THE INVERSE OF A
TOTALLY POSITIVE BI-INFINITE BAND MATRIX[1]

BY

CARL DE BOOR

ABSTRACT. It is shown that a bounded bi-infinite banded totally positive matrix A is boundedly invertible iff there is one and only one bounded sequence mapped by A to the sequence $((-)^i)$. The argument shows that such a matrix has a main diagonal, i.e., the inverse of A is the bounded pointwise limit of inverses of finite sections of A principal with respect to a particular diagonal; hence $((-)^{i+j}A^{-1}(i, j))$ or its inverse is again totally positive.

0. Introduction. This paper is a further step in a continuing effort to understand certain linear spline approximation schemes. Convergence of such processes is intimately tied to their stability, i.e., to their boundedness, as maps on C, say. Use of the B-spline basis shows this question to be equivalent to bounding the inverse of certain totally positive band matrices. The calculation of bounds on the inverse of a given matrix is in general a difficult task. It is hoped that the present investigation into the consequences of bandedness and total positivity for the structure of the inverse may ultimately prove helpful in obtaining such bounds.

The results in this paper were obtained in the study of a conjecture due to C. A. Micchelli [7]. In connection with his work on the specific approximation scheme of interpolation at a (strictly increasing) point sequence τ by elements of $S_{m,t}$, i.e., by splines of some order m with some knot sequence $\mathbf{T} = (t_i)$, Micchelli became convinced that every bounded function has one and only one bounded spline interpolant iff the particular function which satisfies $f(\tau_i) = (-)^i$, all i, has a bounded spline interpolant in $S_{m,t}$. If $(N_i) = (N_{i,m,t})$ denotes the corresponding B-spline basis for $S_{m,t}$, then Micchelli's conjecture can be phrased thus: The matrix $A := (N_j(\tau_i))$ is boundedly invertible iff the linear system

$$(1) \qquad\qquad Ax = ((-)^i)$$

has a bounded solution. Micchelli points out that, for a finite A, this conjecture is indeed true and can be established using the total positivity of A.

Whether $A = (N_j(\tau_i))$ is finite or not, it is not difficult to see that A fails to be invertible unless A is m-banded, i.e., unless τ and $\mathbf{t}$ so harmonize that at most $m + 1$ consecutive bands of A are not identically zero. It is then a small step to the conjecture

"A totally positive m-banded matrix A is boundedly invertible iff (1) has a bounded solution"

whose truth Micchelli demonstrated to me for the case that A is a (bi-infinite) Toeplitz matrix and which was proved recently in Cavaretta, Dahmen, Micchelli and Smith [8] for the case that A is a block banded Toeplitz matrix.

As it turns out, this conjecture is incorrect; it fails unless one assumes, more strongly, that (1) has a *unique* bounded solution. This uniqueness plays a crucial role in the proof of the (corrected) conjecture given below. The proof is first given for strictly m-banded matrices and is then extended to a general m-banded matrix by a limit argument, using a 'smoothing' result from [4].

In outline, our argument is as follows: We show that the nullspace

$$\mathfrak{N} := \mathfrak{N}_A := \{ f \in \mathbf{R}^{\mathbf{Z}} : Af = 0 \}$$

of a strictly m-banded matrix A is Haar' in the sense that, for every $f \in \mathfrak{N} \setminus 0$, f' has less than $m = \dim \mathfrak{N}$ weak sign changes. Here, f' is obtained from f by changing the sign of every other entry, $f'(i) := (-)^i f(i)$, all i. This makes it possible to interpolate the bounded solution x to (1) at any m-set $I := \{i_1, \ldots, i_m\}$ by some $y_I \in \mathfrak{N}$. The next (and hardest) step consists in showing that, for $k = m^+ := \dim \mathfrak{N}^+$ with

$$\mathfrak{N}^+ := \left\{ f \in \mathfrak{N} : \lim_{i \to \infty} |f(i)| < \infty \right\},$$

y_I lies between x and 0 on the interval $I^{(k)} :=]i_k, i_{k+1}[$, if we assume that $i_1 < \cdots < i_m$. This implies that $x_I := x - y_I$ satisfies $\|x_{I|I^{(k)}}\| \leq \|x\|$. We use this fact as follows. If J is any integer interval, and $A_{J, J+k}$ is the corresponding section of A having the kth band as its main diagonal, then

$$\left((-)^i \right)_{|J} = (Ax_I)_{|J} = A_{J, J+k}(x_{I|J+k})$$

provided we choose $I = [J \cup (J + m)] \setminus (J + k)$. But then, because of the total positivity of A, $\|(A_{J, J+k})^{-1}\| = \|x_{I|J+k}\|$ and this is bounded by $\|x\|$ since $I^{(k)} = J + k$ in this case. This uniform boundedness of the inverses of all sections which are principal with respect to the kth band is sufficient for the bounded invertibility of A itself.

In this way, we show not only that existence and uniqueness of a bounded solution for (1) implies bounded invertibility of A, but gain structure information about the inverse: The inverse is the pointwise bounded limit of the inverses of finite sections principal with respect to one particular band. In the terms of [1, 2], A has a main diagonal. This, in turn, permits the conclusion that the inverse of a totally positive band matrix is checkerboard, a statement conjectured in [5, p. 319].

1. Preliminaries. In this section, we list certain notational conventions for easy reference.

We use lower case letters to denote elements of $\mathbf{R}^I$, i.e., real functions on some integer set I, with $f(i)$ the value at i of the function (or sequence) f. If f never vanishes, then $S(f)$ denotes the number of *sign changes* in f, i.e.,

$$S(f) := | \{i \in I : f(i)f(s(i)) < 0\} | ,$$

with $s(i) := \min\{j \in I: j > i\}$ the successor to i if we think of I as an ordered sequence. Here, $|J|$ denotes the cardinality of the set J. If f vanishes somewhere, then it is customary to distinguish between *strong* and *weak* sign changes. These are given by

$$S^-(f) := \inf\{S(v): v \in \text{sign } f\}, \qquad S^+(f) := \sup\{S(v): v \in \text{sign } f\}$$

respectively, with sign $f := \{v \in \{-1, 1\}^I: v(i)f(i) = |f(i)|, \text{ all } i\}$. In the sequel, an unqualified "sign change" will always mean "weak sign change". It is convenient to supplement the definition of S^- and S^+ by setting $S^-(f) := -1$, $S^+(f) := |I|$ if $f = 0$. It is then easy to check that

$$S^-(f) + S^+(f') = |I| - 1, \quad \text{for all } f \in \mathbf{R}^I,$$

with $I = \{i_1, \ldots, i_n\}$, $i_1 < \cdots < i_n$, and $f'(i_s) := (-)^s f(i_s)$, $s = 1, \ldots, n$. We also employ the prime to indicate a signature change in every other entry in case $I = \mathbf{Z}$. To be definite, we set

$$f'(i) := (-)^i f(i), \quad \text{all } i, \text{ all } f \in \mathbf{R}^{\mathbf{Z}}.$$

In particular, $1'$ denotes the bi-infinite sequence given by $1'(i) := (-)^i$, all i.

If J is a subset of I, then $f_{|J}$ denotes the restriction of $f \in \mathbf{R}^I$ to J. In this connection, $\smallsetminus J := I \setminus J$ denotes the complement of J in I. We write $\smallsetminus j$ instead of $\setminus\{j\}$. Further, $J + k := \{j + k: j \in J\}$, and $[i, j] := \{k \in \mathbf{Z}: i \leqslant k \leqslant j\}$. More generally, $[J]$ denotes the smallest integer interval containing J.

Correspondingly (though not completely in the same way), we denote by $A_{K,L}$ the restriction of the "matrix" $A \in \mathbf{R}^{I \times J}$ to the subset $K \times L$ of $I \times J$. If $K = \{k_1, \ldots, k_p\}$ and $L = \{l_1, \ldots, l_p\}$, with $k_1 < \cdots < k_p$, $l_1 < \cdots < l_p$, then

$$\det A_{K,L} := \det\big(A(k_i, l_j)\big)_{i,j=1}^p.$$

All norms are sup-norms. Explicitly, $\|f\| := \sup_{i \in I}|f(i)|$, all $f \in \mathbf{R}^I$ and

$$\|A\| := \sup\{\|Af\|/\|f\|: f \in \mathbf{R}^J, \|f\| < \infty\} = \sup_{i \in I} \sum_{j \in J} |A(i, j)|,$$

$$\text{for } A \in \mathbf{R}^{I \times J}.$$

Further, $l_\infty := l_\infty(\mathbf{Z}) := \{f \in \mathbf{R}^{\mathbf{Z}}: \|f\| < \infty\}$. We call $A \in \mathbf{R}^{\mathbf{Z}}$ *bounded* if A maps l_∞ to itself, or, equivalently, if $\|A\| < \infty$. If need be, we distinguish between the bounded matrix A and the linear map induced by it on l_∞ by calling the latter $A_{|l_\infty}$.

The matrix $A \in \mathbf{R}^{I \times J}$ is *totally positive* ($=:$ tp) if

$$\det A_{K,L} \geqslant 0 \quad \text{for all } K \times L \subseteq I \times J \text{ with } |K| = |L|.$$

See Karlin's comprehensive book [6] for details.

Finally, if $i_1, \ldots, i_p$ and $j_1, \ldots, j_p$ are sequences in I and J, respectively, and $A \in \mathbf{R}^{I \times J}$, then we use the customary abbreviations

$$A\begin{bmatrix} i_1, \ldots, i_p \\ j_1, \ldots, j_p \end{bmatrix} := \big(A(i_s, j_t)\big)_{s,t=1}^p$$

and

$$A\begin{pmatrix} i_1, \ldots, i_p \\ j_1, \ldots, j_p \end{pmatrix} := \det\left(A(i_s, j_t)\right)_{s,t=1}^p.$$

2. The nullspace of a strictly m-banded bi-infinite matrix.

The rth *diagonal* or *band* of a matrix A is, by definition, the sequence $(A(i, i + r))$. As in [1], we call a matrix A *m-banded* if all nonzero entries of A can be found in at most $m + 1$ consecutive bands. Explicitly, the matrix A is m-banded if, for some l, $A(i + l, j) \neq 0$ implies $i \leq j \leq i + m$. Unless otherwise indicated (e.g., by context), we will always assume that $l = 0$. For a bi-infinite matrix A, this is merely a normalization achieved by considering $E^{-l}A$ instead of A, with E the shift,

$$(Ef)(i) := f(i + 1), \quad \text{all } i, \text{ all } f \in \mathbf{R}^{\mathbf{Z}},$$

an invertible operator which preserves more or less all interesting structures in $\mathbf{R}^{\mathbf{Z}}$.

The m-banded matrix A is called *strictly m-banded* if

$$A(i, i)A(i, i + m) \neq 0, \quad \text{all } i,$$

i.e., the first and last nontrivial band is never zero. In case of a bi-infinite matrix A. this nontrivial assumption insures that, for every m-interval $I = \{i + 1, \ldots, i + m\}$, every $a \in \mathbf{R}^I$ gives rise to one and only one sequence f with $Af = 0$ and $f_{|I} = a$. To put it differently, with $\mathfrak{N} := \mathfrak{N}_A := \{f \in \mathbf{R}^{\mathbf{Z}}: Af = 0\}$ denoting the kernel or *nullspace* of A, strict m-bandedness insures that, for every $I := [i + 1, i + m]$, the map $\mathfrak{N} \to \mathbf{R}^m: f \mapsto f_{|I}$ is 1-1 and onto.

We now prove this statement to be true for *every* m-set I in case A is also tp. We begin with the following

LEMMA 2.1. *If A is strictly m-banded and tp, and $i_1 < \cdots < i_p, j_1 < \cdots < j_p$, then*

$$A\begin{pmatrix} i_1, \ldots, i_p \\ j_1, \ldots, j_p \end{pmatrix} > 0 \quad \text{iff} \quad i_r \leq j_r \leq i_r + m, \text{ all } r.$$

PROOF. Proof of the 'if' part is by induction on p, it being true for $p = 0$ with the customary convention that $A\binom{\varnothing}{\varnothing} = 1$. If $i_1 = j_1$, then, A being m-banded, we have

$$A\begin{pmatrix} i_1, \ldots, i_p \\ j_1, \ldots, j_p \end{pmatrix} = A\begin{pmatrix} i_1 \\ j_1 \end{pmatrix} A\begin{pmatrix} i_2, \ldots, i_p \\ j_2, \ldots, j_p \end{pmatrix}$$

and this is strictly positive, the first factor by the strict m-bandedness of A and the second by induction. The corresponding argument applies when $j_1 = i_1 + m$.

Otherwise $i_1 < j_1 < i_1 + m$. But then, A being m-banded and tp, we have

$$0 \leqslant A \begin{pmatrix} j_1 - m, & i_1, \ldots, i_p \\ i_1, & j_1, \ldots, j_p \end{pmatrix}$$

$$= A \begin{pmatrix} j_1 - m \\ i_1 \end{pmatrix} A \begin{pmatrix} i_1, \ldots, i_p \\ j_1, \ldots, j_p \end{pmatrix}$$

$$- A \begin{pmatrix} i_1 \\ i_1 \end{pmatrix} A \begin{pmatrix} j_1 - m \\ j_1 \end{pmatrix} A \begin{pmatrix} i_2, \ldots, i_p \\ j_2, \ldots, j_p \end{pmatrix}$$

with the subtrahend strictly positive by the strict m-bandedness of A and the induction hypothesis, and this implies that the factors of the minuend must be strictly positive, too.

As to the 'only if' part,

$$A \begin{bmatrix} i_1, \ldots, i_p \\ j_1, \ldots, j_p \end{bmatrix}$$

has zeros in columns (rows) $1, \ldots, r$ and rows (columns) $r, \ldots, p$ in case $i_r > j_r$ ($j_r > i_r + m$), hence is then singular. $\square$

COROLLARY. *If A is strictly m-banded and tp, and J, K are integer intervals with $K = [J \cup (J + m)]$, then $x_{|K} \neq 0$ and $(Ax)_{|J} = 0$ imply $S^-(x_{|K}) \geqslant |J|$.*

PROOF (an adaptation of the argument for Theorem 5.1.2 in [6, p. 219f]). The assumption that $p := S^-(x_{|K}) < |J|$ leads to a contradiction as follows.

Let $K_0, \ldots, K_p$ be a corresponding partition of K, i.e., (without loss of generality)

$$(2) \qquad 0 \neq (-)^i x_{|K_i} \geqslant 0, \qquad i = 0, \ldots, p.$$

Then $0 = \sum_{i=0}^p (-)^i v_i$ with

$$(3) \qquad v_i := \sum_{k \in K_i} |x(k)| A(\cdot, k)_{|J}, \qquad i = 0, \ldots, p,$$

showing that $(v_i)_0^p$ is linearly dependent. This implies, with $V := [v_0 | \cdots | v_p]$, that, for any $i_0 < \cdots < i_p$ in J,

$$0 = V \begin{pmatrix} i_0, \ldots, i_p \\ 0, \ldots, p \end{pmatrix} = \sum_{k_0 \in K_0} \cdots \sum_{k_p \in K_p} |x(k_0)| \cdots |x(k_p)| A \begin{pmatrix} i_0, \ldots, i_p \\ k_0, \ldots, k_p \end{pmatrix}$$

and all summands on the right are nonnegative by the tp of A. On the other hand, we can, by (2), choose $k_i \in K_i$, all i, so that $|x(k_0)| \cdots |x(k_p)| \neq 0$ and will therefore have reached a contradiction as soon as we exhibit a corresponding choice for $i_0 < \cdots < i_p$ in J for which

$$(4) \qquad A \begin{pmatrix} i_0, \ldots, i_p \\ k_0, \ldots, k_p \end{pmatrix} > 0.$$

This we can do as follows. Define $(i_r)_{-1}^{p+1}$ by

$$[i_{-1}, i_{p+1}] := J, \quad i_r := \max\{i_{-1} + r, k_r - m\}, \quad r = 0,\ldots,p.$$

Then $i_0 < \cdots < i_p$ since both sequences $(i_{-1} + r)_r$ and $(k_r - m)_r$ are strictly increasing. Also, $i_0,\ldots,i_p \in J$ since trivially $i_{-1} \leqslant i_r$, while

$$i_r = \max\{i_{-1} + r, k_r - m\} \leqslant i_{p+1}$$

since $r \leqslant p < |J|$ and $k_r \in K = [i_{-1}, i_{p+1} + m]$. Finally, $k_r - m \leqslant i_r \leqslant k_r$, since $i_{-1} \leqslant k_0 < \cdots < k_r$ implies that $i_{-1} + r \leqslant k_r$; hence

$$k_r - m \leqslant \max\{i_{-1} + r, k_r - m\} \leqslant k_r.$$

The lemma now gives (4) and thereby the desired contradiction. $\square$

REMARK. We have proved here a particular instance of the statement: "If B is tp and of full rank, then $x \neq 0$ implies $S^-(x) \geqslant S^+(Bx)$", provided we define $S^+(0) :=$ length of 0, as we did earlier.

We conclude that if $y \in \mathfrak{R}$ and $y_{|K} \neq 0$, then

$$S^+(y'_{|K}) = |K| - 1 - S^-(y_{|K}) \leqslant |K| - 1 - |J| = m - 1;$$

hence

PROPOSITION 2.5. *If A is strictly m-banded and tp, then $y \in \mathfrak{R} \setminus 0$ implies $S^+(y') \leqslant m - 1$.*

Since $S^+(f) \geqslant |\{i: f(i) = 0\}|$, this shows that $\mathfrak{R}$ is then a Haar space, i.e., $\mathfrak{R} \to R^m: y \mapsto y_{|I}$ is 1-1 and onto whenever $|I| = m$. Further, this shows that any nontrivial $y' \in \mathfrak{R}'$ with $m - 1$ zeros changes sign across each of these zeros and nowhere else, i.e.,

COROLLARY. *If A is strictly m-banded and tp, and $y \in \mathfrak{R} \setminus 0$ vanishes at $i_1 < \cdots < i_{m-1}$, then (with $i_0 := -\infty$, $i_m := \infty$),*

$$(6) \qquad y'(i_1 - 1)(-)^r y'(i) > 0 \quad \text{for } i_r < i < i_{r+1}, r = 0,\ldots,m-1.$$

We make use below of the two subspaces $\mathfrak{R}^+$ and $\mathfrak{R}^-$ of $\mathfrak{R}$. These are defined by

$$\mathfrak{R}^* := \left\{ f \in \mathfrak{R}: \varlimsup_{i \to *\infty} |f(i)| < \infty \right\}, \qquad * = +, -.$$

Their intersection consists of all bounded solutions to the homogeneous problem $Ay = 0$. The intersection is therefore trivial iff (0.1) has at most one bounded solution. See [2] for conditions on A equivalent to having $\mathfrak{R} = \mathfrak{R}^+ \oplus \mathfrak{R}^-$.

3. The algorithms L and R. Let x be a bounded sequence satisfying $Ax = 1'$. In this section, we investigate y_I, the sequence in $\mathfrak{R} = \mathfrak{R}_A$ which matches x at the m-set $I = \{i_1,\ldots,i_m\}$ with $i_1 < \cdots < i_m$. For $k = 0,\ldots,m$, the *interval k* of such an m-set is, by definition, the integer interval $I^{(k)} :=]i_k, i_{k+1}[$, with $i_0 := -\infty$, $i_{m+1} := \infty$.

Our ultimate goal is to show that, for $k = m^+ := \dim \mathfrak{R}^+$, y'_I on $I^{(k)}$ lies between x' and 0. The proof of this fact involves certain manipulations which are conveniently described in terms of two algorithms, given below. For their analysis, the following fact is useful.

PROPOSITION 3.1. *Let A be strictly m-banded and tp, and let $z = x - y_I$ for some m-set I. Then*

(2) $$z(i) \neq 0 \quad \text{for all } i \notin I.$$

More precisely,

(3) $$(-)^k z'(i) > 0 \quad \text{for } i \in I^{(k)}, k = 0, \ldots, m.$$

PROOF. Let J be an integer interval containing I in its 'interior'. Since A is m-banded, we have

$$1'_{|J} = (Az)_{|J} = A_{J,L}(z_{|L}), \quad \text{with } L := [J \cup (J + m)] \setminus I,$$

and, since A is strictly banded, $A_{J,L}$ is invertible, by Lemma 2.1. Thus $(A_{J,L})^{-1}$ is checkerboard, and $z_{|L} = (A_{J,L})^{-1}(1'_{|J})$. It follows that $z_{|L}$ changes sign strongly at every slot, with $z(j)(-)^j > 0$ for all $j \in J$ and to the left of I. Since J is essentially arbitrary, we conclude that $z' > 0$ near $-\infty$ and that $z_{|\setminus I}$ strongly changes sign at every slot. $\square$

COROLLARY. *If the m-set $J = \{j_1, \ldots, j_m\}$ is obtained from the m-set $I = \{i_1, \ldots, i_m\}$ by moving the leftmost k points to the left, i.e., $j_r \leqslant i_r$, $r = 1, \ldots, k$, while $j_r = i_r$, $r = k + 1, \ldots, m$, then, on $[i_k, \infty[$, y'_I lies between x' and y'_J, with equality only at the points $i_{k+1}, \ldots, i_m$ (unless $J = I$).*

PROOF. The assertion follows by repeated application of the special case

$$j_r = i_r \quad \text{for } r \neq k, \; i_{k-1} < j_k < i_k.$$
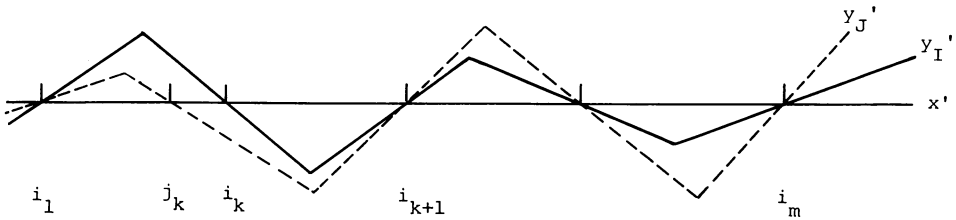
The proof for this special case goes as follows. By (3),

$$(-)^k (y'_J - y'_I)(j_k) = (-)^k (x' - y'_I)(j_k) < 0.$$

On the other hand, y_J and y_I agree at the $m-1$ points of $I \setminus i_k$; hence their difference changes sign strongly across each point of $I \setminus i_k$ and nowhere else, by the corollary to Proposition 2.1. Consequently,

(4a) $$(-)^r (y'_I - y'_J)(i) > 0 \quad \text{for } i_r < i < i_{r+1}, r = k, \ldots, m,$$

while, again by (3),

(4b) $$(-)^r (x' - y'_J)(i) > 0 \quad \text{for } i_r < i < i_{r+1}, r = k, \ldots, m. \quad \square$$

What is to follow is based on the proposition and its corollary and the following attempt at constructing a nontrivial element of $\mathfrak{N}^-$.

By the proposition, $S^+(x') \leqslant m$; hence x' has constant sign near $-\infty$. Let

$$(5) \qquad\qquad \varepsilon := \operatorname{sign} x'(i) \quad \text{for } i \text{ near } -\infty$$

and let $k \in [1, m]$ be even or odd depending on whether ε is 1 or -1. Choose an m-set I so that i_{k+1} lies to the left of all sign changes of x' (if any). Then, on $I^{(k)}$, $\varepsilon y_I'$ lies below $\varepsilon x'$, by (3). We then consider what happens to y_I on $I^{(k)}$ as we move $i_1, \ldots, i_k$ to the left. By the corollary, $\varepsilon y_I'$ must then decrease. There are two possibilities:

(i) No matter how far we move $i_1, \ldots, i_k$ to the left, $\varepsilon y_I'$ remains positive on $I^{(k)}$. Then we obtain as a limit point some $y \in \mathfrak{N}$ which agrees with x at $i_{k+1}, \ldots, i_m$, hence is not just 0, and for which $\varepsilon y'$ lies between $\varepsilon x'$ and 0 on $]-\infty, i_{k+1}[$. But this implies that $y \in \mathfrak{N}^- \backslash 0$ since x is bounded by assumption. A refinement of the limit process actually gives $\dim \mathfrak{N}^- \geqslant m - k$.

(ii) Eventually, $\varepsilon y_I'$ becomes nonpositive somewhere in $I^{(k)}$, hence has (at least) two sign changes there. We would then decrease k by 2 and try again. By the corollary, the two sign changes of y_I' just acquired would not be affected by subsequent moves.

In this way, we either obtain some nontrivial element of $\mathfrak{N}^-$ or else find ourselves once again at (ii) but with $k = 1$ or 2, making further decreases in k impossible. The current y_I' must then have two sign changes for every time we passed through (ii). Since $S^+(y_I') < m$, this limits the number of times we can pass through (ii). In particular, if we start with $k = m$ or $m - 1$, we must eventually reach (i).

This allows us to talk about the smallest k we manage to arrive with at (i) as we vary the initial k and I in the above procedure; call it k_L. Analogously, we define k_R as the largest k we manage to arrive with at (i) as we play the game to the right rather than the left. The extremality of k_L and k_R and the fact that then $\dim \mathfrak{N}^- \geqslant m - k_L$, $\dim \mathfrak{N}^+ \geqslant k_R$ lead to the desired conclusion in ways to be made precise below.

We now give a formal description of the game just played.

ALGORITHM L. *Input.* The integer $k_{\text{in}} \in [0, m]$ and the m-set $I_{\text{in}} = \{i_1, \ldots, i_m\}$ with $i_1 < \cdots < i_m$. Also, recall $\varepsilon := \operatorname{sign} x'(i)$ for i near $-\infty$.

Step 0. $k := k_{\text{in}}$, $I := I_{\text{in}}$.

Step 1. If $(-)^k \varepsilon < 0$, or if x' changes sign to the left of $I^{(k)}$, then EXIT 1.

Step 2. Let $I_r := \{i_1 - r, \ldots, i_k - r, i_{k+1}, \ldots, i_m\}$, $r = 0, 1, 2, \ldots$.

Step 3. If, for all r, $\varepsilon y_{I_r}' > 0$ on $I^{(k)}$, then EXIT 2.

Step 4. Pick an r for which $\varepsilon y_{I_r}' \leqslant 0$ somewhere in $I_r^{(k)}$ and replace I by I_r.

Step 5. If $k < 2$, then EXIT 3.

Step 6. Decrease k by 2 and return to Step 2.

We now analyse the output from this algorithm.

EXIT 1 is a failure exit which allows us to be less careful about the input than we might otherwise have to be. In the applications of the algorithm, it will be obvious that we do not exit via EXIT 1.

EXIT 2 is the most interesting of the three, because of the following

LEMMA L. *If k is as on exit from Algorithm* L *via* EXIT 2, *then*

$$\dim \mathfrak{N}^- \geq m - k.$$

PROOF. For $k = m$, there is nothing to prove, so assume $k < m$. Then we have in hand a sequence $(y_r) := (y_{I_r})$ in $\mathfrak{N}$ for which $0 < \varepsilon y'_{r+1}(i) < \varepsilon y'_r(i) < \varepsilon x'(i)$ for all $i \in]i_k - r, i_{k+1}[$. Now let $I_{r,k} := I_r$ and, for $r > 0, j > k$, consider

$$I_{r,j} := \{i_1 - r, \ldots, i_k - r, i_k, \ldots, i_{j-1}, i_{j+1}, \ldots, i_m\}.$$

Let $y_{r,j} := y_{I_{r,j}}$ be the corresponding interpolants to x from $\mathfrak{N}$; see figure.
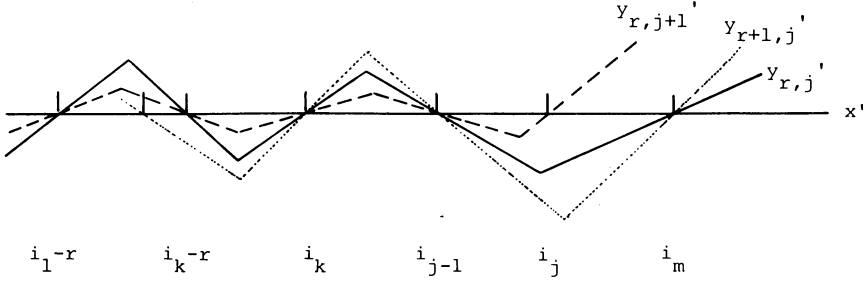
Then, by the corollary to the proposition,

$$0 < \varepsilon y'_{r,k}(i) < \varepsilon y'_{r,k+1}(i) < \cdots < \varepsilon y'_{r,m}(i) < \varepsilon x'(i) \quad \text{for } i_k - r < i < i_k$$

while $y_{r,j}(i) = x(i)$ for $i = i_{j+1}, \ldots, i_m$, and $y_{r,j}(i_j)$ moves away from $x(i_j)$ as r increases. Since $\mathfrak{N}$ is finite dimensional and independent over any m-set (by Proposition 2.1), it now follows that $(y_{r,j})_r$ has limit points in $\mathfrak{N}$ and any such limit point y_j satisfies $0 \leq \varepsilon y'_j < \varepsilon x'$ on $]-\infty, i_k[$, and $y_j(i) = x(i), i = i_{j+1}, \ldots, i_m$. Since x is bounded, this implies that $z_s := y_s - y_{s-1}$ is in $\mathfrak{N}^-$, vanishes at $i_{s+1}, \ldots, i_m$, but does not vanish at i_s, i.e., the matrix $(z_s(i_t))_{s,t=k+1}^m$ is triangular with nonzero diagonal, hence invertible. This shows $(z_s)_{k+1}^m$ to be independent; hence $\dim \mathfrak{N}^- \geq m - k$. □

Finally, if I is obtained via EXIT 3, then we are now certain that y_I has a weak sign change in each of the intervals $0, 2, \ldots, k_{in}$ of I in case $\varepsilon = 1$, or in each of the intervals $1, 3, \ldots, k_{in}$ in case $\varepsilon = -1$. This is so because once a sign change is obtained, in Step 4, in the current interval k, this sign change persists, by the corollary to the proposition. Further, since $\varepsilon x' > 0$ on that interval (we would have exited via EXIT 1 otherwise), it follows that y'_I has two sign changes in each of the intervals $k_{in} - 2j$ with $k_{in} - 2j > 0$, and one in interval 0 if k_{in} is even, for a total of $k_{in} + 1$ sign changes. Since a nontrivial $y \in \mathfrak{N}$ can have at most $m - 1$ sign changes, *an input of $k_{in} = m$ or $m - 1$ (depending on the sign of ε) together with an I_{in} which lies to the left of all sign changes of x' (to avoid* EXIT 1) *is guaranteed to bring us to* EXIT 2. In particular, it makes sense to define

$$k_L := \min\{k: k \text{ obtained as output via EXIT 2 from Algorithm L}\}$$

and then

(6) $$\dim \mathfrak{N}^- \geq m - k_L$$

follows.

Algorithm R is constructed just as Algorithm L, except that all moves are made toward the right rather than the left. For completeness, we give the full description.

ALGORITHM R. *Input.* The integer $k_{in} \in [0, m]$ and the m-set $I_{in} = \{i_1, \ldots, i_m\}$ with $i_1 < \cdots < i_m$. Also, set $\varepsilon := \text{sign } x'(i)$ for i near ∞.

Step 0. $k := k_{in}$, $I := I_{in}$.

Step 1. If $(-)^k \varepsilon < 0$, or if x' changes sign to the right of $I^{(k)}$, then EXIT 1.

Step 2. Let $I_r := \{i_1, \ldots, i_k, i_{k+1} + r, \ldots, i_m + r\}$, $r = 0, 1, 2, \ldots$.

Step 3. If, for all r, $\varepsilon y'_{I_r} > 0$ on $I_r^{(k)}$, then EXIT 2.

Step 4. Pick an r for which $\varepsilon y'_{I_r} \leq 0$ somewhere in $I_r^{(k)}$ and replace I by I_r.

Step 5. If $k > m - 2$, then EXIT 3.

Step 6. Increase k by 2 and return to Step 2.

A discussion very close to that following Algorithm L would establish the following facts.

LEMMA R. *If k is as on exit from Algorithm R via EXIT 2, then* $\dim \mathfrak{R}^+ \geqslant k$. *Further, the number*

$$k_R := \max\{k : k \text{ obtained as output via EXIT 2 from Algorithm R}\}$$

is well defined, and

(7) $$\dim \mathfrak{R}^+ \geqslant k_R$$

follows.

We are now ready to prove the main result of this section.

THEOREM 3.8. *If x is the unique bounded solution to the linear system $Ax = 1'$, with A a strictly m-banded bi-infinite tp matrix, then, for $k = m^+ := \dim \mathfrak{R}^+$ and for any m-set $I = \{i_1, \ldots, i_m\}$ with $i_1 < \cdots < i_m$, y'_I lies between 0 and x' on the interval $I^{(k)} =]i_k, i_{k+1}[$. In particular, then*

(9) $$|x - y_I| \leqslant |x| \quad \text{on } I^{(k)}.$$

PROOF. Let $j_0 < \cdots < j_{S^+(x')}$ be points on which x' alternates in sign, with $x(j_0) \neq 0$. Let I be any m-set with $i_{k_L + r} = j_r$, $r = 0, 1, \ldots, s$, and

$$s := \min\{S^+(x'), m - k_L\}.$$

If $k_L > 1$, then, because of the minimality of k_L, an application of Algorithm L to the input $k_L - 2$, I is bound to end via EXIT 3; hence y'_I (with a possibly changed I) has two sign changes in each of the intervals $k_L - 2, k_L - 4, \ldots$, and one sign change in the interval 0 in case k_L is even, for a total of $k_L - 1$ sign changes. In addition, y'_I alternates in sign on the points $i_{k_L}, \ldots, i_{k_L + s}$ since it agrees there with x' and x' does, giving an additional s sign changes. We conclude that

(10) y'_I has at least $k_L - 1 + s$ sign changes to the left of $i_{k_L + s}$,

and this conclusion holds trivially in case $k_L \leqslant 1$.

We now prove that $s = S^+(x')$. Suppose that $s < S^+(x')$. Then $s = m - k_L$, and we now know that y'_I has $m - 1$ sign changes on $]-\infty, i_{k_L + s}[$, hence does not change

sign on $[i_{k_L+s}, \infty[$, yet matches x' at the points $i_{k_L+s}, \ldots, i_{k_L+S^+(x')}$ on which x' alternates in sign, a contradiction.

We conclude that x' has no sign changes to the right of i_{k_L+s}; hence an application of Algorithm R to the input $k_L + s$, I is bound to terminate via EXIT 2 (because of (10)) with some $k =: k_I$ which is at least as big as $k_L + s$, yet no bigger than k_R by the maximality of k_R. In symbols,

(11) $$k_L + s \leqslant k_I \leqslant k_R$$

and therefore, with (6) and (7),

(12) $$\dim \mathfrak{N}^- + \dim \mathfrak{N}^+ \geqslant (m - k_L) + k_R \geqslant m + s = \dim \mathfrak{N} + S^+(x').$$

This proves that

$$\dim \mathfrak{N}^- \cap \mathfrak{N}^+ \geqslant S^+(x').$$

Thus, if now x is the only bounded solution of $Ax = 1'$, then $\mathfrak{N}^- \cap \mathfrak{N}^+ = \{0\}$; hence then $s = S^+(x') = 0$, and $\dim \mathfrak{N} \geqslant \dim \mathfrak{N}^- + \dim \mathfrak{N}^+$. This shows that there must be equality throughout (12). This implies equality in (7), i.e., $m^+ = k_R$ and equality in (11) (with $s = 0$), i.e.,

$$k_L = k_I = k_R,$$

and thus shows (9) to hold for the original I. But now, since $S^+(x') = 0$, this could have been any m-set I. $\square$

COROLLARY 1. *If A is strictly m-banded and tp, and x is a bounded sequence satisfying $Ax = 1'$, then $S^+(x') \leqslant \dim \mathfrak{N}^- \cap \mathfrak{N}^+$.*

COROLLARY 2. *The conclusions of Theorem 3.8 and Corollary 1 remain valid if, in the hypotheses, $1'$ is replaced by any strictly alternating sequence u.*

Here, we call u *strictly alternating* if $u(i)u(i + 1) < 0$ for all i.

4. The main result.

THEOREM 4.1. *If A is a bounded strictly m-banded bi-infinite tp matrix, and x is the unique bounded sequence mapped by A to $1'$, then A is boundedly invertible on l_∞ and $\|A^{-1}\| = \|x\|$.*

PROOF. For any integer interval J, let $I := [J \cup (J + m)] \setminus (J + k)$. Then I is an m-set, $I = \{i_1, \ldots, i_m\}$ with $i_1 < \cdots < i_m$, say. Let $z_J := x - y_I$, with $y_I \in \mathfrak{N} = \mathfrak{N}_A$ and $x = y_I$ on I. Then we conclude from Theorem 3.8 that

(2) $$|z_J(j)| \leqslant |x(j)|, \quad \text{all } j \in J + k$$

while

$$1'_{|J} = (A(x - y_I))_{|J} = A_J(z_{J|J+k})$$

with $A_J := A_{J,J+k}$. Since A_J is tp, this implies that A_J is invertible (as a general result, though the invertibility of A_J could in the present circumstance be derived directly from Lemma 2.1); hence its inverse is checkerboard and so takes on its norm on the vector $1'_{|J}$, i.e.,

$$\|A_J^{-1}\| = \|A_J^{-1}(1'_{|J})\| = \|z_{J|J+k}\|.$$

Combine this with (2) to get

$$(3) \qquad \|A_J^{-1}\| \le \|x_{|J+k}\| \le \|x\|, \quad \text{for all intervals } J.$$

Since A is bounded and banded, it carries $c_0 := c_0(\mathbf{Z}) := \{ \in \mathbf{R}^{\mathbf{Z}} : \lim_{|i| \to \infty} |f(i)| = 0\}$ to itself, and the bounded invertibility of $A_{|c_0}$ follows now by a standard argument: Let P_J be the truncation projector,

$$(P_J f)(i) := \begin{cases} f(i), & i \in J, \\ 0, & \text{otherwise.} \end{cases}$$

Then $P_J \to 1$ pointwise on c_0; therefore $P_J A P_{J+k} \to A$ pointwise on c_0 as $J \to \mathbf{Z}$. Now $A_J = A_{J,J+k}$ represents the interesting part of $P_J A P_{J+k}$, i.e., the map $P_J A_{|\text{ran} P_{J+k}}$. Therefore, for $u \in c_0$ and $u_J := A_J^{-1} P_J A u \in \text{ran } P_{J+k}$, we have

$$\|u_J - u\| = \|A_J^{-1}(P_J A u - P_J A P_{J+k} u)\|$$
$$\le \|x\| \|A\| \|u - P_{J+k} u\| \underset{J \to \mathbf{Z}}{\to} 0$$

since $\|A_J^{-1}\| \le \|x\|$ and $\|P_J A\| \le \|A\|$. Thus $A_J^{-1} P_J$ converges pointwise on ran $A_{|c_0}$ to a left inverse of $A_{|c_0}$. Further,

$$\|u\| = \lim \|u_J\| \le \overline{\lim} \|A_J^{-1}\| \|Au\| \le \|x\| \|Au\|,$$

i.e., $A_{|c_0}$ is bounded below; hence ran $A_{|c_0}$ is closed. The same argument shows that also $(A_{|c_0})^* = A^T_{|l_1}$ is bounded below; hence ran $A_{|c_0}$ is also dense. We conclude that $A_{|c_0}$ is 1-1 and onto c_0, hence boundedly invertible. Its inverse is therefore again (representable as) a matrix, i.e., $(A_{|c_0})^{-1} = A^{-1}_{|c_0}$ for some matrix A^{-1} whose rows are uniformly in l_1, and $(A_{|c_0})^{-1} = \lim_{J \to \mathbf{Z}} (P_J A P_{J+k})^{-1} P_J$ pointwise on c_0; hence

$$(4) \qquad A^{-1} = \lim_{J \to \mathbf{Z}} A_J^{-1} \quad \text{entrywise.}$$

But then A^{-1} provides the inverse of A on $l_\infty = (c_0)^{**}$, and $\|A^{-1}\| = \|x\|$ since $\|A^{-1}\| \le \|x\|$ from (3) and (4), while $A^{-1}(1') = x$. $\square$

The assumptions of Theorem 4.1 can be weakened in two ways.

As already pointed out earlier, the results of §3 do not depend on having a bounded sequence x satisfying $Ax = 1'$. It is sufficient to consider bounded sequences for which $u := Ax$ is strictly alternating, i.e., $u(i)u(i + 1) < 0$, all i. For the results of Theorem 4.1, we need, more strongly, that u is *uniformly* alternating, i.e., strictly alternating and with $\inf |u(i)| > 0$. In that case, the diagonal matrix

$$D := \lceil \cdots, (-)^i u(i), \cdots \rceil$$

is bounded (since $u = Ax$ is) and boundedly invertible, while $D^{-1}A$ is still strictly m-banded and tp and carries the bounded sequence x to $D^{-1}u = 1'$; hence $D^{-1}A$ is invertible on l_∞ and $\|(D^{-1}A)^{-1}\| = \|x\|$. Therefore $A = D(D^{-1}A)$ is invertible and $\|A^{-1}\| \le \|(D^{-1}A)^{-1}\| \|D^{-1}\| = \sup_{i,j} |x(i)/u(j)|$.

Secondly, the assumption of *strict* m-bandedness, though essential for part of the argument, is not essential for the conclusion. For, according to [4], a bounded m-banded tp matrix A, whose rows and columns are linearly independent, is the uniform limit of strictly m-banded tp matrices A_ε (as $\varepsilon \to 0$, say). In our case, the

linear independence of the columns follows from the assumed *uniqueness* of the bounded solution to $Ax = u$, while the linear independence of the rows follows from the assumed total positivity of A and the assumed existence of x with Ax strictly alternating. Existence and uniqueness of a bounded solution to the equation $Ax = u$ (with a uniformly alternating u) therefore implies existence and uniqueness of a bounded solution x to the equation $A_\varepsilon x = u_\varepsilon$, with $u_\varepsilon := u - (A - A_\varepsilon)x$ again uniformly alternating for all sufficiently small ε. Consequently, A_ε is then boundedly invertible on l_∞ and

$$\left\| A_\varepsilon^{-1} \right\| \leq \sup_{i,j} \left| x(i)/u_\varepsilon(j) \right| \underset{\varepsilon \to 0}{\to} \sup_{i,j} \left| x(i)/u(j) \right|.$$

Thus A must be boundedly invertible on l_∞ and $\| A^{-1} \| \leq \sup_{i,j} \left| x(i)/u(j) \right|$.

COROLLARY. *The conclusions of Theorem* 4.1 *remain true if A is only m-banded and* 1' *is replaced by a uniformly alternating sequence u.*

The proof of Theorem 4.1 shows more than just the invertibility of A on l_∞. It shows that A has its kth diagonal as *main diagonal* in the sense introduced in [1]: The sections $A_J = A_{J,J+k}$ are invertible as $J \to \mathbf{Z}$ and the corresponding set (A_J^{-1}) is bounded. Hence A^{-1} is the bounded entrywise limit of these finite matrices A_J^{-1}. Again, this conclusion persists if A is only m-banded since it is then the uniform limit of strictly m-banded tp matrices.

THEOREM 4.5. *Let A be an m-banded bi-infinite tp matrix which is bounded and boundedly invertible. Then A has a main diagonal, i.e., for some k and all intervals J, $A_{J,J+k}$ is invertible and A^{-1} is the bounded entrywise limit of $(A_{J,J+k})^{-1}$.*

Consequently, with D the diagonal matrix

$$D := \lceil \cdots, (-)^i, \cdots \rceil,$$

$(-)^k D^{-1} A^{-1} D$ *is again tp. In particular, A^{-1} is checkerboard,*

$$(-)^{i+j+k} A^{-1}(i, j) \geq 0, \quad \text{all } i, j.$$

5. Concluding remarks. S. Friedland, in reaction to a presentation of these results, suggested that a tp matrix, whether banded or not, must map l_∞ onto itself if its range on l_∞ contains 1', since it is then possible to generate a preimage for every $u \in l_\infty$ as a limit point of minimal solutions of $(P_J A)y = P_J u$, using the checkerboard nature of the inverses of finite sections of A. Further, A. Pinkus showed how to establish the sign regularity of $DA^{-1}D$, with A^{-1} the bounded inverse of a tp matrix A, without assuming that the inverse is the limit of inverses of finite sections. These matters are made precise in [3].

REFERENCES

1. C. de Boor, *What is the main diagonal of a biifinite band matrix?*, MRC TSR 2049 (1980); in [5, 11–23].

2. _____, *Dichotomies for band matrices*, MRC TSR 2057 (1980); SIAM J. Numer. Anal. **17** (1980), 894–907.

3. C. de Boor, S. Friedland and A. Pinkus, *Inverses of infinite sign regular matrices*, MRC TSR 2159 (1980); Trans. Amer. Math. Soc. **274 (1982), 59–68.**

4. C. de Boor and A. Pinkus, *The approximation of a totally positive band matrix by a strictly totally positive one*, Linear Algebra and Appl. **42** (1982), 81–98.

5. R. DeVore and K. Scherer (eds.), *Quantitative approximation*, Academic Press, New York, 1980.

6. S. Karlin, *Total positivity*. I, Stanford University Press, Stanford, Calif., 1968.

7. C. A. Micchelli, *Infinite spline interpolation*, Approximation in Theorie und Praxis. Ein Symposiumsbericht (G. Meinardus, ed.), Bibliographisches Institut, Mannheim, 1979, pp. 209–238.

8. A. Cavaretta, W. Dahmen, C. A. Micchelli and P. Smith, *On the solvability of certain systems of linear difference equations*, RC 8329, IBM Research Report, 1980; Siam J. Math. Anal. **12** (1981), 833–841.

MATHEMATICS RESEARCH CENTER, 610 WALNUT ST., MADISON, WISCONSIN 53706