\qquad + \sum_{i=0}^{r-1} a_i \int_0^1 x^i |f_n(x)|^{p-1} \operatorname{sgn}(f_n(x))\, dx \\
&\qquad + \sum_{i=1}^{n-r} b_i \int_0^1 (x-\xi_i)_+^{r-1} |f_n(x)|^{p-1} \operatorname{sgn}(f_n(x))\, dx.
\end{aligned}$$

The latter two terms of the last expression are zero by (2.1), (2.2), and the definition of the $\{\xi_i\}_{i=1}^{n-r}$. Again by (2.2) we obtain the desired result.

The quantity

$$\sup_{h \in L^p} \frac{\left\| \int_0^1 M(\cdot, y)h(y)\, dy \right\|_p}{\|h\|_p}$$

is attained and occurs at a critical point, i.e., by $\hat{h}$ and $\hat{\mu}$ where

$$\hat{g}(x) = \int_0^1 M(x, y)\hat{h}(y)\, dy$$

satisfies

$$\int_0^1 M(x, y)|\hat{g}(x)|^{p-1} \operatorname{sgn}(\hat{g}(x))\, dx = \hat{\mu}^p|\hat{h}(y))|^{p-1} \operatorname{sgn}(\hat{h}(y)), \tag{2.11}$$

for all $y \in [0, 1]$, and where

$$\hat{\mu} = \frac{\|\hat{g}\|_p}{\|\hat{h}\|_p} = \sup_{h \in L^p} \frac{\left\| \int_0^1 M(\cdot, y)h(y)\, dy \right\|_p}{\|h\|_p}.$$

The sign pattern of $M(x, y)$ is well-defined [see (ii)], and it therefore follows, multiplying by -1 if necessary, that we may assume $\hat{h}(x)f_n^{(r)}(x) \geq 0$ and $\hat{g}(x)f_n(x) \geq 0$ for all $x \in [0, 1]$.

From (iii) we see that (2.11) holds where $f_n^{(r)}$, f_n, and λ_n replace $\hat{h}$, $\hat{g}$, and $\hat{\mu}$, respectively. We claim that $\lambda_n = \hat{\mu}$. Assume $\lambda_n < \hat{\mu}$. The function $|f_n^{(r)}(x)|^{p-1} \operatorname{sgn}(f_n^{(r)}(x)) \in C^r[0, 1]$, and has simple zeros at $\xi_1, \ldots, \xi_{n-r}$ and zeros of multiplicity exactly r at 0 and at 1. (This function is $\lambda_n^{-p}G_n$, where G_n is as in Proposition 2.3.) From the form of $M(x, y)$ and from (2.11), it follows that $|\hat{h}(x)|^{p-1} \operatorname{sgn}(\hat{h}(x))$ is at least a C^{r-1} function (recall that $r \geq 2$), vanishes at $\xi_1, \ldots, \xi_{n-r}$, and is C^r at $x = 0$ and $x = 1$, with zeros of multiplicity at least r at 0 and at 1.

We claim that there exists a $\beta < \infty$ such that $|\hat{h}(x)| \leq \beta |f_n^{(r)}(x)|$ for all $x \in [0, 1]$. To find such a β, we must consider what happens in neighborhoods of zeros of $f_n^{(r)}(x)$. Set

$$\gamma_i = \lim_{x \to \xi_i} \frac{|\hat{h}(x)|^{p-1} \operatorname{sgn}(\hat{h}(x))}{|f_n^{(r)}(x)|^{p-1} \operatorname{sgn}(f_n^{(r)}(x))}, \qquad i = 1, \ldots, n-r.$$

This limit exists by L'Hospital's rule ($\gamma_i \geq 0$) and because $|f_n^{(r)}(x)|^{p-1} \operatorname{sgn}(f_n^{(r)}(x))$ has a *simple zero* at ξ_i. For $\epsilon > 0$, small, there exist $\delta_i > 0$ such that for all $x \in (\xi_i - \delta_i, \xi_i + \delta_i)$

$$\frac{|\hat{h}(x)|^{p-1}}{|f_n^{(r)}(x)|^{p-1}} \leq \gamma_i + \epsilon, \qquad i = 1, \ldots, n-r,$$

i.e., $|\hat{h}(x)| \leq (\gamma_i + \epsilon)^{1/(p-1)}|f_n^{(r)}(x)|$. This same result holds at $x = 0$ and $x = 1$, since both $|\hat{h}(x)|^{p-1} \operatorname{sgn}(\hat{h}(x))$ and $|f_n^{(r)}(x)|^{p-1} \operatorname{sgn}(f_n^{(r)}(x))$ have zeros of multiplicity r there, are C^r functions in the appropriate neighborhoods and $|f_n^{(r)}(x)|^{p-1}$ sgn $(f_n^{(r)}(x))$ has zeros of exact multiplicity r at 0 and 1. Thus, there exists a β such that

$$|\hat{h}(x)| \le \beta |f_n^{(r)}(x)|$$

for all $x \in [0, 1]$.

Set

$$\beta_0 = \inf \{\beta: |\hat{h}(x)| \le \beta |f_n^{(r)}(x)|, \text{ all } x \in [0, 1]\}.$$

Thus $0 < \beta_0 < \infty$ and $|\hat{h}(x)| \le \beta_0 |f_n^{(r)}(x)|$ for all $x \in [0, 1]$. Let $L(x, y) = |M(x, y)|$. From the sign patterns of M, $f_n^{(r)}$, f_n, $\hat{h}$, and $\hat{g}$, and from (2.11) and (iii),

$$\int_0^1 L(x, z)\left(\int_0^1 L(x, y)|\hat{h}(y)|\, dy\right)^{p-1} dx = \hat{\mu}^p |\hat{h}(z)|^{p-1} \tag{2.12}$$

and

$$\int_0^1 L(x, z)\left(\int_0^1 L(x, y)|f_n^{(r)}(y)|\, dy\right)^{p-1} dx = \lambda_n^p |f_n^{(r)}(z)|^{p-1}. \tag{2.13}$$

Now,

$$\beta_0 \int_0^1 L(x, y)|f_n^{(r)}(y)|\, dy \ge \int_0^1 L(x, y)|\hat{h}(y)|\, dy$$

and therefore

$$\beta_0^{p-1} \int_0^1 L(x, z)\left(\int_0^1 L(x, y)|f_n^{(r)}(y)|\, dy\right)^{p-1} dx$$
$$\ge \int_0^1 L(x, z)\left(\int_0^1 L(x, y)|\hat{h}(y)|\, dy\right)^{p-1} dx.$$

From (2.12) and (2.13) this translates into

$$\beta_0^{p-1}\lambda_n^p |f_n^{(r)}(x)|^{p-1} \ge \hat{\mu}^p |\hat{h}(x)|^{p-1}$$

for all $x \in [0, 1]$. Thus $\beta_0(\lambda_n/\mu)^{p/(p-1)} |f_n^{(r)}(x)| \ge |\hat{h}(x)|$ for all $x \in [0, 1]$, and from the definition of β_0 we obtain $\lambda_n \ge \hat{\mu}$. ∎

This proves Theorem 2.4.

As a matter of interest, f_n is unique (up to multiplication by constants).

Proposition 2.8. *Fix $p \in (1, \infty)$ and n, $n \ge r \ge 2$. Let f, $g \in W_p^{(r)}$ satisfy* (2.1) *and* (2.2) *with $S(f) = S(g) = n$. Then $f = \alpha g$ for some $\alpha \in \mathbf{R}$.*

Proof. From Theorem 2.4 it follows that f and g satisfy (2.1) and (2.2) with the same λ, $\lambda > 0$. Thus for all α, $\beta \in \mathbf{R}$,

$$\int_0^1 x^i [|\alpha f(x)|^{p-1} \operatorname{sgn}(\alpha f(x)) + |\beta g(x)|^{p-1} \operatorname{sgn}(\beta g(x))]\, dx = 0, \tag{2.14}$$

$$i = 0, 1, \ldots, r-1$$

and

$$(2.15)\quad \frac{1}{(r-1)!}\int_0^1 (x-y)_+^{r-1}[|\alpha f(x)|^{p-1}\operatorname{sgn}(\alpha f(x)) + |\beta g(x)|^{p-1}\operatorname{sgn}(\beta g(x))]\,dx$$
$$= \lambda^p[|\alpha f^{(r)}(y)|^{p-1}\operatorname{sgn}(\alpha f^{(r)}(y)) + |\beta g^{(r)}(y)|^{p-1}\operatorname{sgn}(\beta g^{(r)}(y))]$$

for all $y \in [0, 1]$.

We now parallel the proofs of Propositions 2.2 and 2.3 (with slightly more care). Use the fact that $\operatorname{sgn}(|\alpha f(x)|^{p-1}\operatorname{sgn}(\alpha f(x)) + |\beta g(x)|^{p-1}\operatorname{sgn}(\beta g(x))) = \operatorname{sgn}(\alpha f(x) + \beta g(x))$ for all $x \in [0, 1]$, and the analogous result where $f^{(r)}$ and $g^{(r)}$ replace f and g, respectively. The resulting analysis proves that $S(\alpha f + \beta g) = Z^\star(\alpha f + \beta g)$ for all $\alpha, \beta \in \mathbf{R}$ unless $\alpha f + \beta g \equiv 0$. Assuming f and g are linearly independent, a contradiction ensues simply by choosing α, β ($\alpha^2 + \beta^2 > 0$) such that, for example, $(\alpha f + \beta g)(0) = 0$, in which case $S(\alpha f + \beta g) < Z^\star(\alpha f + \beta g)$. ■

These same ideas may be used to prove that the $\{\lambda_n\}_{n=r}^{\infty}$ are a strictly decreasing sequence.

Proposition 2.9. *Fix $p \in (1, \infty)$ and let $n \geq r \geq 2$. Assume Theorem 2.1 holds. Then $\lambda_n > \lambda_{n+1}$.*

Proof. Since, by definition, the n-widths are a nonincreasing sequence of n, it follows that $\lambda_n \geq \lambda_{n+1}$. If $\lambda_n = \lambda_{n+1}$, then we may apply the proof of Proposition 2.8 to f_n and f_{n+1} (of Theorem 2.1) to obtain $S(\alpha f_n + \beta f_{n+1}) = Z^\star(\alpha f_n + \beta f_{n+1})$ for all α, β ($\alpha^2 + \beta^2 > 0$). A contradiction follows as above. ■

The class $W_p^{(r)}$ is defined via the semi-norm $\|f^{(r)}\|_p$. For some purposes it is more convenient to consider the following subset rather than $B_p^{(r)}$. Set

$$B_{p,p}^{(r)} = \{f : f \in W_p^{(r)},\ \|f\|_p + \|f^{(r)}\|_p \leq 1\}.$$

Thus $B_{p,p}^{(r)}$ is a bounded subset of L^p. On the basis of Theorem 2.4, it is an easy matter to deduce the n-widths of $B_{p,p}^{(r)}$ in L^p. $X_n^\star$, $L_n^\star$, $P_n^\star$, $Y_{n+1}^\star$, and λ_n will be as in Theorem 2.4. We also assume $r \geq 2$. First note that the value of the n-widths for $n \leq r-1$ is easily seen to be equal to 1. For $n \geq r$ we have:

Proposition 2.10. *For $B_{p,p}^{(r)}$ as above, and assuming Theorem 2.1,*

$$\delta_n(B_{p,p}^{(r)};L^p) = d_n(B_{p,p}^{(r)};L^p) = d^n(B_{p,p}^{(r)};L^p) = b_n(B_{p,p}^{(r)};L^p) = \lambda_n/(1+\lambda_n).$$

Furthermore,

a) *$X_n^\star$ is optimal for $d_n(B_{p,p}^{(r)}; L^p)$,*
b) *$L_n^\star$ is optimal for $d^n(B_{p,p}^{(r)};L^p)$,*
c) *$P_n^\star/(1+\lambda_n)$ is optimal for $\delta_n(B_{p,p}^{(r)};L^p)$, and*
d) *$Y_{n+1}^\star$ is optimal for $b_n(B_{p,p}^{(r)};L^p)$.*

Proof. The upper bound is proved as follows. Set $\mu_n = \dfrac{\lambda_n}{1+\lambda_n}$. From Theorem 2.4,

$\|f - P_n^\star f\|_p \le \|f^{(r)}\|_p \lambda_n$ for all $f \in W_p^{(r)}$. Thus

$$\begin{aligned} \delta_n(B_{p,p}^{(r)};L^p) &\le \sup_{f\in B_{p,p}^{(r)}} \|f - (1-\mu_n)P_n^\star f\|_p \\ &\le \sup_{f\in B_{p,p}^{(r)}} \{\mu_n\|f\|_p + (1-\mu_n)\|f - P_n^\star f\|_p\} \\ &\le \sup_{f\in B_{p,p}^{(r)}} \{\mu_n\|f\|_p + (1-\mu_n)\|f^{(r)}\|_p\lambda_n\} \\ &= \sup_{f\in B_{p,p}^{(r)}} (\|f\|_p + \|f^{(r)}\|_p)\lambda_n/(1+\lambda_n) \\ &= \lambda_n/(1+\lambda_n). \end{aligned}$$

The lower bound argument is the following. From Theorem 2.4 it follows that $\|g\|_p \ge \|g^{(r)}\|_p\lambda_n$ for all $g \in Y_{n+1}^\star$. Thus

$$b_n(B_{p,p}^{(r)};L^p) \ge \inf_{\substack{g\in Y_{n+1}^\star \\ g\neq 0}} \|g\|_p/(\|g\|_p + \|g^{(r)}\|_p).$$

If $g^{(r)} \equiv 0$, then the above ratio is one. It therefore suffices to assume that $\|g^{(r)}\|_p = 1$. Thus

$$b_n(B_{p,p}^{(r)};L^p) \ge \inf_{\substack{g\in Y_{n+1}^\star \\ \|g^{(r)}\|_p = 1}} \|g\|_p/(\|g\|_p + 1).$$

Since $\|g\|_p \ge \lambda_n$ and $x/(1+x)$ is an increasing function of x, we have

$$b_n(B_{p,p}^{(r)};L^p) \ge \lambda_n/(1+\lambda_n). \qquad \blacksquare$$

The analysis of this section allows us to determine the Kolmogorov n-widths of the following class of functions. This class is considered because of its relationship with the K-functional used in interpolation theory and because of the simple form of the result. We concern ourselves with the evaluation of the n-widths of this set, regarding this as an application of the methods of this section. The idea behind the proof of this result derives from the work of Saates [14].

Assume $\omega(t)$ is a nonnegative, nondecreasing function on $[0, \infty)$ for which $\omega(t)/t$ is nonincreasing. Note that if ω is a concave modulus of continuity then it satisfies these assumptions. Let $V(W_p^{(r)}, \omega)$ denote those $f \in C[0, 1]$ for which

$$\inf_{g\in W_p^{(r)}} \{\|f-g\|_p + t\|g^{(r)}\|_p\} \le \omega(t)$$

for all t.

Proposition 2.11. *Under the assumptions of Theorem* 2.4,

$$d_n(V(W_p^{(r)}, \omega);L^p) = \omega(d_n(B_p^{(r)};L^p))$$

and $X_n^\star$ *is optimal for* $n \ge r$.

Proof. The upper bound is simple and uses none of the properties of ω. Let $X_n^\star$ be as above. Then for each $g \in W_p^{(r)}$,

$$\inf_{h \in X_n^\star} \|g - h\|_p \leq \|g^{(r)}\|_p d_n(B_p^{(r)};L^p).$$

Thus

$$\begin{aligned} d_n(V(W_p^{(r)};\omega);L^p) &\leq \sup_{f \in V(W_p^{(r)},\,\omega)} \inf_{h \in X_n^\star} \|f-h\|_p \\ &\leq \sup_{f \in V(W_p^{(r)},\omega)} \inf_{h \in X_n^\star} \inf_{g \in W_p^{(r)}} \{\|f-g\|_p + \|g-h\|_p\} \\ &= \sup_{f \in V(W_p^{(r)},\omega)} \inf_{g \in W_p^{(r)}} \{\|f-g\|_p + \inf_{h \in X_n^\star} \|g-h\|_p\} \\ &\leq \sup_{f \in V(W_p^{(r)},\,\omega)} \inf_{g \in W_p^{(r)}} \{(\|f-g\|_p + \|g^{(r)}\|_p d_n(B_p^{(r)};L^p)\} \\ &\leq \omega(d_n(B_p^{(r)};L^p)). \end{aligned}$$

The lower bound is less general and depends on the following consequence of Theorem 2.4. Let $Y_{n+1}^\star$ be as in Theorem 2.4. Then for every $f \in Y_{n+1}^\star$

$$\|f\|_p \geq \|f^{(r)}\|_p d_n(B_p^{(r)};L^p).$$

Set

$$M_{n+1} = \{f\colon f \in Y_{n+1},\ \|f\|_p \leq \omega(d_n(B_p^{(r)};L^p))\}.$$

From Theorem 1.1, it suffices to prove that $M_{n+1} \subseteq V(W_p^{(r)}, \omega)$, i.e.,

$$\inf_{g \in W_p^{(r)}} \{\|f-g\|_p + t\|g^{(r)}\|_p\} \leq \omega(t)$$

for all $t \geq 0$ and all $f \in M_{n+1}$.

First assume that $t \geq d_n(B_p^{(r)};L^p)$. Then

$$\begin{aligned} \inf_{g \in W_p^{(r)}} \{\|f-g\|_p + t\|g^{(r)}\|_p\} &\leq \|f\|_p \\ &\leq \omega(d_n(B_p^{(r)};L^p)) \\ &\leq \omega(t) \end{aligned}$$

for all $f \in M_{n+1}$, since ω is a nondecreasing function of t.

For $t \leq d_n(B_p^{(r)};L^p)$ and $f \in M_{n+1}$,

$$\begin{aligned} \inf_{g \in W_p^{(r)}} \{\|f-g\|_p + t\|g^{(r)}\|_p\} &\leq t\|f^{(r)}\|_p \\ &\leq t\|f\|_p / d_n(B_p^{(r)};L^p) \\ &\leq t\omega(d_n(B_p^{(r)};L^p)) / d_n(B_p^{(r)};L^p) \\ &\leq \omega(t) \end{aligned}$$

since $\omega(t)/t$ is a nonincreasing function of t. ∎

Remark. The methods of this section and those of the following sections may be used to deduce the n-widths of

$$B_p = \{f\colon \|Lf\|_p \le 1\}$$

in L^p, where L is an rth order disconjugate ordinary differential equation. The analogues of (2.4) and (2.5) may also be written in the form

$$g^{(i)}(0) = g^{(i)}(1) = 0, \qquad i = 0, 1, \ldots, r-1,$$
$$g = L(|f|^{p-1} \operatorname{sgn}(f)),$$
$$\lambda^q f = L^\star(|g|^{q-1} \operatorname{sgn}(g)),$$

where $(1/p + 1/q) = 1$, $1 < p < \infty$, and $L^\star$ is the adjoint of L.

3. Total Positivity and Tchebycheff Systems

In the remaining sections of this paper, extensive use will be made of the concept of total positivity and related matters. As such, we have collected in this section those definitions and results that are necessary for the subsequent analysis. We start first with the matrix case.

Let $A = (a_{ij})_{i=1}^{M}{}_{j=1}^{N}$, $a_{ij} \in \mathbf{R}$, and $R = \min\{M, N\}$.

Definition 3.1. *The matrix A is said to be totally positive* (TP) *if*

$$A\begin{pmatrix} i_1, & \ldots, & i_k \\ j_1, & \ldots, & j_k \end{pmatrix} = \det (a_{i_m j_n})_{m,n=1}^{k} \ge 0 \tag{3.1}$$

for all choices of integers $1 \le i_1 < \ldots < i_k \le M$, $1 < j_1 < \ldots < j_k \le N$, *and all* $k = 1, \ldots, R$. *It is said to be strictly totally positive* (STP) *if strict inequality replaces the inequality in* (3.1). *If strict inequality holds only for* $k = 1, \ldots, r$ *in* (3.1), *with no restrictions for* $k > r$, *then A is said to be strictly totally positive of order r* (STP_r).

One important property of STP matrices is that of variation diminishing. To explain this property we need the following.

Definition 3.2. *Let* $\mathbf{x} \in \mathbf{R}^m$. *Then for* $\mathbf{x} \ne \mathbf{0}$,

i) $S^-(\mathbf{x})$ *counts the number of ordered sign changes in the vector* $\mathbf{x} = (x_1, \ldots, x_m)$ *where zero entries are discarded.*

ii) $S^+(\mathbf{x})$ *counts the maximum number of ordered sign changes in the vector* $\mathbf{x} = (x_1, \ldots, x_m)$ *where zero entries are given arbitrary values.*

For convenience we set $S^+(\mathbf{0}) = m$.

Thus, for example, if $\mathbf{x} = (1, 0, 1, 0, -1)$, then $S^-(\mathbf{x}) = 1$ and $S^+(\mathbf{x}) = 3$. Note that if $S^+(\mathbf{x}) = S^-(\mathbf{x})$, then

1) $x_1, x_m \ne 0$

and

2) $x_i = 0$ implies $x_{i-1}x_{i+1} < 0$.

Proposition 3.1. *Let A be an $M \times N$ STP matrix and $\mathbf{x} \neq \mathbf{0}$. Then*

$$S^+(A\mathbf{x}) \le S^-(\mathbf{x}).$$

In particular, if $A\mathbf{x} = \mathbf{0}$ then $S^-(\mathbf{x}) \ge M$.

The following facts concerning the spectral and oscillation properties of STP matrices will be used.

Proposition 3.2. *Let A be an $N \times N$ STP_r matrix of rank r. If $|\lambda_1| \ge \ldots \ge |\lambda_N|$ denote the eigenvalues of A listed to their algebraic multiplicity, then*

$$\lambda_1 > \ldots > \lambda_r > \lambda_{r+1} = \ldots = \lambda_N = 0.$$

Furthermore, if $A\mathbf{x}^i = \lambda_i\mathbf{x}^i$, ($\mathbf{x}^i \neq \mathbf{0}$), then

$$S^+(\mathbf{x}^i) = S^-(\mathbf{x}^i) = i - 1, \qquad i = 1, \ldots, r.$$

Finally, the following fact will also be used.

Proposition 3.3. *Let $\mathbf{x}^k \in \mathbf{R}^m$, $k = 1, 2, \ldots$. Assume that $\lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}$, $\mathbf{x} \neq \mathbf{0}$. Then $S^-(\mathbf{x}) \le \varliminf_{k\to\infty} S^+(\mathbf{x}^k)$ and $\varlimsup_{k\to\infty} S^-(\mathbf{x}^k) \le S^+(\mathbf{x})$.*

The continuous version of these results deals with kernels. Let $K(x, y) \in C([0, 1] \times [0, 1])$.

Definition 3.3.

i) *The kernel $K(\cdot, \cdot)$ is said to be totally positive* (TP) *if*

$$K\begin{pmatrix} x_1, \ldots, x_m \\ y_1, \ldots, y_m \end{pmatrix} = \det\,(K(x_i, y_j))_{i,j=1}^m \ge 0 \tag{3.2}$$

for all choices of $0 \le x_1 < \ldots < x_m \le 1$, $0 \le y_1 < \ldots < y_m \le 1$, and $m = 1, 2, \ldots$.

ii) *The kernel $K(\cdot, \cdot)$ is said to be strictly totally positive* (STP) *if strict inequality holds in* (3.2).

iii) *The kernel $K(\cdot, \cdot)$ is said to be extended totally positive* (ETP) *if $K \in C^\infty([0, 1] \times [0, 1])$ and*

$$K^\star\begin{pmatrix} x_1, \ldots, x_m \\ y_1, \ldots, y_m \end{pmatrix} > 0$$

for all choices of $0 \le x_1 \le \ldots \le x_m \le 1$, $0 \le y_1 \le \ldots \le y_m \le 1$, and $m = 1, 2, \ldots$. If both the x_i's and the y_i's are distinct, then

$$K^\star\begin{pmatrix} x_1, \ldots, x_m \\ y_1, \ldots, y_m \end{pmatrix} = K\begin{pmatrix} x_1, \ldots, x_m \\ y_1, \ldots, y_m \end{pmatrix}.$$

When a block of n equal x's occurs, then the corresponding rows are determined by successive derivatives with respect to x, i.e.,

$$\left(\frac{\partial^r K(x, y_1)}{\partial x^r}, \ldots, \frac{\partial^r K(x, y_m)}{\partial x^r}\right), \qquad r = 0, 1, \ldots, n-1.$$

This same rule is applied to equal blocks of y's.

Thus, for example,

$$K^\star\binom{x, x}{y, y} = \begin{vmatrix} K(x, y) & \dfrac{\partial K(x, y)}{\partial y} \\ \dfrac{\partial K(x, y)}{\partial x} & \dfrac{\partial^2 K(x, y)}{\partial x \partial y} \end{vmatrix}$$

Associated with each of the above definitions is a variation diminishing property. To explain this property we need various sign changes and zero definitions.

Definition 3.4.

i) *Let* $f \in L^1[0, 1]$. *We say that* $S(f) = n$ *if there exist* $n+1$ *disjoint ordered intervals* $I_1, \ldots, I_{n+1}$ *(by ordered we mean that* $x < y$ *for all* $x \in I_j$, $y \in I_{j+1}$, $j = 1, \ldots, n$*) whose union is* $[0, 1]$, *and such that* $f(x)(-1)^{-j}\epsilon \geq 0$ *a.e. on* I_j, $j = 1, \ldots, n+1$, *where* $\epsilon \in \{-1, 1\}$, *fixed, and where* meas $\{x: x \in I_j, f(x) \neq 0\} > 0$, $j = 1, \ldots, n+1$. *If no such* n *exists, we set* $S(f) = \infty$.

ii) *Let* $f \in C[0, 1]$. $Z(f)$ *will denote the number of distinct zeros of* f *on* $[0, 1]$.

iii) *Let* $f \in C^\infty[0, 1]$. $Z^\star(f)$ *will denote the number of zeros of* f *on* $[0, 1]$, *counting multiplicities.*

The following holds:

Proposition 3.4. *Let* $K \in C([0, 1] \times [0, 1])$. *For* $h \in L^1[0, 1]$, $h \not\equiv 0$, *set*

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

i) *If* K *is TP, then* $S(Kh) \leq S(h)$.

ii) *If* K *is STP, then* $Z(Kh) \leq S(h)$.

iii) *If* K *is ETP, then* $Z^\star(Kh) \leq S(h)$.

Intermediate between matrices and kernels are sets of functions.

Definition 3.5. *Let* $f_i \in C[0, 1]$, $i = 1, \ldots, m$.

i) *We say that* $\{f_i\}_{i=1}^m$ *is a weak Tchebycheff (WT-) system on* $[0, 1]$ *if*

$$F\binom{x_1, \ldots, x_m}{1, \ldots, m} = \det\,(f_j(x_i))_{i,j=1}^m \geq 0 \tag{3.3}$$

for all choices of $0 \leq x_1 < \ldots < x_m \leq 1$.

ii) *We say that* $\{f_i\}_{i=1}^m$ *is a Tchebycheff (T-) system on* $[0, 1]$ *if strict inequality holds in* (3.3).

iii) *Let* $f_i \in C^\infty[0, 1]$, $i = 1, \ldots, m$. *Then* $\{f_i\}_{i=1}^m$ *is said to be an extended*

Descartes (ED-) system on [0, 1] *if*

$$F^\star\begin{pmatrix} x_1, \ldots, x_k \\ i_1, \ldots, i_k \end{pmatrix} > 0$$

for all $0 \le x_1 \le \ldots \le x_k \le 1$, $1 \le i_1 < \ldots < i_k \le m$ (i_k *integer) and* $k = 1, \ldots, m$.

Remark. If $x_1 < \ldots < x_k$ then

$$F^\star\begin{pmatrix} x_1, \ldots, x_k \\ i_1, \ldots, i_k \end{pmatrix} = F\begin{pmatrix} x_1, \ldots, x_k \\ i_1, \ldots, i_k \end{pmatrix}.$$

In case of blocks of equal x's we apply the convention of Definition 3.3 (iii).

Proposition 3.5.
i) *If* $\{f_i\}_{i=1}^m$ *is a WT-system, then*

$$S\left(\sum_{i=1}^m a_i f_i\right) \le m - 1.$$

ii) *If* $\{f_i\}_{i=1}^m$ *is a T-system and* $\mathbf{a} = (a_1, \ldots, a_m) \ne \mathbf{0}$, *then*

$$Z\left(\sum_{i=1}^m a_i f_i\right) \le m - 1.$$

iii) *If* $\{f_i\}_{i=1}^m$ *is an ED-system, and* $\mathbf{a} = (a_1, \ldots, a_m) \ne \mathbf{0}$, *then*

$$Z^\star\left(\sum_{i=1}^m a_i f_i\right) \le S^-(a_1, \ldots, a_m).$$

The following two results will also be used.

Proposition 3.6. *Assume* $\{f_i\}_{i=1}^m$ *is a T-system and* $g \in L^1[0, 1]$. *Set*

$$b_k = \int_0^1 f_k(x) g(x)\, dx, \qquad k = 1, \ldots, m.$$

If $b_k b_{k+1} < 0$, $k = 1, \ldots, m-1$, *then* $S(g) \ge m - 1$.

Proposition 3.7. *Let* $g_n, g \in C^\infty[0, 1]$, $n = 1, 2, \ldots$, *and* $g \ne 0$. *Assume* $g_n \to g$ *uniformly on* [0, 1]. *Then*

$$Z^\star(g) \ge \overline{\lim_{n\to\infty}}\, S(g_n) \ge S(g).$$

Proofs of the results of this section may be found in Gantmacher and Krein [2], Karlin [3], or Karlin and Studden [4].

4. n-Widths of Other Classes of Functions

In this section we develop results analogous to those obtained in Section 2, and again defer to Sections 5 and 6 one of the major steps of the proof. We consider six distinct classes of functions, two of which are presented as an aid in the proofs of Section 5. However, we only give a proof for the first class. The other cases may be proven in a similar manner. The first class is the following.

A. Continuous Kernels

Let $p \in (1, \infty)$ be fixed. Define

$$K_p = \left\{ Kh(x) = \int_0^1 K(x, y)h(y)\, dy : \|h\|_p \leq 1 \right\}.$$

We wish to determine the n-widths of K_p in L^p where K is an ETP kernel. (We always assume the underlying interval to be $[0, 1]$.)

We defer to Section 5 the proof of the following result.

Theorem 4.1. *Fix $p \in (1, \infty)$. Assume K is an* ETP *kernel. For each nonnegative integer n there exist an $h_n \in L^p$, $\|h_n\|_p = 1$, and λ_n such that $S(h_n) = n$ and*

$$\int_0^1 K(x, y)\,|Kh_n(x)|^{p-1} \operatorname{sgn}(Kh_n(x))\, dx = \lambda_n^p |h_n(y)|^{p-1} \operatorname{sgn}(h_n(y)), \tag{4.1}$$

for all $y \in [0, 1]$.

Multiplying both sides of (4.1) by h_n and integrating, we obtain $\|Kh_n\|_p = \lambda_n \|h_n\|_p$. Since $\|h_n\|_p = 1$ and K is ETP, we have $\|Kh_n\|_p > 0$ and thus $\lambda_n > 0$. Thus (4.1) actually implies that $h_n \in C[0, 1]$, and since $K \in C^\infty([0, 1])^2$, $|h_n|^{p-1} \operatorname{sgn}(h_n) \in C^\infty[0, 1]$.

The following facts will prove useful.

Lemma 4.2. *Assume (h_n, λ_n) satisfy (4.1) with $S(h_n) = n$. Then*

$$n = S(Kh_n) = Z^\star(Kh_n) = Z^\star(|h_n|^{p-1} \operatorname{sgn}(h_n)).$$

Remark. This immediately implies that $S(h_n) = Z(h_n) = n$.

Proof. $n = S(h_n) \leq Z(h_n) = Z(\lambda_n |h_n|^{p-1} \operatorname{sgn}(h_n)) \leq Z^\star(\lambda_n |h_n|^{p-1} \operatorname{sgn}(h_n))$, since $\lambda_n > 0$. By (4.1) the last quantity equals

$$Z^\star\left(\int_0^1 K(x, \cdot)\,|Kh_n(x)|^{p-1} \operatorname{sgn}(Kh_n(x))\, dx \right).$$

Applying Proposition 3.4 (iii), the above quantity is seen to be less than

$$\begin{aligned} S(|Kh_n|^{p-1} \operatorname{sgn}(Kh_n)) &= S(Kh_n) \\ &\leq Z^\star(Kh_n) \leq S(h_n). \end{aligned}$$

Equality holds throughout. ∎

For h_n as above, let

$$0 < \xi_1 < \ldots < \xi_n < 1$$

denote its zeros (sign changes), and let

$$0 < \eta_1 < \ldots < \eta_n < 1$$

denote the zeros (sign changes) of Kh_n. Let $P_n^\star$ denote the rank n operator that interpolates to $f \in C[0, 1]$ at $\eta_1, \ldots, \eta_n$ from $X_n^\star = \text{span}\,\{K(\cdot, \xi_1), \ldots, K(\cdot, \xi_n)\}$. Such an operator is well defined, since $K\begin{pmatrix}\eta_1, \ldots, \eta_n \\ \xi_1, \ldots, \xi_n\end{pmatrix} > 0$. Set

$$f_i(x) = \begin{cases} |h_n(x)|, & \xi_{i-1} < x < \xi_i, \\ 0, & \text{otherwise}, \end{cases}$$

$i = 1, \ldots, n+1$, where $\xi_0 = 0$ and $\xi_{n+1} = 1$. Define $g_i(x) = (Kf_i)(x)$, $i = 1, \ldots, n+1$.

Theorem 4.3. *Fix $p \in (1, \infty)$. Assume K is ETP and (h_n, λ_n) satisfy* (4.1) *with $S(h_n) = n$. Let $\{\xi_i\}_{i=1}^n$, $\{\eta_i\}_{i=1}^n$, $\{g_i\}_{i=1}^{n+1}$, and $P_n^\star$ be as defined above. Then*

$$\delta_n(K_p;L^p) = d_n(K_p;L^p) = d^n(K_p;L^p) = b_n(K_p;L^p) = \|Kh_n\|_p/\|h_n\|_p = \lambda_n.$$

Furthermore,

a) *$P_n^\star$ is an optimal rank n operator for $\delta_n(K_p;L^p)$;*
b) *$X_n^\star = \text{span}\,\{K(\cdot, \xi_1), \ldots, K(\cdot, \xi_n)\}$ is optimal for $d_n(K_p;L^p)$;*
c) *$L^n_\star = \{f: f \in K_p, f(\eta_i) = 0, i = 1, \ldots, n\}$ is optimal for $d^n(K_p;L^p)$;* and
d) *$Y_{n+1}^\star = \text{span}\,\{g_1, \ldots, g_{n+1}\}$ is optimal for $b_n(K_p;L^p)$.*

Proof. To prove all the above results it suffices to show that
1) $\|Kh - P_n^\star(Kh)\|_p \le \lambda_n \|h\|_p$ for all $h \in L^p$, and
2) $\|\Sigma_{i=1}^{n+1} \alpha_i g_i\|_p \le \lambda_n$ implies $\|\Sigma_{i=1}^{n+1} \alpha_i f_i\|_p \le 1$.

Here we are proving a result analogous to Theorem 2.4. The proof of this result is even easier.

Proposition 4.4. *If $\|\Sigma_{i=1}^{n+1} \alpha_i g_i\|_p \le \lambda_n$, then $\|\Sigma_{i=1}^{n+1} \alpha_i f_i\|_p \le 1$.*

Proof. We must prove that

$$\min_{\alpha \neq 0} \frac{\left\| \sum_{i=1}^{n+1} \alpha_i g_i \right\|_p}{\left\| \sum_{i=1}^{n+1} \alpha_i f_i \right\|_p} \ge \lambda_n.$$

(The minimum is easily seen to be obtained.) From the definitions of f_i, we see that they have essentially distinct support and

$$\left\| \sum_{i=1}^{n+1} \alpha_i f_i \right\|_p = \left(\sum_{i=1}^{n+1} |\alpha_i|^p c_i \right)^{1/p},$$

where $c_i = \int_{\xi_{i-1}}^{\xi_i} |h_n(x)|^p \, dx > 0$, $i = 1, \ldots, n+1$. We must therefore prove that

$$\min_{\alpha \neq 0} \frac{\left\| \sum_{i=1}^{n+1} \alpha_i g_i \right\|_p}{\left(\sum_{i=1}^{n+1} |\alpha_i|^p c_i \right)^{1/p}} \geq \lambda_n.$$

This minimum is attained for some $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{n+1}) \neq \mathbf{0}$ and $\hat{\mu}$, i.e.,

$$\hat{\mu} = \frac{\left\| \sum_{i=1}^{n+1} \hat{\alpha}_i g_i \right\|_p}{\left(\sum_{i=1}^{n+1} |\hat{\alpha}_i|^p c_i \right)^{1/p}} = \min_{\alpha \neq 0} \frac{\left\| \sum_{i=1}^{n+1} \alpha_i g_i \right\|_p}{\left(\sum_{i=1}^{n+1} |\alpha_i|^p c_i \right)^{1/p}}.$$

Since we are at a minimum value ($\hat{\mu} > 0$) and $p \in (1, \infty)$, differentiating with respect to each α_k, we obtain

$$(4.2) \qquad \int_0^1 g_k(x) \left| \sum_{i=1}^{n+1} \hat{\alpha}_i g_i(x) \right|^{p-1} \operatorname{sgn} \left(\sum_{i=1}^{n+1} \hat{\alpha}_i g_i(x) \right) dx = \hat{\mu}^p |\hat{\alpha}_k|^{p-1} \operatorname{sgn} (\hat{\alpha}_k) c_k,$$

$$k = 1, \ldots, n+1.$$

Equation (4.2) is also satisfied for $\alpha_i = (-1)^{i+1}$, $i = 1, \ldots, n+1$ and $\mu = \lambda_n$, i.e.,

$$(4.3) \qquad \int_0^1 g_k(x) \left| \sum_{i=1}^{n+1} (-1)^{i+1} g_i(x) \right|^{p-1} \operatorname{sgn} \left(\sum_{i=1}^{n+1} (-1)^{i+1} g_i(x) \right) dx = \lambda_n^p (-1)^{k+1} c_k,$$

$$k = 1, \ldots, n+1.$$

To verify (4.3), note that choosing h_n to be positive on $(0, \xi_1)$, we have $Kh_n = \sum_{i=1}^{n+1} (-1)^{i+1} g_i$. Equation (4.3) now follows from (4.1) by multiplying both sides of of (4.1) by $|h_n|$ and integrating over (ξ_{k-1}, ξ_k), $k = 1, \ldots, n+1$.

By definition, $\lambda_n \geq \hat{\mu}$. We claim that from (4.2) and (4.3) it follows that $\lambda_n \leq \hat{\mu}$, proving the proposition.

Normalize $\{\hat{\alpha}_k\}_{k=1}^{n+1}$ so that $|\hat{\alpha}_k| \leq 1$ for all k and $\hat{\alpha}_j = (-1)^{j+1}$ for some j. If $\hat{\alpha}_k = (-1)^{k+1}$ for all k, then $\hat{\mu} = \lambda_n$. We therefore assume this not to be the case.

For ease of notation set $\sum_{i=1}^{n+1} \hat{\alpha}_i g_i = K\hat{h}$. The function

$$Kh_n(x) - K\hat{h}(x) = \sum_{\substack{i=1 \\ i \neq j}}^{n+1} ((-1)^{i+1} - \hat{\alpha}_i) \int_{\xi_{i-1}}^{\xi_i} K(x, y) |h_n(y)| \, dy$$

satisfies $Z^*(Kh_n - K\hat{h}) \leq n-1$ as a consequence of Proposition 3.4 (iii). Set $F(x) = |Kh_n(x)|^{p-1} \operatorname{sgn} (Kh_n(x)) - |K\hat{h}(x)|^{p-1} \operatorname{sgn} (K\hat{h}(x))$. Since $\operatorname{sgn} (a-b) = \operatorname{sgn} (|a|^{p-1} \operatorname{sgn} a - |b|^{p-1} \operatorname{sgn} b)$ for all $a, b \in \mathbf{R}$, it follows that $S(F) = S(Kh_n - K\hat{h}) \leq Z^*(Kh_n - K\hat{h})$

$\leq n-1$. Furthermore, from (4.2) and (4.3), $\int_0^1 g_k(x)F(x)\,dx = (\lambda_n^p(-1)^{k+1} - \hat{\mu}^p|\hat{\alpha}_k|^{p-1}\operatorname{sgn}(\hat{\alpha}_k))c_k$, for $k = 1, \ldots, n+1$. If $\lambda_n > \hat{\mu}$, then since $|\hat{\alpha}_k| \leq 1$ for all k, we have $\int_0^1 g_k(x)F(x)\,dx \cdot \int_0^1 g_{k+1}(x)F(x)\,dx < 0$, $k = 1, \ldots, n$. The $\{g_k\}_{k=1}^{n+1}$ are a T-system on [0, 1]. Thus from Proposition 3.6, $S(F) \geq n$. A contradiction has been reached. Therefore $\lambda_n \leq \hat{\mu}$. ■

Proposition 4.5. $\sup\{\|Kh - P_n^\star(Kh)\|_p : \|h\|_p \leq 1\} \leq \lambda_n$.

Proof. Set

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

Then $(Kh)(x) - P_n^\star(Kh)(x) = Mh(x) = \int_0^1 M(x, y)h(y)\,dy$. The following facts will be used:

i) $Mh_n(x) = Kh_n(x)$, since $Kh_n(\eta_i) = 0$, $i = 1, \ldots, n$.

ii) $\operatorname{sgn}(Kh_n(x))M(x, y)\operatorname{sgn}(h_n(y)) \geq 0$ with equality if and only if $x = \eta_i$ or $y = \xi_j$.

iii) $\int_0^1 M(x, y)|Mh_n(x)|^{p-1}\operatorname{sgn}(Mh_n(x))\,dx = \lambda_n^p|h_n(y)|^{p-1}\operatorname{sgn}(h_n(y))$ for all $y \in [0, 1]$. This is a result of (i) and (4.1).

Set $L(x, y) = |M(x, y)|$. Thus $L(x, y) \geq 0$ with equality if and only if $x = \eta_i$ or $y = \xi_j$ by (ii). Furthermore, from (ii) and (iii) it follows that for $H_n(y) = |h_n(y)|$

$$\int_0^1 L(x, y)(LH_n(x))^{p-1}\,dx = \lambda_n^p(H_n(y))^{p-1} \tag{4.4}$$

for all $y \in [0, 1]$.

From (ii) we see that

$$\sup_{h \neq 0} \frac{\|Mh\|_p}{\|h\|_p} = \sup_{h \neq 0} \frac{\|Lh\|_p}{\|h\|_p}.$$

This latter supremum is attained by $\hat{H}$, $\hat{\mu}$, where we may assume that $\hat{H} \geq 0$ and satisfies

$$\int_0^1 L(x, y)(L\hat{H}(x))^{p-1}\,dx = \hat{\mu}(\hat{H}(y))^{p-1}, \tag{4.5}$$

for all $y \in [0, 1]$, and

$$\hat{\mu} = \frac{\|L\hat{H}\|_p}{\|\hat{H}\|_p} = \sup_{h \neq 0} \frac{\|Lh\|_p}{\|h\|_p}.$$

We claim that there exists a β, finite, such that $\beta H_n(x) \geq \hat{H}(x)$ for all $x \in [0, 1]$. Now $H_n(x) \geq 0$ with equality only at $x = \xi_i$, $i = 1, \ldots, n$. From (4.5), $\hat{H}(\xi_i) = 0$, $i = 1, \ldots, n$. Set $\hat{h}(x) = \hat{H}(x)\operatorname{sgn}(h_n(x))$. It is easily seen that $|\hat{h}(x)|^{p-1}\operatorname{sgn}(\hat{h}(x))$,

$|h_n(x)|^{p-1}$ sgn $(h_n(x)) \in C^\infty[0, 1]$, since $M \in C^\infty([0, 1] \times [0, 1])$. Furthermore, $|h_n(x)|^{p-1}$ sgn $(h_n(x))$ has a simple zero at ξ_i, $i = 1, \ldots, n$. Set

$$\gamma_i = \lim_{x \to \xi_i} \frac{|\hat{h}(x)|^{p-1} \operatorname{sgn}(\hat{h}(x))}{|h_n(x)|^{p-1} \operatorname{sgn}(h_n(x))}, \qquad i = 1, \ldots, n.$$

These limits exist by L'Hospital's rule and $\gamma_i \geq 0$. For $\epsilon > 0$, small, there exist $\delta_i > 0$, $i = 1, \ldots, n$, such that for $x \in (\xi_i - \delta_i, \xi_i + \delta_i)$,

$$\hat{H}(x) = |\hat{h}(x)| \leq (\gamma_i + \epsilon)^{1/(p-1)} |h_n(x)| = (\gamma_i + \epsilon)^{1/(p-1)} H_n(x).$$

There therefore exists a β such that $\hat{H}(x) \leq \beta H_n(x)$ for all $x \in [0, 1]$. Let

$$\beta_0 = \inf \{\beta: \hat{H}(x) \leq \beta H_n(x)\}.$$

Thus $0 < \beta_0 < \infty$ and $\hat{H}(x) \leq \beta_0 H_n(x)$. From (4.4) and (4.5) it now follows that

$$\hat{\mu}^p(\hat{H}(y))^{p-1} \leq \beta_0^{p-1} \lambda_n^p (H_n(y))^{p-1}$$

for all $y \in [0, 1]$. From our definition of β_0 we obtain $\lambda_n \geq \hat{\mu}$. This proves the proposition. ■

Theorem 4.3 has been proved.

The uniqueness of the above h_n is a consequence of Theorem 4.3.

Proposition 4.6. *Fix $p \in (1, \infty)$ and n, a nonnegative integer. Let $f, g \in L^p$ satisfy* (4.1) *with $S(f) = S(g) = n$. Then $f = \alpha g$ for some $\alpha \in \mathbf{R}$.*

Proof. From Theorem 4.3 we see that f and g satisfy (4.1) with the same λ, $\lambda > 0$. Thus

$$\int_0^1 K(x, y)[|\alpha Kf(x)|^{p-1} \operatorname{sgn}(\alpha Kf(x)) + |\beta Kg(x)|^{p-1} \operatorname{sgn}(\beta Kg(x))]\, dx$$

$$= \lambda^p[|\alpha f(y)|^{p-1} \operatorname{sgn}(\alpha f(y)) + |\beta g(y)|^{p-1} \operatorname{sgn}(\beta g(y))]$$

for all $y \in [0, 1]$ and $\alpha, \beta \in \mathbf{R}$. We now parallel the proof of Proposition 2.8 to obtain, in this case, $Z^*(\alpha f + \beta g) = S(\alpha f + \beta g)$ for all $\alpha, \beta \in \mathbf{R}$. Obviously such an inequality cannot hold for all $\alpha, \beta \in \mathbf{R}$ unless $f = \alpha g$ for some $\alpha \in \mathbf{R}$. ■

In this same way we may parallel the proof of Proposition 2.9 to obtain $\lambda_n > \lambda_{n+1}$ for all n.

We presented Theorem 4.3 and its full proof to emphasize two points which bear consideration vis-à-vis Theorem 2.4. Firstly, the result (and this will be seen again in Sections 5 and 6) depends on the total positivity properties and only on these properties. The proof of Theorem 4.3 is technically simpler and the ideas are, we hope, more self-evident. Secondly, the existence of the free $\{x^i\}_{i=0}^{r-1}$ in Theorem 2.4 is a technical problem and may be overcome because together with the kernel $(x - y)_+^{r-1}$ certain total positivity properties still hold.

The following result is also valid. Its proof is a combination of the methods of proof of Section 2 and the results so far presented in this section. We state it because it is, we hope, of interest, and also because it is necessary in the proof of Theorem 2.1.

Let $p \in (1,\infty)$, fixed, and set

$$K_p^r = \left\{ \sum_{i=1}^r a_i k_i(x) + \int_0^1 K(x, y)h(y)\, dy : \|h\|_p \le 1,\ a_i \in \mathbf{R},\ i = 1, \ldots, r \right\}.$$

We wish to determine the n-widths of K_p^r in L^p under the following assumptions.

Property A.

1) $K \in C^\infty([0, 1] \times [0, 1])$, and $k_i \in C^\infty[0, 1]$, $i = 1, \ldots, r$.

2) $K^*\begin{pmatrix} x_1, & \ldots & , x_{s+k} \\ i_1, \ldots, i_s, & y_1, \ldots, & y_k \end{pmatrix} > 0$ for all $0 \le x_1 \le \ldots \le x_{s+k} \le 1$, $0 \le y_1 \le \ldots \le y_k \le 1$, $1 \le i_1 < \ldots < i_s \le r$, and all $s = 0, 1, \ldots, r$, $k = 0, 1, 2, \ldots$.

For distinct $\{x_i\}_{i=1}^{s+k}$ and $\{y_i\}_{i=1}^k$,

$$K^*\begin{pmatrix} x_1, & \ldots & , x_{s+k} \\ i_1, \ldots, i_s, & y_1, \ldots, & y_k \end{pmatrix} = K\begin{pmatrix} x_1, & \ldots & , x_{s+k} \\ i_1, \ldots, i_s, & y_1, \ldots, & y_k \end{pmatrix}$$
$$= \begin{vmatrix} k_{i_1}(x_1) & \ldots & k_{i_s}(x_1) & K(x_1, y_1) & \ldots & K(x_1, y_k) \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ k_{i_1}(x_{s+k}) & \ldots & k_{i_s}(x_{s+k}) & K(x_{s+k}, y_1) & \ldots & K(x_{s+k}, y_k) \end{vmatrix}.$$

When we have a block of equal x's or y's then we use the definition corresponding to Definition 3.3 (iii), i.e., consecutive derivatives.

The following variation diminishing property holds.

Proposition 4.7. *For $\{k_i\}_{i=1}^r$ and K satisfying Property A,*

$$Z^*\left(\sum_{i=1}^r a_i k_i + Kh\right) \le S(h) + S^-(a_1, \ldots, a_r)$$

if $\Sigma_{i=1}^r a_i k_i + Kh \not\equiv 0$.

In fact, to be precise, equality can hold only if the last nonzero sign in $(a_1, \ldots, a_r)$ and the sign of h near 0 (i.e., its first sign as we count from the left) are not equal.

The proof of the following result is deferred to Section 5.

Theorem 4.8. *Fix $p \in (1, \infty)$. Assume $\{k_i\}_{i=1}^r$ and K satisfy Property* A. *For each integer n, $n \ge r$, there exists an $f_n(x) = \Sigma_{i=1}^r a_i^* k_i(x) + \int_0^1 K(x, y)h_n(y)\, dy$ and a λ_n such that $S(f_n) = n$ and*

$$\int_0^1 k_i(x)\,|f_n(x)|^{p-1}\,\mathrm{sgn}\,(f_n(x))\,dx = 0, \qquad i = 1, \ldots, r,$$

$$\int_0^1 K(x, y)\,|f_n(x)|^{p-1}\,\mathrm{sgn}\,(f_n(x))\,dx = \lambda_n^p\,|h_n(y)|^{p-1}\,\mathrm{sgn}\,(h_n(y))$$

for all $y \in [0, 1]$.

On the basis of Theorem 4.8 we easily obtain $\lambda_n > 0$ and the next two results.

Proposition 4.9. *Under the assumptions of Theorem 4.8,*

$$n = S(f_n) = Z^\star(f_n) = S(h_n) + r = Z^\star(|h_n|^{p-1}\,\mathrm{sgn}\,(h_n)) + r.$$

Let $\{\eta_i\}_{i=1}^n$ denote the zeros of f_n and $\{\xi_i\}_{i=1}^{n-r}$ the zeros of h_n. Let $P_n^\star$ be the rank n operator determined by interpolation from $X_n^\star = \mathrm{span}\,\{k_1, \ldots, k_r, K(\cdot, \xi_1), \ldots, K(\cdot, \xi_{n-r})\}$ to $f \in K_p^r$ at the $\{\eta_i\}_{i=1}^n$, and set

$$g_i(x) = \int_{\xi_{i-1}}^{\xi_i} K(x, y)\,|h_n(y)|\,dy, \; i = 1, \ldots, n-r+1,$$

where $\xi_0 = 0$, $\xi_{n-r+1} = 1$.

Theorem 4.10. *Under the assumptions of Theorem 4.8,*

$$\delta_n(K_p^r;L^p) = d_n(K_p^r;L^p) = d^n(K_p^r;L^p) = b_n(K_p^r;L^p) = \|f_n\|_p/\|h_n\|_p = \lambda_n.$$

Furthermore,

a) $P_n^\star$ *is optimal for* $\delta_n(K_p^r;L^p)$;
b) $X_n^\star$ *is optimal for* $d_n(K_p^r;L^p)$;
c) $L_n^\star = \{f: f \in K_p^r, f(\eta_i) = 0, i = 1, \ldots, n\}$ *is optimal for* $d^n(K_p^r;L^p)$; *and*
d) $Y_{n+1}^\star = \mathrm{span}\,\{k_1, \ldots, k_r, g_1, \ldots, g_{n-r+1}\}$ *is optimal for* $b_n(K_p^r;L^p)$.

B. Matrices

In the proofs of Theorems 4.1 and 4.8, a matrix analogue of these results will be used.
Let $A = (a_{ij})_{i=1,\ j=1}^{M,\ N}$ be an $M \times N$ STP matrix. For $p \in (1, \infty)$, set

$$A_p = \{A\mathbf{x}: \|\mathbf{x}\|_p \le 1\}.$$

We consider the n-widths of A_p in ℓ_p^M. This problem was studied and solved in Pinkus [12], although in that paper there is little reference to n-widths, per se. This problem in a matrix setting has an additional meaning. The n-widths in the sense of Gel'fand and Bernstein are, for $p \neq 2$, generalizations of the minmax and maxmin, respectively, characterizations of the eigenvalues of A^TA (the s-numbers of A). The following results are proved in [12].

Theorem 4.11. *Fix* $p \in (1, \infty)$. *Let A be an* $M \times N$ STP *matrix and* $R = \min\,\{M, N\}$.

For each integer n, $0 \le n \le R-1$, there exists an $\mathbf{x}^n \in \mathbf{R}^N$ for which $S^-(\mathbf{x}^n) = n$ and a λ_n such that

$$\sum_{i=1}^{M} a_{ik} |(A\mathbf{x}^n)_i|^{p-1} \operatorname{sgn}((A\mathbf{x}^n)_i) = \lambda_n^p |x_k^n|^{p-1} \operatorname{sgn}(x_k^n), \qquad k = 1, \ldots, N.$$

Proposition 4.12. *Under the assumptions of Theorem 4.11, $\lambda_n > 0$ and*

$$n = S^-(\mathbf{x}^n) = S^+(\mathbf{x}^n) = S^-(A\mathbf{x}^n) = S^+(A\mathbf{x}^n).$$

The following result may also be found in [12], although this exact statement is not given.

Theorem 4.13. *Under the assumptions of Theorem 4.11,*

$$\delta_n(A_p;\ell_p^M) = d_n(A_p;\ell_p^M) = d^n(A_p;\ell_p^M) = b_n(A_p;\ell_p^M) = \|A\mathbf{x}^n\|_p / \|\mathbf{x}^n\|_p = \lambda_n.$$

The exact form of the optimal subspaces and operator for these n-widths is rather complicated. The interested reader should consult [12].

Let A be as above and $p \in (1, \infty)$. For $1 \le r \le N-1$, set

$$A_p^r = \left\{ A\mathbf{x}: \sum_{i=r+1}^{N} |x_i|^p \le 1 \right\}.$$

Note that there is no restriction on $x_1, \ldots, x_r$.

The following will be proved in full detail in Section 6.

Theorem 4.14. *Fix $p \in (1, \infty)$. Let A be an $M \times N$ STP matrix, $R = \min\{M, N\}$ and $1 \le r \le R-1$. For each integer n, $r \le n \le R-1$, there exists an $\mathbf{x}^n \in \mathbf{R}^N$ for which $S^-(\mathbf{x}^n) = n$ and a λ_n such that*

$$\sum_{i=1}^{M} a_{ik} |(A\mathbf{x}^n)_i|^{p-1} \operatorname{sgn}((A\mathbf{x}^n)_i) = 0, \qquad k = 1, \ldots, r$$

and

$$\sum_{i=1}^{M} a_{ik} |(A\mathbf{x}^n)_i|^{p-1} \operatorname{sgn}((A\mathbf{x}^n)_i) = \lambda_n^p |x_k^n|^{p-1} \operatorname{sgn}(x_k^n), \qquad k = r+1, \ldots, N.$$

From Theorem 4.14, the following results hold.

Proposition 4.15. *Under the assumptions of Theorem 4.14, $\lambda_n > 0$ and*

$$n = S^-(\mathbf{x}^n) = S^+(\mathbf{x}^n) = S^-(A\mathbf{x}^n) = S^+(A\mathbf{x}^n)$$
$$= r + S^-(x_{r+1}^n, \ldots, x_N^n) = r + S^+(x_{r+1}^n, \ldots, x_N^n).$$

Theorem 4.16. *Under the assumptions of Theorem 4.14,*

$$\delta_n(A_p^r;\ell_p^M) = d_n(A_p^r;\ell_p^M) = d^n(A_p^r;\ell_p^M) = b_n(A_p^r;\ell_p^M)$$

$$= \|A\mathbf{x}^n\|_p \Big/ \left(\sum_{i=r+1}^{N} |x_i^n|^p\right)^{1/p} = \lambda_n.$$

C. Systems of Functions

Intermediate between kernels and matrices are systems of functions. The next few results will be used as intermediate results in Section 5 and are used to simplify the proofs therein.

Let $\{g_i\}_{i=1}^N$ be an ED-system on $[0, 1]$. Fix $p \in (1, \infty)$ and set

$$G_p = \left\{\sum_{i=1}^{N} a_i g_i(x)\colon \mathbf{a} = (a_1, \ldots, a_N),\ \|\mathbf{a}\|_p \le 1\right\}.$$

In Section 5 we prove

Theorem 4.17. *Fix $p \in (1, \infty)$. Let G_p be as above. For each integer n, $0 \le n < N$, there exists an $\mathbf{a}^n \in \mathbf{R}^N$ for which $S^-(\mathbf{a}^n) = n$ and a λ_n such that*

$$\int_0^1 g_k(x) \left|\sum_{i=1}^{N} a_i^n g_i(x)\right|^{p-1} \operatorname{sgn}\left(\sum_{i=1}^{N} a_i^n g_i(x)\right) dx = \lambda_n^p |a_k^n|^{p-1} \operatorname{sgn}(a_k^n), k = 1, \ldots, N.$$

On the basis of Theorem 4.17, we have:

Proposition 4.18. *Under the assumptions of Theorem 4.17, $\lambda_n > 0$ and*

$$n = S^-(\mathbf{a}^n) = S^+(\mathbf{a}^n) = S\left(\sum_{i=1}^{N} a_i^n g_i\right) = Z^\star\left(\sum_{i=1}^{N} a_i^n g_i\right).$$

Theorem 4.19. *Under the assumptions of Theorem 4.17,*

$$\delta_n(G_p;L^p) = d_n(G_p;L^p) = d^n(G_p;L^p) = b_n(G_p;L^p)$$

$$= \left\|\sum_{i=1}^{N} a_i^n g_i\right\|_p \Big/ \|\mathbf{a}^n\|_p = \lambda_n.$$

Set

$$G_p^r = \left\{\sum_{i=1}^{N} a_i g_i(x);\ \mathbf{a} = (a_1, \ldots, a_N),\ \sum_{i=r+1}^{N} |a_i|^p \le 1\right\}$$

for some fixed r, $1 \le r < N$. In Section 5, we also prove:

Theorem 4.20. *Fix $p \in (1, \infty)$. Let $\{g_i\}_{i=1}^N$ be an ED-system and G_p^r as above. For each integer n, $r \le n < N$, there exists an $\mathbf{a}^n \in \mathbf{R}^N$ for which $S^-(\mathbf{a}^n) = n$ and a λ_n such that*

$$\int_0^1 g_k(x) \left| \sum_{i=1}^N a_i^n g_i(x) \right|^{p-1} \operatorname{sgn}\left(\sum_{i=1}^N a_i^n g_i(x) \right) dx = 0, \qquad k = 1, \ldots, r,$$

and

$$\int_0^1 g_k(x) \left| \sum_{i=1}^N a_i^n g_i(x) \right|^{p-1} \operatorname{sgn}\left(\sum_{i=1}^N a_i^n g_i(x) \right) dx = \lambda_n^p |a_k^n|^{p-1} \operatorname{sgn}(a_k^n),$$

$$k = r+1, \ldots, N.$$

On the basis of Theorem 4.20, we have:

Proposition 4.21. *Under the assumptions of Theorem 4.20, $\lambda_n > 0$ and*

$$n = S^-(\mathbf{a}^n) = S^+(\mathbf{a}^n) = S\left(\sum_{i=1}^N a_i^n g_i \right) = Z^\star\left(\sum_{i=1}^N a_i^n g_i \right)$$

$$= r + S^-(a_{r+1}^n, \ldots, a_N^n) = r + S^+(a_{r+1}^n, \ldots, a_N^n).$$

Theorem 4.22. *Under the assumptions of Theorem 4.20,*

$$\delta_n(G_p^r;L^p) = d_n(G_p^r;L^p) = d^n(G_p^r;L^p) = b_n(G_p^r;L^p)$$

$$= \left\| \sum_{i=1}^N a_i^n g_i \right\|_p \Big/ \left(\sum_{i=r+1}^N |a_i^n|^p \right)^{1/p} = \lambda_n.$$

5. Existence Theorems: Part A

We are concerned in this section with proving the existence of functions and vectors satisfying Theorem 2.1, Theorem 4.1, etc. There are two parallel lines of proof to be given. On the basis of Theorem 4.11 we prove Theorem 4.17, and from that prove Theorem 4.1. Similarly, we shall prove Theorem 4.14 and on the basis of Theorem 4.14 obtain Theorems 4.20, 4.8, and 2.1, in that order. Because of its length, we defer the proof of Theorem 4.14 to the next section. We start with the first of these cases.

Proof of Theorem 4.17. Theorem 4.17 is based on Theorem 4.11, which was proved in [12]. Let M be large ($M > N$). Set $x_i^M = \frac{i}{M}$, $i = 1, \ldots, M$, and $g_j(x_i^M) = b_{ij}^M$, $i = 1, \ldots, M, j = 1, \ldots, N$. Then $B^M = (b_{ij}^M)$ is an $M \times N$ STP matrix. From Theorem 4.11 we have, for $p \in (1, \infty)$, fixed, and each n, $0 \le n < N$, the existence of an $\mathbf{a}^M = (a_1^M, \ldots, a_N^M) \in \mathbf{R}^N$ and a λ_M for which $S^-(\mathbf{a}^M) = n$ and

$$\frac{1}{M}\sum_{i=1}^{M} b_{ik}^M |(B^M\mathbf{a}^M)_i|^{p-1} \operatorname{sgn}((B^M\mathbf{a}^M)_i) = \lambda_M^p |a_k^M|^{p-1} \operatorname{sgn}(a_k^M),\ k = 1, \ldots, N.$$

In other words,

$$(5.1)\quad \frac{1}{M}\sum_{i=1}^{M} g_k\left(\frac{i}{M}\right)\left|\sum_{j=1}^{N} a_j^M g_j\left(\frac{i}{M}\right)\right|^{p-1} \operatorname{sgn}\left[\sum_{j=1}^{N} a_j^M g_j\left(\frac{i}{M}\right)\right] = \lambda_M^p |a_k^M|^{p-1} \operatorname{sgn}(a_k^M),$$

$$k = 1, \ldots, N.$$

Furthermore, from Proposition 4.12, $\lambda_M > 0$ and $n = S^-(\mathbf{a}^M) = S^+(\mathbf{a}^M)$ for all M.

Normalize $\mathbf{a}^M$ so that $\|\mathbf{a}^M\|_p = 1$. ($\mathbf{a}^M$ is in fact uniquely determined up to multiplication by a constant.) As a function of M it is easily seen that $\lambda_M > 0$ is bounded above. There therefore exists a subsequence (again denoted by M) such that $\mathbf{a}^M \to \mathbf{a}$ and $\lambda_M \to \lambda$ as $M \uparrow \infty$. Thus $0 \le \lambda < \infty$ and $\|\mathbf{a}\|_p = 1$.

From (5.1) we obtain the desired equalities, namely

$$(5.2)\qquad \int_0^1 g_k(x)\left|\sum_{j=1}^{N} a_j g_j(x)\right|^{p-1} \operatorname{sgn}\left[\sum_{j=1}^{N} a_j g_j(x)\right] dx = \lambda^p |a_k|^{p-1} \operatorname{sgn}(a_k),$$

$$k = 1, \ldots, N.$$

It remains to prove that $S^-(\mathbf{a}) = n$. Now, $S^+(\mathbf{a}^M) = S^-(\mathbf{a}^M) = n$ for all M. From Proposition 3.3 it follows that

$$S^+(\mathbf{a}) \ge \overline{\lim_{M\to\infty}}\, S^-(\mathbf{a}^M) = n = \underline{\lim_{M\to\infty}}\, S^+(\mathbf{a}^M) \ge S^-(\mathbf{a}).$$

Since $\mathbf{a}$ satisfies (5.2) and $S^-(\mathbf{a}) \le n$, then from Proposition 4.18, $S^+(\mathbf{a}) = S^-(\mathbf{a})$ and $\lambda > 0$. Thus $S^-(\mathbf{a}) = n$. This proves Theorem 4.17. ■

Proof of Theorem 4.1. We want, for each fixed $p \in (1, \infty)$ and nonnegative integer n, to prove the existence of an $h \in L^p$ and $\lambda \in \mathbf{R}$ such that $S(h) = n$, and

$$\int_0^1 K(x, y)|Kh(x)|^{p-1} \operatorname{sgn}(Kh(x))\, dx = \lambda^p |h(y)|^{p-1} \operatorname{sgn}(h(y)),$$

for all $y \in [0, 1]$.

Set $g_i(x) = \int_{(i-1)/N}^{i/N} K(x, y)\, dy$, $i = 1, \ldots, N$. Then $\{g_i\}_{i=1}^N$ is an ED-system on $[0, 1]$ as a consequence of the ETP property of K. Assume $N > n$. From Theorem 4.17 there exists a vector (actually a unique vector) $\mathbf{a}^N \in {}^N$ for which $S^-(\mathbf{a}^N) = n$, $\|\mathbf{a}^N\|_p = 1$, $a_1^N > 0$, and a $\lambda_N > 0$ such that

$$\int_0^1 g_k(x)\left|\sum_{i=1}^{N} a_i^N g(x)\right|^{p-1} \operatorname{sgn}\left[\sum_{i=1}^{N} a_i^N g_i(x)\right] dx$$

$$= \lambda_N^p |a_k^N|^{p-1} \operatorname{sgn}(a_k^N), \qquad k = 1, \ldots, N.$$

Set $h_n(y) = a_i^N N^{1/p}$ for $y \in [(i-1)/N, i/N)$, $i = 1, \ldots, N$. Thus $\|h_N\|_p = 1$, and

$$(5.3) \qquad N \int_{(k-1)/N}^{k/N} \left(\int_0^1 K(x, z) |Kh_N(x)|^{p-1} \operatorname{sgn}(Kh_N(x))\, dx \right) dz$$
$$= \hat{\lambda}_N^p |h_N(y)|^{p-1} \operatorname{sgn}(h_N(y)),$$

for $y \in [(k-1)/N, k/N)$, $k = 1, \ldots, N$, where $\hat{\lambda}_N^p = N\lambda_N^p$. Note that $\|Kh_N\|_p = \hat{\lambda}_N \|h_N\|_p = \hat{\lambda}_N$. Thus $\hat{\lambda}_N$ is bounded above by sup $\{\|Kh\|_p : \|h\|_p \le 1\}$. N will be used as a general index. We will be repeatedly taking subsequences of N as $N \uparrow \infty$. From the fact that $\hat{\lambda}_N$ is equal to certain n-widths, it follows, at least on the subsequence 2^N, that $\hat{\lambda}_{2^N}$ is a nondecreasing sequence. Thus there exists a subsequence for which $\hat{\lambda}_N \to \lambda$ as $N \uparrow \infty$ and $0 < \lambda < \infty$.

Now $\|h_N\|_p = 1$ for all N. Taking subsequences, there exists an $h \in L^p$, $\|h\|_p \le 1$, such that $h_N \to h$ *weakly* on $[0, 1]$ as $N \uparrow \infty$. Since K induces a compact operator from L^p to C, it follows that

$$\int_0^1 K(x, y) h_N(y)\, dy \to \int_0^1 K(x, y) h(y)\, dy$$

uniformly on $[0, 1]$ as $N \uparrow \infty$, and thus

$$\int_0^1 K(x, z) |Kh_N(x)|^{p-1} \operatorname{sgn}(Kh_N(x))\, dx \to \int_0^1 K(x, y) |Kh(x)|^{p-1} \operatorname{sgn}(Kh(x))\, dx$$

uniformly on $[0, 1]$ as $N \to \infty$. Therefore

$$N \int_{(k_N-1)/N}^{k_N/N} \left(\int_0^1 K(x, w) |Kh_N(x)|^{p-1} \operatorname{sgn}(Kh_N(x))\, dx \right) dw$$
$$\to \int_0^1 K(x, z) |Kh(x)|^{p-1} \operatorname{sgn}(Kh(x))\, dx$$

uniformly on $[0, 1]$, where $\frac{k_N - 1}{N} \le z < \frac{k_N}{N}$ for all N. From (5.3) we obtain that

$$\hat{\lambda}_N^p |h_N(z)|^{p-1} \operatorname{sgn}(h_N(z)) \to \int_0^1 K(x, z) |Kh(x)|^{p-1} \operatorname{sgn}(Kh(x))\, dx$$

uniformly on $[0, 1]$ as $N \to \infty$.

Now $\hat{\lambda}_N \to \lambda$ and $h_N \to h$ weakly on $[0, 1]$. Thus

$$\hat{\lambda}_N^p |h_N(z)|^{p-1} \operatorname{sgn}(h_N(z)) \to \lambda^p |h(z)|^{p-1} \operatorname{sgn}(h(z)),$$

uniformly on $[0, 1]$, and

$$(5.4) \qquad \int_0^1 K(x, z) |Kh(x)|^{p-1} \operatorname{sgn}(Kh(x))\, dx = \lambda^p |h(z)|^{p-1} \operatorname{sgn}(h(z)),$$

for all $z \in [0, 1]$. Since $\lambda > 0$, we have proved that $h_N \to h$ uniformly on $[0, 1]$ and $\|h\|_p = 1$.

It remains to prove that $S(h) = n$. From Proposition 4.18, $S(Kh_N) = Z^\star(Kh_N) = n$

for all N. Thus, from Proposition 3.7, $S(Kh) \leq n \leq Z^*(Kh)$. However, from Lemma 4.2, $S(h) = S(Kh) = Z^*(Kh)$. Thus $S(h) = n$ and we have proved Theorem 4.1. ∎

If Theorem 4.14 is true then the above arguments, with minor modifications, prove Theorems 4.20 and 4.8. We assume the validity of Theorem 4.8 and on that basis prove Theorem 2.1. The final step of the proof will be to prove Theorem 4.14.

Proof of Theorem 2.1. We are assuming the validity of Theorem 4.8. Our aim is to prove, for fixed $p \in (1, \infty)$ and $r \geq 2$, the existence, for each positive integer n, $n \geq r$, of an $f \in W_p^{(r)}$ and $\lambda \in \mathbf{R}$ for which $S(f) = n$ and

$$(5.5) \quad \begin{cases} \int_0^1 x^i |f(x)|^{p-1} \operatorname{sgn}(f(x))\, dx = 0, \qquad i = 0, 1, \ldots, r-1, \\ \dfrac{1}{(r-1)!} \int_0^1 (x-y)_+^{r-1} |f(x)|^{p-1} \operatorname{sgn}(f(x))\, dx = \lambda^p |f^{(r)}(y)|^{p-1} \operatorname{sgn}(f^{(r)}(y)), \end{cases}$$

for all $y \in [0, 1]$.

We first simultaneously "smooth" $\{x^{i-1}\}_{i=1}^r$ and $(x-y)_+^{r-1}/(r-1)!$ (i.e., convolute with the kernel $M_\epsilon(x, y) = \dfrac{1}{\epsilon\sqrt{2\pi}} \exp\{-(x-y)^2/2\epsilon^2\}$) to obtain $\{k_i^\epsilon\}_{i=1}^r$ and $K^\epsilon (\epsilon > 0)$ which satisfy Property A of Section 4. Note that $k_i^\epsilon(x) \to x^{i-1}$, $i = 1, \ldots, r$, and $K^\epsilon(x, y) \to (x-y)_+^{r-1}/(r-1)!$ uniformly as $\epsilon \downarrow 0$. From Theorem 4.8 there exists an

$$f^\epsilon(x) = \sum_{i=1}^r a_i^\epsilon k_i^\epsilon(x) + \int_0^1 K^\epsilon(x, y) h^\epsilon(y)\, dy$$

and a $\lambda_\epsilon > 0$ such that $S(f^\epsilon) = n$, $\|h^\epsilon\|_p = 1$, and

$$(5.6) \quad \begin{cases} \int_0^1 k_i^\epsilon(x) |f^\epsilon(x)|^{p-1} \operatorname{sgn}(f^\epsilon(x))\, dx = 0, \qquad i = 1, \ldots, r, \\ \int_0^1 K^\epsilon(x, y) |f^\epsilon(x)|^{p-1} \operatorname{sgn}(f^\epsilon(x))\, dx = \lambda_\epsilon^p |h^\epsilon(y)|^{p-1} \operatorname{sgn}(h^\epsilon(y)), \end{cases}$$

for all $y \in [0, 1]$.

It is an easy matter to show that the a_i^ϵ remain bounded as does λ^ϵ as $\epsilon \downarrow 0$. Furthermore, a Birman and Solomjak [1] type argument proves that $\lambda^\epsilon \geq An^{-r}$ for some $A > 0$ independent of n, r, and ϵ. Thus λ^ϵ is bounded below away from zero. Therefore, along some subsequence (always indexed by ϵ), $a_i^\epsilon \to a_i$, $i = 1, \ldots, r$, $\lambda^\epsilon \to \lambda$, $0 < \lambda < \infty$, and $h^\epsilon \to h$ weakly on $[0, 1]$. We now apply the reasoning of the previous proof. Thus $f^\epsilon \to f$ uniformly on $[0, 1]$, which in turn implies that $|f^\epsilon|^{p-1} \operatorname{sgn}(f^\epsilon) \to |f|^{p-1} \operatorname{sgn}(f)$ uniformly on $[0, 1]$. From (5.6) it follows that $\lambda^\epsilon |h^\epsilon|^{p-1} \operatorname{sgn}(h^\epsilon) \to$

$\lambda|h|^{p-1}$ sgn (h) uniformly on $[0, 1]$, and since $\lambda > 0$, $h^\epsilon \to h$ uniformly on $[0, 1]$. We therefore obtain from (5.6) the existence of an $f \in W_p^{(r)}$ (note that $f^{(r)} = h$) that satisfies (5.5) and such that $\|f^{(r)}\|_p = 1$.

It remains to prove that $S(f) = n$. Since $S(f^\epsilon) = n$ for all $\epsilon > 0$ and $f^\epsilon \to f$ uniformly on $[0, 1]$, it is a simple matter to prove that $S(f) \leq n$. Since f satisfies (5.5) it follows from Proposition 2.3 that $S(f) = Z^*(f)$, where here Z^* counts zeros up to multiplicity at most r. Recall that $r \geq 2$, so that all the zeros of f are in $(0, 1)$ and are simple. Now $S(f^\epsilon) = Z^*(f^\epsilon) = n$ for all $\epsilon > 0$ by Proposition 4.9, i.e., f^ϵ has n simple zeros and no others. If $S(f) < n$, then as $\epsilon \downarrow 0$ some of the zeros of f^ϵ either go to the end points or collapse on each other (coalesce). In either case we would obtain $S(f) < Z^*(f)$. This contradicts Proposition 2.3, so $S(f) = n$. ■

6. Existence Theorems: Part B

It remains to prove Theorem 4.14. This is a generalization of Theorem 4.11, which was proved in [12]. Theorem 4.11 is a special case of Theorem 4.14, i.e., for $r = 0$. While Theorem 4.11 was proven by a fairly short elegant proof using the Liusternik-Schnirelman Theorem on critical points, we were unable to extend that method to this case. We do believe, however, that this shorter proof will extend, and a proof along these lines would be of interest. The idea of the proof of Theorem 4.14 presented here is to start at $p = 2$, where the theorem may be directly verified, and then to extend the result via the Implicit Function Theorem to all $p \in (2, \infty)$. For the case $r = 0$ (Theorem 4.11) this suffices, from duality considerations, to prove the result for all $p \in (1, \infty)$. For $r > 0$, the dual problem is different in character and we must substantially rework the proof to verify Theorem 4.14 for $p \in (1, 2)$. Our proof of Theorem 4.14 is rather lengthy and is therefore divided into a series of steps. Before proceeding with the proof, we restate, for convenience, our final goal.

Fix $p \in (1, \infty)$. Let A be an $M \times N$ STP matrix and $R = \min\{M, N\}$. Fix r, $1 \leq r \leq R-1$. For each integer n, $r \leq n \leq R-1$, we claim that there exists an $\mathbf{x} \in \mathbf{R}^N$ for which $S^-(\mathbf{x}) = n$ and a σ such that

$$\sum_{i=1}^{M} a_{ik}|(A\mathbf{x})_i|^{p-1} \operatorname{sgn}((A\mathbf{x})_i) = 0, \qquad k = 1, \ldots, r, \tag{6.1}$$

and

$$\sum_{i=1}^{M} a_{ik}|(A\mathbf{x})_i|^{p-1} \operatorname{sgn}((A\mathbf{x})_i) = \sigma|x_k|^{p-1} \operatorname{sgn}(x_k), \qquad k = r+1, \ldots, N. \tag{6.2}$$

(For convenience we have replaced λ^p by σ). Note that $\|A\mathbf{x}\|_p^p = \sigma(\Sigma_{k=r+1}^N |x_k|^p)$ from (6.1) and (6.2). It therefore follows from Proposition 3.1 that if $\Sigma_{k=r+1}^N |x_k|^p = 1$, then since $r \leq S^-(\mathbf{x}) = n \leq R-1$ we have $A\mathbf{x} \neq \mathbf{0}$, so that $\sigma > 0$. We start by first proving the case $p = 2$.

Proposition 6.1. *Theorem 4.14 holds for $p = 2$.*

Proof. Let $B = A^TA$. B is a symmetric $N \times N$ matrix and is STP of order rank B, where rank B = rank $A = R = \min\{M, N\}$. Equations (6.1) and (6.2) are equivalent, for $p = 2$, to

$$(B\mathbf{x})_k = \begin{cases} 0, & k = 1, \ldots, r, \\ \sigma x_k, & k = r+1, \ldots, N. \end{cases}$$

Set

$$c_{ij} = \frac{B\begin{pmatrix} 1, \ldots, r, i \\ 1, \ldots, r, j \end{pmatrix}}{B\begin{pmatrix} 1, \ldots, r \\ 1, \ldots, r \end{pmatrix}}, \qquad i,j = r+1, \ldots, N,$$

and $C = (c_{ij})_{i,j=r+1}^N$. From Sylvester's Determinant Identity (see, e.g., Karlin [3], p. 3), it follows that C is an STP $(N-r) \times (N-r)$ matrix of order rank $C = R - r$. Thus from Proposition 3.2, C has $R - r$ positive distinct eigenvalues

$$\sigma_{r+1} > \sigma_{r+2} > \ldots > \sigma_R > 0.$$

Let $\hat{\mathbf{x}}^i = (x_{r+1}^i, \ldots, x_N^i)$, $i = r+1, \ldots, R$, denote the associated eigenvectors, i.e., $C\hat{\mathbf{x}}^i = \sigma_i \hat{\mathbf{x}}^i$, $i = r+1, \ldots, R$. From Proposition 3.2 we have $S^+(\hat{\mathbf{x}}^i) = S^-(\hat{\mathbf{x}}^i) = i - 1 - r$, $i = r+1, \ldots, R$.

Define $\mathbf{x}^i = (x_1^i, \ldots, x_r^i, x_{r+1}^i, \ldots, x_N^i)$, $i = r+1, \ldots, R$, where the $x_1^i, \ldots, x_r^i$ are uniquely determined by the conditions

$$\sum_{j=1}^{r} b_{kj} x_j^i + \sum_{j=r+1}^{N} b_{kj} x_j^i = 0, \qquad k = 1, \ldots, r,$$

for each $i = r+1, \ldots, R$. We first claim that

$$(B\mathbf{x}^i)_k = \begin{cases} 0, & k = 1, \ldots, r, \\ \sigma_i x_k^i, & k = r+1, \ldots, N. \end{cases}$$

For $k = 1, \ldots, r$, this holds by construction. For $k = r+1, \ldots, N$,

$$\sigma_i x_k^i = \sum_{j=r+1}^{N} \frac{B\begin{pmatrix} 1, \ldots, r, k \\ 1, \ldots, r, j \end{pmatrix}}{B\begin{pmatrix} 1, \ldots, r \\ 1, \ldots, r \end{pmatrix}} x_j^i$$

$$= \sum_{j=r+1}^{N} b_{kj} x_j^i + \sum_{j=r+1}^{N} \left[\sum_{\ell=1}^{r} (-1)^{\ell+r-1} \frac{B\begin{pmatrix} 1, \ldots, \hat{\ell}, \ldots, r, k \\ 1, \quad \ldots \quad, r \end{pmatrix}}{B\begin{pmatrix} 1, \ldots, r \\ 1, \ldots, r \end{pmatrix}} b_{\ell j} x_j^i \right]$$

$$= \sum_{j=r+1}^{N} b_{kj}x_j^i + \sum_{\ell=1}^{r} (-1)^{\ell+r-1} \frac{B\begin{pmatrix} 1, \dots, \hat{\ell}, \dots, r, k \\ 1, \quad . \quad . \quad . \quad , \quad r \end{pmatrix}}{B\begin{pmatrix} 1, \dots, r \\ 1, \dots, r \end{pmatrix}} \sum_{j=r+1}^{N} b_{\ell j} z_j^i,$$

where the ^ on the index ℓ indicates that that row is deleted. Since $\ell \in \{1, \dots, r\}$,

$$\sum_{j=r+1}^{N} b_{\ell j} x_j^i = -\sum_{j=1}^{r} b_{\ell j} x_j^i$$

Thus

$$\sum_{\ell=1}^{r} (-1)^{\ell+r-1} \frac{B\begin{pmatrix} 1, \dots, \hat{\ell}, \dots, r, k \\ 1, \quad . \quad . \quad . \quad , \quad r \end{pmatrix}}{B\begin{pmatrix} 1, \dots, r \\ 1, \dots, r \end{pmatrix}} \sum_{j=r+1}^{N} b_{\ell j} x_j^i$$

$$= \sum_{\ell=1}^{r} (-1)^{\ell+r} \frac{B\begin{pmatrix} 1, \dots, \hat{\ell}, \dots, r, k \\ 1, \quad . \quad . \quad . \quad , \quad r \end{pmatrix}}{B\begin{pmatrix} 1, \dots, r \\ 1, \dots, r \end{pmatrix}} \sum_{j=1}^{r} b_{\ell j} x_j^i$$

$$= \sum_{j=1}^{r} \left(\sum_{\ell=1}^{r} (-1)^{\ell+r} \frac{B\begin{pmatrix} 1, \dots, \hat{\ell}, \dots, r, k \\ 1, \quad . \quad . \quad . \quad , \quad r \end{pmatrix}}{B\begin{pmatrix} 1, \dots, r \\ 1, \dots, r \end{pmatrix}} b_{\ell j} \right) x_j^i.$$

For $j = 1, \dots, r$,

$$0 = \frac{B\begin{pmatrix} 1, \dots, r, k \\ 1, \dots, r, j \end{pmatrix}}{B\begin{pmatrix} 1, \dots, r \\ 1, \dots, r \end{pmatrix}} = b_{kj} + \sum_{\ell=1}^{r} (-1)^{\ell+r-1} \frac{B\begin{pmatrix} 1, \dots, \hat{\ell}, \dots, r, k \\ 1, \quad . \quad . \quad . \quad , \quad r \end{pmatrix}}{B\begin{pmatrix} 1, \dots, r \\ 1, \dots, r \end{pmatrix}} b_{\ell j}.$$

Substituting, we obtain

$$\sigma_i x_k^i = \sum_{j=r+1}^{N} b_{kj} x_j^i + \sum_{j=1}^{r} b_{kj} x_j^i = (B\mathbf{x})_k .$$

It remains to prove that $S^-(\mathbf{x}^i) = i-1$, for $i = r+1, \dots, R$. *Since* $S^-(\hat{\mathbf{x}}^i) = S^+(\hat{\mathbf{x}}^i) = i-1-r$, we have $S^+(0, \dots, 0, \sigma_i x_{r+1}^i, \dots, \sigma_i x_N^i) \geq i-1$. Thus $i-1 \leq S^+(B\mathbf{x}^i) \leq S^-(\mathbf{x}^i) \leq S^+(\mathbf{x}^i) \leq r + S^+(\hat{\mathbf{x}}^i) \leq i-1$, implying that $S^-(\mathbf{x}^i) = S^+(\mathbf{x}^i) = i-1$, $i = r+1, \dots, R$. ∎

For $\mathbf{x} = (x_1, \dots, x_N) \in \mathbf{R}^N$, $\sigma \in \mathbf{R}$, and $p \in (1, \infty)$, let

$$G(\mathbf{x}, \sigma, p) = (G_1(\mathbf{x}, \sigma, p), \ldots, G_{N+1}(\mathbf{x}, \sigma, p)) \in \mathbf{R}^{N+1},$$

where

$$G_k(\mathbf{x}, \sigma, p) = \sum_{i=1}^{M} a_{ik} |(A\mathbf{x})_i|^{p-1} \operatorname{sgn}((A\mathbf{x})_i), \qquad k = 1, \ldots, r,$$

$$G_k(\mathbf{x}, \sigma, p) = \sum_{i=1}^{M} a_{ik} |(A\mathbf{x})_i|^{p-1} \operatorname{sgn}((A\mathbf{x})_i) - \sigma |x_k|^{p-1} \operatorname{sgn}(x_k), \qquad k = r+1, \ldots, N,$$

and

$$G_{N+1}(\mathbf{x}, \sigma, p) = \sum_{k=r+1}^{N} |x_k|^p - 1.$$

For ease of notation we shall sometimes write x_{N+1} for σ.

To use the Implicit Function Theorem (to vary p), we must prove that a certain Jacobian is nonzero. This is the crux of the next series of results.

Lemma 6.2. *For* $p \in (2, \infty)$, $G(\mathbf{x}, \sigma, p) \in C^1(\mathbf{R}^{N+2}, \mathbf{R}^{N+1})$.

Proof. The continuity of $G(\mathbf{x}, \sigma, p)$ as a function of $x_1, \ldots, x_N$, σ and $p \in (1, \infty)$ is obvious. Since we shall need the partial derivatives for later analysis, we simply evaluate them to show that they too are continuous.

For $k = 1, \ldots, r$,

$$\frac{\partial G_k(\mathbf{x}, \sigma, p)}{\partial x_\ell} = (p-1)\left[\sum_{i=1}^{M} a_{ik} a_{i\ell} |(A\mathbf{x})_i|^{p-2}\right], \qquad \ell = 1, \ldots, N,$$

is continuous for $p > 2$. The function $\partial G_k(\mathbf{x}, \sigma, p)/\partial\sigma = 0$, while $\partial G_k(\mathbf{x}, \sigma, p)/\partial p = \Sigma_{i=1}^{M} a_{ik} |(A\mathbf{x})_i|^{p-1} \operatorname{sgn}((A\mathbf{x})_i) \ln(|(A\mathbf{x})_i|)$. This latter expression is continuous for $p > 1$, since $\lim_{x \to 0} |x|^{p-1} \ln(|x|) = 0$ for all $p > 1$. For $k = r+1, \ldots, N$,

$$\frac{\partial G_k(\mathbf{x}, \sigma, p)}{\partial x_\ell} = (p-1)\left[\sum_{\ell=1}^{M} a_{ik} a_{i\ell} |(A\mathbf{x})_i|^{p-2} - \sigma |x_\ell|^{p-2} \delta_{\ell k}\right], \qquad \ell = 1, \ldots, N,$$

$$\frac{\partial G_k(\mathbf{x}, \sigma, p)}{\partial \sigma} = -|x_k|^{p-1} \operatorname{sgn}(x_k),$$

and

$$\frac{\partial G_k(\mathbf{x}, \sigma, p)}{\partial p} = \sum_{i=1}^{M} a_{ik} |(A\mathbf{x})_i|^{p-1} \operatorname{sgn}((A\mathbf{x})_i) \ln(|(A\mathbf{x})_i|) - \sigma |x_k|^{p-1} \operatorname{sgn}(x_k) \ln(|x_k|).$$

The second and third expressions are continuous for $p > 1$, while the first expression is continuous for $p > 2$.

For $k = N+1$,

$$\frac{\partial G_{N+1}(\mathbf{x}, \sigma, p)}{\partial x_\ell} = 0, \qquad \ell = 1, \ldots, r,$$

$$\frac{\partial G_{N+1}(\mathbf{x}, \sigma, p)}{\partial x_\ell} = p|x_\ell|^{p-1} \operatorname{sgn}(x_\ell), \qquad \ell = r+1, \ldots, N,$$

$$\frac{\partial G_{N+1}(\mathbf{x}, \sigma, p)}{\partial \sigma} = 0,$$

and

$$\frac{\partial G_{N+1}(\mathbf{x}, \sigma, p)}{\partial p} = \sum_{k=r+1}^{N} |x_k|^p \ln(|x_k|),$$

are all continuous for $p > 1$.

One of our problems is that of starting the analysis at some point p. The only information we presently possess is at $p = 2$, where $G(\mathbf{x}, \sigma, p)$ is not necessarily a C^1 function. Let $\hat{\mathbf{x}} = (x_{r+1}, \ldots, x_N)$. A reading of the proof of Lemma 6.2 proves:

Lemma 6.3. *If $\hat{\mathbf{x}}$ and $A\mathbf{x}$ have no zero entries, then $G(\mathbf{x}, \sigma, p)$ for $p = 2$ is locally a C^1 function.*

It is to be understood that we are taking right limits.

The crux of the proof of Theorem 4.14 for $p \in (2, \infty)$ is contained in the next result.

Proposition 6.4. *Let $(\mathbf{x}^\star, \sigma^\star, p^\star)$, $p^\star \geq 2$, satisfy $G(\mathbf{x}^\star, \sigma^\star, p^\star) = \mathbf{0}$, $\sigma^\star > 0$. In addition assume that if $p^\star = 2$, then $\hat{\mathbf{x}}^\star$ and $A\mathbf{x}^\star$ have no zero entries. Then*

$$\det \left[\frac{\partial G_k}{\partial x_\ell}\right]_{k,\ell=1}^{N+1} \Bigg|_{(\mathbf{x}^\star, \sigma^\star, p^\star)} \neq 0.$$

Proof. Recall that for notational ease we have set $x_{N+1} = \sigma$. Assume that the determinant is zero. There therefore exists a vector $\mathbf{z} \in \mathbf{R}^{N+1}$, $\mathbf{z} \neq \mathbf{0}$, such that

$$\sum_{\ell=1}^{N+1} \frac{\partial G_k}{\partial x_\ell} z_\ell \Bigg|_{(\mathbf{x}^\star, \sigma^\star, p^\star)} = 0, \qquad k = 1, \ldots, N+1.$$

This translates into

$$(6.3) \qquad (p^\star - 1)\left[\sum_{\ell=1}^{N} \sum_{i=1}^{M} a_{ik} a_{i\ell} |(A\mathbf{x}^\star)_i|^{p^\star-2} z_\ell\right] = 0, \qquad k = 1, \ldots, r,$$

$$(6.4) \qquad (p^\star - 1)\left[\sum_{\ell=1}^{N} \sum_{i=1}^{M} a_{ik} a_{i\ell} |(A\mathbf{x}^\star)_i|^{p^\star-2} z_\ell - \sigma^\star |x_k^\star|^{p^\star-2} z_k\right]$$

$$- |x_k^\star|^{p^\star-1} \operatorname{sgn}(x_k^\star) z_{N+1} = 0, \qquad k = r+1, \ldots, N,$$

and

$$(6.5) \qquad p^\star \sum_{\ell=r+1}^{N} |x_\ell^\star|^{p^\star-1} \operatorname{sgn}(x_\ell^\star) z_\ell = 0.$$

We multiply (6.3) and (6.4) by $x_k^\star$, sum over k, and use the fact that $G(\mathbf{x}^\star, \sigma^\star, p^\star) = \mathbf{0}$ to obtain $z_{N+1} = 0$.

Since $z_{N+1} = 0$, we now consider $\mathbf{z} = (z_1, \ldots, z_N)$, $\mathbf{z} \neq \mathbf{0}$, in $\mathbf{R}^N$ satisfying (6.3)-(6.5). These equations may be rewritten as

$$(6.6) \qquad \sum_{\ell=1}^{N} \sum_{i=1}^{M} a_{ik} a_{i\ell} |(A\mathbf{x}^\star)_i|^{p^\star-2} z_\ell = 0, \qquad k = 1, \ldots, r,$$

$$(6.7) \quad \sum_{\ell=1}^{N} \sum_{i=1}^{M} a_{ik} a_{i\ell} |(A\mathbf{x}^\star)_i|^{p^\star-2} z_\ell = \sigma^\star |x_k^\star|^{p^\star-2} z_k, \qquad k = r+1, \ldots, N,$$

and

$$(6.8) \qquad \sum_{\ell=r+1}^{N} |x_\ell^\star|^{p^\star-1} \operatorname{sgn}(x_\ell^\star) z_\ell = 0.$$

From $G(\mathbf{x}^\star, \sigma^\star, p^\star) = \mathbf{0}$, we see that (6.6) and (6.7) also hold where $\mathbf{x}^\star$ replaces $\mathbf{z}$. Since $\mathbf{x}^\star$ satisfies (6.6), it follows that $S^-(\mathbf{x}^\star) \geq r$. Furthermore, since $\sigma^\star > 0$, it is easily proven that $S^-(\mathbf{x}^\star) \leq R-1$. Now $\mathbf{x}^\star$, $\mathbf{z}$ are linearly independent solutions of (6.6) and (6.7). We prove that $\mathbf{z} = \mathbf{0}$, which implies the proposition.

Case I. Assume $\mathbf{x}^\star$, $\mathbf{z}$ *satisfy* (6.6) *and* (6.7), *and* $x_k^\star z_k = 0$ *for all* $k = r+1, \ldots, N$.

Set $B = (b_{k\ell})_{k,\ell=1}^N$ where

$$b_{k\ell} = \sum_{i=1}^{M} a_{ik} a_{i\ell} |(A\mathbf{x}^\star)_i|^{p^\star-2}.$$

From (6.6), (6.7), and our assumption, it follows that $B\mathbf{z} = \mathbf{0}$. B is an $N \times N$ matrix. Set $d_i = |(A\mathbf{x}^\star)_i|^{p^\star-2}$. By the Cauchy-Binet formula,

$$B\begin{pmatrix} i_1, \ldots, i_k \\ j_1, \ldots, j_k \end{pmatrix} = \sum_{1 \leq \alpha_1 < \ldots < \alpha_k \leq M} A\begin{pmatrix} a_1, \ldots, \alpha_k \\ i_1, \ldots, i_k \end{pmatrix} A\begin{pmatrix} \alpha_1, \ldots, \alpha_k \\ j_1, \ldots, j_k \end{pmatrix} d_{\alpha_1} \ldots d_{\alpha_k}.$$

Thus B is STP of order rank B, where rank $B = \min\{N, \#\{i, d_i \neq 0\}\}$. If rank $B = N$, then since $B\mathbf{z} = \mathbf{0}$, it follows that $\mathbf{z} = \mathbf{0}$. Assume therefore that rank $B = \#\{i: d_i \neq 0\} < N$. From Proposition 3.1 it follows that $S^-(\mathbf{z}) \geq$ rank B. Let $m = S^-(\mathbf{x}^\star) = S^+(\mathbf{x}^\star) = S^-(A\mathbf{x}^\star) = S^+(A\mathbf{x}^\star) = r + S^+(x_{r+1}^\star, \ldots, x_N^\star)$ (Proposition 4.15). Thus $m = S^-(A\mathbf{x}^\star) = S^-(d_1, \ldots, d_M) \leq \#\{i: d_i \neq 0\} - 1 =$ rank $B - 1$, i.e., rank $B \geq M+1$. On the other hand, since $S^-(\mathbf{z}) \geq$ rank B, it follows that $\#\{i: z_i \neq 0\} \geq$ rank $B + 1$. From our assumption, $z_i \neq 0$ implies $x_i^\star = 0$ for $i =$

$r+1, \ldots, N$. Therefore, $\#\{i: x_i^\star = 0,\ i = r+1, \ldots, N\} \geq \#\{i: z_i \neq 0,\ i = r+1, \ldots, N\} \geq \operatorname{rank} B + 1 - r$, and thus $m = r + S^+(x_{r+1}^\star, \ldots, x_N^\star) \geq \operatorname{rank} B + 1$. This is a contradiction.

Case II. Assume $\mathbf{x}^\star$, $\mathbf{z}$ *satisfy* (6.6) and (6.7) *and* $\mathbf{z}$ *satisfies* (6.8). *Assume also that there exists a* j, $r+1 \leq j \leq N$, *such that* $x_j^\star z_j \neq 0$.

$\mathbf{x}^\star$ and $\mathbf{z}$ are linearly independent. Hence for every $\alpha \in \mathbf{R}$, $\alpha\mathbf{x}^\star + \mathbf{z} \neq \mathbf{0}$. Thus $S^-(A(\alpha\mathbf{x}^\star + \mathbf{z})) \leq S^+(A(\alpha\mathbf{x}^\star + \mathbf{z})) \leq S^-(\alpha\mathbf{x}^\star + \mathbf{z}) \leq S^+(\alpha\mathbf{x}^\star + \mathbf{z}) \leq S^+(0, \ldots, 0, \sigma^\star|x_{r+1}^\star|^{p^\star-2}(\alpha x_{r+1}^\star + z_{r+1}), \ldots, \sigma^\star|x_N^\star|^{p^\star-2}(\alpha x_N^\star + z_N))$. If the vector $\{\sigma^\star|x_k^\star|^{p^\star-2}(\alpha x_k^\star + z_k)\}_{k=r+1}^N$ is identically zero, then $|x_k^\star| \neq 0$ implies $z_k = -\alpha x_k^\star$. Substituting in (6.8), we obtain

$$\begin{aligned} 0 &= \sum_{\ell=r+1}^{N} |x_\ell^\star|^{p^\star+1} \operatorname{sgn}(x_\ell^\star) z_\ell \\ &= -\alpha \sum_{\ell=r+1}^{N} |x_\ell^\star|^{p^\star} \\ &= -\alpha. \end{aligned}$$

Since $\alpha = 0$ we see that $x_k^\star \neq 0$ implies $z_k = 0$, $k = r+1, \ldots, N$. This contradicts our assumption. The above vector is therefore not identically zero. From (6.6), (6.7), and Proposition 3.1,

$$\begin{aligned} S^+(0, \ldots, 0, \sigma^\star|x_{r+1}^\star|^{p^\star-2}&(\alpha x^\star_{r+1} + z_{r+1}), \ldots, \sigma^\star|x_N^\star|^{p^\star-2}(\alpha x_N^\star + z_N)) \\ &= S^+\left(\left\{\sum_{i=1}^{M} a_{ik} \sum_{\ell=1}^{N} a_{i\ell}|(A\mathbf{x}^\star)_i|^{p^\star-2}(\alpha x_\ell^\star + z_\ell)\right\}_{k=1}^{N}\right) \\ &\leq S^-\left(\left\{\sum_{\ell=1}^{N} a_{i\ell}|(A\mathbf{x}^\star)_i|^{p^\star-2}(\alpha x_\ell^\star + z_\ell)\right\}_{i=1}^{M}\right) \\ &\leq S^-\left(\left\{\sum_{\ell=1}^{N} a_{i\ell}(\alpha x_\ell^\star + z_\ell)\right\}_{i=1}^{M}\right) \\ &= S^-(A(\alpha\mathbf{x}^\star + \mathbf{z})). \end{aligned}$$

We have come full circle so that equality holds throughout. In particular, $S^-(\alpha\mathbf{x}^\star + \mathbf{z}) = S^+(\alpha\mathbf{x}^\star + \mathbf{z})$ for all $\alpha \in \mathbf{R}$. This is a contradiction, since we may choose $\alpha^\star \in \mathbf{R}$ for which $(\alpha^\star\mathbf{x}^\star + \mathbf{z})_N = 0$. (From Proposition 4.15, $x_N^\star \neq 0$.) ■

Fix n, $r \leq n \leq R-1$. We know from Proposition 6.1 that Theorem 4.14 is valid for $p = 2$. We now extend this result to $p \in (2, \infty)$ under the assumptions of Lemma 6.3.

Proposition 6.5. *Set* $r \leq n \leq R-1$. *Let* $G(\mathbf{x}, \sigma, 2) = \mathbf{0}$ *with* $S^-(\mathbf{x}) = n$, *and* *assume that* $\hat{\mathbf{x}}$ *and* $A\mathbf{x}$ *have no zero entries. Then Theorem 4.14 is valid for this n and all* $p \in (2, \infty)$.

Proof. Starting from $p = 2$, there exists, by the Implicit Function Theorem and Proposition 6.4, a $\hat{p} > 2$ such that for all $p \in [2, \hat{p})$ there are continuously differentiable functions of p, $\mathbf{x}(p) = (x_1(p), \ldots, x_N(p))$ and $\sigma(p)$, for which $G(\mathbf{x}(p), \sigma(p), p) = \mathbf{0}$. Furthermore, since $S^+(\mathbf{x}) = S^-(\mathbf{x}) = n$ and $\sigma > 0$ for $p = 2$, we have $n = S^+(\mathbf{x}(p)) = S^-(\mathbf{x}(p))$ for p near 2.

Let $p^\star > 2$ be the largest value for which for all $p \in [2, p^\star)$ there exist $\mathbf{x}(p)$ and $\sigma(p)$ such that $G(\mathbf{x}(p), \sigma(p), p) = \mathbf{0}$ and $S^-(\mathbf{x}(p)) = n$. We claim that $p^\star = \infty$. Assume $p^\star < \infty$. Then there exists a subsequence $p^k \uparrow p^\star$ such that along this subsequence $\mathbf{x}(p^k) \to \mathbf{x}^\star$ and $\sigma(p^k) \to \sigma^\star$. (The boundedness of the $\mathbf{x}(p^k)$ and $\sigma(p^k)$ is easily proven.) From Proposition 4.15, $S^-(\mathbf{x}(p^k)) = S^+(\mathbf{x}(p^k)) = n$ and $\sigma(p^k) > 0$. Thus, from Proposition 3.3, $S^-(\mathbf{x}^\star) \le n \le S^+(\mathbf{x}^\star)$, and $G(\mathbf{x}^\star, \sigma^\star, p^\star) = \mathbf{0}$. Since $G_k(\mathbf{x}^\star, \sigma^\star, p^\star) = 0$, $k = 1, \ldots, r$, it follows from Proposition 3.1 that $S^-(\mathbf{x}^\star) \ge r$. Applying Proposition 4.15 once again, we see that $S^-(\mathbf{x}^\star) = S^+(\mathbf{x}^\star)$ and $\sigma^\star > 0$. In particular, this implies that $S^-(\mathbf{x}^\star) = n$. For $p = p^\star$ we may again apply the Implicit Function Theorem and Proposition 6.4 to contradict the definition of $p^\star$. Thus $p^\star = \infty$. ■

It remains to consider the case where $\mathbf{x}$ satisfies $G(\mathbf{x}, \sigma, 2) = \mathbf{0}$, $S^-(\mathbf{x}) = n$, but the vectors $\hat{\mathbf{x}}$ and/or $A\mathbf{x}$ have zero entries. This is a technical problem, which we overcome by perturbation.

Lemma 6.6. *Let* $G(\mathbf{x}, \sigma, 2) = \mathbf{0}$ *and* $S^-(\mathbf{x}) = n$, $r \le n \le R-1$. *For* $\epsilon > 0$, *small, there exists* A_ϵ *STP such that* $A_\epsilon \to A$ *as* $\epsilon \downarrow 0$, *and* $\mathbf{x}^\epsilon \in \mathbf{R}^N$ *such that* $\mathbf{x}^\epsilon \to \mathbf{x}$ as $\epsilon \downarrow 0$. *Furthermore,*

1) $S^-(\mathbf{x}^\epsilon) = n$,
2) $A_\epsilon^T A_\epsilon \mathbf{x}^\epsilon = \sigma(0, \ldots, 0, x_{r+1}^\epsilon, \ldots, x_N^\epsilon)^T$, *and*
3) *neither* $\hat{\mathbf{x}}^\epsilon$ *nor* $A_\epsilon \mathbf{x}^\epsilon$ *has zero entries.*

Proof. Assume $x_k = 0$, $r+1 \le k \le N$. From Proposition 4.15, $r+2 \le k \le N-1$. Set $B_\epsilon = (b_{ij}^\epsilon)_{i,j=1}^N$, where

$$b_{ij}^\epsilon = \begin{cases} 1, & i = j,\ i \ne k, k+1, \\ 1-\epsilon, & i = j,\ i = k, k+1, \\ \sqrt{2\epsilon - \epsilon^2}, & i = k+1, j = k, \\ -\sqrt{2\epsilon - \epsilon^2}, & i = k, j = k+1, \\ 0, & \text{otherwise}, \end{cases}$$

for $\epsilon > 0$, small. Since B_ϵ is close to I, $A_\epsilon = AB_\epsilon$ is STP for ϵ sufficiently small. Furthermore, B_ϵ is unitary; i.e., $B_\epsilon^{-1} = B_\epsilon^T$. Thus $A_\epsilon^T A_\epsilon = B_\epsilon^{-1} A^T A B_\epsilon$. Let

$$(\mathbf{x}^\epsilon)_i = \begin{cases} x_i &, \quad i \neq k,k+1, \\ \sqrt{2\epsilon - \epsilon^2}x_{k+1}, & \quad i = k, \\ (1-\epsilon)x_{k+1} &, \quad i = k+1. \end{cases}$$

Then $B_\epsilon \mathbf{x}^\epsilon = \mathbf{x}$ and $S^-(\mathbf{x}^\epsilon) = n$. Furthermore, $A_\epsilon^T A_\epsilon \mathbf{x}^\epsilon$ satisfies (2). We apply this result as many times as is necessary so that $\hat{\mathbf{x}}^\epsilon$ has no zero entries. The perturbation used to obtain $A_\epsilon \mathbf{x}^\epsilon$ with no zero entries is similar. We construct an $M \times M$ STP matrix C_ϵ similar to B_ϵ and set $A_\epsilon = C_\epsilon A$. (The vector $\mathbf{x}$ is left unchanged.) The rest of the proof now follows. ∎

In this manner we perturb away from our problem at $p = 2$. We can now apply Proposition 6.5 to obtain the desired result for all $p \in (2, \infty)$, but for A_ϵ rather than A. We must then perturb back.

Lemma 6.7. *Assume that A_ϵ is STP, and $A_\epsilon \to A$ as $\epsilon \downarrow 0$. Let $G(\mathbf{x}^\epsilon, \sigma^\epsilon, p) = \mathbf{0}$, where A_ϵ replaces A, and assume $S^-(\mathbf{x}^\epsilon) = n$, $r \leq n \leq R-1$, for all $\epsilon > 0$. Then there exists $(\mathbf{x}, \sigma, p)$ such that $G(\mathbf{x}, \sigma, p) = \mathbf{0}$ and $S^-(\mathbf{x}) = n$.*

The proof is a reworking of a part of the proof of Proposition 6.5. The details are left undone.

We have proven Theorem 4.14 for all $p \in [2, \infty)$. It remains to consider the case $p \in (1, 2)$. This case depends on a dual version of our problem. This is easily done if $r = 0$, and basically reduces to the same problem as that already considered except that A^T replaces A and q replaces p, where $1/p + 1/q = 1$. However, because of the conditions $G_k(\mathbf{x}, \sigma, p) = 0$, $k = 1, \ldots, r$, the dual version for $r > 0$ is more complicated. We prove the following directly, although it might well be obtained by means of Lagrange multipliers.

Proposition 6.8 *Assume that there exists $\mathbf{y} \in \mathbf{R}^M$ and numbers $\mu_1, \ldots, \mu_r$ such that for some $q \in (1, \infty)$,*

(6.9) $$(\mathbf{y}A)_i = 0, \; i = 1, \ldots, r$$

and

(6.10) $$\sum_{i=r+1}^{N} a_{ki} |(\mathbf{y}A)_i|^{q-1} \operatorname{sgn}((\mathbf{y}A)_i) + \sum_{i=1}^{r} a_{ki}\mu_i$$
$$= \lambda^q |y_k|^{q-1} \operatorname{sgn}(y_k), \qquad k = 1, \ldots, M,$$

where $S^-(\mathbf{y}) = n$, $r \leq n \leq R-1$. *Then for $p \in (1, \infty)$, $1/p + 1/q = 1$, there exists an $\mathbf{x} \in \mathbf{R}^N$ for which $S^-(\mathbf{x}) = n$ and $G(\mathbf{x}, \lambda^p, p) = \mathbf{0}$. The converse is also true.*

Proof. Set $x_i = \mu_i$, $i = 1, \ldots, r$, and $x_i = |(\mathbf{y}A)_i|^{q-1} \operatorname{sgn}((\mathbf{y}A)_i)$, $i = r+1, \ldots, N$. From (6.10), $\Sigma_{i=1}^N a_{ki}x_i = \lambda^q |y_k|^{q-1} \operatorname{sgn}(y_k)$, $k = 1, \ldots, M$, which implies that $|(A\mathbf{x})_i|^{p-1} \operatorname{sgn}((A\mathbf{x})_i) = \lambda^{q(p-1)}y_i = \lambda^p y_i$, $i = 1, \ldots, M$. Therefore,

$$\sum_{i=1}^{M} a_{ik}|(A\mathbf{x})_i|^{p-1}\operatorname{sgn}((A\mathbf{x})_i) = \lambda^p \sum_{i=1}^{M} a_{ik}y_i$$

$$= \lambda^p(\mathbf{y}A)_k,\ k = 1, \ldots, N.$$

From (6.9), $(\mathbf{y}A)_k = 0$, $k = 1, \ldots, r$. From the definition of x_k, $|x_k|^{p-1}\operatorname{sgn}(x_k) = (\mathbf{y}A)_k$, $k = r+1, \ldots, N$. Thus $G(\mathbf{x}, \lambda^p, p) = \mathbf{0}$.

It remains to prove that $S^-(\mathbf{x}) = n$. If $\lambda = 0$, then from Proposition 3.1, $S^-(\mu_1, \ldots, \mu_r, (\mathbf{y}A)_{r+1}, \ldots, (\mathbf{y}A)_N) \geq R$ unless the vector is identically zero. If the vector is identically zero, then $(\mathbf{y}A)_i = 0$, $i = r+1, \ldots, N$, together with (6.9) implies that $\mathbf{y}A = \mathbf{0}$. From Proposition 3.1, $S^-(\mathbf{y}) \geq R$, contradicting our assumption. If $\lambda = 0$, then

$$\begin{aligned} S^+(\mathbf{y}A) &= S^+(\overbrace{0, \ldots, 0}^{r}, (\mathbf{y}A)_{r+1}, \ldots, (\mathbf{y}A)_N) \\ &\geq S^+(\mu_1, \ldots, \mu_r, (\mathbf{y}A)_{r+1}, \ldots, (\mathbf{y}A)_N) \\ &\geq R. \end{aligned}$$

This, together with Proposition 3.1, again implies that $S^-(\mathbf{y}) \geq R$. Thus $\lambda > 0$.

Since $\operatorname{sgn}(A\mathbf{x})_i = \operatorname{sgn} y_i$, $i = 1, \ldots, M$, it follows that $S^-(A\mathbf{x}) = S^-(\mathbf{y}) = n$. A variant of Proposition 4.15 shows that if $S^-(A\mathbf{x}) = n$, then $S^-(\mathbf{x}) = S^+(\mathbf{x}) = S^-(A\mathbf{x}) = S^+(A\mathbf{x}) = r + S^-(x_{r+1}, \ldots, x_N) = r + S^+(x_{r+1}, \ldots, x_N)$. In particular, $S^-(\mathbf{x}) = n$. The converse is proven in a similar manner. ■

In the course of the above proof we also obtained:

Proposition 6.9. *Let* $\mathbf{y}\in\mathbf{R}^M$ *and* $\mu_1, \ldots, \mu_r$ *satisfy* (6.9) *and* (6.10) *with* $S^-(\mathbf{y}) = n$, $r \leq n \leq R-1$. *Then*

$$\begin{aligned} n &= S^-(\mathbf{y}) = S^+(\mathbf{y}) = S^-(\mu_1, \ldots, \mu_r, (\mathbf{y}A)_{r+1}, \ldots, (\mathbf{y}A)_N) \\ &= S^+(\mu_1, \ldots, \mu_r, (\mathbf{y}A)_{r+1}, \ldots, (\mathbf{y}A)_N) = S^+(\mathbf{y}A) \\ &= r + S^+((\mathbf{y}A)_{r+1}, \ldots, (\mathbf{y}A)_N) \\ &= r + S^-((\mathbf{y}A)_{r+1}, \ldots, (\mathbf{y}A)_N). \end{aligned}$$

Set $H(\mathbf{y}, \mu, \sigma, q) = (H_1(\mathbf{y}, \mu, \sigma, q), \ldots, H_{r+M+1}(\mathbf{y}, \mu, \sigma, q))\in\mathbf{R}^{r+M+1}$, where

$$H_k(\mathbf{y}, \mu, \sigma, q) = \sum_{i=r+1}^{N} a_{ki}|(\mathbf{y}A)_i|^{q-1}\operatorname{sgn}((\mathbf{y}A)_i)$$

$$+ \sum_{i=1}^{r} a_{ki}\mu_i - \sigma|y_k|^{q-1}\operatorname{sgn}(y_k), \qquad k = 1, \ldots, M,$$

$$H_{k+M}(\mathbf{y}, \mu, \sigma, q) = \sum_{i=1}^{M} y_i a_{ik}, \qquad k = 1, \ldots, r,$$

and

$$H_{r+M+1}(\mathbf{y}, \mu, \sigma, q) = \sum_{i=1}^{M} |y_i|^q - 1.$$

To prove Theorem 4.14 for $p \in (1, 2)$ we set $q \in (2, \infty)$, $1/p + 1/q = 1$, and prove for each such q the existence of $\mathbf{y}$, μ, and σ satisfying $H(\mathbf{y}, \mu, \sigma, q) = \mathbf{0}$ and $S^-(\mathbf{y}) = n$, $r \le n \le R-1$. From Propositions 6.1 and 6.8 we know this result to be true for $q = 2$. The remaining proof is similar to that used for $p \in (2, \infty)$. The analogues of Lemmas 6.2, 6.3, 6.6, and 6.7 and Proposition 6.5 are easily proven. The key to the proof is the analogue of Proposition 6.4, which we now prove.

Proposition 6.10. *Let* $(\mathbf{y}^\star, \mu^\star, \sigma^\star, q^\star)$, $q^\star \ge 2$, *satisfy* $H(\mathbf{y}^\star, \mu^\star\ \sigma^\star, q^\star) = \mathbf{0}$, $\sigma^\star > 0$. *In addition, assume that if* $q^\star = 2$, *then* $\mathbf{y}^\star$ *and* $\widehat{\mathbf{y}^\star A} = ((\mathbf{y}^\star A)_{r+1}, \ldots, (\mathbf{y}^\star A)_N)$ *have no zero entries. Then*

$$\det \left[\frac{\partial H_k}{\partial y_\ell}\right]_{k,\ell=1}^{r+M+1} \Bigg|_{(\mathbf{y}^\star,\, \mu^\star,\, \sigma^\star,\, q^\star)} \neq 0,$$

where for convenience we have set $y_{k+M} = \mu_k$, $k = 1, \ldots, r$, *and* $y_{r+M+1} = \sigma$.

Proof. Assume that the determinant is zero. There therefore exists a $\mathbf{z} = (z_1, \ldots, z_{r+M+1}) \in \mathbf{R}^{r+M+1}$ for which

$$\sum_{\ell=1}^{r+M+1} \frac{\partial H_k}{\partial y_\ell} z_\ell \Bigg|_{(\mathbf{y}^\star,\, \mu^\star,\, \sigma^\star,\, q^\star)} = 0,\ k = 1, \ldots, r+M+1.$$

This translates into

$$(6.11)\qquad (q^\star - 1)\left[\sum_{\ell=1}^{M} \sum_{i=r+1}^{N} a_{ki}\, a_{\ell i}\, |(\mathbf{y}^\star A)_i|^{q^\star-2} z_\ell - \sigma^\star |y_k^\star|^{q^\star-2} z_k\right]$$

$$+ \sum_{\ell=M+1}^{M+r} a_{k,\ell-M} z_\ell - |y_k^\star|^{q^\star-1} \operatorname{sgn}(y_k^\star) z_{r+M+1} = 0,\ k = 1, \ldots, M,$$

$$(6.12)\qquad \sum_{\ell=1}^{M} z_\ell a_{\ell k} = 0,\ k = 1, \ldots, r,$$

and

$$(6.13)\qquad q^\star \sum_{\ell=1}^{M} |y_\ell^\star|^{q^\star-1} \operatorname{sgn}(y_\ell^\star) z_\ell = 0.$$

We first claim that $z_{r+M+1} = 0$. To prove this fact, multiply Eq. (6.11) by $y_k^\star$ and sum over k to obtain

$$(q^\star - 1)\left[\sum_{\ell=1}^{M}\sum_{i=r+1}^{N} a_{\ell i}|(\mathbf{y}^\star A)_i|^{q^\star-1} \operatorname{sgn}((\mathbf{y}^\star A)_i)z_\ell - \sigma^\star \sum_{k=1}^{M} |y_k^\star|^{q^\star-1} \operatorname{sgn}(y_k^\star)z_k\right]$$

$$+ \sum_{\ell=M+1}^{M+r}\sum_{k=1}^{M} z_\ell a_{k,\ell-M}y_k^\star - \sum_{k=1}^{M} |y_k^\star|^{q}z_{r+M+1} = 0.$$

Repeated application of $H(\mathbf{y}^\star, \mu^\star, \sigma^\star, q^\star) = \mathbf{0}$ reduces this to

$$-(q^\star - 1)\sum_{\ell=1}^{M}\left[\sum_{i=1}^{r} a_{\ell i}\mu_i^\star\right]z_\ell - z_{r+M+1} = 0.$$

We now apply (6.12) to obtain $z_{r+M+1} = 0$.

For ease of notation, set $\mathbf{z} = (z_1, \ldots, z_M)$ and $\lambda = (\lambda_1, \ldots, \lambda_r)$, where $\lambda_i = z_{i+M}/(q^\star - 1)$, $i = 1, \ldots, r$. From (6.11)–(6.13) we have

$$(6.14)\quad \sum_{\ell=1}^{M}\sum_{i=r+1}^{N} a_{ki}a_{\ell i}|(\mathbf{y}^\star A)_i|^{q^\star-2}z_\ell - \sigma^\star|y_k^\star|^{q^\star-2}z_k + \sum_{i=1}^{r} a_{ki}\lambda_i = 0, k = 1, \ldots, M,$$

$$(6.15)\quad \sum_{\ell=1}^{M} z_\ell a_{\ell k} = 0, \qquad k = 1, \ldots, r,$$

and

$$(6.16)\quad \sum_{\ell=1}^{M} |y_\ell^\star|^{q^\star-1} \operatorname{sgn}(y_\ell^\star)z_\ell = 0.$$

For convenience we say that $(\mathbf{z}, \lambda)$ satisfies (6.14)–(6.16). Now $(\mathbf{y}^\star, \mu^\star)$ also satisfies (6.14) and (6.15), since $H(y^\star, \mu^\star, \sigma^\star, q^\star) = \mathbf{0}$. We prove that $(\mathbf{z}, \lambda) = \alpha(\mathbf{y}^\star, \mu^\star)$ for some $\alpha \in \mathbf{R}$. It then follows that $\alpha = 0$, proving the proposition. Note that if $\mathbf{z} = \mathbf{0}$, then, from (6.14), we have $\lambda = \mathbf{0}$. We therefore assume that $\mathbf{z} \neq \mathbf{0}$.

Case I. Assume that $(\mathbf{z}, \lambda)$ *satisfies* (6.14) *and* (6.15), *and* $y_k^\star z_k = 0$ *for* $k = 1, \ldots, M$.

From (6.14),

$$\sum_{i=r+1}^{N} a_{ki}\sum_{\ell=1}^{M} a_{\ell i}z_\ell|(\mathbf{y}^\star A)_i|^{q^\star-2} + \sum_{i=1}^{r} a_{ki}\lambda_i = 0, \qquad k = 1, \ldots, M.$$

From Proposition 3.1, either

i) $S^-(\lambda_1, \ldots, \lambda_r, \{\Sigma_{\ell=1}^{M} a_{\ell i}z_\ell|(\mathbf{y}^\star A)_i|^{q^\star-2}\}_{i=r+1}^{N}) \geq R$, or

ii) $\lambda_i = 0$, $i = 1, \ldots, r$, and $(\mathbf{z}A)_i((\mathbf{y}^\star A)_i)^{q^\star-2} = 0$, $i = r+1, \ldots, N$.

Assume (i) holds. If R ($=\min\{M, N\}$) $= N$, then (i) cannot hold. (The vector is simply not of sufficient length.) We therefore assume that $R = M < N$. Thus

$$S^-((\mathbf{z}A)_{r+1}|(\mathbf{y}^\star A)_{r+1}|^{q^\star-2}, \ldots, (\mathbf{z}A)_N|(\mathbf{y}^\star A)_{r+1}|^{q^\star-2}) \geq M-r,$$

which implies that $S^+((\mathbf{z}A)_{r+1}, \ldots, (\mathbf{z}A)_N) \geq M-r$. From (6.15), $(\mathbf{z}A)_i = 0$, $i = 1, \ldots, r$. Therefore, $S^+(\mathbf{z}A) \geq M$. From Proposition 3.1, if $\mathbf{z} \neq \mathbf{0}$, then $S^-(\mathbf{z}) \geq M$. Since $\mathbf{z} \in \mathbf{R}^M$, this is impossible.

Assume that (ii) holds. Thus $(\mathbf{z}A)_i|(\mathbf{y}^\star A)_i|^{q^\star-2} = 0$, $i = r+1, \ldots, N$. If $(\mathbf{y}^\star A)_i \neq 0$ then $(\mathbf{z}A)_i = 0$ for $i = r+1, \ldots, N$. In addition, $(\mathbf{z}A)_i = 0$, $i = 1, \ldots, r$, from (6.15). Therefore

$$\begin{aligned}\#\{i: (\mathbf{z}A)_i = 0\} &\geq r + \#\{i: (\mathbf{y}^\star A)_i \neq 0,\ i = r+1, \ldots, N\}\\ &\geq r + S^-((\mathbf{y}^\star A)_{r+1}, \ldots, (\mathbf{y}^\star A)_N) + 1\\ &= S^+(\mathbf{y}^\star) + 1\end{aligned}$$

from Proposition 6.9. Since $S^-(\mathbf{z}) \geq S^+(\mathbf{z}A) \geq \#\{i: (\mathbf{z}A)_i = 0\}$, it follows that $S^-(\mathbf{z}) \geq S^+(\mathbf{y}^\star) + 1$.

Now, $S^-(\mathbf{z}) \leq \#\{i: z_i \neq 0\} - 1$, and by hypothesis $z_i \neq 0$ implies $y_i^\star = 0$. Thus

$$S^-(\mathbf{z}) \leq \#\{i: y_i^\star = 0\} - 1 \leq S^+(\mathbf{y}^\star) - 1.$$

This, together with the previous inequality, leads to a contradiction.

Case II. Assume that $(\mathbf{z}, \lambda)$ *satisfies* (6.14)–(6.16) *and there exists a* $j \in \{1, \ldots, M\}$ *for which* $y_j^\star z_j \neq 0$.

For every $\alpha \in \mathbf{R}$,

$$S^+(\alpha\mathbf{y}^\star + \mathbf{z}) \leq S^+(\{\sigma^\star |\, y_k^\star|^{q^\star-2}(\alpha y_k^\star + z_k)\}_{k=1}^M).$$

From our hypothesis, (6.16), and $\sigma^\star > 0$, it follows that this latter vector is not identically zero unless $\mathbf{z} = \mathbf{0}$. We therefore have, from (6.14) and Proposition 3.1,

$$\begin{aligned}S^+(\alpha\mathbf{y}^\star + \mathbf{z}) &\leq S^+\left(\left\{\sum_{i=1}^r a_{ki}(\alpha\mu_i^\star + \lambda_i) + \sum_{i=r+1}^N a_{ki}((\alpha\mathbf{y}^\star + \mathbf{z})A)_i|(\mathbf{y}^\star A)_i|^{q^\star-2}\right\}_{k=1}^M\right)\\ &\leq S^-(\{\alpha\mu_i^\star + \lambda_i\}_{i=1}^r,\ \{((\alpha\mathbf{y}^\star + \mathbf{z})A)_i|(\mathbf{y}^\star A)_i|^{q^\star-2}\}_{i=r+1}^N)\\ &\leq S^-(\{\alpha\mu_i^\star + \lambda_i\}_{i=1}^r,\ \{((\alpha\mathbf{y}^\star + \mathbf{z})A)_i\}_{i=r+1}^N)\\ &\leq S^+(\overbrace{0, \ldots, 0}^{r}, ((\alpha\mathbf{y}^\star + \mathbf{z})A)_{r+1}, \ldots, ((\alpha\mathbf{y}^\star + \mathbf{z})A)_N).\end{aligned}$$

From (6.15), $((\alpha\mathbf{y}^\star + \mathbf{z})A)_i = 0$, $i = 1, \ldots, r$. Thus

$$\begin{aligned}S^+(\alpha\mathbf{y}^\star + \mathbf{z}) &\leq S^+((\alpha\mathbf{y}^\star + \mathbf{z})A)\\ &\leq S^-(\alpha\mathbf{y}^\star + \mathbf{z})\\ &\leq S^+(\alpha\mathbf{y}^\star + \mathbf{z}).\end{aligned}$$

We have come full circle and therefore $S^-(\alpha\mathbf{y}^\star + z) = S^+(\alpha\mathbf{y}^\star + z)$ for all $\alpha \in \mathbf{R}$. This cannot hold unless $\mathbf{z} = -\alpha\mathbf{y}^\star$ for some $\alpha \in \mathbf{R}$. But this implies, from (6.16), that $\mathbf{z} = \mathbf{0}$. ■

Acknowledgment. This work was supported in part by a grant from the U.S. Army through its European Research Office under contract no. DAJA 37-81-C-0234, and in part by the fund for the promotion of research at the Technion.

References

1. M. S. Birman, M. Z. Solomjak (1966): *Approximation of the functions of the class W_p^α by piecewise polynomial functions.* Dokl. Akad. Nauk SSSR, **171**: 1015–1018; see also Soviet Math. Dokl., 7(1966):1573-1577.
2. F. R. Gantmacher, M. G. Krein (1960): Oszillationsmatrizen, Oscillationskerne und kleine Schwingungen mechanischer Systeme. Berlin: Akademie-Verlag.
3. S. Karlin (1968): Total Positivity, Volume I. Stanford, California: Stanford University Press.
4. S. Karlin, W. J. Studden (1966): Tchebycheff Systems: With Applications in Analysis and Statistics. New York: Wiley Interscience.
5. A. Kolmogorov (1936): *Über die beste Annaherung von Funktionen einer gegebenen Funktionenklasse.* Ann. Math., **37**: 107–110.
6. M. G. Krein, M. A. Krasnosel'ski, D. P. Milman (1948): *On deficiency numbers of linear operators in Banach spaces and on some geometric problems.* Sb. Trudov Inst. Mat. Akad. Nauk SSSR, **11**: 97–112.
7. Yu. I. Makovoz (1972): *On a method for estimation from below of diameters of sets in Banach spaces.* Mat. Sb. (N.S.), **87**: 136–142; see also Math. USSR Sb., **16**(1972): 139–146.
8. A. A. Melkman, C. A. Micchelli. (1978): *Spline spaces are optimal for L^2 n-widths.* Illinois J. Math., **22**:541–564.
9. C. A. Micchelli, A. Pinkus (1977): *On n-widths in L^∞.* Trans. Amer. Math. Soc., **234**:139–174.
10. C. A. Micchelli, A. Pinkus (1978): *Some problems on the approximation of functions of two variables and n-widths of integral operators.* J. Approx. Theory, **24**:51–77.
11. C. A. Micchelli, T. J. Rivlin (1977): *A survey of optimal recovery.* In: C. A. Micchelli, T. J. Rivlin, (eds.) Optimal Estimation in Approximation Theory. New York: Plenum Press, pp. 1–54.
12. A. Pinkus (in press): *Some extremal problems for strictly totally positive matrices.* Lin. Alg. Appl.
13. A. Pinkus (in press): n-Widths in Approximation Theory. Berlin Heidelberg New York: Springer-Verlag.
14. U. Saates (1980): *Beste Approximation durch glatte Funktionen und Anwendungen in der intermediaren Approximation*, Dissertation, Universität Erlangen-Nürnberg.
15. V. M. Tihomirov (1969): *Best methods of approximation and interpolation of differentiable functions in the space C[−1, 1].* Mat. Sbornik, **80**:290–304; see also Math. USSR Sbornik, **9**(1969):275–289.
16. V. M. Tihomirov, S. B. Babadjanov (1967): *On the width of a functional class in the space L_p ($p \geq 1$).* Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk, **2**:24–30.
17. J. F. Traub, H. Wozniakowski (1980): A General Theory of Optimal Algorithms. New York: Academic Press.

A.P., Department of Mathematics, Technion—Israel Institute of Technology, Haifa 32000, Israel.