TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 234, Number 1, 1977

ON n-WIDTHS IN L^∞

BY

CHARLES A. MICCHELLI AND ALLAN PINKUS

ABSTRACT. The n-width in L^∞ of certain sets determined by matrices and integral operators is determined. The notion of total positivity is essential in the analysis.

1. **Introduction.** Let $X = (X, \|\cdot\|)$ be a normal linear space, $\mathcal{C}$ a subset of X and X_n any n-dimensional linear subspace of X. Then the n-width of $\mathcal{C}$ relative to X (in the sense of Kolmogorov) is defined to be

$$(1.1) \qquad d_n(\mathcal{C}; X) = \inf_{X_n} \sup_{x \in \mathcal{C}} \inf_{y \in X_n} \|x - y\|.$$

Whenever there is no ambiguity in the choice of X we will denote the n-width of $\mathcal{C}$ by $d_n(\mathcal{C})$. X_n is called an optimal subspace for $\mathcal{C}$ provided that

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

A typical choice for $\mathcal{C}$ is the image of the unit ball of some normed linear space $Y = (Y, \|\cdot\|)$ under a compact mapping A of Y into X,

$$(1.2) \qquad \mathcal{C} = \{Ay : \|y\| \leq 1, y \in Y\}.$$

When $Y = X$ and X is a Hilbert space it is possible to obtain an expression for $d_n(\mathcal{C}; X)$ and identify optimal subspaces (see §2). These facts originated with the example given by Kolmogorov in his paper [5] in which the concept of n-width of a set was introduced. Subsequently, a number of papers appeared treating n-widths in Hilbert spaces. However, there still does not exist a corresponding complete theory for n-widths of $\mathcal{C}$ when $\|\cdot\|$ is an L^∞-norm.

In this paper we present several results of a general nature concerning the n-width of $\mathcal{C}$ in the max-norm. Our main tool is the use of the notion of total positivity. Total positivity seems to have direct bearing on computing n-widths of $\mathcal{C}$ in L^∞. Very little is known to us when this hypothesis is not satisfied.

The paper is organized into eight sections. §2 is a preliminary section containing some useful results for computing n-widths. We include in this

Received by the editors June 30, 1975 and, in revised form, June 4, 1976.

AMS (MOS) subject classifications (1970). Primary 41A45, 41A35; Secondary 15A48, 45B05, 45L05.

© American Mathematical Society 1977

section a brief discussion of the n-width of $\mathcal{A}$ when A is an $N \times M$ matrix and $X = l_2^N$, Euclidean N-space, as well as results and terminology concerning totally positive matrices.

§3 contains some results on the n-width of $\mathcal{A}$ when A is an arbitrary $N \times M$ matrix and $X = l_\infty^N$. We show that when $\min(M, N) = n + 1$ and A has rank $n + 1$ then there exists n linearly independent column vectors of A which span an optimal subspace for $\mathcal{A}$. Also, for a given $N \times M$ matrix we introduce a real-valued function $G_n(\mathbf{i}, \mathbf{j})$ where $\mathbf{i} = (i_1, \ldots, i_{n+1})$, $\mathbf{j} = (j_1, \ldots, j_n)$ are vectors of integers with $1 \leqslant i_1 < \cdots < i_{n+1} \leqslant N$, $1 \leqslant j_1 < \cdots < j_n \leqslant M$ and show that $d_n(\mathcal{A})$ lies in the interval $[\mu_n(A), \lambda_n(A)]$, where

$$\mu_n(A) = \max_{\mathbf{i}} \min_{\mathbf{j}} G_n(\mathbf{i}, \mathbf{j}), \quad \lambda_n(A) = \min_{\mathbf{j}} \max_{\mathbf{i}} G_n(\mathbf{i}, \mathbf{j}).$$

In §4 we prove that $\mu_n(A) = \lambda_n(A) = d_n(\mathcal{A})$ when A is a strictly totally positive matrix and that some n column vectors of A form an optimal subspace for the n-width of $\mathcal{A}$. Furthermore, we find the Gel'fand n-width of $\mathcal{A}$ and the best approximation to A by rank n matrices. Next, in §5 we relax the hypothesis of §4 to include totally positive matrices, and in addition consider sets of the form $\mathcal{A} + X_r$ where X_r is some fixed linear subspace of dimension r.

§6 contains some examples of our previous results. In particular, we give an optimal procedure, based upon n-widths, for compressing a large table of numerical data when only information on the consecutive divided differences of some order is available.

In §7, we apply some standard estimates for approximating an integral by a Riemann sum and extend to integral operators our previous results on totally positive matrices.

Finally, in §8, we describe how, again by using our results on matrices, to compute the n-width of restricted moment spaces as well as the n-width of a subset of $L^\infty[0, 1]$ of the form

$$\mathfrak{M}_M = \left\{ \sum_{j=1}^{M} a_j u_j(t) \colon (a_1, \ldots, a_M) \in R^M,\ \max_{1 \leqslant j \leqslant M} |a_j| \leqslant 1, 0 \leqslant t \leqslant 1 \right\}.$$

Our results, both in the discrete and continuous form, were motivated by and are dependent upon the methods employed by V. M. Tihomirov [10]. Our approach to the continuous width problem (§7) has the advantage that, by means of a "discretization", we avoid the necessity of the nonlinear variational arguments given in [10]. This matter occupies a large portion of Tihomirov's paper.

2. Preliminaries. In this section we collect some facts about widths and totally positive matrices of which we will later make use.

Perhaps the most important theorem is computing widths is the following result of V. M. Tihomirov, cf. Singer [9, p. 277].

THEOREM 2.1. *Let X_{n+1} be an $n + 1$-dimensional subspace of X and suppose U_{n+1} is the closed unit ball of X_{n+1}. Then $d_n(U_{n+1}; X) = 1$.*

Theorem 2.1 is an effective tool in obtaining lower bounds for the n-width of a set $\mathcal{Q}$. The usual method of application is to find a ball of some $n + 1$-dimensional subspace of X which is contained in $\mathcal{Q}$. Then the radius of the ball is a lower bound for the n-width.

If $X \subseteq Y$ then clearly $d_n(\mathcal{Q}; Y) \leqslant d_n(\mathcal{Q}; X)$. Moreover, there are examples where strict inequality prevails. The depth of Theorem 2.1 lies in the fact that the n-dimensional width of $\mathcal{Q} = U_{n+1}$ does not decrease when U_{n+1} is embedded in a space of higher dimension. Although, in general this is not the case, it is nevertheless true that the *n-width* of any bounded set is an $n + 1$-*dimensional* space has this property, cf. Singer [9, p. 279].

THEOREM 2.2 (A. L. BROWN). *Let X be a normed linear space and X_{n+1} an $n + 1$-dimensional linear subspace of X. Then for every bounded set $\mathcal{Q} \subseteq X_{n+1}$ we have $d_n(\mathcal{Q}; X) = d_n(\mathcal{Q}; X_{n+1})$.*

Theorem 2.2 is an easy consequence of Theorem 2.1 and the following elementary but useful fact which is also due to A. L. Brown, cf. Singer [9, p. 276].

LEMMA 2.1. *Let $\mathcal{Q}$ be a convex, centrally symmetric, closed subset of an $n + 1$-dimensional normed linear space X_{n+1} then*

$$d_n(\mathcal{Q}; X_{n+1}) = \inf\{\|x\|: x \in \text{Boundary } \mathcal{Q}\}.$$

Furthermore, an n-dimensional subspace X_n is optimal for $\mathcal{Q}$ if and only if the hyperplane $x + X_n$ is a support for both $\mathcal{Q}$ and the ball of radius $d_n(\mathcal{Q}; X_{n+1})$ in X_{n+1}.

Let us give an application of Lemma 2.1. Suppose A is an $(n + 1) \times (n + 1)$ nonsingular matrix. Let $\|\cdot\|$ be *any* vector norm on R^{n+1} and let $\|A\|$ be the induced matrix norm of A. We define the set

$$\mathcal{Q}_0 = \{Ax: \|x\| \leqslant 1, x \in R^{n+1}\} \tag{2.1}$$

and denote the n-width of $\mathcal{Q}_0$ in R^{n+1} by $d_n(A)$.

LEMMA 2.2. $d_n(A) = \|A^{-1}\|^{-1}$.

PROOF. According to Lemma 2.1 we have

$$\begin{aligned} d_n(A) &= \inf\{\|y\| : y \in \text{Boundary } \mathcal{Q}\} \\ &= \inf\{\|y\| : \|A^{-1}y\| = 1\} \\ &= \left[\sup\{\|A^{-1}y\| : \|y\| = 1\}\right]^{-1} \\ &= \|A^{-1}\|^{-1}. \end{aligned}$$

This completes the proof.

In particular, when we choose $\|\cdot\|$ to be the l^2-norm on R^{n+1} we obtain $d_n(A) = \lambda_{n+1}^{1/2}$ where λ_{n+1} is the smallest eigenvalue of the matrix A^TA (A^T = transpose of A). This fact is a special case of the following well-known theorem on the n-width of $\mathcal{Q}_0$ when $\|x\|$ is the Euclidean norm of x. We include it here so that it may be compared to our results for the n-width of $\mathcal{Q}_0$ for the max-norm.

For an $N \times M$ matrix $A = \|a_{ij}\|$ we let $0 \leqslant \lambda_M \leqslant \lambda_{M-1} \leqslant \cdots \leqslant \lambda_1$ be the eigenvalues of the positive semidefinite matrix A^TA and $x^1, x^2, \ldots, x^M$ the corresponding set of orthonormal eigenvectors. (We will use superscripts to enumerate vectors while a subscript will be used to refer to a particular component of a vector, except when otherwise indicated.) If $r = \operatorname{rank} A$ then $0 = \lambda_M = \cdots = \lambda_{r+1} < \lambda_r \leqslant \cdots \leqslant \lambda_1$ and $d_n(A) = 0$ when $n \geqslant r$.

THEOREM 2.3. *Let A be any $N \times M$ matrix. Then the n-width of the set $\mathcal{Q}_0$, defined by* (2.1), *corresponding to the l^2-norm $\|x\|_2^2 = \sum_{j=1}^M |x_j|^2$, is given by*

$$d_n(A) = \begin{cases} \lambda_{n+1}^{1/2}, & n < r, \\ 0, & n \geqslant r. \end{cases}$$

Furthermore, when $n \leqslant r-1$, an optimal subspace for $\mathcal{Q}_0$ is spanned by the vectors Ax^i, $i = 1, 2, \ldots, n$.

PROOF. We assume $n < r$ and define the vectors $y^i = Ax^i$, $i = 1, 2, \ldots, r$, and denote by Y_n the subspace spanned by $y^1, \ldots, y^n$. Since $A^TAx^i = \lambda_i x^i$, $i = 1, \ldots, r$, it may be easily verified that $(y^i, y^j) = \lambda_i \delta_{ij}$, $i, j = 1, \ldots, r$. Let $S_{n+1} = \{y : y \in Y_{n+1}, \|y\|_2 \leqslant \lambda_{n+1}^{1/2}\}$. If $y = \sum_{j=1}^{n+1} c_j y^j$ and $x = \sum_{j=1}^{n+1} c_j x^j$, then

$$\left\|\sum_{j=1}^{n+1} c_j y^j\right\|_2^2 = \left\|A\left(\sum_{j=1}^{n+1} c_j x^j\right)\right\|_2^2 = \left(\sum_{j=1}^{n+1} c_j x^j, \sum_{j=1}^{n+1} \lambda_j c_j x^j\right)$$
$$= \sum_{j=1}^{n+1} \lambda_j c_j^2 \geqslant \lambda_{n+1}\|x\|_2^2.$$

Thus the ball S_{n+1} is contained in $\mathcal{Q}_0$ and we conclude from Theorem 2.1 that $d_n(A) \geqslant \lambda_{n+1}^{1/2}$ (actually this may be argued directly because there is clearly a $y \in \partial S_{n+1}$ such that $y \perp X_n$). The reverse inequality is proven in the following manner.

$$d_n^2(A) \leqslant \sup_{\|x\|_2 \leqslant 1} \inf_{y \in Y_n} \|Ax - y\|_2^2$$
$$= \sup\left\{ \sum_{j=n+1}^{M} \lambda_j |(x, x^j)|^2 : \|x\|_2 \leqslant 1 \right\}$$
$$= \lambda_{n+1}.$$

Thus we have proven Theorem 2.3.

Finally, we end this section by introducing some further notation and results concerning totally positive matrices.

We will use the following notation for the minors of A.

$$A\begin{pmatrix} i_1, \ldots, i_k \\ j_1, \ldots, j_k \end{pmatrix} = \begin{vmatrix} a_{i_1j_1} & \cdots & a_{i_1j_k} \\ \vdots & & \vdots \\ a_{i_kj_1} & \cdots & a_{i_kj_k} \end{vmatrix} = \det_{l,m}(a_{i_lj_m}).$$

We shall also use the notation

$$A\begin{pmatrix} i_1, \ldots, \hat{i}_s, \ldots, i_k \\ j_1, \ldots, j_{k-1} \end{pmatrix} = A\begin{pmatrix} i_1, \ldots, i_{s-1}, i_{s+1}, \ldots, i_k \\ j_1, \ldots, j_{k-1} \end{pmatrix}$$

and shall denote the jth column vector of A by a^j. Recall that superscripts are used to enumerate vectors while subscripts are used for a particular component of a vector.

DEFINITION 2.1. A is said to be sign consistent of order k (SC_k) if

$$A\begin{pmatrix} i_1, \ldots, i_k \\ j_1, \ldots, j_k \end{pmatrix}\sigma_k \geqslant 0 \tag{2.2}$$

for all $1 \leqslant i_1 < \cdots < i_k \leqslant N$, $1 \leqslant j_1 < \cdots < j_k \leqslant M$, and $\sigma_k^2 = 1$. A is said to be strict sign consistent of order k (SSC_k), if (2.2) holds with strict inequality.

DEFINITION 2.2. A is said to be totally positive of order $n+1$ (TP_{n+1}) if A is SC_k, $k = 1, \ldots, n+1$, and $\sigma_k = 1$, $k = 1, \ldots, n+1$. A is said to be strictly totally positive of order $n+1$ (STP_{n+1}) if A is SSC_k, $k = 1, \ldots, n+1$, and $\sigma_k = 1$, $k = 1, \ldots, n+1$.

DEFINITION 2.3. Let $x = (x_1, \ldots, x_l)$ be a real vector of l components

(i) $S^-(x)$ denotes the number of actual sign changes in the sequence $x_1, \ldots, x_l$, with zero terms discarded.

(ii) $S^+(x)$ counts the maximum number of sign changes in the sequence $x_1, \ldots, x_l$, where zero terms are assigned values $+1$ or -1, arbitrarily.

For example,

$$S^-(-1, 0, 1, -1, 0, -1) = 2, \quad S^+(-1, 0, 1, -1, 0, -1) = 4.$$

THEOREM 2.4. *If A is an $N \times M$ matrix which is STP_{n+1}, and if x is a nontrivial vector such that $S^-(x) \leq n$, then*

(i) $S^+(Ax) \leq S^-(x)$;

(ii) *If $S^+(Ax) = S^-(x)$, then the first (and last) component of Ax (if zero, then the sign given in determining $S^+(Ax)$) agrees in sign with the first (and last) nonzero component of x.*

REMARK 2.1. The above theorem is to be found in Karlin [2, p. 223] in a slightly different form. The proof of the above theorem is found in Karlin and Pinkus [3].

3. **The n-width of an aribtrary matrix.** Let A be any real $N \times M$ matrix, $A = \|a_{ij}\|$, and let

$$d_n(A; R^N) = \inf_{X_n} \sup_{\|x\|_\infty \leq 1} \inf_{y \in X_n} \|Ax - y\|_\infty,$$

where X_n denotes any n-dimensional linear space of N-vectors, $x \in R^M$ and $\|x\|_\infty = \max_i |x_i|$, be the n-width of the set $\mathcal{Q}_0 = \{Ax: \|x\|_\infty \leq 1, x \in R^M\}$. Whenever there is no ambiguity, we shall denote $d_n(A; R^N)$ by $d_n(A)$.

Let $I = \{\mathbf{i} = (i_1, \ldots, i_{n+1}): 1 \leq i_1 < \cdots < i_{n+1} \leq N\}$ and $J = \{\mathbf{j} = (j_1, \ldots, j_n): 1 \leq j_1 < \cdots < j_n \leq M\}$. Boldface $\mathbf{i}$ or $\mathbf{j}$ will always be used to denote vectors whose components are positive integers. Also, we shall denote by $\{a^j: j \in \mathbf{j}\}$ the set of N-vectors $\{a^{j_1}, \ldots, a^{j_n}\}$ and by $\sum_{j \in \mathbf{j}} \alpha_j a^j$ we mean $\sum_{k=1}^n \alpha_{j_k} a^{j_k}$, where $\mathbf{j} = (j_1, \ldots, j_n)$. Generally, when we write $j \in \mathbf{j}$ we will interpret the n-tuple $\mathbf{j}$ as the set $\{j_1, \ldots, j_n\}$ and mean that $j \in \{j_1, \ldots, j_n\}$. Otherwise $\mathbf{j}$ has its usual meaning as an ordered n-tuple.

We begin by defining two ancillary functions related to the n-width $d_n(A)$. Let

$$\lambda_n(A) = \inf_{Y_n} \sup_{\|x\|_\infty \leq 1} \inf_{y \in Y_n} \|Ax - y\|_\infty, \tag{3.1}$$

where Y_n denotes any n-dimensional linear space spanned by n *column vectors of A*. Thus, by definition, $d_n(A) \leq \lambda_n(A)$.

Let $A(\mathbf{i})$, where $\mathbf{i} = (i_1, \ldots, i_{n+1}) \in I$, be the $(n+1) \times M$ submatrix of A consisting of the $n+1$ rows $\{i_1, \ldots, i_{n+1}\}$ of A. Define

$$\mu_n(A) = \max_{\mathbf{i} \in I} \lambda_n(A(\mathbf{i})). \tag{3.2}$$

Thus $\mu_n(A) \leq \lambda_n(A)$.

If A has rank at most n, then $\mu_n(A) = d_n(A) = \lambda_n(A) = 0$, so we shall assume throughout this section that A has rank at least $n+1$. Define

$$H = \{\mathbf{j} \in J: \text{the set } \{a^j: j \in \mathbf{j}\} \text{ is linearly independent}\}$$

and for $\mathbf{i} \in I$,

$$H(\mathbf{i}) = \{\mathbf{j} \in J: \text{the set } \{a^j(\mathbf{i}): j \in \mathbf{j}\} \text{ is linearly independent}\}$$

where $a^j(\mathbf{i})$ is the restriction of the column vector a^j to the rows $\{i_1, \ldots, i_{n+1}\}$. Thus $a^j(\mathbf{i})$ is the $n+1$-vector whose components are $a_{i_1 j}, \ldots, a_{i_{n+1} j}$. Also, we define

$$G(\mathbf{i}, \mathbf{j}) = \frac{\sum_{j=1}^{M} \left| A\begin{pmatrix} i_1, \ldots, i_{n+1} \\ j_1, \ldots, j_n, j \end{pmatrix} \right|}{\sum_{s=1}^{n+1} \left| A\begin{pmatrix} i_1, \ldots, \hat{i}_s, \ldots, i_{n+1} \\ j_1, \ldots, j_n \end{pmatrix} \right|}, \qquad \mathbf{i} \in I, \mathbf{j} \in J,$$

with the convention that if the denominator of the above expression is zero, then $G(\mathbf{i}, \mathbf{j}) = 0$. (Note that if the denominator is zero, then so is the numerator.)

PROPOSITION 3.1. *If A has rank at least $n + 1$, then for $n \geqslant 1$,*

$$\lambda_n(A) = \min_{\mathbf{j} \in H} \max_{\mathbf{i} \in I} G(\mathbf{i}, \mathbf{j}), \tag{3.3}$$

$$\mu_n(A) = \max_{\mathbf{i} \in I} \min_{\mathbf{j} \in H(\mathbf{i})} G(\mathbf{i}, \mathbf{j}). \tag{3.4}$$

For $n = 0$, $\mu_0(A) = d_0(A) = \lambda_0(A) = \max_i \sum_{j=1}^{M} |a_{ij}|$.

PROOF. The case $n = 0$ follows by definition. For $n \geqslant 1$, the characterization (3.4) of $\mu_n(A)$ is a result of (3.3) and the definition (3.2) of $\mu_n(A)$. Thus, it remains to prove (3.3). First note that from the definition of $\lambda_n(A)$, (3.1), we need only consider n linearly independent columns of A.

Fix $\mathbf{j} \in H$ and let x be any N-dimensional vector. It is well known, Singer [9, p. 179], that the distance between x and the subspace spanned by $\{a^j: j \in \mathbf{j}\}$ is given by

$$\inf_{\alpha_j} \left\| x - \sum_{j \in \mathbf{j}} \alpha_j a^j \right\|_\infty = \max_{\mathbf{i} \in I} \frac{\left\| \begin{matrix} a_{i_1 j_1} & \cdots & a_{i_1 j_n} & x_{i_1} \\ \vdots & & \vdots & \vdots \\ a_{i_{n+1} j_1} & \cdots & a_{i_{n+1} j_n} & x_{i_{n+1}} \end{matrix} \right\|}{\sum_{s=1}^{n+1} \left| A\begin{pmatrix} i_1, \ldots, \hat{i}_s, \ldots, i_{n+1} \\ j_1, \ldots, j_n \end{pmatrix} \right|}$$

(double vertical bars means that we are taking the absolute value of the determinant). Furthermore, if the above maximum is achieved at $\mathbf{i} = (i_1, \ldots, i_{n+1})$ and the columns $\{a^j: j \in \mathbf{j}\}$ form a Haar system (that is, the matrix $\|a_{ij}\|$, $i = 1, 2, \ldots, N$, $j = j_1, \ldots, j_n$ is SC_n), then the minimum error equioscillates on $\{i_1, \ldots, i_{n+1}\}$ (cf. Singer [9, p. 183] and Definition 4.1 for the meaning of the phrase "equioscillates on $\{i_1, \ldots, i_{n+1}\}$").

Thus, by a multilinear expansion of the last column of the numerator,

$$\text{(3.5)} \qquad \inf_{\alpha_j} \left\| Ax - \sum_{j \in \mathbf{j}} \alpha_j a^j \right\|_\infty = \max_{\mathbf{i} \in I} \frac{\left| \sum_{j=1}^{M} x_j A\begin{pmatrix} i_1, \ldots, i_{n+1} \\ j_1, \ldots, j_n, j \end{pmatrix} \right|}{\sum_{s=1}^{n+1} \left| A\begin{pmatrix} i_1, \ldots, \hat{i}_s, \ldots, i_{n+1} \\ j_1, \ldots, j_n \end{pmatrix} \right|}.$$

For any choice of $\mathbf{i} \in I$, the supremum of (3.5) over $\|x\|_\infty \leqslant 1$ is obviously achieved by choosing

$$\text{(3.6)} \qquad x_j = \operatorname{sgn} A\begin{pmatrix} i_1, \ldots, i_{n+1} \\ j_1, \ldots, j_n, j \end{pmatrix}\sigma, \qquad j \notin \mathbf{j},$$

$\sigma^2 = 1$, where $i_1, \ldots, i_{n+1}$ are the rows which yield the maximum of the expression in (3.5). Therefore,

$$\text{(3.7)} \qquad \sup_{\|x\|_\infty \leqslant 1} \inf_{\alpha_j} \left\| Ax - \sum_{j \in \mathbf{j}} \alpha_j a^j \right\|_\infty = \max_{\mathbf{i} \in I} G(\mathbf{i}, \mathbf{j}).$$

The identity (3.3) now follows from (3.7) and the definition (3.1) of $\lambda_n(A)$. The proof is completed.

Proposition 3.1 leads us to

PROPOSITION 3.2. *If A is an $(n + 1) \times M$ real matrix of rank $n + 1$, then* $\mu_n(A) = \lambda_n(A) = d_n(A)$.

PROOF. Since A has $n + 1$ rows we may conclude from our definitions that $\mu_n(A) = \lambda_n(A) \geqslant d_n(A)$. But $\mathcal{Q}_0 = \{Ax\colon \|x\|_\infty \leqslant 1\} \subseteq R^{n+1}$ and so the hypothesis of Lemma 2.1 is satisfied. Hence $d_n(A) = \inf\{\|Ax\|_\infty\colon Ax \in \partial \mathcal{Q}_0\}$ where $\partial \mathcal{Q}_0$ denotes the boundary of $\mathcal{Q}_0$.

Our problem is thus one of determining $\partial \mathcal{Q}_0$. To this end let us assume without loss of generality, that each set of $n + 1$ columns of A is linearly independent. For each $\mathbf{j} \in J$, let

$$x_j = \operatorname{sgn} A\begin{pmatrix} 1, \ldots, n+1 \\ j_1, \ldots, j_n, j \end{pmatrix}, \qquad j \notin \mathbf{j},$$

and consider the n-dimensional hyperplanes

$$\text{(3.8)} \qquad \sum_{j \in \mathbf{j}} \alpha_j a^j \pm \left(\sum_{j \notin \mathbf{j}} x_j a^j \right), \qquad (\alpha_{j_1}, \ldots, \alpha_{j_n}) \in R^n.$$

Then,

LEMMA 3.1. *Each hyperplane of the form* (3.8) *contains an n-dimensional face of $\mathcal{Q}_0$, and each n-dimensional face of $\mathcal{Q}_0$ is contained in a hyperplane of the form* (3.8) *for some* $\mathbf{j} \in J$.

REMARK 3.1. According to Lemma 3.1 the faces of $\mathcal{Q}_0$ are the restrictions of the hyperplanes (3.8) to the cube $|\alpha_j| \leq 1, j \in \mathbf{j}$.

PROOF. Let $\mathbf{j} \in J$ be fixed and choose a nonzero $n+1$-dimensional vector y such that $(y, a^j) = 0, j \in \mathbf{j}$. Since $y \neq 0$, and by assumption each set of $n+1$ columns of A is linearly independent, $(y, a^j) \neq 0$, $j \notin \mathbf{j}$. Now, $\max\{(y, Ax): \|x\|_\infty \leq 1\}$ is achieved by vectors on the boundary of $\mathcal{Q}$ and

$$(y, Ax) = \sum_{j \notin \mathbf{j}} (y, a^j) x_j \leq \sum_{j \notin \mathbf{j}} |(y, a^j)|. \tag{3.9}$$

The maximum in (3.9) is achieved by taking $x_j = \operatorname{sgn}(y, a^j), j \notin \mathbf{j}$, while the components $x_j, j \in \mathbf{j}$, may be chosen arbitrarily, under the restriction $|x_j| \leq 1$, $j \in \mathbf{j}$. Thus, the vectors x which maximize (y, Ax) form an n-dimensional face of $\mathcal{Q}_0$, and any n-dimensional face of $\mathcal{Q}_0$ may be determined in this manner. Moreover, since $(y, a^j) = 0, j \in \mathbf{j}$, y is determined up to a nonzero constant. In fact, it is easily seen that

$$y_i = dA\begin{pmatrix} 1, \ldots, \hat{i}, \ldots, n+1 \\ j_1, \ldots, j_n \end{pmatrix}(-1)^i$$

where d is some nonzero constant. For $j \notin \mathbf{j}$,

$$(y, a^j) = \sum_{i=1}^{n+1} y_i a_{ij} = d \sum_{i=1}^{n+1} A\begin{pmatrix} 1, \ldots, \hat{i}, \ldots, n+1 \\ j_1, \ldots, j_n \end{pmatrix}(-1)^i a_{ij}$$
$$= d(-1)^{n+1} A\begin{pmatrix} 1, \ldots, n+1 \\ j_1, \ldots, j_n, j \end{pmatrix}.$$

Thus

$$x_j = \operatorname{sgn} A\begin{pmatrix} 1, \ldots, n+1 \\ j_1, \ldots, j_n, j \end{pmatrix}\sigma, \qquad j \notin \mathbf{j}, \sigma^2 = 1.$$

This proves our lemma.

REMARK 3.2. A similar argument shows that each of the points $\sum_{j \in \mathbf{j}} \alpha_j a^j \pm \sum_{j \notin \mathbf{j}} x_j a^j$ where $|\alpha_j| = 1$, for all $j \in \mathbf{j}$ is an extreme point of $\mathcal{Q}_0$.

Now, returning to the proof of Proposition 3.2 we see from Lemma 3.1 that

$$d_n(A) = \inf\{\|Ax\|_\infty : Ax \in \partial \mathcal{Q}_0\}$$
$$= \inf\left\{\left\| \sum_{j \in \mathbf{j}} \alpha_j a^j \pm \left(\sum_{j \notin \mathbf{j}} x_j a^j\right)\right\|_\infty : |\alpha_j| \leq 1, j \in \mathbf{j}, \mathbf{j} \in J\right\}$$

where x_j is defined by (3.6). Thus, according to Proposition 3.1,

$$\lambda_n(A) = \inf\left\{\left\| \sum_{j \in \mathbf{j}} \alpha_j a^j \pm \left(\sum_{j \notin \mathbf{j}} x_j a^j\right)\right\|_\infty : \alpha_j \in R, j \in \mathbf{j}, \mathbf{j} \in J\right\} \leq d_n(A).$$

We have already observed, by the definition of $\lambda_n(A)$, that $\lambda_n(A) \geqslant d_n(A)$ and therefore $\mu_n(A) = d_n(A) = \lambda_n(A)$. Thus the proof of Proposition 3.2 is complete.

Another application of Lemma 2.1 gives us the following result.

PROPOSITION 3.3. *If A is an $N \times (n+1)$ real matrix of rank $n+1$ then* $d_n(A) = \lambda_n(A)$.

PROOF. Since A is an $N \times (n+1)$ real matrix of rank $n+1$, the boundary of $\mathcal{Q}_0$ is characterized by $\partial \mathcal{Q}_0 = \{Ax: \|x\|_\infty = 1\}$ and $\mathcal{Q}_0$ is a subset of an $n+1$-dimensional subspace X_{n+1} spanned by the column vectors of A. From Theorem 2.2 and Lemma 2.1,

$$d_n(A) = d_n(A; X_{n+1}) = \inf\{\|Ax\|_\infty: \|x\|_\infty = 1\}.$$

Let $d_n(A) = \|Ax^0\|_\infty$, $\|x^0\|_\infty = 1$, then there exists an i, $1 \leqslant i \leqslant n+1$, such that $|x_i| = 1$. Furthermore, from Lemma 2.1 we see that the n-dimensional subspace spanned by the columns $\{a^1, \ldots, a^{i-1}, a^{i+1}, \ldots, a^{n+1}\}$ is an optimal subspace and

$$d_n(A) = \inf_{|\alpha_j| \leqslant 1} \left\| a^i - \sum_{j \neq i} \alpha_j a^j \right\|_\infty .$$

By definition,

$$\lambda_n(A) = \inf_{1 \leqslant i \leqslant n+1} \inf_{\alpha_j} \left\| a^i - \sum_{j \neq i} \alpha_j a^j \right\|_\infty .$$

Thus $\lambda_n(A) \leqslant d_n(A)$. However we also know that $d_n(A) \leqslant \lambda_n(A)$ by the definition (3.1) of $\lambda_n(A)$. Thus Proposition 3.3 is proven.

On the basis of Propositions 3.2 and 3.3 it seems plausible to conjecture that $\mu_n(A) = d_n(A) = \lambda_n(A)$ for any matrix A. In the next section, we will prove that this is the case for the class of totally positive matrices. However, in general this is not true as the following examples demonstrate.

EXAMPLE 1. When

$$A = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix},$$

a simple computation shows that $\mu_1(A) = 1$, while $\lambda_1(A) = d_1(A) = 2$.

EXAMPLE 2. If

$$B = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}$$

then $\mu_1(B) = d_1(B) = 1$ and $\lambda_1(B) = 2$. (Note that $B = 2A^{-1}$.)

In general, we have the following result.

THEOREM 3.1. *If A is any $N \times M$ real matrix then*

$$\mu_n(A) \leqslant d_n(A) \leqslant \lambda_n(A). \tag{3.10}$$

PROOF. We have already observed that if A has rank less than or equal to n, the above inequalities are equalities. Hence we assume A has rank at least $n+1$. The right-hand inequality in (3.10) follows from the definitions of $\lambda_n(A)$ and $d_n(A)$. To prove that $\mu_n(A) \leqslant d_n(A)$, we appeal to Proposition 3.2 which tells us that

$$\max_{\mathbf{i}\in I} d_n(A(\mathbf{i})) = \max_{\mathbf{i}\in I} \lambda_n(A(\mathbf{i})) = \mu_n(A).$$

The last equation follows from the definition (3.2) of $\mu_n(A)$. Moreover, it is easily seen, by definition, that $d_n(A(\mathbf{i})) \leqslant d_n(A)$ for all choices of $\mathbf{i} \in I$. Thus $\mu_n(A) \leqslant d_n(A)$ and the proof is complete.

Before we turn to the proof that $\mu_n(A) = d_n(A) = \lambda_n(A)$ for strictly totally positive matrices, we note the following corollary of Proposition 3.1.

PROPOSITION 3.4. *If A is an $N \times N$ nonsingular matrix, then $\lambda_n(A)\mu_{N-n-1}(A^{-1}) = 1$, $n = 0, 1, \ldots, N-1$.*

PROOF. First, let us recall that

$$A\begin{pmatrix} l_1, \ldots, l_k \\ m_1, \ldots, m_k \end{pmatrix} = \frac{A^{-1}\begin{pmatrix} m'_1, \ldots, m'_{N-k} \\ l'_1, \ldots, l'_{N-k} \end{pmatrix}(-1)^{\sum_{i=1}^k (l_i+m_i)}}{\det A^{-1}}, \tag{3.11}$$

where $l_1 < \cdots < l_k$ and $l'_1 < \cdots < l'_{N-k}$ are complementary sets of indices in $\{1, \ldots, N\}$ as are $m_1 < \cdots < m_k$ and $m'_1 < \cdots < m'_{N-k}$. We shall assume, without loss of generality, that all $n \times n$ and $(n+1) \times (n+1)$ minors of A are nonzero. Thus for $1 \leqslant n \leqslant N-2$,

$$\lambda_n(A) = \min_{1\leqslant j_1<\cdots<j_n\leqslant N}\ \max_{1\leqslant i_1<\cdots<i_{n+1}\leqslant N} \frac{\sum_{j=1}^{N}\left|A\begin{pmatrix} i_1, \ldots, i_{n+1} \\ j_1, \ldots, j_n, j \end{pmatrix}\right|}{\sum_{s=1}^{n+1}\left|A\begin{pmatrix} i_1, \ldots, \hat{i}_s, \ldots, i_{n+1} \\ j_1, \ldots, j_n \end{pmatrix}\right|}$$

$$= \min_{1\leqslant j_1<\cdots<j_n\leqslant N}\ \max_{1\leqslant i_1<\cdots<i_{n+1}\leqslant N} \frac{\sum_{s=1}^{N-n}\left|A^{-1}\begin{pmatrix} j'_1, \ldots, \hat{j}'_s, \ldots, j'_{N-n} \\ i'_1, \ldots, i'_{N-n-1} \end{pmatrix}\right|}{\sum_{s=1}^{n+1}\left|A^{-1}\begin{pmatrix} j'_1, \ldots, j'_{N-n} \\ i'_1, \ldots, i'_{N-n-1}, i_s \end{pmatrix}\right|}$$

$$= \min_{1\leqslant j_1<\cdots<j_n\leqslant N}\ \max_{1\leqslant i_1<\cdots<i_{n+1}\leqslant N} \left[\frac{\sum_{j=1}^{N}\left|A^{-1}\begin{pmatrix} j'_1, \ldots, j'_{N-n} \\ i'_1, \ldots, i'_{N-n-1}, j \end{pmatrix}\right|}{\sum_{s=1}^{N-n}\left|A^{-1}\begin{pmatrix} j'_1, \ldots, \hat{j}'_s, \ldots, j'_{N-n} \\ i'_1, \ldots, i'_{N-n-1} \end{pmatrix}\right|}\right]^{-1}$$

where $\{j'_k\}_{k=1}^{N-n}$ is the set of complementary (ordered) indices to $\{j_k\}_{k=1}^{n}$, and $\{i'_k\}_{k=1}^{N-n-1}$ is the set of complementary (ordered) indices to $\{i_k\}_{k=1}^{n+1}$ in $\{1, \ldots, N\}$. From (3.11), it follows that

$$\lambda_n(A) = \min_{1 \le j'_1 < \cdots < j'_{N-n} \le N} \; \max_{1 \le i'_1 < \cdots < i'_{N-n-1} \le N}$$

$$\cdot \left[\frac{\sum_{j=1}^{N} \left| A^{-1}\begin{pmatrix} j'_1, \ldots, j'_{N-n} \\ i'_1, \ldots, i'_{N-n-1}, j \end{pmatrix} \right|}{\sum_{s=1}^{N-n} \left| A^{-1}\begin{pmatrix} j'_1, \ldots, \hat{j}'_s, \ldots, j'_{N-n} \\ i'_i, \ldots, i'_{N-n-1} \end{pmatrix} \right|} \right]^{-1}$$

$$= \left(\mu_{N-n-1}(A^{-1})\right)^{-1}.$$

The cases $n = 0, N-1$ are even simpler and may be easily proven as above. Alternatively, $\mu_0(A) = \lambda_0(A) = d_0(A)$, by definition, and $\lambda_{N-1}(A) = \mu_{N-1}(A) = d_{N-1}(A)$ by Proposition 3.2 while from Lemma 2.2 $d_0(A)d_{N-1}(A^{-1}) = 1$.

4. The n-width of totally positive matrices. In this section we shall assume that $A = \|a_{ij}\|$ is an $N \times M$ STP_{n+1} matrix.

DEFINITION 4.1. Given $0 = j_0 < j_1 < \cdots < j_n < j_{n+1} = M + 1$, and a vector $x \in R^M$, we will say that x alternates between $j_1, \ldots, j_n$ provided that there exists a sign σ, $\sigma^2 = 1$, such that $x_l = (-1)^i\sigma$, $j_{i-1} < l < j_i$, $i = 1, 2, \ldots, n+1$. (Note that if x alternates between $j_1, \ldots, j_n$ no requirement is placed on the components $x_{j_1}, \ldots, x_{j_n}$.)

Also, we will use the terminology that a vector $y \in R^N$ equioscillates on $i_1, \ldots, i_{n+1}$, $1 \le i_1 < \cdots < i_{n+1} \le N$, if there exists a sign σ, $\sigma^2 = 1$, such that $y_{i_l} = \sigma(-1)^l\|y\|_\infty$, $l = 1, \ldots, n+1$.

Let us introduce some further notation. Recall that, according to Proposition 3.1,

$$\lambda_n(A) = \min_{\mathbf{j} \in J} \max_{\mathbf{i} \in I} G(\mathbf{i}, \mathbf{j}). \tag{4.1}$$

For any $\mathbf{j} \in J$, we define

$$\begin{aligned} \sup_{\|x\|_\infty \le 1} \inf_{\alpha_j} \left\| Ax - \sum_{j \in \mathbf{j}} \alpha_j a^j \right\|_\infty &= \inf_{\alpha_j} \left\| Ax^*_{\mathbf{j}} - \sum_{j \in \mathbf{j}} \alpha_j a^j \right\|_\infty \\ &= \left\| Ax^*_{\mathbf{j}} - \sum_{j \in \mathbf{j}} \alpha^*_j a^j \right\|_\infty \end{aligned} \tag{4.2}$$

and

$$(4.3)\qquad (x_{\mathbf{j}})_l = \begin{cases} (x_{\mathbf{j}}^*)_l - \alpha_l^*, & \text{if } l = j_k, k = 1, 2, \ldots, n, \\ (x_{\mathbf{j}}^*)_l, & l \notin \mathbf{j}. \end{cases}$$

Let $\mathbf{i}^0 = (i_1^0, \ldots, i_{n+1}^0)$, $\mathbf{j}^0 = (j_1^0, \ldots, j_n^0)$ denote the rows and columns, respectively, for which (4.1) is attained and, in addition, denote the vector $x_{\mathbf{j}^0}$ by x^0.

Since A is STP_{n+1}, the proof in Proposition 3.1 (see (3.6) and the preceding remarks) shows that

(4.4)
(i) $x_{\mathbf{j}}$ alternates between $j_1, \ldots, j_n$, and
(ii) $Ax_{\mathbf{j}}$ equioscillates on some $n + 1$ components.

In particular, Ax^0 equioscillates on $i_1^0, \ldots, i_{n+1}^0$. Finally, according to our definitions,

$$(4.5)\qquad \|Ax^0\|_\infty \leq \|Ax_{\mathbf{j}}\|_\infty, \quad \text{for all } \mathbf{j} \in J.$$

THEOREM 4.1. *Let A be a strictly totally positive matrix of order $n + 1$ and $0 < n < \min(N, M)$. Then*

(i) $\mu_n(A) = d_n(A) = \lambda_n(A)$.

Furthermore,

(ii) *The linear space X_n^0 spanned by the columns $\{a^j: j \in \mathbf{j}^0\}$ is an optimal subspace for the n-width $d_n(A)$.*

(iii) *The vector x^0 defined above alternates between $j_1^0, \ldots, j_n^0$, Ax^0 equioscillates on $i_1^0, \ldots, i_{n+1}^0$ and $\|x^0\|_\infty = 1$.*

Since the second assertion of Theorem 4.1 follows from (i), and the first two statements of (iii) are a result of (4.4), it remains for us to prove (i) and $\|x^0\|_\infty = 1$. We will prove these facts in a series of lemmas.

We begin with

LEMMA 4.1. $\|x^0\|_\infty = 1$.

PROOF. For any $\mathbf{j} \in J$, the definition of $x_{\mathbf{j}}$ implies $\|x_{\mathbf{j}}\|_\infty \geq 1$, and $S^-(x_{\mathbf{j}}) \leq n$ by (4.4). However, from (4.4), and by the definition of equioscillation we have $S^+(Ax_{\mathbf{j}}) \geq n$. Therefore, applying Theorem 2.4, we obtain $S^+(Ax_{\mathbf{j}}) = S^-(x_{\mathbf{j}}) = n$ for all $\mathbf{j} \in J$. Theorem 2.4, (ii) tells us that the sign patterns of the vectors $Ax_{\mathbf{j}}$ and $x_{\mathbf{j}}$ agree. Multiplying $x_{\mathbf{j}}$ by -1, if necessary, we assume that the sign pattern of $x_{\mathbf{j}}$ begins with a plus. Thus $\operatorname{sgn}(x_{\mathbf{j}})_l = (-1)^{i+1}, j_{i-1} < l < j_i, i = 1, \ldots, n+1, j_0 = 0, j_{n+1} = M + 1$.

Our object is to show that $|(x^0)_{j_k^0}| \leq 1, k = 1, \ldots, n$.

First, let us suppose that $j_k^0 > k$. Let l be the largest integer less than j_k^0 such that $l \neq j_i^0$, $i = 1, 2, \ldots, k-1$. Define $\mathbf{j} = (j_1, \ldots, j_n)$ where $\{j_1, \ldots, j_n\}$ is the set of indices $\{j_1^0, \ldots, j_{k-1}^0, j_{k+1}^0, \ldots, j_n^0, l\}$ rearranged in increasing order. Due to our convention of arranging for the sign of the first component of x_j to be positive we have $(x_j)_{j_k^0} = (-1)^k$. Now, consider the vector $Ax_j - Ax^0 = A(x_j - x^0)$. If $x_j - x^0 = 0$ then $|(x^0)_{j_k^0}| = 1$. If $x_j - x^0 \neq 0$ then from (4.5) and the fact that Ax_j equioscillates on some $n+1$ rows, $i_1, \ldots, i_{n+1}$, $S^+(Ax_j - Ax^0) \geq n$. Since $x_j - x^0$ has, by construction, at most $n+1$ nonzero components, that is, the components corresponding to the columns $j_1^0, \ldots, j_n^0, l$ we conclude that $S^-(x_j - x^0) \leq n$. From Theorem 2.4, $S^+(A(x_j - x^0)) = S^-(x_j - x^0) = n$ and the sign patterns must agree.

Since A is STP_{n+1}, $A(x_j - x^0)$ cannot have $n+1$ zero components. Because the sign pattern in $S^+(Ax_j)$ begins with a plus, and $\|Ax^0\|_\infty \leq \|Ax_j\|_\infty$ it follows that the sign pattern in $S^+(A(x_j - x^0))$ begins with a plus. Applying Theorem 2.4 (ii), we see that $\operatorname{sgn}((x_j)_{j_k^0} - (x^0)_{j_k^0}) = (-1)^k$. Since $(x_j)_{j_k^0} = (-1)^k$ we conclude that $1 > (x^0)_{j_k^0}(-1)^k$. Similarly, if $j_k^0 < M - n + k$, we let l be the smallest integer greater than j_k^0 such that $l \neq j_i^0$, $i = k+1, \ldots, n$, then as before, we may show that $1 > (x^0)_{j_k^0}(-1)^{k+1}$. Hence, if both $j_k^0 > k$ and $j_k^0 < M - n + k$ we obtain the desired conclusion that $|(x^0)_{j_k^0}| \leq 1$.

In the case that $j_k^0 = k$ we have $j_i^0 = i$, $i = 1, \ldots, k-1$ and thus $\operatorname{sgn}(x^0)_{j_k^0} = (-1)^{k+1}$. However, $n < M$, which implies $j_k^0 < M - n + k$. Thus by our above remarks $|(x^0)_{j_k^0}| = (-1)^{k+1}(x^0)_{j_k^0} \leq 1$. Similarly, if $j_k^0 = M - n + k$ then $j_k^0 > k$ and $|(x^0)_{j_k^0}| = (-1)^k(x^0)_{j_k^0} \leq 1$. Thus in all cases we arrive at the desired conclusion.

LEMMA 4.2. $d_n(A) = \lambda_n(A)$.

PROOF. Since $S^-(x^0) = n$ and due to the orientation given x^0, there exists $\{l_1, \ldots, l_n\}$, $0 = l_0 < l_1 < \cdots < l_n < l_{n+1} = M$, such that if $l_{k-1} < i \leq l_k$ then $\operatorname{sgn}(x^0)_i = (-1)^{k-1}$ or zero, $k = 1, \ldots, n+1$, and for each $k = 1, \ldots, n+1$ there exists at least one i, $l_{k-1} < i \leq l_k$, such that $\operatorname{sgn}(x^0)_i = (-1)^{k-1}$.

Let $b^k = \sum_{i=l_{k-1}+1}^{l_k} |(x^0)_i| a^i$, $k = 1, \ldots, n+1$, and denote by B the $N \times (n+1)$ matrix composed of the columns $b^1, \ldots, b^{n+1}$. Then B is STP_{n+1}. Set

$$U_{n+1} = \left\{ \sum_{k=1}^{n+1} \alpha_k b^k : \left\| \sum_{k=1}^{n+1} \alpha_k b^k \right\|_\infty \leq \lambda_n(A) \right\}.$$

We shall show that $U_{n+1} \subseteq \mathcal{C}_0$, where as before $\mathcal{C}_0 = \{Ax : \|x\|_\infty \leq 1\}$. This inclusion in turn implies, by Theorem 2.1, that $\lambda_n(A) \leq d_n(A)$, proving the lemma. By the construction of the vectors $b^1, \ldots, b^{n+1}$ and Lemma 4.1 it is sufficient to prove that $\sum_{k=1}^{n+1} \alpha_k b^k \in U_{n+1}$ implies $|\alpha_k| \leq 1$, $k = 1, \ldots, n+1$.

Assume to the contrary that there exists $\sum_{k=1}^{n+1} \alpha_k b^k \in U_{n+1}$ for which $\max_k |\alpha_k| = c > 1$. Now, $Ax^0 = \sum_{k=1}^{n+1} (-1)^{k-1} b^k$ and therefore there exists an

i, $1 \leqslant i \leqslant n+1$, such that $\alpha_i/c = (-1)^{i-1}\sigma$, where $\sigma^2 = 1$. Since $\lambda_n(A) = \|Ax^0\|_\infty$, Ax^0 equioscillates $n+1$ times on $i_1^0, \ldots, i_{n+1}^0$, and $c^{-1}\|\Sigma_{k=1}^{n+1}\alpha_k b^k\|_\infty < \lambda_n(A)$, we conclude that

$$S^-\left(\sigma \sum_{k=1}^{n+1} (-1)^{k-1} b^k - \sum_{k=1}^{n+1} \frac{\alpha_k}{c} b^k\right) \geqslant n.$$

Moreover, $S^-(\{\sigma(-1)^{k-1} - \alpha_k/c\}) \leqslant n-1$, since the ith component of the vector vanishes. A contradiction now follows from Theorem 2.4. Thus $U_{n+1} \subseteq \mathcal{C}_0$ and the proof is complete.

Finally with our next lemma we finish the proof of Theorem 4.1.

LEMMA 4.3. $d_n(A) = \mu_n(A)$.

PROOF. The proof depends on the following observation. Given any $x \in R^M$ which alternates between some n components, then

$$\|Ax^0\|_\infty \leqslant \max_{i \in \mathbf{i}^0} |(Ax)_i|. \tag{4.6}$$

Assume, to the contrary, that there exists a vector x which alternates between some n components and $|(Ax)_i| < \|Ax^0\|_\infty$, $i \in \mathbf{i}^0$. Then for any sufficiently small $\delta > 0$, $S^-(Ax^0 \pm (1+\delta)Ax) \geqslant n$ because Ax^0 equioscillates on $\mathbf{i}^0$. Moreover, $S^-(x^0 \pm (1+\delta)x) \leqslant n$, since x alternates between some n components. Thus Theorem 2.4 implies that $S^-(x^0 \pm (1+\delta)x) = n$ and the components of the vector $\pm(1+\delta)x$ have a fixed sign orientation determined by the vector Ax^0. This contradiction implies that (4.6) is valid.

Now, to complete the proof we note that for any $\mathbf{j}$, there exists a vector $x_{\mathbf{j}}$ which alternates between $j_1, \ldots, j_n$ such that

$$\max_{i \in \mathbf{i}^0} |(Ax_{\mathbf{j}})_i| = \sup_{\|x\|_\infty < 1} \inf_{\alpha_j} \max_{i \in \mathbf{i}^0} \left|\left(Ax - \sum_{j \in \mathbf{j}} \alpha_j a^j\right)_i\right|.$$

This follows from (4.2) and (4.4) when applied to the $(n+1) \times M$ matrix $A(\mathbf{i}^0)$. Hence by the definitions of $\lambda_n(A(\mathbf{i}^0))$ and $\mu_n(A)$ we have

$$d_n(A) = \|Ax^0\|_\infty \leqslant \lambda_n(A(\mathbf{i}^0)) \leqslant \mu_n(A).$$

We conclude from Theorem 3.1 that $d_n(A) = \mu_n(A)$. This is our desired conclusion.

The following result is closely related to the above lemma and is proven in a similar manner.

PROPOSITION 4.1. *If A is an $N \times M$ STP_{n+1} matrix and if x is any vector such that $\|x\|_\infty \leqslant 1$, and $(Ax)_{i_k}(-1)^k\sigma \geqslant 0$, $k = 1, \ldots, n+1$, where $\sigma^2 = 1$ for some $n+1$ rows, $\mathbf{i} = (i_1, \ldots, i_{n+1})$, $1 \leqslant i_1 < \cdots < i_{n+1} \leqslant N$, then*

$$\min_{i \in \mathbf{i}} |(Ax)_i| \leqslant \|Ax^0\|_\infty.$$

The following two quantities are closely related to the n-width of $\mathscr{Q}_0$,

$$\gamma_n(A) = \inf_{\operatorname{rank} B = n} \sup_{\|x\|_\infty < 1} \|Ax - Bx\|_\infty \tag{4.7}$$

where B is an $N \times M$ matrix, and

$$d^n(A) = d^n(\mathscr{Q}_0) = \inf_{e^1, \ldots, e^n} \sup_{\substack{(y, e^i) = 0,\ i = 1, \ldots, n \\ y \in \mathscr{Q}_0}} \|y\|_\infty. \tag{4.8}$$

The first quantity is the error in the best approximation to A, relative to the induced matrix norm, by rank n matrices, while $d^n(A)$ is the n-width of $\mathscr{Q}_0$ in the sense of Gel'fand (the infimum in (4.8) is taken over all sets of n vectors in R^N).

Given any vectors $e^1, \ldots, e^n$ in R^n, let U_{n+1} be the $n + 1$-dimensional ball defined in Lemma 4.2. Then there exists a $y \in U_{n+1}$ with $\|y\|_\infty = d_n(A)$ and $(y, e^i) = 0$, $i = 1, \ldots, n$. This is a consequence of the fact that any n homogeneous linear equations in $n + 1$ unknowns always have a nonzero solution. Hence we conclude from Theorem 4.1 that $d_n(A) \leqslant d^n(A)$. Also, by definition we have $d_n(A) \leqslant \gamma_n(A)$. We claim that

$$d^n(A) = \gamma_n(A) = d_n(A).$$

This conclusion will follow from our next lemma.

In preparation for this lemma let us observe that since $S^+(Ax^0) = S^-(Ax^0) = S^-(x^0) = n$ then $(Ax^0)_1(Ax^0)_N \neq 0$ and if $(Ax^0)_i = 0$, $1 < i < N$, then $(Ax^0)_{i-1}(Ax^0)_{i+1} < 0$. Furthermore, at the ith sign change (S^+ or S^-) of Ax^0, $i = 1, \ldots, n$, one of two possibilities occurs. Either

(a) $(Ax^0)_{k_i}(Ax^0)_{k_i+1} < 0$, or
(b) $(Ax^0)_{k_i} = 0$, and $(Ax^0)_{k_i-1}(Ax^0)_{k_i+1} < 0$,

where $1 \leqslant k_1 < \cdots < k_n < N$.

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

$$(e^i)_l = \begin{cases} |(Ax^0)_l|^{-1}, & l = k_i, k_i + 1, \\ 0, & \text{otherwise.} \end{cases}$$

If (b) arises let $(e^i)_l = \delta_{k_i l}$, $l = 1, \ldots, N$. Thus in either case $(Ax^0, e^i) = 0$, $i = 1, \ldots, n$.

Now, we define an $M \times M$ matrix P by the condition that for any $x \in R^M$ the vector $y = Px$ has the property

$$(x - Px, A^T e^i) = 0, \qquad i = 1, \ldots, n,$$

and

$$(Px)_l = 0, \qquad l \notin \{j_1^0, \ldots, j_n^0\}.$$

Px exists, since otherwise there is a nonzero vector $y \in R^M$ such that $(Ay, e^i) = (y, A^Te^i) = 0$, $i = 1, \ldots, n$, and $y_l = 0$, $l \notin \{j_1^0, \ldots, j_n^0\}$. From the equations, $(Ay, e^i) = 0, i = 1, \ldots, n$, it follows that $S^+(Ay) \geqslant n$. But the vector y has at most n nonzero components and thus we conclude that $S^-(y) \leqslant n - 1$. Now, we may apply Theorem 2.4 and arrive at a contradiction.

Let $B = AP$ and note that B is an $N \times M$ matrix of rank n whose column space is spanned by the set of vectors $\{a^j: j \in \mathbf{j}^0\}$.

LEMMA 4.4. $|(Ax - Bx)_i| \leqslant |(Ax^0)_i|$, $i = 1, 2, \ldots, N$, *for* $\|x\|_\infty \leqslant 1$.

PROOF. We prove this lemma by contradiction. Suppose there is an i such that $|(Ax - Bx)_i| > |(Ax^0)_i|$.

Let $d = (Ax^0)_i/(Ax - Bx)_i$, then the ith component of the vector $z = d(Ax - Bx) - Ax^0$ is zero. Also, $(Ax^0)_i \neq 0$, otherwise, by the definition of the vector e^i and the fact that $(Ax - Bx, e^l) = 0$, $l = 1, \ldots, n$, we conclude that $(Ax - Bx)_i = 0$. This contradicts our assumption that $|(Ax - Bx)_i| > |(Ax^0)_i|$. Since $(z, e^j) = 0, j = 1, \ldots, n$, we conclude that $S^+(z) \geqslant n + 1$, as before. However,

$$S^+(z) \leqslant S^-\big(d(x - Px) - x^0\big) \leqslant n.$$

This contradiction proves the lemma.

THEOREM 4.2. *If A is an $N \times M$ STP_{n+1} matrix then*

$$d_n(A) = d^n(A) = \gamma_n(A).$$

PROOF. According to Lemma 4.4, $\gamma_n(A) \leqslant \|Ax^0\|_\infty = d_n(A)$. Also, if $(Ax, e^i) = 0$, $i = 1, \ldots, n$, then $Px = 0$. Therefore Lemma 4.4 implies $\|Ax\|_\infty \leqslant \|Ax^0\|_\infty$, if $\|x\|_\infty \leqslant 1$. Thus $d^n(A) \leqslant d_n(A)$. Since we already observed that $d_n(A) \leqslant \gamma_n(A)$ and $d_n(A) \leqslant d^n(A)$ the theorem is proven.

REMARK 4.1 By a variation on the method used in Lemma 4.3, it may be proven that for A STP_{n+1}, if x alternates between some n components, and $x \neq x^0$, then $\|Ax\|_\infty > \|Ax^0\|_\infty$. Thus we conclude that the vector x^0 is unique. However, this fact does not imply the uniqueness of the n optimal columns $j_1^0, \ldots, j_n^0$. Nonuniqueness may occur only if $|(x^0)_{j_k^0}| = 1$ for some $k = 1, \ldots, n$.

5. **Some extensions.** In this section we include some useful extensions of our results of §4.

We begin by observing that the maximum norm of a vector is unchanged by permuting any of its components or multiplying them by ± 1. Thus Theorem 4.1, parts (i), (ii) and an appropriate formulation of (iii), remains valid for any matrix A which after row and column interchange and/or multiplication by ± 1 is STP_{n+1}.

An important example of matrices of this form are inverses of STP matrices. If A is an $N \times N$ STP_N matrix, then it follows from (3.11) that if $A^{-1} = \|b_{ij}\|$, then the matrix $B = \|b_{ij}(-1)^{i+j}\|$ is STP_N. As a result of this observation we have

COROLLARY 5.1. *If A is an $N \times N$ STP_N matrix, then* $d_n(A)d_{N-n-1}(A^{-1}) = 1, n = 0, 1, \ldots, N-1$.

PROOF. From Proposition 3.4, $\lambda_n(A)\mu_{N-n-1}(A^{-1}) = 1$, $n = 0, 1, \ldots, N-1$, for any $N \times N$ nonsingular matrix. For an STP_N matrix we have from Theorem 4.1 that $\lambda_n(A) = d_n(A)$ and $\mu_n(A^{-1}) = d_n(A^{-1})$. Thus we obtain the desired conclusion.

Let A be an $N \times M$ STP_{n+1} matrix. For any r, $0 \leq r < M$, we define

$$\mathcal{C}_r = \{Ax : |x_i| \leq 1, i = r+1, \ldots, M\}.$$

Our objective is to extend the results of §4 to the set $\mathcal{C}_r$. We begin with

PROPOSITION 5.1. *If $n < r$ then $d_n(\mathcal{C}_r) = \infty$. If $n \geq r$ and if X_n is an n-dimensional subspace of R^N such that* $\sup_{y \in \mathcal{C}_r} \inf_{x \in X_n} \|y - x\| < \infty$ *then* $\{a^1, \ldots, a^r\} \subseteq X_n$.

PROOF. Suppose $n < r$ and X_n is any n-dimensional subspace of R^N. Since the vectors $a^1, \ldots, a^{n+1}$ are linearly independent there exists a vector $y = Ax = \sum_{j=1}^{n+1} x_j a^j$ such that $\inf_{x \in X_n} \|y - x\|_\infty = 1$. But $\lambda y \in \mathcal{C}_r$ for all $\lambda > 0$. Thus $d_n(\mathcal{C}_r) \geq \lambda$ and we conclude that $d_n(\mathcal{C}_r) = \infty$.

Similarly, if $n \geq r$ and $\sup_{y \in \mathcal{C}_r} \inf_{x \in X_n} \|y - x\|_\infty = d < \infty$ then, for all $\lambda > 0$,

$$\lambda \inf_{x \in X_n} \|a^j - x\|_\infty \leq d, \qquad j = 1, \ldots, r.$$

Therefore $\{a^1, \ldots, a^r\} \subseteq X_n$, as asserted in the proposition.

Let us now modify the definitions in §4 so that they apply to the set $\mathcal{C}_r$.

We will use the notation $\pi_r x = (0, \ldots, 0, x_{r+1}, \ldots, x_M)$, $d_{n,r}(A) = d_n(\mathcal{C}_r)$ and $d^{n,r}(A) = d^n(\mathcal{C}_r)$. With this notation $\mathcal{C}_r = \{Ax : \|\pi_r x\|_\infty \leq 1\}$.

The quantity $\lambda_{n,r}(A)$ is defined by (3.1) where we replace the condition $\|x\|_\infty \leq 1$ by the weaker requirement $\|\pi_r x\|_\infty \leq 1$. $\mu_{n,r}(A)$ is defined to be $\max_{\mathbf{i} \in I} \lambda_{n,r}(A(\mathbf{i}))$ and $\gamma_{n,r}(A)$ is defined by (4.7) where again the condition $\|x\|_\infty \leq 1$ is replaced by $\|\pi_r x\|_\infty \leq 1$.

Also, we define for any set of integers $\{j_{r+1}, \ldots, j_n\}$, $r + 1 \leq j_{r+1} < \cdots < j_n \leq M$, the vector $\mathbf{j}_r = (1, \ldots, r, j_{r+1}, \ldots, j_n)$. Let

$$\mathbf{i}^0 = (i_1^0, \ldots, i_{n+1}^0), \qquad 1 \leq i_1^0 < \cdots < i_{n+1}^0 \leq N,$$

and

$$\mathbf{j}_r^0 = (1, \ldots, r, j_{r+1}^0, \ldots, j_n^0)$$

denote the rows and columns, respectively, for which $\min_{\mathbf{j}_r \in J} \max_{\mathbf{i} \in I} G(\mathbf{i}, \mathbf{j}_r)$ is achieved.

THEOREM 5.1. *Let A be an $N \times M$ STP_{n+1} matrix with $r \leqslant n < \min(N, M)$. Then*

(1) $\mu_{n,r}(A) = \max_{\mathbf{i} \in I} \min_{\mathbf{j}_r \in J} G(\mathbf{i}, \mathbf{j}_r)$, $\lambda_{n,r}(A) = \min_{\mathbf{j}_r \in J} \max_{\mathbf{i} \in I} G(\mathbf{i}, \mathbf{j}_r)$.

(2) $d_{n,r}(A) = d^{n,r}(A) = \gamma_{n,r}(A) = \mu_{n,r}(A) = \lambda_{n,r}(A)$.

(3) *A best n-dimensional linear subpsace for the width $d_{n,r}(A)$ is attained by the subspace spanned by the columns $\{a^j : j \in \mathbf{j}_r^0\}$.*

(4) $d_{n,r}(A) = \|Ax^0\|_\infty$, *where Ax^0 is a vector which equioscillates on $i_1^0, \ldots, i_{n+1}^0$, $\pi_r x^0$ alternates between $j_{r+1}^0, \ldots, j_n^0$ and $\|\pi_r x^0\|_\infty = 1$.*

The proof of Theorem 5.1 is similar to the proof of Theorem 4.1 since the first r components of the vector x^0 will not interfere with our previous sign change arguments.

In addition to Theorem 5.1 we also know as in the proof of Theorem 4.1 the following facts about the vector x^0

(a) $S^-(Ax^0) = S^+(Ax^0) = S^-(x^0) = n$.

(b) $\|Ax^0\|_\infty \leqslant \mathrm{Max}_{i \in \mathbf{i}^0} |(Ax)_i|$, for any vector $x \in R^M$ for which $\pi_r x$ alternates between some $n - r$ components.

(c) $\min_{i \in \mathbf{i}} |(Ax)_i| \leqslant \|Ax^0\|_\infty$ for any vector $x \in R^M$ such that $\|\pi_r x\|_\infty \leqslant 1$ and $(Ax)_{i_k}(-1)^k \sigma \geqslant 0$, $k = 1, \ldots, n+1$, where $\sigma^2 = 1$ for some $n + 1$ rows $i_1, \ldots, i_{n+1}$, $1 \leqslant i_1 < \cdots < i_{n+1} \leqslant N$.

(d) Let $e^1, \ldots, e^n$ be vectors constructed from Ax^0 as before. Also, we define an $M \times M$ matrix P_r by the condition that for any $x \in R^M$

$$\left(x - P_r x, A^T e^i\right) = 0, \qquad i = 1, \ldots, n.$$

and $(P_r x)_l = 0$, $l \notin \mathbf{j}_r^0$. Let $B_r = AP_r$. Then

$$\left|(Ax - B_r x)_i\right| \leqslant \left|(Ax^0)_i\right|, \qquad i = 1, \ldots, N,$$

for any $x \in R^M$ with $\|\pi_r x\|_\infty \leqslant 1$.

The class of STP matrices is frequently too small to include many important applications of Theorem 5.1. Below we give an extension of part of Theorem 5.1 which is valid for TP matrices. §7 contains a thorough discussion of the continuous analogue of Theorem 5.2 for TP kernels.

THEOREM 5.2. *Let A be an $N \times M$ TP_{n+1} matrix of rank at least $n + 1$ such that any n columns of A are linearly independent and $r \leqslant n < \min(N, M)$. Then there exists a vector $x^0 \in R^M$ such that $\pi_r x^0$ alternates between some $n - r$ columns $j_{r+1}^0, \ldots, j_n^0$, $r + 1 \leqslant j_{r+1}^0 < \cdots < j_n^0 \leqslant M$, $\|\pi_r x^0\|_\infty = 1$ and Ax^0 equioscillates on some $n + 1$ rows.*

Furthermore, for any vector x^0 with these properties we have $d_{n,r}(A) =$

$\|Ax^0\|_\infty$ *and an optimal subspace for the set* $\mathcal{Q}_r$ *is spanned by the vectors* $\{a^j: j \in \mathbf{j}_r^0\}$, $\mathbf{j}_r^0 = (1, \ldots, r, j_{r+1}^0, \ldots, j_n^0)$.

PROOF. The existence of a vector x^0 with the above properties follows from Theorem 5.1 by the following argument.

Since A is TP_{n+1} and of rank at least $n+1$ there exists a sequence of $N \times M$ STP_{n+1} matrices $\{A_t: t \geq 0\}$ such that $A_t \to A$ as $t \to \infty$. According to Theorem 5.1 there exists $x^0(t) \in R^M$ such that $A_t x^0(t)$ equioscillates on some $n+1$ rows $i_1^0(t), \ldots, i_{n+1}^0(t)$, $\|\pi_r x^0(t)\|_\infty = 1$ and $\pi_r x^0(t)$ alternates between $j_{r+1}^0(t), \ldots, j_n^0(t)$, $r+1 \leq j_{r+1}^0(t) < \cdots < j_n^0(t) \leq M$. There exists a subsequence $\{t_s\}$, $t_s \to \infty$, such that $i_l^0(t_s) = i_l^0$, $l = 1, \ldots, n+1$, $j_l^0(t_s) = j_l^0$, $l = r+1, \ldots, n$; $s = 1, 2, \ldots$.

Thus

$$(5.1)\quad \left(A_{t_s} x^0(t_s)\right)_{i_l^0} = (-1)^l \left\|A_{t_s} x^0(t_s)\right\|_\infty, \qquad s = 1, 2, \ldots, ;\ l = 1, \ldots, n+1.$$

We claim that the sequence

$$d_s = |x_1^0(t_s)| + \cdots + |x_r^0(t_s)| + \left\|A_{t_s} x^0(t_s)\right\|_\infty, \qquad s = 1, 2, \ldots,$$

is bounded. For, if $d_s \to \infty$ we may divide both sides of (5.1) by d_s. Then after passing through a subsequence, if necessary, we conclude that there exists numbers $y_1, \ldots, y_{r+1}$, $\sum_{i=1}^{r+1} |y_i| = 1$ such that $y_{r+1} \geq 0$ and

$$(5.2)\qquad \left(\sum_{j=1}^{r} y_j a^j\right)_{i_l^0} = (-1)^l y_{r+1}, \qquad l = 1, \ldots, n+1,$$

$$(5.3)\qquad \left|\left(\sum_{j=1}^{r} y_j a^j\right)_i\right| \leq y_{r+1}, \qquad i \notin \mathbf{i}^0.$$

If $y_{r+1} \neq 0$ then from (5.2) we conclude that $S^-(\sum_{j=1}^r y_j a^j) \geq n$. However, A is TP_{n+1} and we know that $S^-(\sum_{j=1}^r y_j a^j) \leq r - 1 \leq n - 1$. Thus $y_{r+1} = 0$ and we conclude that $\sum_{j=1}^r y_j a^j = 0$. Since $a^1, \ldots, a^r$ are assumed to be linearly independent we conclude that $y_1 = \cdots = y_r = 0$ which is a contradiction.

Now that we know that $\sup_s d_s < \infty$ we may easily pass to the limit in (5.1), perhaps through a subsequence, and conclude that there exists a vector x^0 with the properties demanded by the theorem.

Thus it remains to prove that for any such vector x^0 we have $d_{n,r}(A) = \|Ax^0\|_\infty$. The proof of this fact is very similar to the proofs given in §4, so we will be brief.

From the discussion in Proposition 3.1 we may easily conclude that

$$\sup_{\|\pi_r x\|_\infty < 1} \inf_{\alpha_j} \left\|Ax - \sum_{j \in \mathbf{j}_r^0} \alpha_j a^j\right\|_\infty = \|Ax^0\|_\infty.$$

Thus $d_{n,r}(A) \leq \|Ax^0\|_\infty$. To prove the reverse inequality we follow the proof of Lemma 4.2. We define the vectors $b^1, \ldots, b^{n+1}$ as in Lemma 4.2 and note that since $l_i = i, i = 1, \ldots, r$, b^k is a multiple of a^k, $k = 1, \ldots, r$. Now, it is an easy matter to prove that if $\|\Sigma_{k=1}^{n+1}\alpha_k b^k\|_\infty \leq \|Ax^0\|_\infty$ then $|\alpha_k| \leq 1$, $k = r + 1, \ldots, n$. From this conclusion it follows that the ball U_{n+1}, defined in the proof of Lemma 4.2, is contained in $\mathcal{Q}_r$. Hence by Theorem 2.1 $d_{n,r}(A) \geq \|Ax^0\|_\infty$. This inequality proves the theorem.

6. **Some examples.** Let $A = \|a_{ij}\|$ be the $N \times N$ matrix defined by $a_{ij} = 1$, $i \leq j$, $a_{ij} = 0$, $i > j$. Thus A is the lower triangular matrix with 1's on and below the diagonal and is clearly TP_N.

Let $[x]$ denote the greatest integer less than or equal to x. Then

PROPOSITION 6.1.

$$d_n(A) = \frac{1}{2}\left(N - \left[\frac{N(2n-1)+2n}{2n+1}\right]\right), \qquad n = 0, 1, \ldots, N,$$

and the columns $a^{j_1}, \ldots, a^{j_n}$ *are optimal if and only if*

$$j_l = (N - j_n)\left(l - \tfrac{1}{2}\right) + e_1 + \cdots + e_l, \qquad l = 1, 2, \ldots, n \quad \textit{for some } \{e_i\}_{i=1}^n, \tag{6.1}$$

where

$$e_i = \begin{cases} 0, \frac{1}{2}, \textit{or } 1, & i = 1, \\ 0 \textit{ or } 1, & i = 2, \ldots, n, \end{cases}$$

and

$$j_n = \left[\frac{N(2n-1)+2n}{2n+1}\right].$$

PROOF. Let x^0 be the N-dimensional vector, $x^0 = (1, 1, \ldots, 1, 0, -1, \ldots, -1, 0, 1, \ldots, 1, 0, \ldots)$ where the zeros occur at the components $j_1, \ldots, j_n$.

Also, let $j_0 = 0$, $j_{n+1} = N + 1$, $\gamma_i = j_i - j_{i-1}$, $i = 1, \ldots, n + 1$, and $\beta_i = \Sigma_{j=1}^i(-1)^{j+1}(\gamma_j - 1)$, $i = 1, \ldots, n + 1$. Then $(Ax^0)_{j_k - 1} = (Ax^0)_{j_k} = \beta_k$, $k = 1, \ldots, n$, and $(Ax^0)_N = \beta_{n+1}$.

Also, by their definition, $0 \leq \beta_1 \geq \beta_2 \leq \beta_3 \geq \beta_4, \ldots,$ the difference between consecutive components of Ax^0 is at most one and the components of Ax^0 are monotonically increasing or decreasing between the j_l and j_{l+1} components.

Let $\{\alpha_j\}_{j=1}^n$ satisfy

$$\beta_i - \sum_{j=1}^{i} \alpha_j = (-1)^{i+1}\lambda, \qquad i = 1, \ldots, n,$$

(6.2)

$$\beta_{n+1} - \sum_{j=1}^{n} \alpha_j = (-1)^{n+2}\lambda.$$

Then, from Theorem 5.2, $\lambda = d_n(A)$ if

(a) $|\alpha_j| \leq 1, j = 1, \ldots, n,$

(b) $|\beta_i - \sum_{j=1}^{i-1}\alpha_j| \leq \lambda, i = 1, \ldots, n.$

Now, from (6.2),

$$\begin{aligned} \alpha_i &= (-1)^{i+1}(\gamma_i - 1 - 2\lambda), \qquad i = 2, \ldots, n, \\ \alpha_1 &= \gamma_1 - 1 - \lambda, \\ \lambda &= (N - j_n)/2. \end{aligned} \tag{6.3}$$

Thus (a) is equivalent to

$$\begin{aligned} |\gamma_i - 1 - 2\lambda| &\leq 1, \quad i = 2, \ldots, n, \\ |\gamma_1 - 1 - \lambda| &\leq 1, \end{aligned} \tag{a'}$$

while (b) implies

$$\begin{aligned} |\gamma_i - 1 - \lambda| &\leq \lambda, \quad i = 2, \ldots, n, \\ \gamma_1 - 1 &\leq \lambda. \end{aligned} \tag{b'}$$

From (a′) and (b′) it follows

$$N - j_n \leq \gamma_i \leq N - j_n + 1, \quad i = 2, \ldots, n, \tag{a''}$$

$$(N - j_n)/2 \leq \gamma_1 \leq (N - j_n)/2 + 1. \tag{b''}$$

Since $\sum_{i=1}^{n}\gamma_i = j_n$, we have

$$\frac{N(2n-1)}{2n+1} \leq j_n \leq \frac{N(2n-1) + 2n}{2n+1}$$

or equivalently

$$j_n = \left[\frac{N(2n-1) + 2n}{2n+1} \right].$$

From (6.3)

$$d_n(A) = \lambda = \frac{1}{2}\left(N - \left[\frac{N(2n-1) + 2n}{2n+1} \right]\right),$$

$n = 0, 1, \ldots, N$, and any columns satisfying (a″) and (b″) are optimal. These inequalities are easily seen to be equivalent to (6.1). Thus the proof is complete.

Let us now consider the following problem. Given N fixed points, $0 \leq t_1 < \cdots < t_N \leq 1$, we wish to "store" the values of one or more functions computed at the points $t_1, \ldots, t_N$. We are allowed only n storage locations where $n << N$, but we have some assurance that the rth ($r < n$) consecutive divided differences of the computed data are bounded by some constant which we normalize to one. Then the n-width of the set

$$\mathfrak{B} = \{(f(t_1), \ldots, f(t_N)) \colon |f(t_i, \ldots, t_{i+r})| \leq 1, i = 1, \ldots, N - r\}$$

will give us an estimate of the minimum intrinsic error in storing the data.

As a corollary to Theorem 5.2 we have the following result.

THEOREM 6.1. *$d_n(\mathfrak{B}) = \max_{1 \leq j \leq N} |f_0(t_j)|$, where the vector $(f_0(t_1), \ldots, f_0(t_N))$ equioscillates on some $n + 1$ components and the vector $(f_0(t_1, \ldots, t_{r+1}), f_0(t_2, \ldots, t_{r+2}), \ldots, f_0(t_{N-r}, \ldots, t_N))$ alternates between some $n - r$ columns.*

PROOF. The proof of this theorem consists of showing that the set $\mathfrak{B}$ may be expressed as $\mathcal{Q}_r$ for some totally positive matrix. We begin with the identity

$$\begin{bmatrix} f(t_1) \\ \vdots \\ f(t_N) \end{bmatrix} = \begin{bmatrix} t_1 - t_0 & 0 & 0 & \cdots & 0 \\ t_1 - t_0 & t_2 - t_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ t_1 - t_0 & t_2 - t_1 & & \cdots & t_N - t_{N-1} \end{bmatrix} \begin{bmatrix} f(t_0, t_1) \\ f(t_1, t_2) \\ \vdots \\ f(t_{N-1}, t_N) \end{bmatrix}$$

where we choose any $t_0 < t_1$ and set $f(t_0) = 0$. Let us denote the above $N \times N$ matrix by $B(t_0, \ldots, t_N)$. Then it easily follows by induction on r that

$$f(t_i) = \sum_{j=1}^{N} A_{ij} f(t_{-r+j}, \ldots, t_j), \qquad i = 1, \ldots, N,$$

where $A = B(t_0, \ldots, t_N) B(t_{-1}, \ldots, t_{N-1}) \cdots B(t_{-r+1}, \ldots, t_{N-r})$, $t_{-r} < \cdots < t_0$ and $f(t_{-r}) = \cdots = f(t_0) = 0$. Since $B(t_i, \ldots, t_{i+N})$ is totally positive, $i = 0, -1, \ldots, -r + 1$, (it is just a column scaling of the matrix which appears in Proposition 6.1), it follows that A is totally positive. If we define the vector $x = (x_1, \ldots, x_N)$, $x_i = f(t_{-r+i}, \ldots, t_i)$, $i = 1, \ldots, N$, then $\mathfrak{B} = \{Ax \colon \|\pi_r x\|_\infty \leq 1\}$. Hence Theorem 6.1 follows from Theorem 5.2.

7. **n-widths in $L^\infty[0, 1]$.** Let $X = L^\infty[0, 1]$ and $\|h\|_\infty = \operatorname{ess\,sup}\{|h(x)| \colon 0 \leq x \leq 1\}$. Suppose $k_1(t), \ldots, k_r(t)$ are continuous functions on $[0, 1]$ and $K(t, s)$ is jointly continuous for $t, s \in [0, 1]$.

We define the subset of continuous functions

$$\mathcal{K}_r = \left\{ \sum_{j=1}^{r} x_j k_j(t) + \int_0^1 K(t, s) h(s)\, ds \colon (x_1, \ldots, x_r) \in R^r, \|h\|_\infty \leq 1 \right\}.$$

We will assume throughout this section that for $m \geqslant 0$, $0 \leqslant k \leqslant r$, $0 \leqslant s_1 < \cdots < s_m \leqslant 1$, $0 \leqslant t_1 < \cdots < t_{m+k} \leqslant 1$, $1 \leqslant i_1 < \cdots < i_k \leqslant r$, the determinant

$$K\begin{pmatrix} i_1, \ldots, i_k, s_1, \ldots, s_m \\ t_1, \ldots, t_k, t_{k+1}, \ldots, t_{m+k} \end{pmatrix}$$

$$(7.1) \qquad = \begin{vmatrix} k_{i_1}(t_1) & \cdots & k_{i_1}(t_{m+k}) \\ \vdots & & \vdots \\ k_{i_k}(t_1) & \cdots & k_{i_k}(t_{m+k}) \\ K(t_1, s_1) & \cdots & K(t_{m+k}, s_1) \\ \vdots & & \vdots \\ K(t_1, s_m) & \cdots & K(t_{m+k}, s_m) \end{vmatrix}$$

is nonnegative. In addition to this hypothesis we also require the following assumptions to hold.

The sets of functions

$$(7.2) \qquad \{k_1(t), \ldots, k_r(t), K(t, s_1), \ldots, K(t, s_m)\},$$

are linearly independent on [0, 1], for any $0 < s_1 < \cdots < s_m < 1$, $m = 1, 2, \ldots$.

(7.3) For any $0 < t_1 < \cdots < t_r < 1$, $\det\limits_{i,j=1,\ldots,r} \|k_i(t_j)\| > 0$.

Thus our assumptions on $\mathcal{K}_r$ consist of the essential total positivity assumption (7.1) and, in addition, we require the linear independence hypotheses (7.2) and (7.3).

Lemma 7.1. *For any constants* $\alpha_1, \ldots, \alpha_n$, $0 = s_0 < s_1 < \cdots < s_{n-r} < s_{n-r+1} = 1$, *the function*

$$g(t) = \sum_{j=1}^{r} \alpha_j k_j(t) + \sum_{j=0}^{n-r} (-1)^j \int_{s_j}^{s_{j+1}} K(t, y)\, dy + \sum_{j=r+1}^{n} \alpha_j K(t, s_{j-r})$$

has at most n distinct zeros in (0, 1).

Proof. Suppose to the contrary that $g(t)$ has $n + 1$ distinct zeros $t = z_1, \ldots, z_{n+1}$ in (0, 1). We introduce the functions

$$v_i(y) = K\begin{pmatrix} 1, \ldots, r, y \\ z_1, \ldots, z_r, z_{r+1} \end{pmatrix}, \qquad i = 1, \ldots, n - r + 1.$$

By Sylvester's determinant identity we have

$$\det_{i,j=1,\ldots,n-r+1}\|v_i(y_j)\| = \left(K\begin{pmatrix}1,\ldots,r\\ z_1,\ldots,z_r\end{pmatrix}\right)^{n-r} K\begin{pmatrix}1,\ldots,r,y_1,\ldots,y_{n-r+1}\\ z_1,\ldots,z_r,z_{r+1},\ldots,z_{n+1}\end{pmatrix}.$$

Hence by our hypotheses (7.2) and (7.3) the set $\{v_1(x),\ldots; v_{n-r+1}(x)\}$ is a weak Chebyshev system on [0, 1]. Moreover, from the equations $g(z_i)=0$, $i=1,2,\ldots,n+1$, it follows that

$$\sum_{j=0}^{n-r}(-1)^j\int_{s_j}^{s_{j+1}}v_i(y)\,dy+\sum_{j=r+1}^{n}\alpha_j v_i(s_{j-r})=0, \qquad i=1,\ldots,n-r+1. \tag{7.4}$$

We arrive at a contradiction, as in [7], by constructing a nontrivial function $v(y)=\sum_{j=1}^{n-r+1}\beta_j v_j(y)$ such that $v(y)(-1)^j>0$, $s_j<y<s_{j+1}$, $j=0,1,\ldots,n-r$. From (7.4) we obtain $\int_0^1|v(y)|\,dy=0$ which is impossible.

THEOREM 7.1. *Given any $n>r$, there exists a function of the form*

$$P_0(t)=\sum_{j=1}^{r}b_jk_j(t)+\sum_{j=0}^{n-r}(-1)^j\int_{\xi_j}^{\xi_{j+1}}K(t,s)\,ds$$

with $0=\xi_0<\xi_1<\cdots<\xi_{n-r}<\xi_{n-r+1}=1$ *which equioscillates on* $n+1$ *points of* [0, 1],

$$P_0(e_j)=(-1)^{j+r+1}\|P_0\|_\infty,\qquad j=1,\ldots,n+1,$$

where $0\leqslant e_1<\cdots<e_{n+1}\leqslant 1$.

PROOF. We set

$$s_j=j/(N-1),\qquad j=0,1,\ldots,N-1,$$

and

$$t_j=j/(N+r-1),\qquad j=0,1,\ldots,N+r-1.$$

Then according to (7.1) and Theorem 5.2 there exists a vector $x^0(N)=(x_1^0(N),\ldots,x_{N+r}^0(N))$ such that $\|\pi_r x^0(N)\|_\infty=1$, $\pi_r x^0(N)$ alternates between some $n-r$ columns and, if we define

$$P^N(t)=\sum_{j=1}^{r}x_j^0(N)k_j(t)+\frac{1}{N-1}\sum_{j=1}^{N}x_{j+r}^0(N)K\left(t,\frac{j-1}{N-1}\right),$$

then there exist integers, $0\leqslant i_1^N<\cdots<i_{n+1}^N\leqslant N+r-1$, such that

$$P^N\left(\frac{i_l^N}{N+r-1}\right)=(-1)^{l+r+1}\|P^N\|_{\infty,N} = (-1)^{l+r+1}\max_{0\leqslant i\leqslant N+r-1}\left|P^N\left(\frac{i}{N+r-1}\right)\right|,\qquad l=1,\ldots,n+1.$$

We define $d_N = |x_1^0(N)| + \cdots + |x_r^0(N)| + \|P^N\|_{\infty,N}$. Then just as in Theorem 5.2 we may prove that $\sup_N d_N < \infty$. There exists a subsequence $\{i_l^{N_j}\}$ such that $i_l^{N_j}/(N_j + r - 1) \to e_l$, $l = 1, \ldots, n+1$, $0 \leqslant e_1 \leqslant e_2 \leqslant \cdots \leqslant e_{n+1} \leqslant 1$. Using the fact that $\sup_N d_N < \infty$ and the simple estimate

$$\left| \frac{1}{N-1} \sum_{k=0}^{n-r} (-1)^k \sum_{j_k}^{j_{k+1}-1} K\left(t, \frac{j}{N-1}\right) - \sum_{k=0}^{n-r} (-1)^k \int_{j_k/(N-1)}^{j_{k+1}/(N-1)} K(t,s)\, ds \right|$$

$$\leqslant \max\left\{ |K(t,y_1) - K(t,y_2)| : 0 \leqslant t \leqslant 1, |y_1 - y_2| \leqslant 1/(N-1) \right\}$$

which holds for all $t \in [0, 1]$, $0 = j_0 < j_1 < \cdots < j_{n-r} < j_{n-r+1} = N-1$, we conclude that there exists a function

$$P_0(t) = \sum_{j=1}^{r} b_j k_j(t) + \sum_{j=0}^{n-r} (-1)^j \int_{\xi_j}^{\xi_{j+1}} K(t,s)\, ds$$

such that $0 = \xi_0 \leqslant \xi_1 \leqslant \cdots \leqslant \xi_{n-r} \leqslant \xi_{n-r+1} = 1$ and

$$P_0(e_j) = (-1)^{j+r+1} \|P_0\|_\infty,$$

$0 \leqslant e_1 \leqslant \cdots \leqslant e_{n+1} \leqslant 1$. First let us note that $\|P_0\|_\infty > 0$, for otherwise we would contradict Lemma 7.1. Hence $0 \leqslant e_1 < \cdots < e_{n+1} \leqslant 1$.

Now, using (7.1) we may conclude from Theorem 2.4 that for N sufficiently large the function

$$\bar{P}^N(t) = \sum_{j=1}^{r} b_j k_j(t) + \frac{1}{N} \sum_{j=0}^{n-r} (-1)^j \sum_{k=[N\xi_j]/N}^{[N\xi_{j+1}]/N} K\left(t, \frac{k}{N}\right)$$

has the property that

$$n \leqslant S^-\left(\bar{P}^N(e_1), \ldots, \bar{P}^N(e_{n+1})\right) \leqslant r + \text{number of distinct } \xi_i\text{'s} \in (0, 1).$$

Hence we conclude that $0 < \xi_1 < \cdots < \xi_{n-r} < 1$ which completes the proof.

Let us remark that when the determinants in (7.1) are always strictly positive then the function constructed to Theorem 7.1 is unique. To see this, we let P_1 be any other function with these properties and suppose $\|P_1\|_\infty \leqslant \|P_0\|_\infty$. If $P_1 \not\equiv P_0$ then the function $E = P_0 - P_1$ has the property that $S^+(E(e_1), \ldots, E(e_{n+1})) \geqslant n$ while the strict positivity of the determinants in (7.1) implies that $S^+(E(e_1), \ldots, E(e_{n+1})) \leqslant n-1$, [2]. This contradiction implies $P_0 \equiv P_1$.

Before presenting the main theorem of this section we prove the following important lemma.

LEMMA 7.2. *Let P_0 be any function which has the properties described in Theorem 7.1 and let $P_0(z_i) = 0$, $i = 1, 2, \ldots, n$, $0 < z_1 < \cdots < z_n < 1$. Then*

$$K\begin{pmatrix} 1, \ldots, r, \xi_1, \ldots, \xi_{n-r} \\ z_1, \ldots, z_r, z_{r+1}, \ldots, z_n \end{pmatrix} > 0.$$

PROOF. We will prove the lemma by contradiction. Suppose that the above determinant is zero. Then there exists constants $c_1, \ldots, c_n$ not all zero such that

$$\sum_{j=1}^{r} c_j k_j(z_i) + \sum_{j=r+1}^{n} c_j K(z_i, \xi_{j-r}) = 0, \qquad i = 1, \ldots, n.$$

According to (7.2) there exists a $z \in (0, 1) - \{z_1, \ldots, z_n\}$ such that

$$\sum_{j=1}^{r} c_j k_j(z) + \sum_{j=r+1}^{n} c_j K(z, \xi_{j-r}) \neq 0.$$

Choose a constant d such that

$$P_0(z) + d\left(\sum_{j=1}^{r} c_j k_j(z) + \sum_{j=r+1}^{n} c_j K(z, \xi_{j-r}) \right) = 0,$$

(7.5)

$$\text{where } P_0(t) = \sum_{j=1}^{r} b_j k_j(t) + \sum_{j=0}^{n-r} (-1)^j \int_{\xi_j}^{\xi_{j+1}} K(t, s)\, ds.$$

However, (7.5) together with the equations $P_0(z_i) = 0$, $i = 1, \ldots, n$, contradicts Lemma 7.1 and thus the lemma is proven.

To state the main theorem of the section we define the following quantities

$$d_{n,r}(K) = d_n(\mathcal{K}_r; L^\infty[0, 1]),$$

$$d^{n,r}(K) = d^n(\mathcal{K}_r; C[0, 1]) = \inf_{L_n} \sup_{h \in \mathcal{K}_r \cap L_n} \|h\|$$

where the infimum is taken over all subspaces L_n of $C[0, 1]$ of codimension n,

$$\lambda_{n,r}(K) = \inf_{0 < s_1 < \cdots < s_{n-r} < 1} \sup_{\|h\|_\infty < 1} \inf_{\alpha_1, \ldots, \alpha_n} \max_{0 < t < 1} \cdot \left| \int_0^1 K(t, s)h(s)\, ds - \sum_{j=1}^{r} \alpha_j k_j(t) - \sum_{j=r+1}^{n} \alpha_j K(t, s_{j-r}) \right|,$$

$$\mu_{n,r}(K) = \sup_{0 < t_1 < \cdots < t_{n+1} < 1} \inf_{0 < s_1 < \cdots < s_{n-r} < 1} \sup_{\|h\|_\infty < 1} \inf_{\alpha_1, \ldots, \alpha_n} \max_{1 < i < n+1} \cdot \left| \int_0^1 K(t_i, s)h(s)\, ds - \sum_{j=1}^{r} \alpha_j k_j(t_i) - \sum_{j=r+1}^{n} \alpha_j K(t_i, s_{j-r}) \right|,$$

and

$$\gamma_{n,r}(K) = \inf_T \sup_{\|h\|_\infty < 1} \max_{0 < t < 1} \left| \int_0^1 K(t, s)h(s)\, ds - (Th)(t) \right|$$

where T is any bounded linear operator from $L^\infty[0, 1]$ into $C[0, 1]$ whose range is an n-dimensional subspace containing the functions $k_1(t), \ldots, k_r(t)$. Note that $d_{n,r}(K) \leq \gamma_{n,r}(K)$.

THEOREM 7.2. *Let P_0 be any function having the properties described in Theorem 7.1 and let $P_0(z_i) = 0$, $i = 1, \ldots, n$. Then*

(1) $d_{n,r}(K) = \lambda_{n,r}(K) = \|P_0\|_\infty$ *and an optimal subspace X_n^0 for the n-width of $\mathcal{K}_r$ is spanned by the functions $k_1(t), \ldots, k_r(t), K(t, \xi_1), \ldots, K(t, \xi_{n-r})$.*

(2) $d^{n,r}(K) = \|P_0\|_\infty$ *and a best subspace of codimension n for the n-width in the sense of Gel'fand for $\mathcal{K}_r$ is $\{g: g \in C[0, 1], g(z_i) = 0, i = 1, \ldots, n\}$.*

(3) $\gamma_{n,r}(K) = \|P_0\|_\infty$ *and a best rank n approximation to the operator $\int_0^1 K(t, s)h(s)\, ds$ whose range includes the functions $k_1(t), \ldots, k_r(t)$ is defined by $(Th)(z_i) = \int_0^1 K(z_i, s)h(s)\, ds$, $i = 1, \ldots, n$, and $Th \in X_n^0$.*

(4) $\mu_{n,r}(K) = \lambda_{n,r}(K) = \|P_0\|_\infty$, *and*

(7.6)

$$\mu_{n,r}(K) = \sup_{0<t_1<\cdots<t_{n+1}<1} \inf_{0<s_1<\cdots<s_{n-r}<1} \frac{\int_0^1 \left| K\begin{pmatrix} 1, \ldots, r, s_1, \ldots, s_{n-r}, s \\ t_1, \ldots, t_{n+1} \end{pmatrix} \right| ds}{\sum_{l=1}^{n+1} K\begin{pmatrix} 1, \ldots, r, s_1, \ldots, s_{n-r} \\ t_1, \ldots, \hat{t}_l, \ldots, t_{n+1} \end{pmatrix}},$$

$$\lambda_{n,r}(K) = \inf_{0<s_1<\cdots<s_{n-r}<1} \sup_{0<t_1<\cdots<t_{n+1}<1} \frac{\int_0^1 \left| K\begin{pmatrix} 1, \ldots, r, s_1, \ldots, s_{n-r}, s \\ t_1, \ldots, t_{n+1} \end{pmatrix} \right| ds}{\sum_{l=1}^{n+1} K\begin{pmatrix} 1, \ldots, r, s_1, \ldots, s_{n-r} \\ t_1, \ldots, \hat{t}_l, \ldots, t_{n+1} \end{pmatrix}}$$

and the inf-sup *and* sup-inf *are achieved for $t_i = e_i$, $i = 1, \ldots, n + 1$, and $s_i = \xi_i$, $i = 1, \ldots, n - r$.*

PROOF. The following three facts may be proved by either "discretization" as in Theorem 7.1 or by using arguments which are completely analogous to those used in the proof of Theorem 4.1.

(a) Let

$$g(t) = \sum_{j=1}^{r} a_j k_j(t) + \sum_{j=0}^{n-r} a_{j+r+1} \int_{\xi_j}^{\xi_{j+1}} K(t, s)\, ds.$$

If $\|g\|_\infty \leq \|P_0\|_\infty$ then $|a_j| \leq 1$, $j = r + 1, \ldots, n + 1$.

(b) Given any $h \in L^\infty[0, 1]$ we define linear functionals $\lambda_1(h), \ldots, \lambda_n(h)$ by

$$\int_0^1 K(z_i, s)h(s)\, ds = \sum_{j=1}^{r} \lambda_j(h) k_j(z_i) + \sum_{j=1}^{n-r} \lambda_{j+r}(h) K(z_i, \xi_j), \qquad i = 1, \ldots, n.$$

(This may be done in view of Lemma 7.2.) Then

$$\left|\int_0^1 K(t,s)h(s)\,ds - \sum_{j=1}^{r}\lambda_j(h)k_j(t) - \sum_{j=1}^{n-r}\lambda_{j+r}(h)K(t,\xi_j)\right| \leqslant |P_0(t)|,$$

$$t \in [0,1], \text{ if } \|h\|_\infty \leqslant 1.$$

(c) Let $0 < s_1 < \cdots < s_{n-r} < 1$ and

$$P(t) = \sum_{j=1}^{r} a_j k_j(t) + \sum_{j=0}^{n-r}(-1)^j \int_{s_j}^{s_{j+1}} K(t,s)\,ds + \sum_{j=r+1}^{n} a_j K(t, s_{j-r}).$$

Then $\|P_0\|_\infty \leqslant \max_{1 \leqslant i \leqslant n+1}|P(e_i)|$.

Hence from (a) it follows that $d_{n,r}(K) \geqslant \|P_0\|_\infty$ and $d^{n,r}(K) \geqslant \|P\|_\infty$. Using (b) we see that (1), (2) and (3) are valid. Thus it remains to prove (4).

Let $0 \leqslant t_1 < \cdots < t_{n+1} \leqslant 1$, $0 < s_1 < \cdots < s_{n-r} < 1$ and consider the set of equations

$$\int_0^1 K(t_i,s)h(s)\,ds = \sum_{j=1}^{r} b_j k_j(t_i) + \sum_{j=1}^{n-r} b_{j+r}K(t_i,s_j) + (-1)^i\lambda(h), \tag{7.7}$$

$$i = 1, \ldots, n+1,$$

where $h \in L^\infty[0,1]$. Clearly

$$|\lambda(h)| = \frac{\left|\int_0^1 K\begin{pmatrix}1, \ldots, r, s_1, \ldots, s_{n-r}, s\\ t_1, \ldots, t_{n+1}\end{pmatrix} h(s)\,ds\right|}{\sum_{l=1}^{n+1} K\begin{pmatrix}1, \ldots, r, s_1, \ldots, s_{n-r}\\ t_1, \ldots, \hat{t}_l, \ldots, t_{n+1}\end{pmatrix}} \tag{7.8}$$

$$= \inf_{\alpha_1,\ldots,\alpha_n} \max_{1 \leqslant i \leqslant n+1}\left|\int_0^1 K(t_i,s)h(s)\,ds - \sum_{j=1}^{r}\alpha_j k_j(t_i) - \sum_{j=r+1}^{n}\alpha_j K(t_i, s_{j-r})\right|$$

and thus (7.6) is valid by the definition of $\mu_{n,r}(K)$. Now, from (c) we conclude by choosing $t_i = e_i$, $i = 1, \ldots, n+1$, that

$$\|P_0\|_\infty \leqslant \frac{\int_0^1 \left|K\begin{pmatrix}1, \ldots, r, s_1, \ldots, s_{n-r}, s\\ e_1, \ldots, e_{n+1}\end{pmatrix}\right| ds}{\sum_{l=1}^{n+1} K\begin{pmatrix}1, \ldots, r, s_1, \ldots, s_{n-r}\\ e_1, \ldots, \hat{e}_l, \ldots, e_{n+1}\end{pmatrix}}$$

for all $0 < s_1 < \cdots < s_{n-r} < 1$. Hence $\|P_0\|_\infty \leqslant \mu_{n,r}(K)$. But clearly from (7.8) and the definition of $\lambda_{n,r}(K)$ we have

$$\inf_{0<s_1<\cdots<s_{n-r}<1}\ \sup_{0<t_1<\cdots<t_{n+1}<1} \frac{\displaystyle\int_0^1 \left| K\begin{pmatrix} 1,\ldots,r,s_1,\ldots,s_{n-r},s \\ t_1,\ldots,t_{n+1} \end{pmatrix} \right| ds}{\displaystyle\sum_{l=1}^{n+1} K\begin{pmatrix} 1,\ldots,r,s_1,\ldots,s_{n-r} \\ t_1,\ldots,\hat{t}_l,\ldots,t_{n+1} \end{pmatrix}}$$

$$\leq \lambda_{n,r}(K) = \|P_0\|_\infty.$$

The last inequality follows from (1) which we have already proved. Hence $\mu_{n,r}(K) = \lambda_{n,r}(K) = \|P_0\|_\infty$ and (4) is proven.

We conclude this section with several examples of Theorem 7.1.

The first example, treated in [10], was our motivation for this study.

EXAMPLE 1. $k_i(t) = t^{i-1}$, $i = 1, 2, \ldots, r$, $K(t, s) = (t-s)_+^{r-1}/(r-1)!$, $t, s \in [0, 1]$. In this case

$$\mathcal{K}_r = \left\{ f: f^{(r-1)} \text{ is absolutely continuous}, f^{(r)} \in L^\infty[0,1], \|f^{(r)}\|_\infty \leq 1 \right\}$$
$$= \left\{ f: f \in W^r[0,1], \|f^{(r)}\|_\infty \leq 1 \right\}.$$

Hence $d_{n,r}(K) = \|P_0\|_\infty$ where

$$P_0(t) = \sum_{j=0}^{r-1} a_j t^j + \frac{1}{(r-1)!} \sum_{j=0}^{n-r} (-1)^j \int_{\xi_j}^{\xi_{j+1}} (t-s)_+^{r-1}\, ds$$

is a perfect spline with $n - r$ knots $0 < \xi_1 < \cdots < \xi_{n-r} < 1$ which equioscillates at $n + 1$ points. In this case $0 = e_1 < \cdots < e_{n+1} = 1$ and an optimal n-dimensional subspace is given by

$$X_n^0 = \left\{ \sum_{j=0}^{r-1} \alpha_j t^j + \sum_{j=r+1}^{n} \alpha_j (t - \xi_{j-r})_+^{r-1} : (\alpha_1, \ldots, \alpha_n) \in R^n \right\}.$$

EXAMPLE 2. $K(t, s) = 1/(t + s)$, $t, s \in [a, b]$, $a > 0$. This kernel is known to be *strictly* totally positive. Hence the width of the set

$$\mathcal{K}_0 = \left\{ \int_a^b \frac{h(s)}{t+s}\, ds : \|h\|_\infty \leq 1 \right\}$$

is $d_n(K) = \|P_0\|_\infty$ where

$$P_0(t) = \sum_{j=0}^{n} (-1)^j \int_{\xi_j}^{\xi_{j+1}} \frac{1}{t+s}\, ds, \qquad a < \xi_1 < \cdots < \xi_n < b,$$

and P_0 equioscillates $n + 1$ times on $[a, b]$. $P_0(t)$ is unique and an optimal subspace for $\mathcal{K}_0$ is given by

$$X_n^0 = \left\{ \sum_{j=1}^{n} \alpha_j \frac{1}{t+\xi_j} : (\alpha_1, \ldots, \alpha_n) \in R^n \right\}.$$

EXAMPLE 3. $K(t, s) = e^{ts}$, $t, s \in [0, 1]$.

$$\mathcal{K}_0 = \left\{ \int_0^1 e^{ts}h(s)\,ds \colon \|h\|_\infty \leqslant 1 \right\},$$

$d_n(K) = \|P_0\|_\infty$, where

$$P(t) = \sum_{j=0}^{n} (-1)^j \int_{\xi_j}^{\xi_{j+1}} e^{ts}\,ds, \qquad 0 < \xi_1 < \cdots < \xi_n < 1,$$

and P_0 equioscillates $n+1$ times on $[0, 1]$. $P_0(t)$ is unique and an optimal subspace for $\mathcal{K}_0$ is given by

$$X_n^0 = \left\{ \sum_{j=1}^{n} \alpha_j e^{t\xi_j} \colon (\alpha_1, \ldots, \alpha_n) \in R^n \right\}.$$

EXAMPLE 4. Given any nonnegative real numbers $t_1, \ldots, t_m$ we define the polynomial

$$q_{2m}(x) = \prod_{j=1}^{m} \left(x^2 - t_j^2\right)$$

and the set

$$\mathfrak{D} = \left\{ f \colon f \in W^{2m}[0, 1], f^{(2k)}(1) = f^{(2k)}(0) = 0, \right.$$
$$\left. k = 0, 1, \ldots, m-1, \|q_{2m}(D)f\|_\infty \leqslant 1 \right\}.$$

We may express $\mathfrak{D}$ as a $\mathcal{K}_0$ where $K(t, s)$ is the Green's function for the differential operator

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

(7.9)
$$f^{(2k)}(1) = f^{(2k)}(0) = 0, \qquad k = 0, 1, \ldots, m-1.$$

It may be verified that

$$K(t, s) = 2 \sum_{k=1}^{\infty} \frac{\sin k\pi t \sin k\pi s}{q_{2m}(ik\pi)}. \tag{7.10}$$

Furthermore, it is known that $(-1)^m K(t, s)$ is totally positive on $[0, 1] \times [0, 1]$. To see that this is the case we observe that the differential operator (7.9) is the product of the second order differential operators

(7.11)
$$f''(x) - \alpha^2 f(x) = 0, \qquad \alpha = t_1, \ldots, t_m,$$
$$f(1) = f(0) = 0.$$

Thus if we call the Green's function for (7.11) $K(t, s; \alpha)$ and define

$$K_j(t, s) = \int_0^1 K(t, y; t_j) K_{j-1}(y, s)\,dy$$

where $K_1(t, s) = K(t, s; t_1)$, then $K(t, s) = K_m(t, s)$. Hence it is sufficient to prove $-K(t,s;\alpha)$ is totally positive for any $\alpha > 0$. However a direct calculation shows that

$$K(t,s;\alpha) = \begin{cases} -\alpha^{-1}\sinh \alpha t \sinh \alpha(1-s), & t \leqslant s, \\ -\alpha^{-1}\sinh \alpha(1-t)\sinh \alpha s, & s \leqslant t. \end{cases}$$

Since

$$\frac{d}{dx}\left[\frac{\sinh \alpha x}{\sinh \alpha(1-x)}\right] = \frac{a}{\left[\sinh \alpha(1-x)\right]^2} > 0$$

for $0 < x < 1$ we may conclude from Corollary 3.1, p. 112, of [2] that $-K(t,s;\alpha)$ is totally positive.

We define

$$P_0(t) = \sum_{j=0}^{n}(-1)^j\int_{j/(n+1)}^{(j+1)/(n+1)} K(t,s)\,ds.$$

Then using (7.10) we obtain

$$P_0(t) = \frac{4}{\pi}\sum_{l=0}^{\infty}\frac{\sin(2l+1)(n+1)\pi t}{(2l+1)q_{2m}(i(2l+1)(n+1)\pi)}.$$

Clearly,

$$P_0(t+1/(n+1)) = -P_0(t), \qquad P_0(1/(n+1)-t) = P_0(t),$$
$$P_0(0) = 0.$$

Hence

$$P_0(i/(n+1)) = 0, \qquad i = 0, 1, \ldots, n+1,$$
$$P_0'((2i-1)/2(n+1)) = 0, \qquad i = 1, \ldots, n+1.$$

It is also known that P_0' has no further zeros in (0, 1) and $(-1)^m P_0(t) > 0$, $0 < t < 1/(n+1)$, cf. [8]. Thus we conclude from Theorem 7.2 that

$$d_n(\mathfrak{D}) = \frac{4(-1)^m}{\pi}\sum_{l=0}^{\infty}\frac{(-1)^l}{(2l+1)q_{2m}(i(2l+1)(n+1)\pi)}$$

and an optimal subspace for $\mathfrak{D}$ is given by

$$X_n^0 = \left\{\sum_{j=1}^{n}\alpha_j K\left(t, \frac{j}{n+1}\right) : (\alpha_1, \ldots, \alpha_n) \in R^n\right\}.$$

When $q_{2m}(x) = x^{2m}$,

$$d_n(\mathfrak{D}) = \frac{4}{\pi^{2m+1}(n+1)^{2m}}\sum_{l=0}^{\infty}\frac{(-1)^l}{(2l+1)^{2m+1}}.$$

Furthermore in this case

$$P_0(t) = \frac{1}{(n+1)^{2m}} E_{2m}((n+1)t)$$

where E_{2m} is the $2m$th Euler perfect spline function, normalized so that $|E_{2m}^{(2m)}(x)| = 1$.

As a comparison, let us compute the L^2-width of the set $\mathfrak{D}$. Since (7.9) is a selfadjoint operator with eigenvalues $\lambda_n = q_{2m}^{-1}(in\pi)$, $n = 1, 2, \ldots,$ and corresponding eigenfunction $\sin \pi n x$, $n = 1, 2, \ldots,$ we may use the obvious L^2-analogue of Theorem 2.3 and conclude that

$$d_n(\mathfrak{D}; L^2[0, 1]) = \frac{(-1)^m}{q_{2m}(i(n+1)\pi)}.$$

Hence we have

$$\lim_{n\to\infty} \frac{d_n(\mathfrak{D}; L^\infty[0, 1])}{d_n(\mathfrak{D}; L^2[0, 1])} = \frac{4}{\pi} \sum_{l=0}^{\infty} \frac{(-1)^l}{(2l+1)^{2m+1}}.$$

Finally, we remark that using the results in [4] and [6] we may obtain the n-width of other differential operators whose Green's function is totally positive.

Following the completion of this article the paper by C. K. Chui and P. W. Smith [1] appeared. We shall comment below on the relationship of our results to those obtained by Chui and Smith.

Given any real-valued functions $\lambda_1(x), \ldots, \lambda_r(x)$ defined on the interval $[-1, 1]$ we define the rth order differential operator

$$Lf(x) = \prod_{j=1}^{r} (D - \lambda_j(x)) f(x) \tag{7.12}$$

and consider the class

$$\mathfrak{B}(L) = \{f : f \in W^r[-1, 1], \|Lf\|_\infty \leq 1\} \tag{7.13}$$

where in this case $\|g\|_\infty = \text{ess sup}\{|g(x)| : |x| \leq 1\}$.

In [1], the n-width of $\mathfrak{B}(L)$ was found when the functions $\lambda_1(x), \ldots, \lambda_r(x)$ are constants

$$\lambda_j(x) = b_j, \quad j = 1, \ldots, r, x \in [-1, 1],$$

and $\prod_{j=1}^r b_j = 0$.

In the general case, we may express $\mathfrak{B}(L)$ as $\mathcal{K}_r$ by choosing $k_1(t), \ldots, k_r(t)$ to span the null space of L and $K(x, y)$ to be the Green's function of the initial value problem determined by L. It is well known that our hypotheses (7.1)–(7.3) are satisfied in this case, [2]. Thus Theorem 7.2 applies to the set $\mathfrak{B}(L)$.

In particular, given any *real* polynomial $p_r(x)$ of exact degree r such that the set of zeros of p_r is contained in the vertical strip

$$\{z: |\mathrm{Im} z| \leqslant \pi/2\}$$

then $p_r(D)$ may be factored in the form (7.12) and thus Theorem 7.2 applies.

The reason that the approach in [1] for the case that $p_r(x)$ has real zeros $b_1, \ldots, b_r$ does not apply when $\prod_{j=1}^r b_j \neq 0$ results from the choice of the nonlinear approximation problem studied in [1].

Using their notation we define $v \in W^r[-1, 1]$ **by requiring** $v(t) = 0$ for $t < 0$ and $Lv(t) = (t)_+^0$. The best approximation problem of [1] is

$$\inf_g \|g\|_\infty, \tag{7.14}$$

where $g(t) = v(t+1) + \sum_{i=1}^m \alpha_i v(t - t_i) + \sum_{i=1}^r \beta_i w_i(t)$, $\{w_1, \ldots, w_r\}$ is a basis of the null space of the operator L and the minimum is taken over $-1 \leqslant t_1 < \cdots < t_m \leqslant 1$, $\alpha_i = \pm 1$, $(\beta_1, \ldots, \beta_r) \in R^r$. The function g is in $\mathcal{K}_r$ if and only if $\prod_{j=1}^r b_j = 0$. Hence when $\prod_{j=1}^n b_j \neq 0$ the value of (7.14) does not give the width of $\mathfrak{B}(L)$.

8. Restricted moment spaces. The discussion in §7 corresponds, in a sense, to the results in §§4 and 5 when $N = M = \infty$. In this section we briefly consider some examples of the cases $N < \infty$, $M = \infty$ and $N = \infty$, $M < \infty$. We provide no proofs here since the previous methods used in §7, combined with the results in §4, apply readily.

CASE 1. $N < \infty$, $M = \infty$.

Let $u_1(t), \ldots, u_N(t)$ be a set of continuous functions which form a Markov system on [0, 1], that is,

$$\det_{k,j} |u_{i_k}(t_j)| > 0$$

for all $0 \leqslant t_1 < \cdots < t_k \leqslant 1$, $1 \leqslant i_1 < \cdots < i_k \leqslant N$, $k \leqslant N$.

The restricted moment space generated by $u_1, \ldots, u_N$ is defined as follows. Let $\mathcal{U}$ be the mapping from $L^\infty[0, 1]$ into l_∞^N defined by setting

$$(\mathcal{U}h)_i = \int_0^1 u_i(t) h(t)\, dt, \qquad i = 1, \ldots, N.$$

Then

$$\mathfrak{M}_N = \{\mathcal{U}h: \|h\|_\infty \leqslant 1\}$$

is the restricted moment space generated by $u_1(t), \ldots, u_N(t)$.

For every $x = (x_1, \ldots, x_n) \in R^n$, $0 = x_0 < x_1 < \cdots < x_n < x_{n+1} = 1$ we define

$$h_x(t) = (-1)^j, \qquad x_j < t < x_{j+1}, j = 0, 1, \ldots, n.$$

$\mathfrak{M}_N$ is a subset of R^N and its n-widths is given by

THEOREM 8.1. *For any integer n, $0 < n < N$, there exists a $\xi = (\xi_1, \ldots, \xi_n)$, $0 < \xi_1 < \cdots < \xi_n < 1$, such that the N-vector, $\mathfrak{U}h_\xi$, equioscillates $n + 1$ times. The n-width of $\mathfrak{M}_N$ is given by*

$$d_n(\mathfrak{M}_N; l_N^\infty) = \|Uh_\xi\|_\infty$$

and the vectors $(u_1(\xi_i), \ldots, u_N(\xi_i))$, $i = 1, \ldots, n$, span an optimal subspace for the n-width of $\mathfrak{M}_N$.

Note that when $N = n + 1$ then we also have

$$d_n(\mathfrak{M}_{n+1}; l_{n+1}^\infty) = \min_{a_1, \ldots, a_{n+1}} \left\{ \int_0^1 \left| \sum_{j=1}^{n+1} a_j u_j(t) \right| dt: \sum_{j=1}^{n+1} a_j(-1)^j = 1 \right\}.$$

CASE 2. $N = \infty$, $M < \infty$.

For every $x = (x_1, \ldots, x_M) \in R^M$ we define $u(t; x) = \sum_{j=1}^M x_j u_j(t)$. Again, we assume that $\{u_1(t), \ldots, u_M(t)\}$ is a Markov system on [0, 1]. Let

$$\mathfrak{M}_M = \{u(t; x): \|x\|_\infty \leq 1, x \in R^M\}.$$

$\mathfrak{M}_M$ is a subset of $L^\infty[0, 1]$ and its n-width is given by

THEOREM 8.2. *For any integer n, $0 < n < M$, there exists a vector $x^0 = (x_1^0, \ldots, x_M^0)$ which alternates on some n components $\mathbf{j}^0 = (j_1^0, \ldots, j_n^0)$, $1 \leq j_1^0 < \cdots < j_n^0 \leq M$, and $u_0(t) = u(t; x^0)$ equioscillates $n + 1$ times on [0, 1]. The n-width of $\mathfrak{M}_M$ is given by $d_n(\mathfrak{M}_M; L^\infty[0, 1]) = \|u_0\|_\infty$ and the functions $u_{j_1^0}(t), \ldots, u_{j_n^0}(t)$ span an optimal subspace for the n-width of $\mathfrak{M}_M$.*

Note that when $M = n + 1$,

$$d_n(\mathfrak{M}_{n+1}; L^\infty[0, 1]) = \min\{\|u(t; x)\|_\infty: \|x\|_\infty = 1\}.$$

Acknowledgement. We gratefully acknowledge several helpful conversations with Alan J. Hoffman concerning this paper.

REFERENCES

1. Charles K. Chui and Philip W. Smith, *Some nonlinear spline approximation problems related to N-widths*, J. Approximation Theory **13** (1975), 421–430. MR **51** #6352.

2. S. Karlin, *Total positivity*, Vol. 1, Stanford Univ. Press, Stanford, Calif., 1968. MR **37** #5667.

3. S. Karlin and A. Pinkus, *Oscillation properties of generalized characteristic polynomials for totally positive and positive definite matrices*, Linear Algebra and Appl. **8** (1974), 281–312. MR **49** #7284.

4. ______, *Interpolation by splines with mixed boundary conditions*, Studies in Splines and Approximation Theory (S. Karlin, C. A. Micchelli, A. Pinkus, I. J. Schoenberg), Academic Press, New York, 1976.

5. A. Kolmogoroff, *Über die Beste Annaherung von Funktionen einer Gegebenen Functionenklasse*, Ann. of Math. (2) **37** (1936), 107–110.

6. Avrahan A. Melkman, *Interpolation by splines satisfying mixed boundary conditions*, Israel J. Math. **19** (1974), 369–381. MR **51** #8683.

7. Charles A. Micchelli, *Best L^1-approximation by weak Chebyshev systems and uniqueness of perfect splines*, J. Approximation Theory **19** (1977), 1–14.

8. ______, *Cardinal $\mathcal{L}$-splines*, Studies in Splines and Approximation Theory (S. Karlin, C. A. Micchelli, A. Pinkus, I. J. Schoenberg), Academic Press, New York, 1976.

9. I. Singer, *Best approximation in normed linear spaces by elements of linear subspaces*, Grundlehren math. Wiss., Band 171, Publ. House Acad. S. R. Romania, Bucharest; Springer-Verlag, Berlin and New York, 1970. MR **42** #4937.

10. V. M. Tihomirov, *Best methods of approximation and interpolation of differentiable functions in the space $C_{[-1,1]}$*, Mat. Sb. (N.S.) **80 (122)** (1969), 290–304 = Math. USSR **9** (1969), 255–289. MR **41** #703.

MATHEMATICAL SCIENCES DEPARTMENT, IBM THOMAS J. WATSON RESEARCH CENTER, YORKTOWN HEIGHTS, NEW YORK (Current address of C. A. Micchelli)

Current address (Allan Pinkus): Mathematics Research Center, University of Wisconsin, Madison, Wisconsin 53706