Linear Algebra and its Applications 437 (2012) 2179–2199

Contents lists available at SciVerse ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

On best rank n matrix approximations

Allan Pinkus

Department of Mathematics, Technion, Haifa, Israel

ARTICLE INFO

Article history:
Received 31 October 2011
Accepted 15 May 2012
Available online 28 June 2012

Submitted by Volker Mehrmann

AMS classification:
15A60
15B48
15A03

Keywords:
Totally positive matrices
Rank n approximation
Row and column interpolation

ABSTRACT

We consider the problem of approximating matrices by matrices of rank n in the $|\cdot|_{p,q}$ norm. Among other results, we prove that if A is a totally positive matrix, and $|A|_{1,1}$ is the norm given by the sum of the absolute values of the entries of the matrix, then a best rank n approximation to A in this norm is given by a matrix that agrees with A on n rows and n columns.

1. Introduction

Let $A = (a_{ij})_{i=1}^{N}{}_{j=1}^{M}$ be an $N \times M$ matrix. Consider the mixed norms $|\cdot|_{p,q}$ on A defined by

$$|A|_{p,q} := \left(\sum_{i=1}^{N} \left(\sum_{j=1}^{M} |a_{ij}|^q \right)^{p/q} \right)^{1/p},$$

where $1 \le p \le \infty$, $1 \le q \le \infty$. If $p = \infty$ and/or $q = \infty$, then the usual definitions apply. A problem of interest is that of approximating A by rank n matrices in the above norms. That is, consider

$$E_{p,q}^{n}(A) := \min_{B} |A - B|_{p,q},$$

where B runs over the set of real $N \times M$ matrices of rank at most n.

Except in the case where $p = q = 2$, called the Frobenius norm or Hilbert-Schmidt norm, very little seems to be known concerning this problem of best rank n approximation. (Note that the Frobenius

E-mail address: pinkus@technion.ac.il

0024-3795/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.laa.2012.05.016

norm is also the square root of the trace of AA^*.) The method of approximation in this $p = q = 2$ case, or in the associated operator norm (considering A as an operator from ℓ_2^M to ℓ_2^N), is via singular value decomposition (SVD). Truncated SVD provides a best rank n approximation in both the Frobenius and operator norms. These results go back to E. Schmidt [11] (often incorrectly attributed to Eckart and Young). See Stewart [14] for a detailed history of SVD. While the computation of this decomposition is highly nontrivial, almost no other rank n approximations have been considered, probably because they cannot be easily characterized or calculated. The problem is difficult, and is made more so by the fact that the approximating set is not convex.

A similar question can be asked with regards to best rank n approximation to a kernel $K(x, y)$ defined on $X \times Y$. In Micchelli and Pinkus [4] this problem was considered for a continuous totally positive kernel defined on $[a, b] \times [c, d]$. Define, analogously to the above,

$$|K|_{p,q} = \left(\int_a^b \left(\int_c^d |K(x, y)|^q \, dy \right)^{p/q} dx \right)^{1/p}.$$

We say that an operator with kernel L is of rank n if

$$L(x, y) = \sum_{i=1}^n u_i(x)v_i(y).$$

(We also demand that the $u_i \in L^p[a, b]$ and the $v_i \in L^q[c, d]$.) Set

$$E_{p,q}^n(K) = \inf_L |K - L|_{p,q},$$

where the infimum ranges over the set of kernels L of rank at most n.

The kernel K is said to be *totally positive* on $[a, b] \times [c, d]$ if for all $a \le x_1 < \cdots < x_k \le b$ and $c \le y_1 < \cdots < y_k \le d$, and for all k, we have

$$K \begin{pmatrix} x_1, \ldots, x_k \\ y_1, \ldots, y_k \end{pmatrix} := \det \left(K(x_i, y_j)\right)_{i,j=1}^k \ge 0.$$

The result proven in Micchelli and Pinkus [4], without going into extraneous details, is that for a totally positive continuous kernel K, defined on $[a, b] \times [c, d]$, and for arbitrary $p \in [1, \infty]$ and $q = 1$, we have

$$E_{p,1}^n(K) = |E|_{p,1},$$

where

$$E(x, y) = \frac{K \begin{pmatrix} x, \tau_1, \ldots, \tau_n \\ y, \xi_1, \ldots, \xi_n \end{pmatrix}}{K \begin{pmatrix} \tau_1, \ldots, \tau_n \\ \xi_1, \ldots, \xi_n \end{pmatrix}}$$

for some $a \le \tau_1 < \cdots < \tau_n \le b$ and $c \le \xi_1 < \cdots < \xi_n \le d$. This means that an optimal rank n approximating kernel L, in the above norm, is given by

$$L(x, y) = K(x, y) - \frac{K \begin{pmatrix} x, \tau_1, \ldots, \tau_n \\ y, \xi_1, \ldots, \xi_n \end{pmatrix}}{K \begin{pmatrix} \tau_1, \ldots, \tau_n \\ \xi_1, \ldots, \xi_n \end{pmatrix}} = \sum_{i,j}^n c_{ij} K(x, \xi_j) K(\tau_i, y).$$

In other words, this L is given by interpolation to $K(x, y)$ at (x, ξ_j), $j = 1, \ldots, n$, all $x \in [a, b]$, and (τ_i, y), $i = 1, \ldots, n$, all $y \in [c, d]$, using the slices $K(x, \xi_j)$ and $K(\tau_i, y)$. A generalization of this result may be found in Dyn [1]. The fact that the best rank n approximation is determined by slices of the kernel that interpolate it along those slices is aesthetically pleasing. This type of approximation is sometimes called *cross approximation*, see Schneider [12].

A matrix $A = (a_{ij})_{i=1}^N{}_{j=1}^M$ is said to be *totally positive* (TP) if all its minors are nonnegative. It is said to be *strictly totally positive* (STP) if all its minors are (strictly) positive. A main result of this paper is the matrix analogue of the above result for kernels. Unfortunately, for matrices the full analogous result holds only for one particular norm, namely the norm

$$|A|_{1,1} = \sum_{i=1}^{N} \sum_{j=1}^{M} |a_{ij}|.$$

We consider the best rank n approximation in this norm in some detail.

We start, in Section 2, with a general result bounding $E^n_{p,q}(A)$ from below by the associated operator norm, as a map from $\ell^M_{q'}$ to ℓ^N_p, where $1/q + 1/q' = 1$. We show, in certain cases, and in particular when $n = \text{rank}\, A - 1$, that this lower bound gives the exact error. We then consider the special cases $q = 1, p \in [1, \infty]$, and then $p = q = 1$, where we obtain more explicit results. For example, Theorem 2.5 states that if A is any $N \times N$ non-singular matrix, then the best rank $N - 1$ approximation to A in the $|\cdot|_{1,1}$ norm, and in the operator norm from ℓ^N_∞ to ℓ^N_1, is given by a matrix B which agrees with A on $N - 1$ rows and $N - 1$ columns. In other words, B differs from A in only one entry. In fact,

$$E^{N-1}_{1,1}(A) = \min_{k,\ell=1,\ldots,N} \frac{1}{|a^{-1}_{\ell k}|},$$

where $A^{-1} = (a^{-1}_{ij})^N_{i,j,=1}$.

In Section 3 we consider STP and TP matrices. For $q = 1$, $p \in (1, \infty]$ there exists a best rank n approximation to A in the $|\cdot|_{p,1}$ norm which is given by a matrix B which agrees with A on n columns. When $p = 1$ then, as stated earlier, there exists a best rank n approximation to A given by a matrix B which agrees with A on n columns and n rows.

In what follows we use the following notation. For $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$ we set

$$\|\mathbf{x}\|_r := \begin{cases} \left(\sum_{i=1}^N |x_i|^r\right)^{1/r}, & 1 \le r < \infty, \\ \max_{i=1,\ldots,N} |x_i|, & r = \infty. \end{cases}$$

The space ℓ^N_r is $\mathbb{R}^N$ endowed with this norm. For a matrix $A = (a_{ij})^N_{i=1}{}^M_{j=1}$ we set

$$A\begin{pmatrix} i_1, \ldots, i_k \\ j_1, \ldots, j_k \end{pmatrix} := \det\, (a_{i_r j_s})^k_{r,s=1},$$

while

$$A\begin{pmatrix} i_1, \ldots, \widehat{i_u}, \ldots, i_k \\ j_1, \ldots, \widehat{j_v}, \ldots, j_k \end{pmatrix}$$

will denote the minor, as in the previous line, but with row indexed i_u and column indexed j_v removed.

2. Best approximations from matrices of rank $(A) - 1$

For $r, s \in [1, \infty]$ and an $N \times M$ matrix A we set

$$\mathcal{A}_s = \{A\mathbf{x} : \ \|\mathbf{x}\|_s \le 1\}.$$

The error of the best rank n approximation, in the operator norm, as a map from ℓ^M_s to ℓ^N_r is given by

$$\delta_n(\mathcal{A}_s; \ell^N_r) = \min_B \max_{\|\mathbf{x}\|_s \le 1} \|(A - B)\mathbf{x}\|_r,$$

where the minimum extends over all $N \times M$ matrices of rank at most n. This quantity is also called the *linear n-width* of the set $\mathcal{A}_s$ in ℓ^N_r.

We recall that for $p, q \in [1, \infty]$

$$E^n_{p,q}(A) = \min_B |A - B|_{p,q},$$

where, again, B runs over the set of $N \times M$ matrices of rank at most n. Here

$$|A|_{p,q} = \left(\sum_{i=1}^{N} \left(\sum_{j=1}^{M} |a_{ij}|^q \right)^{p/q} \right)^{1/p}.$$

There is a simple relationship between $\delta_n(\mathcal{A}_s; \ell_r^N)$ and $E_{p,q}^n(A)$, and it is given by:

Proposition 2.1. *For any $N \times M$ matrix A we have*

$$\delta_n(\mathcal{A}_{q'}; \ell_p^N) \le E_{p,q}^n(A),$$

where $1/q + 1/q' = 1$.

Proof. For any $N \times M$ matrix C it is easily shown, as a consequence of Hölder's inequality, that

$$\max_{\|\mathbf{x}\|_{q'} \le 1} \|C\mathbf{x}\|_p \le |C|_{p,q}$$

implying

$$\delta_n(\mathcal{A}_{q'}; \ell_p^N) \le E_{p,q}^n(A). \quad \square$$

In general we do not have equality in the above inequality. A classic example is when $p = q = 2$, where

$$\delta_n(\mathcal{A}_2; \ell_2^N) = \mu_{n+1}$$

with $\mu_1 \ge \cdots \ge \mu_M \ge 0$ the s-numbers of A, while

$$E_{2,2}^n(A) = \left(\sum_{k=n+1}^{M} \mu_k^2 \right)^{1/2}.$$

Note that for $n \ge \text{rank}\,(A) - 1$ we do have equality of these two quantities, which begs the question of whether we have equality in Proposition 2.1 for all $p, q \in [1, \infty]$ if $n = \text{rank}\,(A) - 1$. In fact, the answer is yes. This was proven, in a more general setting, in Micchelli and Pinkus [5]. Part of the analysis also appears in Pinkus [8, p. 188–195].

The main result, as is relevant here, is the following, see Pinkus [8, p. 200].

Theorem 2.2. *For any $N \times M$ matrix A of rank $n + 1$ we have*

$$\delta_n(\mathcal{A}_{q'}; \ell_p^N) = E_{p,q}^n(A) = \min_{A\mathbf{x} \in \partial \mathcal{A}_{q'}} \|A\mathbf{x}\|_p,$$

where ∂C represents the boundary of the set C.

The proof of Theorem 2.2 is non-trivial and, in the general case, does not provide us with either a reasonable alternative way of expressing the above minimum or an insight into how to construct the best rank n approximating matrix B. Assuming that A is an $N \times M$ matrix of rank M, this can be rewritten in a more convenient form as:

Corollary 2.3. *For any $N \times M$ matrix A of rank M we have*

$$\delta_{M-1}(\mathcal{A}_{q'}; \ell_p^N) = E_{p,q}^{M-1}(A) = \min_{\mathbf{x} \ne \mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_{q'}}.$$

We can give an explicit expression for the above, when D is an $N \times N$ non-singular diagonal matrix, with $D = \operatorname{diag}\{D_1, \ldots, D_N\}$. Then

$$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\|D\mathbf{x}\|_p}{\|\mathbf{x}\|_{q'}} = \begin{cases} \left(\sum_{i=1}^N |D_i|^r\right)^{1/r}, & 1 \le q' < p \le \infty, \\ \min_{i=1,\ldots,N} |D_i|, & 1 \le p \le q' \le \infty, \end{cases}$$

where $1/r = 1/p - 1/q'(< 0)$. This result appears in Micchelli and Pinkus [5] and also in Pinkus [8, p. 200].

We can also say something further when $q = 1$.

Proposition 2.4. *Let A be an $N \times M$ matrix of rank M. Then a best rank $M-1$ approximation to A in the $|\cdot|_{p,1}$ norm, for any $p \in [1,\infty]$, is given by a matrix B which agrees with A on $M-1$ columns.*

Proof. Consider

$$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_\infty}.$$

Let $\mathbf{a}^j$ denote the jth column vector of A, and for each $k \in \{1, \ldots, M\}$, set

$$\mathcal{C}_k = \operatorname{span}\{\mathbf{a}^1, \ldots, \mathbf{a}^{k-1}, \mathbf{a}^{k+1}, \ldots, \mathbf{a}^M\}.$$

Define

$$e_k = \min_{\mathbf{c}\in\mathcal{C}_k} \|\mathbf{a}^k - \mathbf{c}\|_p.$$

We first claim that

$$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_\infty} = \min_{k=1,\ldots,M} e_k.$$

To see this, consider any $\mathbf{x}$ with $\|\mathbf{x}\|_\infty = 1$. Let $|x_r| = \|\mathbf{x}\|_\infty = 1$. Then

$$\|A\mathbf{x}\|_p \ge e_r \ge \min_{k=1,\ldots,M} e_k.$$

Now, assume

$$e_\ell = \min_{k=1,\ldots,M} e_k.$$

Thus, for some $\alpha_j, j \neq \ell$,

$$e_\ell = \|\mathbf{a}^\ell - \sum_{j\neq\ell} \alpha_j \mathbf{a}^j\|_p.$$

Since, for $\alpha_j \neq 0$,

$$e_\ell \le e_j \le \frac{1}{|\alpha_j|} e_\ell$$

we have that $|\alpha_j| \le 1$ for all j. Thus there exists an $\mathbf{x}$ with $x_\ell = \|\mathbf{x}\|_\infty = 1$ such that

$$\|A\mathbf{x}\|_p = e_\ell = \min_{k=1,\ldots,M} e_k.$$

This proves the claim.

We construct the desired $N \times M$ matrix B as follows. We define its jth column as $\mathbf{a}^j$ for $j \neq \ell$, while its ℓth column is given by

$$\mathbf{b}^\ell = \sum_{j \neq \ell} \alpha_j \mathbf{a}^j.$$

Thus B is of rank $\leq M-1$. Now

$$\begin{aligned} E_{p,1}^{M-1}(A) \leq |A-B|_{p,1} &= \left[\sum_{i=1}^N \left(\sum_{j=1}^M |a_{ij}-b_{ij}|\right)^p\right]^{1/p} \\ &= \left[\sum_{i=1}^N |a_{i\ell}-b_{i\ell}|^p\right]^{1/p} = e_\ell = \min_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_\infty}. \end{aligned}$$

Thus, the best rank $M-1$ approximation to A in the $|\cdot|_{p,1}$ norm is given by a matrix B which agrees with A on $M-1$ columns. □

There is no reason for this matrix B to agree with A on $M-1$ rows. However, there is one that does so in the case $p=1$.

Theorem 2.5. *Assume A is an $N \times M$ matrix of rank M. Then there exists a best rank $M-1$ approximation to A in the $|\cdot|_{1,1}$ which agrees with A on $M-1$ rows and $M-1$ columns. Furthermore, if $N=M$, i.e., A is non-singular, then*

$$E_{1,1}^{M-1}(A) = \min_{k,\ell=1,\ldots,M} \frac{1}{|a_{\ell k}^{-1}|}$$

where $A^{-1} = (a_{ij}^{-1})$.

Proof. From continuity considerations, it follows from Proposition 2.4 that there exists a matrix B which is a best rank $M-1$ approximation to A in the $|\cdot|_{1,1}$ norm, and agrees with A on $M-1$ columns. Let $\mathbf{a}^j$ and $\mathbf{b}^j$ denote the columns of A and B, respectively. Assume B does not agree with A on column ℓ. Thus $\mathbf{b}^\ell$ may be any best approximation to $\mathbf{a}^\ell$ from

$$\mathcal{C}_\ell = \mathrm{span}\{\mathbf{a}^1, \ldots, \mathbf{a}^{\ell-1}, \mathbf{a}^{\ell+1}, \ldots, \mathbf{a}^M\}.$$

While there need not be a unique best approximation in the ℓ_1^N norm, it is well-known that to each vector in $\mathbb{R}^N$ there is a best ℓ_1^N-approximation from any given $M-1$ dimensional subspace which interpolates to the approximated vector on at least $M-1$ indices. (Any extreme point of the bounded, convex set of best approximations has this property, see e.g. Pinkus [9, p. 136].) Thus there exists a best rank $M-1$ approximation to A in the $|\cdot|_{1,1}$ norm which agrees with A on $M-1$ rows and $M-1$ columns.

If $N=M$, then this implies that $A-B$ has exactly one nonzero entry. That is, if $A=(a_{ij})$ and $B=(b_{ij})$, then there exist indices $k, \ell \in \{1, \ldots, M\}$ such that

$$b_{ij} = a_{ij}, \qquad (i,j) \neq (k,\ell).$$

Furthermore, as B has rank $M-1$, it easily follows that we must have

$$A\begin{pmatrix} 1, \ldots, \widehat{k}, \ldots, M \\ 1, \ldots, \widehat{\ell}, \ldots, M \end{pmatrix} \neq 0$$

and

$$b_{k\ell} = a_{k\ell} - (-1)^{k+\ell} \frac{A\begin{pmatrix} 1,\ldots,M \\ 1,\ldots,M \end{pmatrix}}{A\begin{pmatrix} 1,\ldots,\widehat{k},\ldots,M \\ 1,\ldots,\widehat{\ell},\ldots,M \end{pmatrix}},$$

implying that

$$|A-B|_{1,1} = |a_{k\ell} - b_{k\ell}| = \frac{\left|A\begin{pmatrix}1,\ldots,M\\1,\ldots,M\end{pmatrix}\right|}{\left|A\begin{pmatrix}1,\ldots,\widehat{k},\ldots,M\\1,\ldots,\widehat{\ell},\ldots,M\end{pmatrix}\right|} = \frac{1}{|a_{\ell k}^{-1}|}.$$

Obviously an optimal choice of $k, \ell \in \{1, \ldots, M\}$ is one that minimizes the above quantity.

Alternatively, we can also approach this problem directly, as follows. From Corollary 2.3 we have

$$\delta_{M-1}(\mathcal{A}_\infty; \ell_1^M) = E_{1,1}^{M-1}(A) = \min_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty}.$$

Set $\mathbf{y} = A\mathbf{x}$. As A is non-singular, $\mathbf{x} = A^{-1}\mathbf{y}$ and

$$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_1}{\|\mathbf{x}\|_\infty} = \min_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{y}\|_1}{\|A^{-1}\mathbf{y}\|_\infty} = \frac{1}{\max_{\mathbf{y}\neq\mathbf{0}} \frac{\|A^{-1}\mathbf{y}\|_\infty}{\|\mathbf{y}\|_1}}.$$

The extreme points of the unit ball in ℓ_1^M are the unit vectors $\{\mathbf{e}^k\}_{k=1}^M$. Letting $A^{-1} = (a_{ij}^{-1})$ it therefore follows that

$$\max_{\mathbf{y}\neq\mathbf{0}} \frac{\|A^{-1}\mathbf{y}\|_\infty}{\|\mathbf{y}\|_1} = \max_{k=1,\ldots,M} \|A^{-1}\mathbf{e}^k\|_\infty = \max_{k,\ell=1,\ldots,M} |a_{\ell k}^{-1}|.$$

Thus

$$\min_{k,\ell=1,\ldots,M} \frac{1}{|a_{\ell k}^{-1}|} = E_{1,1}^{M-1}(A) = \min_B |(A-B)|_{1,1},$$

where the right-hand minimum extends over all B of rank at most $M-1$. $\square$

Remark. By interchanging rows and columns we see that the appropriate analogue of Theorem 2.5 also holds if A is of rank N, when approximating by matrices of rank $N-1$.

Remark. It can be proven directly, as in Theorem 2.5, or by considering the necessary and sufficient conditions for best approximation in ℓ_p^M, that if A is an $M \times M$ matrix of rank M, then

$$E_{p,1}^{M-1}(A) = \min_{\ell=1,\ldots,M} \frac{\left|A\begin{pmatrix}1,\ldots,M\\1,\ldots,M\end{pmatrix}\right|}{\left(\sum_{k=1}^M \left|A\begin{pmatrix}1,\ldots,\widehat{k},\ldots,M\\1,\ldots,\widehat{\ell},\ldots,M\end{pmatrix}\right|^{p'}\right)^{1/p'}},$$

where $1/p + 1/p' = 1$.

Remark. One strategy for approximating an $M \times M$ matrix A, in the $|\cdot|_{1,1}$ norm, by matrices of rank n might be to first approximate A by a matrix B of rank $M-1$, then approximate B by a matrix C of rank $M-2$, etc. i.e., at each of the $M-n$ steps approximate a matrix by matrices of rank one less than the approximated matrix, and in this way find a "good", but not "best", approximation to the original matrix. Theorem 2.5 gives us a simple formula for the first step of this process. Unfortunately, and despite the result of Theorem 2.2, we do not know how to approximate an $M \times M$ matrix A of rank K, in the $|\cdot|_{1,1}$, norm by matrices of rank $K-1$, for any $1 < K < M$.

Remark. As we have shown, the error in Theorem 2.5 is the reciprocal of the largest entry (in absolute value) of the inverse matrix. For a symmetric matrix this need not be a diagonal entry. Thus, even though the matrix is symmetric, the best rank $M-1$ approximation to A in the $|\cdot|_{1,1}$ norm is not necessarily symmetric. However, if A is positive semi-definite then so is A^{-1}, and thus its largest entry (in absolute value) lies on the diagonal. That is, for a positive semi-definite matrix of rank M, its best rank $M-1$ approximation to A in the $|\cdot|_{1,1}$ norm is symmetric. In fact this approximating matrix B is also positive semi-definite. This follows from Hadamard's inequality.

For a symmetric matrix A, and $\alpha = \{i_1, \ldots, i_k\}$, let

$$A(\alpha) = A\begin{pmatrix} i_1, \ldots, i_k \\ i_1, \ldots, i_k \end{pmatrix},$$

i.e., this is the principal minor of A with rows and columns $\{i_1, \ldots, i_k\}$. Hadamard's inequality for positive semi-definite matrices is given by:

$$A(\alpha)A(\beta) \geq A(\alpha \cup \beta)A(\alpha \cap \beta)$$

for all α and β. Assume A is positive semi-definite, and B is a best rank $M-1$ approximation to A in the $|\cdot|_{1,1}$ norm such that

$$b_{ij} = a_{ij}, \qquad (i,j) \neq (k,k)$$

and

$$b_{kk} = a_{kk} - \frac{A\begin{pmatrix} 1,\ldots,M \\ 1,\ldots,M \end{pmatrix}}{A\begin{pmatrix} 1,\ldots,\widehat{k},\ldots,M \\ 1,\ldots,\widehat{k},\ldots,M \end{pmatrix}}.$$

Thus B is symmetric, and for $j < k$ we have

$$B\begin{pmatrix} 1, \ldots, j \\ 1, \ldots, j \end{pmatrix} = A\begin{pmatrix} 1, \ldots, j \\ 1, \ldots, j \end{pmatrix} \geq 0,$$

while for $k \leq j$ we have

$$B\begin{pmatrix} 1, \ldots, j \\ 1, \ldots, j \end{pmatrix} = A\begin{pmatrix} 1, \ldots, j \\ 1, \ldots, j \end{pmatrix} - \frac{A\begin{pmatrix} 1,\ldots,M \\ 1,\ldots,M \end{pmatrix}}{A\begin{pmatrix} 1,\ldots,\widehat{k},\ldots,M \\ 1,\ldots,\widehat{k},\ldots,M \end{pmatrix}} A\begin{pmatrix} 1, \ldots, \widehat{k}, \ldots, j \\ 1, \ldots, \widehat{k}, \ldots, j \end{pmatrix}.$$

A simple application of the above Hadamard inequality implies that this principal minor is also non-negative. Thus B is positive semi-definite. Of course, $A - B$ is also trivially positive semi-definite since $b_{kk} \geq 0$, also by Hadamard's inequality.

3. Totally positive matrices

The two main theorems we prove in this section are the following.

Theorem 3.1. *Let A be an $N \times M$ strictly totally positive matrix and $1 \leq n < \min\{N, M\}$. Then for all $p \in (1, \infty]$ we have*

$$\delta_n(\mathcal{A}_\infty; \ell_p^N) = E_{p,1}^n(A).$$

Furthermore, a best rank n approximation to A in the $|\cdot|_{p,1}$ norm is given by a matrix B which agrees with A on n columns.

Theorem 3.2. *Let A be an $N \times M$ strictly totally positive matrix and $1 \leq n < \min\{N, M\}$. Then a best rank n approximation to A in the $|\cdot|_{1,1}$ norm is given by a matrix B which agrees with A on n columns and n rows.*

We need some ancillary results before proving these theorems. We first present a short explanation of the *variation diminishing* properties of STP matrices. We use two counts for the number of sign changes of a vector $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$. These are: $S^-(\mathbf{y})$ – the number of sign changes in the sequence $y_1, \ldots, y_m$, with zero terms discarded, and $S^+(\mathbf{y})$ – the maximum number of sign changes in the sequence $y_1, \ldots, y_m$, where zero terms are arbitrarily assigned values $+1$ or -1. For convenience, we set $S^+(\mathbf{0}) = m$. Thus, for example,

$$S^-(1, 0, 2, -3, 0, 1) = 2, \quad \text{and} \quad S^+(1, 0, 2, -3, 0, 1) = 4\,.$$

The connection between STP matrices and the variation diminishing property is the content of this next theorem.

Theorem 3.3. *Let A be an $N \times M$ STP matrix. Then*

(a) for each vector $\mathbf{x} \in \mathbb{R}^M$, $\mathbf{x} \neq \mathbf{0}$,

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

(b) if $S^+(A\mathbf{x}) = S^-(\mathbf{x})$ and $A\mathbf{x} \neq \mathbf{0}$, then the sign of the first (and last) component of $A\mathbf{x}$ (if zero, the sign given in determining $S^+(A\mathbf{x})$) agrees with the sign of the first (and last) nonzero component of $\mathbf{x}$.

Conversely, if (a) and (b) hold for some $N \times M$ matrix A and every $\mathbf{x} \in \mathbb{R}^M$, $\mathbf{x} \neq \mathbf{0}$, then A is STP.

For an explanation and history of this and similar results, see Pinkus [10, Chapter 3]. In addition, we have this next result which follows from Theorem 3.3 (a).

Proposition 3.4. *If A is an $r \times s$ STP matrix, $r > s$, $\mathbf{w} \neq \mathbf{0}$ and $\mathbf{w}A = \mathbf{0}$, then $S^-(\mathbf{w}) \ge s$.*

In what follows we will always assume that A is an $N \times M$ STP matrix, unless otherwise stated. We also fix n, $1 \le n < \min\{N, M\}$.

Let

$$J = \{\mathbf{j} : \mathbf{j} = (j_1, \ldots, j_n),\ 1 \le j_1 < \cdots < j_n \le M\}.$$

Definition 3.1. We will say that $\mathbf{x} \in \mathbb{R}^M$ *alternates between* $\mathbf{j} \in J$, $\mathbf{j} = (j_1, \ldots, j_n)$, if there exists an $\varepsilon \in \{-1, 1\}$ for which

$$x_k = (-1)^{i+1}\varepsilon, \qquad j_{i-1} < k < j_i,$$

for $i = 1, \ldots, n+1$, where $j_0 = 0$ and $j_{n+1} = M + 1$. We say that $\mathbf{x}$ alternates *positively* between $\mathbf{j} \in J$ if $\varepsilon = (-1)^{n+1}$.

If $\mathbf{x}$ alternates between $\mathbf{j} \in J$, then $S^-(\mathbf{x}) \le n$. Note that in the above definition there is no restriction placed on the values $x_{j_1}, \ldots, x_{j_n}$.

For each $p \in [1, \infty]$, we will consider the problem

$$\min \|A\mathbf{x}\|_p \tag{3.1}$$

where the minimum is taken over all $\mathbf{x}$ that alternate positively between some $\mathbf{j} \in J$. We will prove the following result.

Proposition 3.5. *Assume, as above, that A is an $N \times M$ STP matrix, and $p \in (1, \infty]$. Then there exists an $\mathbf{x}^*$ which attains the minimum in (3.1), satisfies $\|\mathbf{x}^*\|_\infty = 1$, and $S^+(A\mathbf{x}^*) = S^-(A\mathbf{x}^*) = S^-(\mathbf{x}^*) = n$.*

When $p = \infty$ this result may be found in Micchelli and Pinkus [3] and Pinkus [8], Chapter VI, Section 3. The proof as presented here for $p \in (1, \infty)$ is a variation thereof.

Proof. Let

$$\mathcal{X}_\mathbf{j} := \{\mathbf{x} : \mathbf{x} \text{ alternates positively on } \mathbf{j}\}.$$

For each $\mathbf{j} \in J$, let $\mathbf{x}_\mathbf{j} \in \mathcal{X}_\mathbf{j}$ satisfy

$$\min_{\mathbf{x} \in \mathcal{X}_\mathbf{j}} \|A\mathbf{x}\|_p = \|A\mathbf{x}_\mathbf{j}\|_p.$$

Let $\mathbf{a}^k$ denote the kth column of A, $k = 1, \ldots, M$. For each $\mathbf{j} \in J$, set

$$(\mathbf{y_j})_k = \begin{cases} (-1)^{i+n}, & j_{i-1} < k < j_i, \quad i = 1, \ldots, n+1, \\ 0, & k \in \{j_1, \ldots, j_n\}. \end{cases}$$

Then

$$\|A\mathbf{x_j}\|_p = \min_{\mathbf{x} \in \mathcal{X}_\mathbf{j}} \|A\mathbf{x}\|_p = \min_{\alpha_1, \ldots, \alpha_n} \|A\mathbf{y_j} + \sum_{r=1}^n \alpha_r \mathbf{a}^{j_r}\|_p. \tag{3.2}$$

Since $p \in (1, \infty)$ the above minimum (3.2) has a unique solution at some $(\alpha_1^*, \ldots, \alpha_n^*)$. This unique solution is characterized by

$$\sum_{i=1}^N \left| (A\mathbf{y_j} + \sum_{r=1}^n \alpha_r^* \mathbf{a}^{j_r})_i \right|^{p-1} \operatorname{sgn}(A\mathbf{y_j} + \sum_{r=1}^n \alpha_r^* \mathbf{a}^{j_r})_i \, a_{ij_k} = 0, \qquad k = 1, \ldots, n.$$

Note that $\mathbf{x_j}$ is the vector in $\mathbb{R}^M$ which alternates positively on $\mathbf{j}$ and such that $\mathbf{x}_{j_k} = \alpha_{j_k}^*$, $k = 1, \ldots, n$. From Theorem 3.3 we have, since $\mathbf{x_j}$ alternates on $\mathbf{j}$, that

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

From Proposition 3.4 it follows that since the vector $\mathbf{c} = (c_1, \ldots, c_N)$ with

$$c_i = \left| \left(A\mathbf{y_j} + \sum_{r=1}^n \alpha_r^* \mathbf{a}^{j_r} \right)_i \right|^{p-1} \operatorname{sgn} \left(A\mathbf{y_j} + \sum_{r=1}^n \alpha_r^* \mathbf{a}^{j_r} \right)_i, \qquad i = 1, \ldots, N,$$

satisfies

$$(\mathbf{c}, \mathbf{a}^{j_k}) = 0, \qquad k = 1, \ldots, n,$$

then $S^-(\mathbf{c}) \ge n$. Now

$$\operatorname{sgn} c_i = \operatorname{sgn} \left(A\mathbf{y_j} + \sum_{r=1}^n \alpha_r^* \mathbf{a}^{j_r} \right)_i = \operatorname{sgn}(A\mathbf{x_j})_i, \qquad i = 1, \ldots, N.$$

Thus $S^-(A\mathbf{x_j}) \ge n$. We have proven, for the above $\mathbf{x_j}$ that minimizes (3.2), for any $\mathbf{j} \in J$, that $S^+(A\mathbf{x_j}) = S^-(A\mathbf{x_j}) = S^-(\mathbf{x_j}) = n$.

By a compactness argument, the minimum in (3.1) is attained by a $\mathbf{x}^*$ which alternates positively on some $\mathbf{j}^* = (j_1^*, \ldots, j_n^*) \in J$. It remains to prove that $\|\mathbf{x}^*\|_\infty = 1$.

From the above we have

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

These equalities imply that $(A\mathbf{x}^*)_1 \, (A\mathbf{x}^*)_N \neq 0$. Furthermore, if $(A\mathbf{x}^*)_i = 0$, then $(A\mathbf{x}^*)_{i-1}(A\mathbf{x}^*)_{i+1} < 0$. Thus at the ℓth sign change of $A\mathbf{x}^*$ we have one of two possibilities. Either

(a) $(A\mathbf{x}^*)_{i_\ell} = 0$ and $(A\mathbf{x}^*)_{i_\ell - 1} \, (A\mathbf{x}^*)_{i_\ell + 1} < 0$, or
(b) $(A\mathbf{x}^*)_{i_\ell} \, (A\mathbf{x}^*)_{i_\ell + 1} < 0$,

where $1 \le i_1 < \cdots < i_n < N$.

We define vectors $\mathbf{g}^\ell \in \mathbb{R}^N$, $\ell = 1, \ldots, n$, as follows. If (a) holds, then set

$$\mathbf{g}^\ell = \mathbf{e}^{i_\ell},$$

where $\mathbf{e}^{i_\ell}$ is the unit vector with a 1 in the i_ℓ entry. If (b) holds, then set

$$\mathbf{g}^\ell = \frac{1}{|(A\mathbf{x}^*)_{i_\ell}|} \mathbf{e}^{i_\ell} + \frac{1}{|(A\mathbf{x}^*)_{i_\ell + 1}|} \mathbf{e}^{i_\ell + 1}.$$

Note that, by construction,

$$(\mathbf{g}^{\ell}, A\mathbf{x}^*) = 0, \qquad \ell = 1, \ldots, n.$$

To prove that $\|\mathbf{x}^*\|_\infty = 1$, we first note that if $j_k^* = k$ for some k, then since $\mathbf{x}^*$ alternates positively on $\mathbf{j}^*$ and $S^-(\mathbf{x}^*) = n$, then it follows that $j_s^* = s$ and

$$\mathrm{sgn}(x^*_{j_s^*}) = (-1)^{n+s+1},$$

for $s = 1, \ldots, k$. Similarly, if $j_k^* = M - (n-k)$, then $j_s^* = M - (n-s)$ and

$$\mathrm{sgn}(x^*_{j_s^*}) = (-1)^{n-s},$$

for $s = k, \ldots, n$.

Assume that $j_k^* > k$, and let r be any integer less than j_k^* not contained in the set $\{j_1^*, \ldots, j_{k-1}^*\}$. Define

$$\mathbf{j} = \{j_1^*, \ldots, j_{k-1}^*, j_{k+1}^*, \ldots, j_n^*, r\}$$

where the indices are to be rearranged in increasing order. Let $\mathbf{z} \in \mathcal{X}_{\mathbf{j}}$, i.e., $\mathbf{z}$ alternates positively on $\mathbf{j}$, satisfy

$$(\mathbf{g}^{\ell}, A\mathbf{z}) = 0, \qquad \ell = 1, \ldots, n. \tag{3.3}$$

These are n linear conditions, and we have the n unknowns $z_{j_1}, \ldots, z_{j_n}$. It is readily verified that this problem has a unique solution. Since $\mathbf{z}$ alternates on $\mathbf{j}$ we have $S^+(A\mathbf{z}) \le S^-(\mathbf{z}) \le n$. Moreover, from (3.3) we see that $n = S^+(A\mathbf{z})$ and due to the form of these sign changes (induced by (3.3)) it easily follows that

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

As $\mathbf{z}$ alternates positively on $\mathbf{j}$, and $j_k^* \notin \mathbf{j}$, we have

$$z_{j_k^*} = (-1)^{n+k}$$

and

$$(\mathbf{z} - \mathbf{x}^*)_\ell = 0, \qquad \ell \notin \{j_1^*, \ldots, j_n^*, r\}.$$

If $\mathbf{z} = \mathbf{x}^*$, then $|x_{j_k^*}| = 1$. Assume not. Then, since $\mathbf{z} - \mathbf{x}^*$ has at most $n+1$ nonzero components, we have

$$S^-(\mathbf{z} - \mathbf{x}^*) \le n.$$

Thus

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

By our construction we also have

$$(\mathbf{g}^{\ell}, A(\mathbf{z} - \mathbf{x}^*)) = 0, \qquad \ell = 1, \ldots, n,$$

and therefore

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

Since

$$(\mathbf{g}^{\ell}, A\mathbf{z}) = (\mathbf{g}^{\ell}, A\mathbf{x}^*) = (\mathbf{g}^{\ell}, A(\mathbf{z} - \mathbf{x}^*)) = 0, \qquad \ell = 1, \ldots, n,$$

and each of $A\mathbf{z}$, $A\mathbf{x}^*$ and $A(\mathbf{z} - \mathbf{x}^*)$ have exactly n sign changes determined by these $\mathbf{g}^{\ell}$, we see that they all have the same sign patterns, up to multiplication by ± 1. Both $\mathbf{z}$ and $\mathbf{x}^*$ alternate positively and thus from (b) of Theorem 3.3 we have that

$$(A\mathbf{z})_N, (A\mathbf{x}^*)_N > 0.$$

Now, if

$$(A\mathbf{z} - A\mathbf{x}^*)_N < 0,$$

then

$$0 < (A\mathbf{z})_N < (A\mathbf{x}^*)_N$$

and it then follows that

$$|(A\mathbf{z})_s| \le |(A\mathbf{x}^*)|_s$$

for all $s = 1, \ldots, N$, with equality only if both terms vanish. But then

$$\|A\mathbf{z}\|_p < \|A\mathbf{x}^*\|_p,$$

a contradiction, since by definition

$$\|A\mathbf{z}\|_p \ge \|A\mathbf{x}_\mathbf{j}\|_p \ge \|A\mathbf{x}^*\|_p.$$

Thus

$$(A\mathbf{z} - A\mathbf{x}^*)_N > 0,$$

(and, in fact, $|(A\mathbf{z})_s| \ge |(A\mathbf{x}^*)|_s$ for all $s = 1, \ldots, N$, with equality only if both terms vanish). Since

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

this implies, by (b) of Theorem 3.3, that the $n + 1$ nonzero components of $\mathbf{z} - \mathbf{x}^*$ strictly alternate in sign with the last nonzero component being strictly positive. Thus

$$\mathrm{sgn}((\mathbf{z} - \mathbf{x}^*)_{j_k^*}) = (-1)^{k+n}.$$

Since

$$z_{j_k^*} = (-1)^{k+n},$$

this implies that

$$(-1)^{k+n} x^*_{j_k^*} < 1.$$

A totally analogous argument can be applied in the case $j_k^* < M - (n - k)$. Here r is taken to be an integer greater than j_k^* not contained in the set $\{j_{k+1}^*, \ldots, j_n^*\}$. The result then obtained is that

$$(-1)^{k+n+1} x^*_{j_k^*} \le 1.$$

So we see that if $k < j_k^* < M - (n - k)$, then $|x^*_{j_k}| \le 1$. If $j_k^* = k$ then, as was shown,

$$\mathrm{sgn}(x^*_{j_k^*}) = (-1)^{n+k+1}.$$

But we also have $j_k^* < M - (n - k)$, and so

$$0 < (-1)^{n+k+1} x^*_{j_k^*} \le 1.$$

Similarly, if $j_k^* = M - (n - k)$, then $j_k^* > k$ and

$$0 < (-1)^{n+k} x^*_{j_k^*} \le 1.$$

Thus

$$\|\mathbf{x}^*\|_\infty = 1.$$

This proves Proposition 3.5. $\square$

Based on Proposition 3.5, we now prove Theorem 3.1.

Proof of Theorem 3.1. The proof of Theorem 3.1 will divide into two parts. From Proposition 2.1 we have

$$\delta_n(\mathcal{A}_\infty; \ell_p^N) \le E_{p,1}^n(A).$$

In the first part of our proof we will prove that for the $\mathbf{x}^*$ as in Proposition 3.5:

$$\|A\mathbf{x}^*\|_p \le \delta_n(\mathcal{A}_\infty; \ell_p^N).$$

The second part of the proof will be a construction of a rank n matrix B with the desired properties for which

$$|A - B|_{p,1} \le \|A\mathbf{x}^*\|_p.$$

We start with the lower bound

$$\|A\mathbf{x}^*\|_p \le \delta_n(\mathcal{A}_\infty; \ell_p^N).$$

We use a variation on the method of proof in Micchelli and Pinkus [4], and Pinkus [8, p. 145]. Set

$$\Xi_{n+1} := \left\{ \mathbf{y} : \mathbf{y} = (y_1, \ldots, y_{n+1}),\ \sum_{j=1}^{n+1} |y_j| = M \right\}.$$

For each $\mathbf{y} \in \Xi_{n+1}$, define

$$t_0(\mathbf{y}) := 0, \quad t_i(\mathbf{y}) := \sum_{j=1}^{i} |y_j|, \qquad i = 1, \ldots, n+1.$$

Thus

$$0 = t_0(\mathbf{y}) \le t_1(\mathbf{y}) \le \cdots \le t_{n+1}(\mathbf{y}) = M.$$

In addition, set

$$h_{\mathbf{y}}(s) := \operatorname{sgn} y_j, \qquad s \in [t_{j-1}(\mathbf{y}), t_j(\mathbf{y})), \quad j = 1, \ldots, n+1,$$

and

$$x_i(\mathbf{y}) := \int_{i-1}^{i} h_{\mathbf{y}}(s)\,\mathrm{d}s, \qquad i = 1, \ldots, M.$$

Finally, let $\mathbf{x}(\mathbf{y}) := (x_1(\mathbf{y}), \ldots, x_M(\mathbf{y}))$. From these definitions it follows that $\mathbf{x}(\mathbf{y})$ is an odd and continuous function of $\mathbf{y}$ on Ξ_{n+1}, and $\|\mathbf{x}(\mathbf{y})\|_\infty = 1$. It also easily follows that for each $\mathbf{y} \in \Xi_{n+1}$ the vector $\mathbf{x}(\mathbf{y})$ alternates between some $\mathbf{j} \in J$, since $h_{\mathbf{y}}$ has at most n sign changes.

Let B be any $N \times M$ rank n matrix. Then we can express $B\mathbf{x}$ in the form

$$B\mathbf{x} = \sum_{i=1}^{n} (\mathbf{u}^i, \mathbf{x})\,\mathbf{v}^i$$

for some $\mathbf{u}^1, \ldots, \mathbf{u}^n \in \mathbb{R}^M$ and $\mathbf{v}^1, \ldots, \mathbf{v}^n \in \mathbb{R}^N$. Consider the map

$$\Gamma(\mathbf{y}) := \left((\mathbf{u}^1, \mathbf{x}(\mathbf{y})), \ldots, (\mathbf{u}^n, \mathbf{x}(\mathbf{y})) \right).$$

Since $\mathbf{x}(\mathbf{y})$ is an odd and continuous function on Ξ_{n+1}, it follows that Γ is an odd, continuous map from Ξ_{n+1} to $\mathbb{R}^n$. Thus, by the Borsuk Antipodality Theorem, there exists a $\mathbf{y}^* \in \Xi_{n+1}$ for which $\Gamma(\mathbf{y}^*) = \mathbf{0}$, and hence

$$B\mathbf{x}(\mathbf{y}^*) = \mathbf{0}.$$

We therefore have

$$\max_{\|\mathbf{x}\|_\infty \le 1} \|(A-B)\mathbf{x}\|_p \ge \|A\mathbf{x}(\mathbf{y}^*)\|_p.$$

And, since $\mathbf{x}(\mathbf{y}^*)$ alternates between some $\mathbf{j} \in J$, we also have from Proposition 3.5 that

$$\|A\mathbf{x}(\mathbf{y}^*)\|_p \ge \|A\mathbf{x}^*\|_p.$$

This proves the lower bound.

We now prove the upper bound. As in the proof of Proposition 3.5, we recall that $\mathbf{x}^* \in \mathbb{R}^M$ alternates positively between $\mathbf{j}^* = (j_1^*, \ldots, j_n^*) \in J$ and $S^+(A\mathbf{x}^*) = S^-(A\mathbf{x}^*) = n$. These latter equalities imply that $(A\mathbf{x}^*)_1\,(A\mathbf{x}^*)_N \ne 0$, and if $(A\mathbf{x}^*)_i = 0$, then $(A\mathbf{x}^*)_{i-1}(A\mathbf{x}^*)_{i+1} < 0$. Thus at the kth sign change of $A\mathbf{x}^*$ we have one of two possibilities. Either

(a) $(A\mathbf{x}^*)_{i_k} = 0$ and $(A\mathbf{x}^*)_{i_k-1}\,(A\mathbf{x}^*)_{i_k+1} < 0$, or
(b) $(A\mathbf{x}^*)_{i_k}\,(A\mathbf{x}^*)_{i_k+1} < 0$,

where $1 \le i_1 < \cdots < i_n < N$.

Given the $\mathbf{g}^k$ as in the proof of Proposition 3.5, we define vectors $\mathbf{c}^k \in \mathbb{R}^M$, $k = 1, \ldots, n$, by

$$\mathbf{c}^k = A^T\mathbf{g}^k, \qquad k = 1, \ldots, n.$$

Let $\mathbf{r}^i$ denote the ith row of A. Thus, if (a) holds, then

$$\mathbf{c}^k = \mathbf{r}^{i_k},$$

while if (b) holds, then

$$\mathbf{c}^k = \frac{1}{|(A\mathbf{x}^*)_{i_k}|}\mathbf{r}^{i_k} + \frac{1}{|(A\mathbf{x}^*)_{i_k+1}|}\mathbf{r}^{i_k+1}.$$

Note that, by construction,

$$(\mathbf{c}^k, \mathbf{x}^*) = 0, \qquad k = 1, \ldots, n. \tag{3.4}$$

We construct the matrix $B = (b_{ij})_{i=1}^{N}{}_{j=1}^{M}$ as follows. Let $\mathbf{c}^k = (c_{k1}, \ldots, c_{kM})$, $k = 1, \ldots, n$, and

$$-b_{ij} = \frac{\begin{vmatrix} 0 & a_{ij_1^*} & \cdots & a_{ij_n^*} \\ c_{1j} & c_{1j_1^*} & \cdots & c_{1j_n^*} \\ \vdots & \vdots & \ddots & \vdots \\ c_{nj} & c_{nj_1^*} & \cdots & c_{nj_n^*} \end{vmatrix}}{C\begin{pmatrix} 1,\ldots,n \\ j_1,\ldots,j_n \end{pmatrix}}.$$

The denominator is strictly positive, as it is a positive linear combination of minors of A. Furthermore, B is an $N \times M$ matrix of rank at most n. This follows from the fact that

$$b_{ij} = \sum_{k,\ell=1}^{n} a_{ij_\ell^*} c_{kj} m_{k\ell} \tag{3.5}$$

where $m_{k\ell}$ are constants, independent of i and j, and therefore each of the n factors

$$a_{ij_\ell^*} \sum_{k=1}^{n} c_{kj} m_{k\ell}$$

defines a matrix of rank 1. Furthermore, from (3.4) and (3.5), we see that

$$B\mathbf{x}^* = \mathbf{0}. \tag{3.6}$$

Now

$$a_{ij} - b_{ij} = \frac{\begin{vmatrix} a_{ij} & a_{ij_1^*} & \cdots & a_{ij_n^*} \\ c_{1j} & c_{1j_1^*} & \cdots & c_{1j_n^*} \\ \vdots & \vdots & \ddots & \vdots \\ c_{nj} & c_{nj_1^*} & \cdots & c_{nj_n^*} \end{vmatrix}}{C\begin{pmatrix} 1,\ldots,n \\ j_1,\ldots,j_n \end{pmatrix}}. \tag{3.7}$$

What is the sign of each $a_{ij} - b_{ij}$? From the above, it is evident that

$$a_{ij} - b_{ij} = 0, \qquad j = j_1^*, \ldots, j_n^*.$$

From the construction of $\mathbf{c}^k$, as either a row of A or a positive combination of two consecutive rows of A, it follows that for $i_{k-1} + 1 \le i \le i_k$, $k = 1, \ldots, n+1$, and $j_{\ell-1}^* < j < j_\ell^*$, $\ell = 1, \ldots, n+1$, where $i_0 = 0$, $i_{n+1} = N+1$, $j_0^* = 0$, $j_{n+1}^* = M+1$, we have that either $a_{ij} - b_{ij}$ vanishes, or

$$\operatorname{sgn}(a_{ij} - b_{ij}) = (-1)^{k+\ell}.$$

This implies, using (3.6), (3.7) and the fact that $\mathbf{x}^*$ alternates positively between $\mathbf{j}^*$, that

$$\sum_{j=1}^{M} |a_{ij} - b_{ij}| = \left| \sum_{j=1}^{M} (a_{ij} - b_{ij}) x_j^* \right| = \left| \sum_{j=1}^{M} a_{ij} x_j^* \right|.$$

Thus,

$$|A - B|_{p,1} = \left(\sum_{i=1}^{N} \left(\sum_{j=1}^{M} |a_{ij} - b_{ij}| \right)^p \right)^{1/p} = \left(\sum_{i=1}^{N} \left| \sum_{j=1}^{M} a_{ij} x_j^* \right|^p \right)^{1/p} = \|A\mathbf{x}^*\|_p.$$

We have proven that

$$E_{p,1}^n(A) \le \|A\mathbf{x}^*\|_p$$

and this upper bound is attained by a rank n matrix B that agrees with A on the n columns $j_1, \ldots, j_n$. $\square$

Proof of Theorem 3.2. There are various methods of proving Theorem 3.2. We will present a proof that utilizes the results of Theorem 3.1 and Proposition 3.5. These results, however, are valid only for $p > 1$. What do they, nonetheless, tell us about the case $p = 1$?

From continuity considerations it easily follows from Theorem 3.1 and Proposition 3.5 that

$$\delta_n(\mathcal{A}_\infty; \ell_1^N) = E_{1,1}^n(A) = \min \|A\mathbf{x}\|_1,$$

where the minimum is taken over all $\mathbf{x}$ that alternate positively between some $\mathbf{j} \in J$. Furthermore, a best rank n approximation to A in the $|\cdot|_{1,1}$ norm is given by a matrix B which agrees with A on n columns.

What does not hold is that the $\mathbf{x}^*$ attaining the minimum in (3.1) for $p = 1$ necessarily satisfies $S^-(A\mathbf{x}^*) = S^+(A\mathbf{x}^*) = S^-(\mathbf{x}^*) = n$. Rather we obtain, from continuity considerations and the variation diminishing property, that $S^-(A\mathbf{x}^*) \le S^+(A\mathbf{x}^*) = S^-(\mathbf{x}^*) = n$. However we will not make use of these facts in what follows. What we will effectively show is that we can choose the $\mathbf{x}^*$ so that $A\mathbf{x}^*$ has n zeros. Thus the B, as constructed in the proof of Theorem 3.2, agrees with A on n rows and n columns. We will not directly prove anything about the $\mathbf{x}^*$, but will work with the matrix B.

From the above, we have a matrix B which is a best rank n approximation to A, and agrees with A on n columns. Let $\mathbf{b}^j$ and $\mathbf{a}^j$ denote the columns of B and A, respectively. Thus

$$\mathbf{b}^{j_k} = \mathbf{a}^{j_k}, \qquad k = 1, \ldots, n,$$

for some $1 \leq j_1 < \cdots < j_n \leq M$. As B is of rank n, and the $\{\mathbf{a}^{j_k}\}_{k=1}^n$ are linearly independent, $\mathcal{A}_{\mathbf{j}} = \text{span}\{\mathbf{a}^{j_1}, \ldots, \mathbf{a}^{j_n}\}$ is the range of B. Thus, each $\mathbf{b}^j$ is, essentially by definition, a best ℓ_1^N-approximation to $\mathbf{a}^j$ from $\mathcal{A}_{\mathbf{j}}$. It is well-known that there always exists a best ℓ_1^N-approximation from any n-dimensional subspace that interpolates any approximated vector at n indices. What we will prove, but is not immediately evident, is that for A STP these n indices of interpolation can be chosen independent of the columns $\mathbf{a}^j$, $j \notin \{j_1, \ldots, j_n\}$, i.e., they are fixed, depending only upon the choice of $\mathbf{j} = (j_1, \ldots, j_n)$.

Let

$$I = \{\mathbf{i} : \mathbf{i} = (i_1, \ldots, i_n), 1 \leq i_1 < \cdots < i_n \leq N\}.$$

As previously, given $\mathbf{c} \in \mathbb{R}^N$, we will say that $\mathbf{c}$ alternates between $\mathbf{i} \in I$, $\mathbf{i} = (i_1, \ldots, i_n)$, if there exists an $\varepsilon \in \{-1, 1\}$ for which

$$c_i = (-1)^{\ell+1}\varepsilon, \qquad i_{\ell-1} < i < i_\ell,$$

for $\ell = 1, \ldots, n+1$, where $i_0 = 0$ and $i_{n+1} = N+1$. We claim that given any n vectors $\mathbf{c}^1, \ldots, \mathbf{c}^n$ in $\mathbb{R}^N$, $1 \leq n < N$, there exists a $\mathbf{c}$ that alternates between some $\mathbf{i} \in I$, with $\|\mathbf{c}\|_\infty = 1$ and satisfies

$$(\mathbf{c}, \mathbf{c}^k) = 0, \qquad k = 1, \ldots, n.$$

This is a well-known result for continuous functions and is called the Hobby-Rice Theorem. A short proof, in both the continuous and this vector case, can be found in Pinkus [7]. The technique of proof as found therein was used by us here in the proof of the lower bound of Theorem 3.1.

Thus, given $\mathbf{a}^{j_1}, \ldots, \mathbf{a}^{j_n}$, there exists a $\mathbf{c} \in \mathbb{R}^N$ that alternates between some $\mathbf{i} \in I$, with $\|\mathbf{c}\|_\infty = 1$ and satisfies

$$(\mathbf{c}, \mathbf{a}^{j_k}) = 0, \qquad k = 1, \ldots, n.$$

We will use these equalities in the form

$$\sum_{i \notin \{i_1, \ldots, i_n\}} c_i a_{ij_k} = -\sum_{\ell=1}^n c_{i_\ell} a_{i_\ell j_k}, \qquad k = 1, \ldots, n. \tag{3.8}$$

Let us show that to each $\mathbf{a}^j$ the $\mathbf{b}^j \in \mathcal{A}_{\mathbf{j}}$ that interpolates to $\mathbf{a}^j$ at the indices of $\mathbf{i}$ is a best ℓ_1^N-approximation to $\mathbf{a}^j$ from $\mathcal{A}_{\mathbf{j}}$. To this end, note that if we have interpolation at the indices $i_1, \ldots, i_n$, then

$$a_{i_\ell j} = \sum_{k=1}^n \alpha_{jk} a_{i_\ell j_k}, \qquad l = 1, \ldots, n. \tag{3.9}$$

This implies that

$$a_{ij} - \sum_{k=1}^n \alpha_{jk} a_{ij_k} = \frac{A\binom{i, i_1, \ldots, i_n}{j, j_1, \ldots, j_n}}{A\binom{i_1, \ldots, i_n}{j_1, \ldots, j_n}}.$$

From this formula we see that, for each fixed j, the sign of this vector, as a function of i, is exactly εc_i for $i \notin \{i_1, \ldots, i_n\}$ and some fixed $\varepsilon \in \{-1, 1\}$.

Now for any $(\gamma_k)_{k=1}^n$

$$\begin{aligned}
\|\mathbf{a}^j - \sum_{k=1}^n \alpha_{jk}\mathbf{a}^{j_k}\|_1 &= \sum_{i=1}^N \left| a_{ij} - \sum_{k=1}^n \alpha_{jk} a_{ij_k} \right| = \varepsilon \sum_{i \notin \{i_1, \ldots, i_n\}} c_i \left(a_{ij} - \sum_{k=1}^n \alpha_{jk} a_{ij_k} \right) \\
&= \varepsilon \sum_{i \notin \{i_1, \ldots, i_n\}} c_i \left(a_{ij} - \sum_{k=1}^n \gamma_k a_{ij_k} \right) + \varepsilon \sum_{i \notin \{i_1, \ldots, i_n\}} c_i \left(\sum_{k=1}^n \gamma_k a_{ij_k} - \sum_{k=1}^n \alpha_{jk} a_{ij_k} \right).
\end{aligned}$$

From (3.8) this equals

$$= \varepsilon \sum_{i \notin \{i_1,\ldots,i_n\}} c_i \left(a_{ij} - \sum_{k=1}^{n} \gamma_k a_{ij_k} \right) - \varepsilon \sum_{\ell=1}^{n} c_{i_\ell} \left(\sum_{k=1}^{n} \gamma_k a_{i_\ell j_k} - \sum_{k=1}^{n} \alpha_{jk} a_{i_\ell j_k} \right).$$

From (3.9) we obtain

$$= \varepsilon \sum_{i \notin \{i_1,\ldots,i_n\}} c_i \left(a_{ij} - \sum_{k=1}^{n} \gamma_k a_{ij_k} \right) - \varepsilon \sum_{\ell=1}^{n} c_{i_\ell} \left(\sum_{k=1}^{n} \gamma_k a_{i_\ell j_k} - a_{i_\ell j_k} \right).$$

Applying the triangle inequality and since $\|\mathbf{c}\|_\infty = 1$ we have

$$\leq \sum_{i \notin \{i_1,\ldots,i_n\}} \left| a_{ij} - \sum_{k=1}^{n} \gamma_k a_{ij_k} \right| + \sum_{\ell=1}^{n} \left| \sum_{k=1}^{n} \gamma_k a_{i_\ell j_k} - a_{i_\ell j} \right| = \|\mathbf{a}^j - \sum_{k=1}^{n} \gamma_k \mathbf{a}^{j_k}\|_1.$$

Thus for any $(\gamma_k)_{k=1}^n$

$$\|\mathbf{a}^j - \sum_{k=1}^{n} \alpha_{jk} \mathbf{a}^{j_k}\|_1 \leq \|\mathbf{a}^j - \sum_{k=1}^{n} \gamma_k \mathbf{a}^{j_k}\|_1,$$

implying that the $\mathbf{b}^j \in \mathcal{A}_\mathbf{j}$ that interpolates to $\mathbf{a}^j$ at the indices of $\mathbf{i}$ is a best ℓ_1^N-approximation to $\mathbf{a}^j$ from $\mathcal{A}_\mathbf{j}$. (This is actually a convexity cone property, as found in Pinkus [9, Theorem 5, p. 210].) We have proved that there exists a best rank n approximation to A which agrees with A on n columns and n rows. □

Remark. If B interpolates to A on the rows $i_1, \ldots, i_n$ and columns $j_1, \ldots, j_n$, where

$$A \begin{pmatrix} i_1, \ldots, i_n \\ j_1, \ldots, j_n \end{pmatrix} \neq 0,$$

then

$$|A - B|_{1,1} = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\left| A \begin{pmatrix} i, i_1, \ldots, i_n \\ j, j_1, \ldots, j_n \end{pmatrix} \right|}{\left| A \begin{pmatrix} i_1, \ldots, i_n \\ j_1, \ldots, j_n \end{pmatrix} \right|}.$$

Thus, for A STP we have

$$E_{1,1}^n(A) = \min_{\substack{i_1,\ldots,i_n \\ j_1,\ldots,j_n}} \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\left| A \begin{pmatrix} i, i_1, \ldots, i_n \\ j, j_1, \ldots, j_n \end{pmatrix} \right|}{A \begin{pmatrix} i_1, \ldots, i_n \\ j_1, \ldots, j_n \end{pmatrix}}.$$

Remark. By interchanging rows and columns we are led to the norm

$$|A|^{r,1} = \left(\sum_{j=1}^{M} \left(\sum_{i=1}^{N} |a_{ij}| \right)^r \right)^{1/r},$$

where the same results hold, except that rows replace columns.

Theorems 3.1 and 3.2 have been proven for STP matrices. As is well-known, see e.g., Karlin [2, p. 88] or Pinkus [10, p. 42], STP matrices are dense in the set of TP matrices. As such we have

Corollary 3.6. *Let A be an $N \times M$ totally positive matrix, and $1 < n < \operatorname{rank} A$. Then for all $p \in [1, \infty]$ we have*

$$\delta_n(\mathcal{A}_\infty; \ell_p^N) = E_{p,1}^n(A).$$

Furthermore, a best rank n approximation to A in the $|\cdot|_{p,1}$ *norm is given by a matrix B which agrees with A on n columns, and when* $p = 1$*, a best rank n approximation to A is given by a matrix B which agrees with A on n columns and n rows.*

In fact, the above results can be easily seen to hold for a class of matrices larger than totally positive matrices. We can consider matrices obtained from A by D_1AD_2, where D_1 is an $N \times N$ matrix with a single 1 or -1 in each row and column, and D_2 is an $M \times M$ matrix of the same form. As is easily seen, if B is a best rank n approximation to A in the $|\cdot|_{1,1}$ norm, then D_1BD_2 is a best rank n approximation to D_1AD_2 in the $|\cdot|_{1,1}$ norm. Note that if B interpolates to A on n rows and columns, then D_1BD_2 interpolates to D_1AD_2 on n rows and columns.

Assume A is an $N \times M$ TP matrix and B is a matrix which agrees with A on n columns and n rows. Thus

$$a_{ij} - b_{ij} = \frac{A\binom{i,i_1,\ldots,i_n}{j,j_1,\ldots,j_n}}{A\binom{i_1,\ldots,i_n}{j_1,\ldots,j_n}},$$

for some $1 \le i_1 < \cdots < i_n \le N$ and $1 \le j_1 < \cdots < j_n \le M$. (We can assume the denominator is non-zero.) This matrix $A - B$ is, in general, not totally positive. However it is of the form described in the previous paragraph. That is, there exist D_1 and D_2, as above, such that $D_1(A - B)D_2$ is totally positive. In fact, both D_1 and D_2 are diagonal matrices and may be constructed as follows. Set

$$(D_1)_{ii} = (-1)^{\ell}, \qquad i_{\ell-1} < i < i_{\ell}, \quad \ell = 1, \ldots, n+1,$$

where $i_0 = 0$ and $i_{n+1} = N + 1$. (The entries $(D_1)_{i_\ell i_\ell}$, $\ell = 1, \ldots, n$, are immaterial.) D_2 is similarly constructed with respect to the $j_1, \ldots, j_n$. Furthermore, the best rank m approximation to the matrix $A - B$ in the $|\cdot|_{1,1}$ norm is, by Sylvester's Determinant Identity, of the same form. That is, it will agree with A on $n + m$ rows and columns, including the n rows and columns previously obtained. That is, if B is any matrix which agrees with A on the n columns $j_1, \ldots, j_n$ and rows $i_1, \ldots, i_n$, and C is any matrix which agrees with $A - B$ on the m other columns $s_1, \ldots, s_m$ and m other rows $r_1, \ldots, r_m$, then

$$a_{ij} - c_{ij} = \frac{A\binom{i,i_1,\ldots,i_n,r_1,\ldots,r_m}{j,j_1,\ldots,j_n,s_1,\ldots,s_m}}{A\binom{i_1,\ldots,i_n,r_1,\ldots,r_m}{j_1,\ldots,j_n,s_1,\ldots,s_m}},$$

assuming the denominator is non-zero. In other words, the interpolant to the interpolant is also an interpolant.

Remark. If A is totally positive, and B is constructed as above, i.e., for some $1 \le i_1 < \cdots < i_n \le N$ and $1 \le j_1 < \cdots < j_n \le M$ we have

$$b_{ij} = a_{ij} - \frac{A\binom{i,i_1,\ldots,i_n}{j,j_1,\ldots,j_n}}{A\binom{i_1,\ldots,i_n}{j_1,\ldots,j_n}},$$

then B need not be totally positive. Here is a simple example. Let

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 4 & 8 & 1 \\ 0 & 4 & 2 \end{pmatrix}.$$

This is a totally positive matrix. Let $n = 2$, $i_1 = 2$, $i_2 = 3$, $j_1 = 1$ and $j_2 = 2$. Now $b_{ij} = a_{ij}$ except for $(i,j) = (1,3)$ while

$$b_{13} = -1.$$

(This choice makes B singular, i.e., of rank 2). This B is obviously not totally positive. The matrix B is the best rank 2 approximation to A in the $|\cdot|_{1,1}$ norm.

Remark. Let $\mathbf{r}^i$, $i = 1, \ldots, N$, and $\mathbf{c}^j$, $j = 1, \ldots, M$, denote the row and column vectors of A, respectively. A best rank n approximation to the totally positive A is given by a matrix B which agrees with A on n columns $j_1, \ldots, j_n$ and n rows $i_1, \ldots, i_n$. These rows and columns have the property that the best ℓ_1^M approximation from $\mathcal{R} = \text{span}\{\mathbf{r}^{i_1}, \ldots, \mathbf{r}^{i_n}\}$ to $\mathbf{r}^i$ is given by interpolating from $\mathcal{R}$ to $\mathbf{r}^i$ at the indices $j_1, \ldots, j_n$. And the best ℓ_1^N approximation from $\mathcal{C} = \text{span}\{\mathbf{c}^{j_1}, \ldots, \mathbf{c}^{j_n}\}$ to $\mathbf{c}^j$ is given by interpolating from $\mathcal{C}$ to $\mathbf{c}^j$ at the indices $i_1, \ldots, i_n$. A natural question to ask is whether the above property implies the converse. That is, if one has found rows $i_1, \ldots, i_n$ and columns $j_1, \ldots, j_n$ satisfying the above properties, then are they necessarily optimal rows and columns in this minimum rank n approximation problem. The answer, unfortunately, is in the negative. Consider the 3×3 matrix

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 4 & 8 \end{pmatrix}.$$

This matrix is strictly totally positive. For $n = 1$ the (i, j)th entry of the matrix

$$E = \begin{pmatrix} 12.5 & 14.5 & 47 \\ 15 & 10 & 15 \\ 55 & 8.25 & 6.75 \end{pmatrix}$$

is the value of the error in the $|\cdot|_{1,1}$ norm when approximating A using the rank 1 matrix based on the ith row and jth column of A. As is evident, the 3rd row and 3rd column are optimal in this example, and uniquely so among the choices of rows and columns. However consider the 1st row and 1st column. The fact that the value 12.5 is strictly smaller than the other values in the first row and column of E means that the best ℓ_1^3 approximation from $\mathcal{R} = \text{span}\{\mathbf{r}^1\}$ to $\mathbf{r}^i$ is given by interpolating from $\mathcal{R}$ to $\mathbf{r}^i$ at the first index. And similarly the best ℓ_1^3 approximation from $\mathcal{C} = \text{span}\{\mathbf{c}^1\}$ to $\mathbf{c}^j$ is given by interpolating from $\mathcal{C}$ to $\mathbf{c}^j$ at the first index. This is an example of a "stationary" point that is not optimal.

Remark. There is an additional class of $N \times N$ matrices A for which we know how to calculate the best rank n approximation to A in the $|\cdot|_{p,1}$ norm, $p \in [1, \infty]$. And that is when A is a diagonal matrix (equivalently, a matrix with at most one non-zero element in each row and column). Assume that $A = D = \text{diag}\,\{D_1, \ldots, D_N\}$, where $|D_1| \geq |D_2| \geq \cdots \geq |D_N|$. Then for each n we have

$$\delta_n(\mathcal{D}_\infty; \ell_p^N) = E_{p,1}^n(D) = \left(\sum_{k=n+1}^{N} |D_k|^p \right)^{1/p},$$

where $\mathcal{D}_\infty = \{D\mathbf{x} : \|\mathbf{x}\|_\infty \leq 1\}$, and a best rank n approximation to D in the $|\cdot|_{p,1}$ norm is given by

$$B = \text{diag}\,\{D_1, \ldots, D_n, 0, \ldots, 0\}.$$

This result was proved, independently, by Pietsch [6] and Stessin [13] in greater generality, see also Pinkus [8, p. 203].

Remark. In Theorems 3.1 and 3.2 we determined the values of the linear n-widths $\delta_n(\mathcal{A}_\infty; \ell_p^N)$. There are, in the literature, also n-widths in the sense of Kolmogorov (denoted d_n) and n-widths in the sense of Gel'fand (denoted d^n). The interested reader can consult Pinkus [8]. All three n-widths δ_n, d_n and d^n are, in fact, equal in the cases considered above.

The above begs the question of whether it might be true that a best rank n approximation to any $N \times M$ matrix in the $|\cdot|_{1,1}$ norm is always given by a matrix which agrees with A on n columns and n rows. After all, it is true if $N = M$ and $n = N - 1$. Not surprisingly, it is not true in general.

The following is a 3×3 matrix A whose best rank 1 approximation in the $|\cdot|_{1,1}$ norm is NOT gotten by using one row and one column of A. Consider the matrix

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ -1 & 2.9 & 2 \end{pmatrix}.$$

On approximating using one row and one column, the minimal error is obtained when using the 3rd row and 2nd column. The resulting rank 1 approximating matrix is

$$B = \begin{pmatrix} \frac{-2}{2.9} & 2 & \frac{4}{2.9} \\ \frac{-1}{2.9} & 1 & \frac{2}{2.9} \\ -1 & 2.9 & 2 \end{pmatrix},$$

and

$$A - B = \begin{pmatrix} \frac{4.9}{2.9} & 0 & \frac{-1.1}{2.9} \\ \frac{3.9}{2.9} & 0 & \frac{.9}{2.9} \\ 0 & 0 & 0 \end{pmatrix}.$$

The $|\cdot|_{1,1}$ error is thus $10.8/2.9$. However we can do better. Take the second column of B, namely $(2, 1, 2.9)^T$ and use its span to approximate in ℓ_1^3 the columns of A. That is, we will construct a 3×3 rank 1 matrix

$$C = (\alpha_i \beta_j)$$

where $\alpha_1 = 2$, $\alpha_2 = 1$ and $\alpha_3 = 2.9$. The optimal choice of the β_j are $\beta_1 = 1/2$, $\beta_2 = 1$ and $\beta_3 = 2/2.9$. That is,

$$C = \begin{pmatrix} 1 & 2 & \frac{4}{2.9} \\ \frac{1}{2} & 1 & \frac{2}{2.9} \\ \frac{2.9}{2} & 2.9 & 2 \end{pmatrix},$$

and thus

$$A - C = \begin{pmatrix} 0 & 0 & \frac{-1.1}{2.9} \\ \frac{-1}{2} & 0 & \frac{.9}{2.9} \\ \frac{-4.9}{2} & 0 & 0 \end{pmatrix}.$$

The resulting $|\cdot|_{1,1}$ error is thus $2/2.9 + 5.9/2 < 10.8/2.9$. This C is better than B, but is also not the best rank 1 approximation to A.

What is true, in general, is that we can always find a best rank n approximation B to A in the $|\cdot|_{1,1}$ norm such that $A - B$ has at least n zeros in each row and in each column. Again, this is true because in ℓ_1^K when approximating any vector from an n-dimensional subspace, there always exists a best approximation that interpolates the approximated vector on at least n indices.

It would be of interest to determine other classes of matrices, if such exist, for which the result of Theorem 3.2 is valid. For example, what can one say in the case of positive semi-definite matrices?

References

[1] N. Dyn, Perfect splines of minimum norm for monotone norms and norms induced by inner products, with applications to tensor product approximation and n-widths of integral operators, J. Approx. Theory 38 (1983) 105–138.
[2] S. Karlin, Total Positivity, Volume I, Stanford University Press, CA, 1968.
[3] C.A. Micchelli, A. Pinkus, On n-Widths in L^∞, Trans. Amer. Math. Soc. 234 (1977) 139–174.
[4] C.A. Micchelli, A. Pinkus, Some problems in the approximation of functions of two variables and n-widths of integral operators, J. Approx. Theory 24 (1978) 51–77.

[5] C.A. Micchelli, A. Pinkus, The n-widths of rank $n+1$ kernels, J. Integral Equations 1 (1979) 111–130.
[6] A. Pietsch, s-Numbers of operators in Banach spaces, Studia Math. 51 (1974) 201–223.
[7] A. Pinkus, A simple proof of the Hobby-Rice theorem, Proc. Amer. Math. Soc. 60 (1976) 82–84.
[8] A. Pinkus, n-Widths in Approximation Theory, Springer-Verlag, Ergebnisse, 1985.
[9] A. Pinkus, On L^1-Approximation, Cambridge University Press, 1989.
[10] A. Pinkus, Totally Positive Matrices, Cambridge University Press, 2010.
[11] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen, I, Math. Ann. 63 (1907) 433–476.
[12] J. Schneider, Error estimates for two-dimensional cross approximation, J. Approx. Theory 162 (2010) 1685–1700.
[13] M.I. Stessin, Aleksandrov diameters of finite-dimensional sets and classes of smooth functions, Dokl. Akad. Nauk SSSR 220 (1975) 1278–1281, see also Soviet Math. Dokl. 16 (1975), 252–256.
[14] G.W. Stewart, On the early history of the singular value decomposition, SIAM Rev. 35 (1993) 551–566.