EORGIA INSTITUTE OF TECHNOLOGY OFFICE OF CONTRACT ADMINISTRATION

PROJECT ADMINISTRATION DATA SHEET

☒ ORIGINAL ☐ REVISION NO. ______

roject No. G37-638 DATE 9-1-82

roject Director: Dr S.G. DEMKO School/~~Lab~~ Mathematics

ponsor: USAF, Air Force Office of Scientific Research, Building 410, Bolling AFB, D.C. 20332

ype Agreement: Grant No. AFOSR-82-0329

ward Period: From 9-1-82 To 9-30-83 (Performance) 11-30-83 (Reports)

ponsor Amount: $23,843 Contracted through:

ost Sharing: $8871 (G37-341) GTRI/

itle: Conditioning and Stability Properties of Large Band Matrices

ADMINISTRATIVE DATA OCA Contact Don Hasty

) Sponsor Technical Contact:	2) Sponsor Admin/Contractual Matters:
Carl Edward Oliver/NM	Valerie M. Spencer/PK
Program Manager	Negotiator
(202) 767-5025	AFOSR/PKZ
	Building 410
	Bolling AFB, DC 20332
	(202) 767-4945

Defense Priority Rating: N/A Security Classification: N/A

RESTRICTIONS

See Attached AFOSR Supplemental Information Sheet for Additional Requirements.

Travel: Foreign travel must have prior approval — Contact OCA in each case. Domestic travel requires sponsor approval where total will exceed greater of $500 or 125% of approved proposal budget category.

quipment: Title vests with N/A - None proposed. For general info, reference page 23 of AFOSR Administration brochure.

COMMENTS:

SEP 1982
RECEIVED
Research Reports
Office

COPIES TO:

~~Administrative Coordinator~~ RAN
Research Property Management
Accounting
Procurement/EES Supply Services

Research Security Services
~~Reports Coordinator (OCA)~~
Legal Services (OCA)
Library

EES Public Relations (2)
Computer Input
Project File
Other ______

FORM OCA 4:781

EORGIA INSTITUTE OF TECHNOLOGY OFFICE OF CONTRACT ADMINISTRATION

SPONSORED PROJECT TERMINATION/CLOSEOUT SHEET

Date 1/25/85

roject No. G-37-638 School/~~Lab~~ Math

cludes Subproject No.(s)

roject Director(s) S. G. Demko GTRI / ~~GIT~~

ponsor AFOSR - Bolling AFB, DC

itle Conditioning and Stability Properties of Large Band Materices

ffective Completion Date: 9/30/83 (Performance) 5/21/84 (Reports)

rant/Contract Closeout Actions Remaining:

☐ None

☐ Final Invoice or Final Fiscal Report

☑ Closing Documents

☐ Final Report of Inventions

☒ Govt. Property Inventory & Related Certificate 2/5/85

☐ Classified Material Certificate

☐ Other

ntinues Project No. None Continued by Project No. None

PIES TO:

oject Director
search Administrative Network
search Property Management
counting
ocurement/EES Supply Services
search Security Services
ports Coordinator (OCA)
gal Services

Library
GTRI
Research Communications (2)
Project File JB 1/28/85
Other M. Heyser; A. Jones

rm OCA 60:1028

FINAL SCIENTIFIC REPORT FOR AFOSR GRANT 82-0329

Investigations into the spectral and invertibility properties of band matrices and other classes of sparse matrices were carried out. Mathematical techniques used included methods from classical approximation theory, from the spectral theory of self-adjoint operators, and from the theory of Banach spaces. Computer experiments were used to help formulate problems and to search for counterexamples. The main scientific content will appear in five papers. Three of these have been submitted for publication and accompany this report. The other two will be forwarded to AFOSR on completion. We now give short summaries of each of these papers together with a discussion of other results and observations we made which, while not suitable for publication now, might be useful to the sponsor and other readers of this report.

1.Decay Rates for Inverses of Band Matrices (written with W.F. Moss and P.W. Smith), to appear in Mathematics of Computation, 1984.

Spectral theory and classical approximation theory are used to give a new proof of the exponential decay of the entries of the inverses of band matrices. The rate of decay is bounded in terms of the extrema of the spectrum of the absolute value of the matrix. The rate predicted can be attained in some cases. These results are then used to establish the exponential decay of the eigenvectors for a class of generalized eigenvalue problems, to establish exponential decay of inverses of certain sparse but non-banded matrices, and to prove exponential decay for the Moore-Penrose inverse of full rank banded rectangular matrices.

2.Spectral Inequalities for Principal Submatrices, submitted to Linear Algebra and Its Applications.

The problem of obtaining spectral information about a given matrix from its submatrices is studied. Inequalities relating the extremal eigenvalues of Hermitian matrices to like quantities of certain principal submatrices are obtained. The spectral norm of a matrix is bounded in terms of the spectral norms of some of its principal submatrices. The Hausdorff distance between the numerical range of a matrix and the convex hull of the numerical ranges of certain of its principal submatrices is bounded.

3.On the Existence of Interpolating Projections onto Spline Spaces, submitted to Journal of Approximation Theory.

It is shown that for subspaces of the continuous functions on the unit interval that have B-spline like bases there is an interpolating projection whose norm depends on only the condition number of the basis. The relation between the norm of the projection and the condition number of the associated B-spline collocation matrix is exploited.

4.Condition Numbers of Rectangular Matrices and Bounds for Generalized Inverses, in preparation.

A notion of intrinsic condition number of a rectangular matrix is defined and shown to enjoy some properties of the classical condition number. For example, the relative distance to the set of all matrices of smaller rank is the reciprocal of the intrinsic condition number. The question of whether a matrix with a small intrinsic condition number must also have a generalized inverse of small norm is then studied. The answer turns out to be norm dependent. In particular, for the maximum norm the answer is "no" while for the sum norm the answer is "yes." These results are consequences of recent deep results in the geometric theory of Banach spaces.

5.Spectral Bounds for $\|A^{-1}\|_\infty$, in preparation.

The problem of bounding $\|A^{-1}\|_\infty$ in terms of the distance of the spectrum of A from zero is considered. It is shown that if A is banded and positive definite then $\|A^{-1}\|_\infty < c\|A^{-1}\|_2^{5/4}$ where c depends on only the bandwidth of A and the spectral norm of A. The argument relies on a Chebyshev series expansion of the inverse of A. We conjecture that the correct exponent is 1. The case of bi-infinite Toeplitz matrices is also considered.

6.Other Findings.

(a).We observed that a theorem of Shmuel Friedland implies that if A is an N by N positive definite matrix, then there is a diagonal matrix D such that the spectral radius of I-DA is zero. This implies that there is an iterative method for Ax=b based on D which converges in N steps and which is readily vectorizable. The catch is that it is virtually impossible to find such a D. However, the gap between current technology and what is theoretically possible is large and might be worth investigating. Friedland's proof is quasi-constuctive in that it is bases on a variational principal. I have discussed this with him but to date no progress has been made.

(b).We briefly considered an abstract marching technique for banded linear systems. Discussions with other mathematicians revealed that the method of parallel shooting was very similar to a concrete realization of our method. Consequently, we did not feel that further development would be an efficient use of time.

DECAY RATES FOR INVERSES OF BAND MATRICES

Stephen Demko*
Georgia Institute of Technology
Atlanta, Georgia

William F. Moss and Philip W. Smith†
Department of Mathematical Sciences
Old Dominion University
Norfolk, Virginia

*Supported in part by AFOSR grant no. 82-0329.

†Supported in part by the U.S. Army Research Office, Grant DAAG 29-81-K-0013.

Abstract

Spectral theory and classical approximation theory are used to give a new proof of the exponential decay of the entries of the inverses of band matrices. The rate of decay of A^{-1} can be bounded in terms of the (essential) spectrum of AA^* for general A and in terms of the (essential) spectrum of A for positive definite A. In the positive definite case the bound can be attained. These results are used to establish the exponential decay for a class of generalized eigenvalue problems and to establish exponential decay for certain sparse but non-banded matrices. We also establish decay rates for certain generalized inverses.

1. Introduction

The exponential decay of the entries of inverses of band matrices has been of some use in establishing local rates of convergence of spline approximations [K], [Kam], [D] and in bounding the L_∞-norm of the orthogonal projection onto spline spaces [deB2] and [Mit]. Kershaw proved a result of this nature for tridiagonal matrices and Descloux's paper [Des] contains such a result for Grammian matrices arising in finite element approximations although exponential decay is not explicitly mentioned. For general banded invertible matrices the first proof appeared in [D]. A later proof in [deB1,3] gave explicit estimates for the rate of decay. In this paper we use spectral theory and a result of Chebyshev on the best of approximation of $(x-a)^{-1}$ by polynomials to give a new proof. The bounds on the rate of decay obtained from this proof appear to be sharper than those previously known and are actually attained in some cases. In addition, the method of proof easily extends to certain generalized inverses and certain non-banded matrices. We show that the rate of decay for A^{-1} given by our method depends on only the essential spectrum of AA^* and is, thus, stable under banded compact perturbations. This fact is used to establish the exponential decay of the eigenvectors of certain generalized eigenvalue problems.

There are two key ideas. First use spectral theory to write

$$||A - p(A)|| = \max_{x\in\sigma(A)} |1/x - p(x)|$$

for any positive-definite operator A and any real polynomial p. Secondly, use approximation theory to estimate the best error.

2. Exponential Decay

We begin with a few preliminary remarks and definitions. Let H be a complex, separable, Hilbert space and let B(H) denote the Banach algebra of all bounded linear operators on H. If $A \in B(H)$ then we can represent A as a matrix with respect to any complete orthonormal set. Once a particular representation is chosen, we may regard A as an element of $B(\ell^2(S))$ where $S = \{1,\dots,N\}$, Z^+, or Z. In this setting the usual matrix product defines the action of A. Throughout this paper we will assume that the above identifications have been made and will reserve the letter A for a matrix representing a bounded operator in $B(\ell^2(S))$. For such matrices A we will say that A is m-banded if there is an index ℓ so that

$$A(i,j) = 0 \text{ if } j \notin [i-\ell, i-\ell+m].$$

We will say that A is centered and m-banded if m is even and the ℓ above may be chosen to be m/2. Thus, for a centered m-banded matrix one has

$$A(i,j) = 0 \text{ if } |i-j| > m/2 \text{ .}$$

Notice that self-adjoint matrices are naturally centered and, for example, a tridiagonal self-adjoint matrix is centered and 2-banded.

As mentioned in the introduction we combine certain approximation theoretic results with the spectral theorem in order to obtain estimates for the exponential decay rate of A^{-1}. Let π_n denote the polynomials of degree less than or equal to n. If K is a subset of the complex plane $\mathbb{C}$ and f is a fixed complex valued function on K we define

$$||f||_K := \sup\{|f(z)| : z \in K\}$$
$$e_n(K) := \inf\{||f-p||_K : p \in \pi_n\}.$$

We are now ready to state a proposition which is just a corollary of a result of Chebyshev [Mein, p. 33], but which is of fundamental importance to all that follows.

<u>Proposition 2.1</u>. Let $f(x) = 1/x$ and let $0 < a < b$. Set $r = b/a$ and

(2.1) $$q := q(r) := (\sqrt{r} - 1)/(\sqrt{r} + 1) .$$

Then

(2.2) $$e_n([a,b]) = \frac{(1+r^{\frac{1}{2}})^2}{2ar} q^{n+1}.$$

This exponential rate of approximation readily yields exponential decay of the inverse of a banded positive definite matrix as the next proposition shows. We will let $\sigma(A)$ denote the spectrum of the matrix A.

Proposition 2.2. *Let* A *be a positive definite*, m-*banded*, *bounded and boundedly invertible matrix in* $\ell^2(S)$. *Let* $[a,b]$ *be the smallest interval containing* $\sigma(A)$. *Set* $r = b/a$, $q = q(r)$ *as in* (2.1), *and set* $C_0 = (1+r^{\frac{1}{2}})^2/(2ar)$ *and* $\lambda = q^{2/m}$. Then we have

$$|A^{-1}(i,j)| \le C\lambda^{|i-j|} \tag{2.3}$$

where $$C := C(a,r) := \mathrm{MAX}\{a^{-1}, C_0\}$$

Proof: Since A is positive definite and invertible we have $0 < a < b$ and we know that A is centered. Thus A^k is centered and km-banded for $k = 0,1,2,\ldots$. Thus if $p \in \pi_k$ then $p(A)$ is km-banded and centered. From Proposition 2.1 we know there exists a sequence of polynomials $p_n \in \pi_n$ satisfying

$$||1/x - p_n||_{[a,b]} \equiv C_0 q^{n+1} .$$

An application of the spectral theory [R] yields

$$||A^{-1} - p_n(A)|| = ||1/x - p_n||_{\sigma(A)} \leq C_0 q^{n+1} .$$

Now writing $|i-j| = \frac{nm}{2} + k$ for $k = 1,\ldots,m/2$ and $i \neq j$ we see that $|i-j|2/m \leq (n+1)$ and hence

$$|A^{-1}(i,j)| = |A^{-1}(i,j) - p_n(A)(i,j)| \leq ||A^{-1} - p_n(A)|| \leq C_0 \lambda^{|i-j|} .$$

In case $i = j$ note that $1/a = ||A^{-1}||$, and (2.3) follows. This completes the proof of Proposition 2.2.

The phenomenon of exponential decay is certainly not restricted to positive definite band matrices. In fact, we can use the preceeding proposition to prove a more universal result. For any matrix A, we will say that A is quasi-centered if the central diagonal (i.e. (i,i)) is contained within the nontrivial bands. For example if $S = \{1,\ldots,N\}$ or Z^+ then $A \in B(\ell^2(S))$ is invertible only if A is quasi-centered. This is of course not true for $A \in \ell^2(Z)$.

Proposition 2.3. *Let* A *be* m-*banded, bounded and boundedly invertible on* $\ell^2(S)$. *Let* [a,b] *be the smallest interval containing* $\sigma(AA^*)$. *Then setting* $r = b/a$, $q = q(r)$ *as in* (2.1), *and* $\lambda_1 = q^{1/m}$, *there is a constant* C_1 *depending on* A *so that*

$$|A^{-1}(i,j)| \le C_1\lambda_1^{|i-j|} . \tag{2.4}$$

If A is quasi-centered then we may choose $C_1 = (m+1)\,||A||\,\lambda_1^{-m}C(a,r)$.

This result follows immediately from Proposition 2.2, the observation that

$$A^{-1} = A^*(AA^*)^{-1}$$

and the fact that $||A|| = ||A^*||$.

We collect these results in a theorem below which will allow us to compare them with earlier estimates. First let us note that if A is positive definite and invertible then $||A|| = \max\{s\colon s \in \sigma(A)\}$. We will set $\mathrm{cond}(A) := ||A||\,||A^{-1}||$. Note that since $B(\ell^2(S))$ is a B*-algebra, then

$$\mathrm{cond}(AA^*) = [\mathrm{cond}(A)]^2 .$$

Theorem 2.4. Let A and A^{-1} be in $B(\ell^2(S))$. Then if A is positive definite and m-banded we have

$$|A^{-1}(i,j)| \le C\lambda^{|i-j|} \tag{2.5}$$

where

$$\lambda = \left(\frac{\sqrt{\mathrm{cond}(A)}-1}{\sqrt{\mathrm{cond}(A)}+1}\right)^{2/m} \tag{2.6}$$

and

$$C = ||A^{-1}||\max\{1,(1+\sqrt{\mathrm{cond}(A)})^2/(2\mathrm{cond}(A))\} . \tag{2.7}$$

If A *fails to be positive definite but is still* m-*banded, quasi-centered, bounded and boundedly invertible then*

$$A^{-1}(i,j) \leq C_1 \lambda_1^{|i-j|} \tag{2.8}$$

where

$$\lambda_1 = \left(\frac{\mathrm{cond}(A) - 1}{\mathrm{cond}(A) + 1}\right)^{1/m} \tag{2.9}$$

and

$$C_1 = (m+1)\lambda_1^{-m}\,||A^{-1}||\,\mathrm{cond}(A)\max\{1,\left[\frac{1 + \mathrm{cond}(A)}{\mathrm{cond}(A)}\right]^2/2\} \tag{2.10}$$

We remind the reader that if $S = Z^+$ or $\{1,\ldots,N\}$ then the quasi-centered hypothesis is redundant and if $S = Z$ we may reindex A by a shift so that it is quasi-centered.

3. Examples and Comparisons

In the previous section we derived several results concerning exponential decay of the inverse of band matrices, culminating in Theorem 2.4. In this section we will present some examples indicating the precision of the estimates in Theorem 2.4 as well as comparing this result with the earlier estimates of [deB3].

We first show that the exponential rate in (2.5) can indeed be attained. Let $0 < s < 1$ and consider the tridiagonal Toeplitz matrix $A = \mathrm{trid}(s, 1+s^2, s)$ in $\ell^2(Z)$ with symbol

$$A(z) = (z+s)(z^{-1}+s)$$

Now it is well known [Goh] that

$$\sigma(A) = \{A(z): |z| = 1\} = [(1-s)^2, (1+s)^2].$$

Thus from Proposition 2.2 or Theorem 2.4 (2.5) we have for $i \neq j$

$$|A^{-1}(i,j)| \le \frac{2}{(1-s^2)^2} s^{|i-j|} .$$

On the other hand we can compute A^{-1} directly using the Laurent expansion for $1/A(z)$ which converges in a neighborhood of the unit circle yielding

$$A^{-1}(i,j) = (1-s^2)^{-1}(-s)^{|i-j|} .$$

Thus we predict the precise exponential decay rate in this case but our constant is pessimistic. This should come as no surprise since there is clearly something lost in the inequality $|A^{-1}(i,j)| \le ||A^{-1} - p_n(A)||$ which is used in the proof of Proposition 2.2.

Next, consider the positive definite matrix A with symbol $A(z) = (z+\frac{1}{2})(z^{-1}+\frac{1}{2})(z+\frac{1}{4})(z^{-1}+\frac{1}{4})$. In this case one may verify that

$$\sigma(A) = [9/64,\ 225/64] \ .$$

Thus Proposition 2.2 or Theorem 2.4 would predict an exponential decay rate of $\sqrt{2/3} \cong .82$ whereas it follows from the Laurent series for $1/A(z)$ that

$$|A^{-1}(i,j)| \le C\, 2^{-|i-j|}.$$

In this case we see that a conservative estimate has been obtained.

In the $\ell^2(S)$ setting [deB3] obtained the estimate

$$\left(\frac{(\mathrm{cond}(A))^2 - 1}{(\mathrm{cond}(A))^2 + 1}\right)^{1/2m}$$

for the decay rate of the inverse of an arbitrary m-banded matrix. We note that the λ_1 in (2.9) is smaller. However, deBoor also obtained results for all p. In particular, as

Demko has pointed out [D], once it is known that a banded matrix A is bounded and boundedly invertible on some $\ell^p(S)$ $(1 \le p \le \infty)$ then in fact it is bounded and boundedly invertible on all $\ell^p(S)$. Thus setting

$$\mathrm{cond}_p(A) = ||A||_p ||A^{-1}||_p$$

deBoor showed [deB3] that the exponential decay rate for the inverse of an m-banded matrix is bounded by

$$\left(\frac{(\mathrm{cond}_p(A))^p - 1}{(\mathrm{cond}_p(A))^p + 1}\right)^{1/pm} . \tag{3.1}$$

Thus, for instance, if A is totally positive Toeplitz then one can see [H, Thm. 6] by a result of deBoor that $\mathrm{cond}_p(A)$ is constant in p and hence the best choice for p in the above setting would be $p = 1$. Interestingly, this reduces precisely to our λ_1 in (2.9). If we had assumed in addition that A was symmetric then we would have had the better estimate λ in (2.6). It is not easy to compare these results but we feel confident that Theorem 2.4 is superior if for no other reason than finding the best p in (3.1) seems to be a difficult task in general.

4. Extensions and Eigenvectors

This section is devoted to certain refinements of Theorem 2.4. The results in this section were motivated by some eigenvalue computations of the second and third authors which will be reported elsewhere. For the rest of this section we will assume that S is either Z or Z^+.

As mentioned earlier $B := B(\ell^2(S))$ is a B*-algebra and the set $\mathcal{C} := \mathcal{C}(\ell^2(S))$ of compact operators is a two sided ideal in $B(\ell^2(S))$. Then the following facts are known [B] and [F]. The quotient space $B/\mathcal{C}$ is a B*-algebra and the quotient map

$$\nu : B \to B/\mathcal{C}$$

satisfies $$\sigma_e(A) := \sigma(\nu(A)) \subseteq \sigma(A)$$

where we use the symbol σ_e to denote the essential spectrum of A, that is the spectrum of the element $\nu(A)$ in $B/\mathcal{C}$. The object $B/\mathcal{C}$ is known as the Calkin algebra. In addition the norm in the Calkin algebra is called the essential norm and will be denoted by

$$||A||_e := ||\nu(A)|| := \inf\{||A-K|| : K \in \mathcal{C}\}.$$

Of course, the essential condition number will be

$$\mathrm{cond}_e(A) := ||A||_e ||A^{-1}||_e .$$

Finally we remark that if A is self-adjoint and U is an open neighborhood of $\sigma_e(A)$, then the complement of U contains at most finitely many points of $\sigma(A)$.

Recall that in the proof of Proposition 2.2 we used the inequality

$$||A^{-1} - p_n(A)|| = ||1/x - p_n||_{\sigma(A)} \leq ||1/x - p_n||_{[a,b]}$$

where $\sigma(A) \subset [a,b]$. Now it is quite possible that the inequality above does not always yield sharp estimates. This is indeed the case as will be illustrated by the following theorem whose proof is quite similar to the proof of Theorem 2.4.

Theorem 4.1. *Let* A *and* A^{-1} *be in* $B(\ell^2(S))$. *Then if* A *is positive definite and* m-*banded set*

$$\lambda = \left(\frac{\sqrt{\text{cond}_e(A)} - 1}{\sqrt{\text{cond}_e(A)} + 1}\right)^{2/m} . \tag{4.1}$$

For any $\gamma > \lambda$ *there is a constant* $C_2 = C_2(\gamma, A)$ *so that*

$$|A^{-1}(i,j)| \leq C_2 \gamma^{|i-j|} . \tag{4.2}$$

If A *fails to be positive definite but is quasi-centered*, m-*banded, bounded and boundedly invertible set*

$$\lambda_1 = \left(\frac{\text{cond}_e(A) - 1}{\text{cond}_e(A) + 1}\right)^{1/m} . \tag{4.3}$$

For any $\gamma > \lambda_1$ *there is a constant* $C_3 = C_3(\gamma, A)$ *so that*

(4.4) $$|A^{-1}(i,j)| \leq C_3\gamma^{|i-j|} .$$

Proof: Of course, (4.4) follows from (4.2) by considering $A^{-1} = A^*(AA^*)^{-1}$ just as in Proposition (2.3). In order to prove (4.2) we use an approximation lemma.

Lemma (4.2). Let $D = [a,b] \cup \{x_1,\ldots,x_k\}$ be a subset of the positive real axis and let $f(x) = 1/x$. Then there is a constant $R = R(x_1,\ldots,x_k,a,b)$ so that

$$e_n(D) \leq Rq^{n-k+1}$$

where $q = q(b/a)$ as in (2.1).

Proof; Let $r \in \pi_{k-1}$ interpolate f at $x_1,\ldots,x_k$ and let $\psi(x) = (x-x_1)\ldots(x-x_k)$. Note that

$$\frac{(1-xr)}{\psi} = -f[x_1,\ldots,x_k] := F_k .$$

Now compute

$$\inf_{p\in\pi_n} ||f-p||_D \leq \inf_{p\in\pi_{n-k}} ||f-(r+\psi p)||_D$$

$$\leq \inf_{p\in\pi_{n-k}} ||\psi||_{[a,b]}||\frac{f-r}{\psi} - p||_{[a,b]}$$

$$= ||\psi||_{[a,b]}|F_k| \inf_{p\in\pi_{n-k}} ||f-p||_{[a,b]}$$

$$= ||\psi||_{[a,b]}|F_k|e_{n-k}([a,b])$$

$$\leq Rq^{n-k+1}$$

where the last inequality follows from Proposition 2.1. This completes the proof of the Lemma.

Returning to the proof of (4.2), let $\gamma > \lambda$ be given, let $[a,b]$ be the smallest interval containing $\sigma_e(A)$, and choose $\varepsilon > 0$ so that $0 < a - \varepsilon$ and

$$\left(q\left(\frac{b+\varepsilon}{a-\varepsilon}\right)\right)^{2/m} < \gamma \, .$$

Then there exists a number k and $0 < x_1 < \ldots < x_k$ so that

$$\sigma(A) \subset (a-\varepsilon,\ b+\varepsilon) \cup x_1 \cup \ldots \cup x_k := D \, .$$

Thus by Lemma 4.2 we can find polynomials $p_n \in \pi_n$ satisfying

$$||f - p_n||_D \leq (Rq^{-k})q^{n+1} \, .$$

The proof of Theorem 4.1 is now completed in the same manner as Proposition 2.2.

Theorem 4.1 can yield significantly better estimates than Theorem 2.4. For example

Corollary 4.3. *Let* $A = I + K$ *be banded where* K *is compact and* I *is the identity operator. Then if* $I + K$ *is invertible we have for any* $\gamma > 0$ *there is a constant* $C = C(\gamma,A)$ *so that*

$$|A^{-1}(i,j)| \leq C\gamma^{|i-j|}.$$

We refer to such behavior as decay faster than any exponential. This corollary follows from Theorem (4.1) by noting that $\mathrm{cond}_e(A) = 1$ and hence $\lambda_1 = 0$ in (4.3).

Exponential decay can also be observed in certain generalized eigenvalue problems.

Corollary 4.4. *Let* A *be* m-*banded, bounded and boundedly invertible. Let* K *be compact so that* $A+K$ *is* m-*banded. Suppose* $\mu \neq 0$ *and* $\underline{x} \in \ell^2(S)$, $\underline{x} \neq \underline{\theta}$, *satisfy* $(A-\mu K)\underline{x} = \underline{\theta}$. *Then for any* $\gamma > \lambda_1$ (λ_1 *as in* 4.3), *there is a constant* $C = C(\gamma, A, K)$ *so that*

$$|\underline{x}(i)| \leq C\gamma^{|i|} .$$

Proof: Define the projections P_N by

$$P_N\underline{e}_j = \begin{cases} \underline{e}_j & \text{if } |j| \leq N \\ \underline{\theta} & \text{otherwise} \end{cases} .$$

Here $\underline{e}_j$ are the standard unit vectors, $\underline{e}_j(i) = \delta_{ij}$. Set $K_N = P_N K P_N$ and note that K_N is m-banded and $||K-K_N|| \to 0$ as $N \to \infty$. It follows that for some N, $A - \mu(K-K_N)$ is boundedly invertible and hence

$$\underline{x} = \mu(A-\mu(K-K_N))^{-1}(K_N\underline{x}) .$$

This yields the result since

$$\mathrm{cond}_e(A-\mu(K-K_N)) = \mathrm{cond}_e(A)$$

and $K_N\underline{x}$ is finitely supported so that $\underline{x}$ can be seen to be a finite linear combination of the columns of $(A - \mu(K - K_N))^{-1}$.

We close this section with a result on generalized inverses of band matrices which is valid in both finite and infinite dimensions.

Corollary 4.5. *Let* A *be an* m-*banded, bounded linear operator from* $\ell^2(S_1)$ *onto* $\ell^2(S_2)$ *where* S_1 and S_2 *are chosen from* $\{1,\ldots,N\}$, Z^+, *or* Z *and let* $A^\dagger$ *denote the Moore-Penrose inverse.* *Let* λ_1 *be as in* (4.3) *if* S_2 *is infinite dimensional or as in* (2.9) *otherwise. Then for any* $\gamma > \lambda_1$ *there is a constant* $C = C(\gamma,A)$ *so that*

$$|A^\dagger(i,j)| \le C\gamma^{|i-j|} .$$

Proof: By the open mapping theorem AA^* is boundedly invertible on $\ell^2(S_2)$. Now $A^\dagger = A^*(AA^*)^{-1}$ and the result follows just as in Proposition 2.3. For more information on generalized inverses the reader may consult [Ben].

5. Remarks

It has been recognized for some time that exponential decay is manifested in the inverses of sparse (but not necessarily banded) matrices, see for example [deB1]. The technique of Section 2 (and its extensions in section 4) yields such a result in a simple way. For any matrix $A \in \ell^2(S)$ let us define the support sets

$$S_n(A) := \bigcup_{k=0}^{n} \{(i,j) : A^k(i,j) \neq 0\}$$

and the decay sets

$$D_n(A) := (S \times S) \setminus S_n(A) .$$

Proposition 5.1. Let A be positive definite bounded and boundedly invertible on $\ell^2(S)$. Then

$$\sup\{|A^{-1}(i,j)| : (i,j) \in D_n(A)\} \leq C_0 q^{n+1} \tag{5.1}$$

where C_0 and q are as in Proposition 2.2.

The proof of this proposition follows the lines of Proposition 2.2 and we omit the details. The following examples illustrate the generality of Proposition 5.1.

Example 1. Suppose A is positive definite and block banded where each block is itself banded. Then the sets $S_n(A)$ and hence $D_n(A)$ would give a much better idea of the decay of the entries of A^{-1} than would the results of Section 2.

Example 2. A is positive definite and tridiagonal with the exception that (1,n) and (n,1) entries are nonzero. Then A^{-1} decays away from the maindiagonal and away from the corners (at least until $D_n = \emptyset$). Such matrices arise in spline interpolation of periodic data.

Example 3. If $\{f_i\}_{i\in S}$ is a local well conditioned basis in L_2, for a space of finite element approximations, then taking $A = (\langle f_i, f_j\rangle)$ where $\langle\cdot,\cdot\rangle$ denotes the inner product in L_2, we see that the biorthogonal functions

$$\phi_i(x) = \sum_j A^{-1}(i,j) f_j$$

satisfy

$$\langle\phi_i, f_\ell\rangle = \delta_{i\ell}$$

and decay exponentially to 0 as x moves away from the support of f_i, provided that the supports of the f_j satisfy a global mesh ratio restriction. A precise formulation can be found in a paper of Descloux [Des], in which the idea of exponential decay is already implicit.

We conclude this paper with some remarks concerning the generality of Propositions 2.1 and 2.2. There are two key ideas here. First use spectral theory to write $(f(x) = 1/x)$

$$||A^{-1} - p(A)|| = ||f - p||_{\sigma(A)}$$

and secondly use approximation theory to estimate the best error $||f - p||_{\sigma(A)}$. In all our propositions we arranged it so that $\sigma(A)$ was a subset of the positive real axis (ie A (or AA*) positive definite). The spectral theory applies more generally to normal matrices and Bernstein's theorem (see [M], p. 114) can be used in place of Proposition 2.1, but we were unable to see how to use this extra freedom quantitatively.

References

[B] F. F. Bonsall and J. Duncan, Complete Normed Algebras, Springer-Verlag, New York, 1973.

[Ben] A. Ben-Israel and T.N.E. Greville, Generalized inverses theory and applications, Wiley-Interscience, New York (1973).

[deB1] C. deBoor, Odd degree spline interpolation at a biinfinite knot sequence, in Approximation Theory, Bonn, 1976, R. Schaback and K. Scherer, eds., Lecture Notes in Math. 556, Springer, Heidelberg 1976, 30-53.

[deB2] C. deBoor, A bound on the L_∞-norm of the L_2-approximation by splines in terms of a global mesh ratio, Math. Comp. 30, 1976, 687-694.

[deB3] C. deBoor, Dichotomies for band matrices, MRC TSR 2057 (1980); SIAM J. Numer. Anal., Vol. 17, 1980, 894-907.

[D] S. Demko, Inverses of band matrices and local convergence of spline projections, SIAM J. Numer. Anal. 14, 1977, 616-619.

[Des] J. Descloux, On finite element matrices, SIAM J. Numer. Anal. 9, 1972, 260-265.

[F] P. A. Fillmore, J. G. Stampfli, and I. P. Williams, On the essential numerical range, the essential spectrum, and a problem of Halmos, Acta. Sci. Math. (Szeged) 33, 1972, 179-192.

[Goh] I. C. Gohberg and I. A. Feldman, Convolution Equations and Projection Methods for Their Solution, Vol. 41, Translation of Mathematical Monographs, American Mathematical Society, Providence, R.I., 1974.

[H] K. Hollig, L_∞-boundedness of L_2 projections on splines for a geometric mesh, J. of Approx. Th., 33, 1981, 318-333.

[Kam] W. J. Kamerer and G. W. Reddien, Jr., Local convergence of smooth cubic spline interpolants, SIAM J. Numer. Anal. 9, 1972, 687-694.

[K] D. Kershaw, Inequalities on the elements of the inverse of a certain tridiagonal matrix, Math. Comp. 24, 1970, 155-158.

[M] A. I. Markushevich, Theory of Functions of a Complex Variable, Vol. III, Prentice Hall, Inc., Englewood Cliffs, N.J., 1967.

[Mein] G. Meinardus, Approximation of Functions, Theory and Numerical Methods, translated by L. L. Schumaker, Springer-Verlag, New York, 1967.

[Mit] B. Mityagin, Quadratic pencils and least-squares piecewise-polynomial approximation, Math. Comp. 40, 1983, 283-300.

[R] W. Rudin, Functional Analysis, McGraw-Hill Book Company, New York, 1973.

1980 Mathematics Subject Classification: Primary 15A09, 65F99; Secondary 41A10, 47A05.

Spectral Inequalities for Principal Submatrices*

by

Stephen Demko
School of Mathematics
Georgia Institute of Technology
Atlanta, Georgia 30332

Abstract

The problem of obtaining spectral information about a given matrix from its submatrices is addressed. The Hausdorff distance between the numerical range of a matrix and the convex hull of the numerical ranges of some of its principal submatrices is bounded. Similar results for the Bauer field of values subordinate to an ℓ_p norm are also found. The spectral norm of a matrix is bounded in terms of the spectral norms of certain principal submatrices. Inequalities relating the extremal eigenvalues of Hermitian matrices to like quantitites of principal submatrices are also derived. Particular attention is paid to the case of banded matrices. The main technique is a simple decomposition method.

---

*Research supported in part by AFOSR grant no. 82-0329.

§1. Introduction

In this paper we exploit a simple decomposition technique to obtain spectral information about a given matrix in terms of similar information about some of its principal submatrices. The basic method is presented in Section 2 and is there used to bound the Hausdorff distance between the numerical range of a matrix and the convex hull of the numerical ranges of some of its principal submatrices. Analogous results for the Bauer field of values with respect to an ℓ_p norm are also discussed. Using a different method, Johnson [5] obtained bounds on the ratio of the area of the numerical range of an N × N matrix to the area of the convex hull of the numerical ranges of the (N-1) × (N-1) principal submatrices. When specialized to this case our results do not give as good an estimate as Johnson's, but nevertheless do tend to 1 as N goes to infinity. In Section 3, we consider extremal eigenvalues of Hermitian matrices with special emphasis on band matrices. The main results along this line can be interpreted to show that the extremal eigenvalues of m × m principal sections of infinite Hermitian band matrices converge to the extrema of the spectrum at least like $O(m^{-1/2})$. By extremal eigenvalues here we mean the supremum (infimum) of the largest (smallest) eigenvalues of all m × m contiguous principal submatrices. In the Toeplitz case results of this type go back to Szegö cf. [4, p. 3].

We consider complex, square matrices of arbitrary countable dimension. If A is a matrix indexed by a set I and if $J \subseteq I$, then A_J denotes the submatrix indexed by J. For a set J, $|J|$ denotes the cardinality of J and $\tilde{J}$ denotes the complement of J (with respect to some previously defined superset of J). Except for a diversion in Section 2 we use the usual ℓ_2 norm: $\|x\|^2 := x^*x$, $\|A\| := \sup\{\|Ax\| : \|x\| = 1\}$. We always assume $\|A\| < \infty$. The spectrum of A is $\sigma(A)$. $\lambda_{max}(A) := \sup\{\lambda : \lambda$ is in the spectrum of $A\} = \sup\{x^*Ax : \|x\| = 1\}$ is defined for Hermitian A. $\lambda_{min}(A)$ is defined similarly. The spectral radius of a matrix A will be denoted by $\rho(A)$. The <u>support of a vector</u> x is <u>supp x</u> $:= \{i : x(i) \neq 0\}$. Finally, when dealing with infinite matrices we make believe that all suprema and infima are attained; this seems neater than introducing ε's that eventually disappear.

§2. The Basic Method and Applications to Fields of Values

Let A be a matrix indexed by a set I. A collection of subsets $\{C_k\}_{k=1}^{m+r}$ of I will be called an <u>(r,m)-regular cover of I</u> if $0 \leq r < m$ and every $i \in I$ is in at least m of the C_k's. For example, if $I = \{1,2,\ldots,N\}$ and $k < N/2$, then we can take our cover to consist of all subsets of I of cardinality N-k. Every $i \in I$ is in exactly $\binom{N-1}{N-k-1}$ of these sets, so the choices $m = \binom{N-1}{N-k-1}$ and $m+r = \binom{N}{N-k}$ give an (r,m)-regular cover. The case $k = 1$ will correspond to investigations of $(N-1) \times (N-1)$ principal minors of A. For banded matrices, the following covers are useful (cf. [2], [3]).

Let $0 \leq r < m$ and for $1 \leq j \leq m+r$ define

$$C_j := \bigcup_{n=-\infty}^{\infty} \{i: j + n(m+r) \leq i \leq j + m-1 + n(m+r)\}.$$

Simply stated, C_j consists of integer intervals of length m separated by intervals of length r, and $[j,j+m-1] \subseteq C_j$. Furthermore, every integer is contained in exactly m of the C_j's. The cover $\{C_1,C_2,\ldots,C_{m+r}\}$ will be denoted by $\mathcal{C}(m,r)$. For a given set of integers J, the cover of J determined by $\mathcal{C}(m,r)$ is simply $\{C_1 \cap J,\ldots,C_{m+r} \cap J\}$. Some useful information is contained in

Lemma 2.1. Let $\{c_j\}$ be an (r,m) regular cover of I. Let x be a vector indexed by I and A a matrix indexed by I. Then,

(a) There is a c_j so that $\|x_{|c_j}\|^2 \geq \frac{m}{m+r}$ and $\|x-x_{|c_j}\|^2 \leq \frac{r}{m+r}$.

If, in addition, the cover determined by $\mathcal{C}(m,r)$ is used, then

(b) $\operatorname{supp}(x-x_{|c_j}) \subseteq c_{j+r} \cap \cdots \cap c_{j+m}$

and if in addition A is r-banded for some $r \leq m/2$, then

(c) $\operatorname{supp}(A(x-x_{|c_j})) \subseteq c_{j+r+[r/2]}$ hence by (b), both $x-x_{|c_j}$ and $A(x-x_{|c_j})$ are supported on a common c_j.

Proof. (a) As in [2], we have

$$m\sum_i |x(i)|^2 = \sum_{j=1}^{m+r} \sum_{i\in c_j} |x(i)|^2 \leq (m+r) \max_{1\leq j\leq m+r} \sum_{i\in c_j} |x(i)|^2$$

from which (a) follows.

For parts (b) and (c) we need only consider $c_1 = \{\ldots, [-m-r+1, -r], [1,m], [m+r+1, 2m+r],\ldots\}$ and note that $x_{|c_1}$

and $x-x_{|c_1}$ have disjoint supports and that due to the bandedness of A the supp $Ae_i \subseteq [i-\frac{r}{2}, i+\frac{r}{2}]$. ▲

Our first use of Lemma 2.1 will be to bound the error in approximating the value of the quadratic form determined by a matrix with the values of two quadratic forms coming from complementary principal submatrices. Let A be indexed by I and let $||x|| = 1$. Let C be an (r,m) regular cover of I, say $C = \{c_j\}$. By Lemma 2.1 we have $x = x_1 + x_2$ with $||x_1||^2 \geq \frac{m}{m+r}$, supp $x_1 \subseteq c_j$ for some j and (supp x_2) $\cap$ $c_j = \emptyset$. Now,

$$x^*Ax = (x_1+x_2)^*A(x_1+x_2) = x_1^*Ax_1 + x_2^*Ax_2 + x_1^*Ax_2 + x_2^*Ax_1.$$

Using the elementary bound

$$|x_1^*Ax_2| \leq ||x_1||\,||x_2||\,||A|| \leq ||A|| \max\{ab\colon a^2+b^2=1,\ b^2 \geq \frac{m}{m+r}\}$$
$$= \frac{\sqrt{mr}}{m+r}\ ||A||\ ,$$

we have

$$|x^*Ax - (x^*A_{c_j}x + x^*A_{c_j})| \leq \frac{2\sqrt{mr}}{m+r}\ ||A||\ .$$

Recalling the definition of <u>the numerical range of a matrix</u> B, $W(B) := \{x^*Bx\colon x^*x = 1\}$, we have

<u>Proposition 2.2</u>. Let $C = \{c_j\}$ be an (r,m) regular cover of I, the index set of A. Then, for every $z \in W(A)$ there is a c_j and numbers $0 \leq \lambda \leq 1$, $z_1 \in W(A_{c_j})$, $z_2 \in W(A_{\tilde{c}_j})$ such that

$$|z - (\lambda z_1 + (1-\lambda)z_2)| \leq \frac{2\sqrt{mr}}{m+r} \ ||A|| \ . \qquad (2.1)$$

Proof. Let $z = x^*Ax$ with $||x|| = 1$ and decompose $x = x_1 + x_2$ as above with supp $x_1 \subseteq c_j$. Since

$$x^*A_{c_j}x = x_1^*Ax_1 = ||x_1||^2 \ (\frac{x_1^*Ax_1}{||x_1||^2})$$

and

$$\frac{x_1^*Ax_1}{||x_1||^2} \in W(A_{c_j}),$$

we see that $\lambda := ||x_1||^2$ works. ▲

Because of the $||A||$ on the right hand side of (2.1), these inequalities should be considered as "relative" error estimates for the approximation of x^*Ax by convex combinations of elements of numerical ranges of certain (relative to C) complementary principal submatrices of A. We can repharse this in terms of Hausdorff distance. The Hausdorff distance between sets S_1, S_2 of the complex plane is

$$\mathcal{H}(S_1,S_2) := \max\{\sup_{x \in S_1} d(x,S_2), \ \sup_{x \in S_2} d(x,S_1)\}$$

where $d(x,T) := \inf_{t \in T} |x-t|$. For an (r,m)-regular cover C of I and a matrix A indexed by I, define

$$W_C(A) := \bigcup_{i=1}^{m+r} \{\text{convex hull of } \{W(A_{c_i}), W(A_{\tilde{c}_i})\}\}.$$

Corollary 2.3. Let C be an (r,m) regular cover of I, the index set of A. Then, $\mathcal{H}(W(A), W_C(A)) \leq \frac{2\sqrt{mr}}{m+r} \ ||A||$.

If C has the property that for every $c_i \in C$ there is $c_j \in C$ with $c_j \supseteq \tilde{c}_i$, then the bound given above is also a bound for $H(W(A)$, convex hull of $\{W(A_{c_1}, \ldots, W(A_{c_{m+r}})\})$. In [5], Johnson obtained the nice bound

$$\frac{\text{Area } (\hat{W}_C(A))}{\text{Area } (W(A))} \geq \frac{N-2}{N+4}$$

where C is the cover of $\{1,2,\ldots,N\}$ given by $c_i = \{1,\ldots,N\}\backslash\{i\}$ and $\hat{W}_C(A)$:= convex hull of $\{W(A_{c_1}),\ldots, W(A_{c_N})\}$. If we use Corollary 2.3 to try to get a bound on these areas, we end up with a lower bound of the form $1 - O(1/\sqrt{N})$ which tends to 1 as $N \to \infty$ but not as quickly as Johnson's rate. On the other hand the results presented here will apply to many more decompositions of matrices as well as to infinite matrices. We'll assume that $c_i \in C$ implies that for some j $c_i \subseteq c_j \in C$, and we use the notation

$$\hat{W}_C(A) := \text{convex hull of } \{W(A_{c_1}),\ldots,W(A_{c_{m+r}})\}$$

for a given (r,m)-regular cover C of the index set of A.

First note that if $\hat{W}_C(A) \subseteq \mathbb{R}$, then $A = A^*$ because $A^*_{c_i} = A_{c_i}$ for all i. More generally, if $\hat{W}_C(A)$ is contained in a line segment, then $W(A)$ is contained in the straight line containing the given segment. This is because there will be numbers of α,β with $(\alpha A + \beta I)^* = \alpha A + \beta I$. In either case Proposition 2.2 gives that $W(A)$ is contained in a line segment no longer than $\frac{4\sqrt{mr}}{m+r}$ $||A||$ plus the length of the segment containing $\hat{W}_C(A)$. That is,

$$\frac{\text{length } \hat{W}_C(A)}{\text{length } W(A)} \geq 1 - \frac{4\sqrt{mr}}{m+r} \,\|A\| .$$

In the case that $W_C(A)$ has interior points we use

Lemma 2.4. Let $S_1 \subseteq S_2$ be convex subsets of $\mathbb{C}$ with S_1 having interior. If for every $z \in S_2$, there is $w \in S_1$ with $|w - z| \leq \varepsilon$, then

$$\text{area}(S_2) - \text{area}(S_1) \leq \varepsilon \cdot \text{perimeter of } S_2;$$

Consequently,

$$\frac{\text{area}(S_1)}{\text{area}(S_2)} \geq 1 - \varepsilon \, \frac{\text{perimeter}(S_2)}{\text{area}(S_2)}.$$

Proof. We may assume S_2 is a polygon with sides of length δ and vertices $z_1,\ldots,z_k$. Let $w_1,\ldots,w_k$ be the points on S_1 closest to $z_1,\ldots,z_k$ respectively. Since S_1 is convex, the area between S_1 and S_2 is no greater than the sum of the areas of the quadrilaterals determined by consecutive vertices of S_2 and their closest points on S_1. This number is $\varepsilon \cdot \delta \cdot k = \varepsilon \cdot \text{perimeter of } S_2$. ▲

Using Corollary 2.3, Lemma 2.4, and the fact that $\text{perimeter}(W(A)) \leq 2\pi \,\|A\|$, we have

Corollary 2.5.

$$\frac{\text{area}(\hat{W}_C(A))}{\text{area}(W(A))} \geq 1 - \frac{4\sqrt{mr}}{m+r} \, \frac{\pi \,\|A\|^2}{\text{area}(W(A))}$$

Proof. The unicity of y is well known. The existence of $c_j \in C$ with $\|x_{|c_j}\|_p^p$ follows as in Lemma (2.1(a). The facts that $p^{-1} + q^{-1} = 1$ and that $y(i)$ must be given by $y(i) = \operatorname{sgn} x(i)|x(i)|^{p-1}$ give the rest. ▲

Now let $z \in F_p(A)$, say $z = y^*Ax$. With C an (r,m)-regular cover of the index set of A we have $x = x_1 + x_2$, $y = y_1 + y_2$ where $x_1 = x_{|c_i}$ and $y_1 = y_{|c_i}$ for some $c_i \in C$ satisfiying (2.2). Let $\alpha := \|x_1\|_p^p$ and note that

$$\left(\frac{1}{\alpha^{1/p}} x_i\right)^*\left(\frac{1}{\alpha^{1/q}} y_1\right) = 1 = \frac{\|x_1\|_p}{\alpha^{1/p}} = \frac{\|y_1\|_q}{\alpha^{1/q}} .$$

Similarly, $1-\alpha$ works for x_2 and y_2. Writing

$$z = y_1^*Ax_1 + y_2^*Ax_2 + y_1^*Ax_2 + y_2^*Ax_1,$$

we see that the distance of z to the convex hull of $F_p(A_{|c_i})$ and $F_p(A_{|\tilde{c}_i})$ is no more than $|y_1^*Ax_2 + y_2^*Ax_1|$. This can be bounded in the standard way by

$$(\|y_1\|_q \ \|x_2\|_p + \|y_2\|_q \ \|x_1\|_p)\ \|A\|_p$$

where

$$\|A\|_p := \sup\{\|Ax\|_p : \ \|x\|_p = 1\}.$$

Using

$$\|x_2\|_p \le \left(\frac{r}{m+r}\right)^{1/p}, \qquad \|x_1\|_p \le 1$$

and similar inequalities for y_1 and y_2, we have

Proposition 2.8. Let A be an $N \times N$ matrix and let C be an (r,m)-regular cover of $\{1,2,\ldots,N\}$. Let $1 < p < \infty$ and let $F_{p,C}(A)$ be the union of the convex hulls of the sets $F_p(A|_{c_i})$, $F_p(A|_{\tilde{c}_i})$ for $c_i \in C$. Then,

$$\mathcal{H}(F_p(A),F_{p,C}(A)) \leq \|A\|_p \{(\frac{r}{m+r})^{1/p} + (\frac{r}{m+r})^{1/q}\}.$$

§3. Bounds for Norms and Extremal Eigenvalues of Banded Matrices

In this section we concentrate on Hermitian banded matrices and refine the estimates of Section 2 to obtain sharper results. We also consider extremal eigenvalues of finite non-banded Hermitian matrices and their relation to extremal eigenvalues of principal submatrices.

Lemma 3.1. Let $A = A^*$ be r-banded and let $m \geq 2r$. Define

$$S(m) := \{E: E = [i+1,i+m] \cap I \text{ for some } i\}. \tag{3.1}$$

Then,

$$\|A\| \leq \{1 + \frac{2\sqrt{mr}}{m+r}\} \sup\{\|A_E\| : E \in S(m)\}. \tag{3.2}$$

Proof. Let $C_1,\ldots,C_{m+r}$ be the cover of I determined by $C(m,r)$. Suppose $\|A\| = x^*Ax$ where $\|x\| = 1$, $x = x_1 + x_2$, $\operatorname{supp} x_1 \subseteq C_j$, $\operatorname{supp} x_2 \subseteq C_i$, $\operatorname{supp} Ax_2 \in C_i$, $\|x_1\|^2 \geq \frac{m}{m+r}$ and $\|x_2\|^2 \leq \frac{r}{m+r}$. We have used Lemma 2.1(c) here. Then, since

$$A = A^* \quad \text{and} \quad A_{C_k} = \sum_{E \in C_k} \oplus A_E,$$

$$\begin{aligned} \|A\| &= x^*Ax + x_1^*Ax_1 + x_2^*Ax_2 + 2\mathrm{Re}(x_1^*Ax_2) \\ &\leq (\|x_1\|^2 + \|x_2\|^2) \sup\{\|A_E\| : E \in S(m)\} \\ &\quad + 2\,\|x_1\|\,\|Ax_2\| . \end{aligned}$$

Now, since x_2 and Ax_2 are both supported in C_i, there is z with $\|z\| = 1$, $\mathrm{supp}\, z \subseteq C_i$, and $\|Ax_2\| = z^*Ax_2 \leq \|x_2\|\,\|A_{\;i}\|$. Hence

$$\begin{aligned} \|A\| &\leq \sup\{\|A_E\| : E \in S(m)\}\{\|x_1\|^2 + 2\,\|x_1\|\,\|x_2\| + \|x_2\|^2\} \\ &\leq \sup\{\|A_E\| : E \in S(m)\}\max\{(a+b)^2 : a^2+b^2 = 1, a^2 \geq \frac{m}{m+r}\} \\ &= (1 + \frac{2\sqrt{mr}}{m+r})\sup\{\|A_E\| : E \in S(m)\}. \ \blacktriangle \end{aligned}$$

If A is, in addition to the above assumptions, non-negative entry-wise, or more generally if $\|A_E\| = \rho(A_E)$ holds for all $E \in S(m)$ and if $\|A\| = p(A)$, we get the following:

$$\rho(A) \leq \{1 + \frac{2\sqrt{mr}}{m+r}\} \sup\{\rho(A_E) : E \in S(m)\}. \tag{3.3}$$

We state one simple corollary.

Corollary 3.2. Let A be N × N, tridiagonal with non-negative entries. Then, for every $m \geq 4$

$$\rho(A) \leq \{1 + \frac{2\sqrt{2m}}{m+2}\} \sup\{\rho(A_E) : E \in S(m)\}. \tag{3.4}$$

For example with $m = 4$, the constant is less than 1.94281.

Proof. There is a non-negative diagonal matrix D such that $B := DAD^{-1}$ is Hermitian, in addition, $\rho(B_E) = \rho(A_E)$ and $\rho(B) = \rho(A)$ so we need only use (3.3) on B. ▲

Remarks: (1) C. Johnson has obtained a result similar to this one.

(2) The collection S(m) is considerably smaller than the collection of all m × m principal submatrices: S(m) has N + m-1 members and there are $\binom{N}{m}$ m × m principal submatrices.

(3) As $m \to \infty$, the constants in (3.2)-(3.4) tend to 1 like $O(m^{-1/2})$ so there is some degree of sharpness, We don't know, however, if the rate is sharp.

Perhaps, the best way to view many of the results here is to think of them as providing rates of convergence in the case that A is infinite.

Now, if A is not self-adjoint but is r-banded we have $A = S_1 + iS_2$ where S_1 and S_2 are self-adjoint: $S_1 = \frac{1}{2}(A + A^*)$, $S_2 = \frac{1}{2}(A - A^*)$. Since the function $B \to B_E$ is linear on the space of matrices, $(S_1)_E = \frac{1}{2} A_E + \frac{1}{2} A_E^*$ so $\|(S_1)_E\| \le \|A_E\|$ and $\|(S_2)_E\| \le \|A_E\|$. Therefore,

$$\|A\| \le \|S_1\| + \|S_2\| \le c(m)\{\sup_E \|(S_1)_E\| + \sup \|(S_2)_E\|\}$$

$$\le 2c(m) \sup_E \|A_E\| ,$$

where $c(m) = 1 + \frac{2\sqrt{mr}}{m+r}$ for $m \ge 2r$.

Thus, we have the following estimates for banded A.

Proposition 3.3. Let A be r-banded and let $m \geq 2r$. Then,

$$||A|| \leq 2\{1 + \frac{2\sqrt{mr}}{m+r}\} \sup\{||A_E|| : E \in S(m)\}.$$

The method of Lemma 3.1 readily gives inequalities for extremal eigenvalues of Hermitian matrices. Let $||x|| = 1$ with $x = x_1 + x_2$ and x_1, x_2 satisfying the norm and support conditions stated in the proof of Lemma 3.1. Also assume that A is r-banded and $m \geq 2r$. Recall, $x^*Ax = x_1^*Ax_1 + x_2{}^*Ax_2 + 2\mathrm{Re}(x_1^*Ax_2)$. With S(m) as in Lemma 3.1, we have choosing x so that $x^*Ax = \lambda_{\min}(A)$:

$$\lambda_{\min}(A) \geq ||x_1||^2 \lambda_{\min}(A_{c_j}) + ||x_2||^2 \lambda_{\min}(A_{c_i})$$

$$- \frac{2\sqrt{mr}}{m+r} \max_{E \in S(m)} ||A_E|| \geq \inf_{E \in S(m)} \lambda_{\min}(A_E) - \frac{2\sqrt{mr}}{m+r} \max_{E \in S(m)} ||A_E|| .$$

A similar inequality holds for $\lambda_{\max}(A)$, we summarize this in:

Proposition 3.4. Let $A = A^*$ be r-banded and let $m \geq 2r$, then

(a) $$0 \leq \lambda_{\max}(A) - \sup_{E \in S(m)} \lambda_{\max}(A_E) \leq \frac{2\sqrt{mr}}{m+r} ||A|| = O(m^{-1/2})$$

and

(b) $$0 \leq \inf_{E \in S(m)} \lambda_{\min}(A_E) - \lambda_{\min}(A) \leq \frac{2\sqrt{mr}}{m+r} ||A|| = O(m^{-1/2}).$$

The following test for positive definiteness of banded matrices now follows.

Corollary 3.5. Let $A = A^*$ be r-banded. If there is $m \geq 2r$ such that all contiguous $m \times m$ submatrices of A have minimal eigenvalue strictly longer than $2 \frac{\sqrt{mr}}{m+r} \|A\|$, then A is positive definite.

Remarks: (1) In the case that A is Toeplitz all A_E's are the same. With some regularity theorems on the symbol of A, Kac, Murdock, and Szegö [7] proved that $\lambda_{min}(A_E) - \lambda_{min}(A) = O(|E|^{-2})$. This line of though was further developed by Widom and Parter in the 1960's. The monographs of Hirschmann and Hughes [4] and Wilf [8] contain more details and complete references.

(2) Based on the results of [2] for the ℓ_∞ norm we suspect that the correct rate in Proposition 3.4 is $O(m^{-1})$.

(3) The results of this section can be extended to non-banded matrices simply by using the estimates used in Section 2. For example, the following result is similar to one of Johnson and Robinson [6; Theorem 3] but not as sharp; their bounds are of order 1/N and they also have inequalities for the intermediate eigenvalues.

Proposition 3.6. Let $A = A^*$ be an $N \times N$ matrix. Define A_i to be the principal minor indexed by $\{1,2,\ldots,N\}\backslash\{i\}$. Then

$$\text{(a)} \quad 0 \le \lambda_{max}(A) - \max_i \lambda_{max}(A_i) \le 2/\sqrt{N} \; \max_i \sqrt{\sum_{j=1}^{N} |A(i,j)|^2} + \frac{1}{N} \max_i |A(i,i)|$$

and

(b) $$0 \le \min_i \lambda_{\min}(A_i) - \lambda_{\min}(A) \le 2/\sqrt{N} \; \max_i \sqrt{\sum_{j=1}^{N} |A(i,j)|^2}$$

$$+ \frac{1}{N} \max_i |A(i,i)|.$$

<u>Proof</u>. Consider the $C(1,N-1)$ cover of $\{1,2,\ldots,N\}$ where $c_i = \{1,2,\ldots,N\}\backslash\{i\}$, for $1 \le i \le N$. Let $\|x\| = 1$ satisfy $x^*Ax = \lambda_{\max}(A)$ with $|x_i|^2 \le \frac{1}{N}$ for some i. Let $x_2 = x|_{\{i\}}$. So, $x = x_1 + x_2$ and $x^*Ax = x_1^*Ax_1 + x_2^*Ax_2 + 2\mathrm{Re}(x_1^*Ax_2)$. Thus,

$$\lambda_{\max}(A) \le \|x_1\|^2 \lambda_{\max}(A_{c_i}) + \|x_2\|^2 |A(i,i)| + 2\mathrm{Re}(x_1^*Ax_2)$$

$$\le \lambda_{\max}(A_i + \frac{1}{N} A(i,i) + 2 \|x_1\| \; \|Ax_2\|$$

$$\le \lambda_{\max}(A_i) + \frac{1}{N} |A(i,i)| + 2/\sqrt{N} \sqrt{\sum_{j=1}^{N} |A(i,j)|^2} .$$

We've used the fact that $x_2(j) = 0$ if $j = i$ to get a good estimate of $\|Ax_2\|$. The estimates for (b) are similar.

References

1. F. L. Bauer, On the field of values subordinate to a norm, Numer. Math. 4 (1962), 103-113.

2. S. Demko, Surjectivity and invertibility properties of totally positive matrices, Lin. Alg. and Its Applications 45 (1982), 13-20.

3. ———, Surjectivity and invertibility properties of certain matrix operators, preprint.

4. I. I. Hirschmann, Jr. and D. E. Hughes, Extreme eigenvalues of Toeplitz operators, Lecture Notes in Mathematics, No. 618, Springer, New York, 1977.

5. C. R. Johnson, Numerical ranges of principal submatrices, Lin. Alg. and Appl., 37 (1981), 23-34.

6. ——— and H. A. Robinson, Eigenvalue inequalities for principal submatrices, Lin. Alg. and Appl. 37 (1981), 11-22.

7. M. Kac, W. L. Murdock, and G. Szegö, On the eigenvalues of certain Hermitian forms, J. Rat. Mech. Anal., 2 (1953), 767-800.

8. H. S. Wilf, Finite Sections of Some Classical Inequalities, Ergebnisse der Mathematik und Ihrer Grenzgebiete, Band 52, Springer, New York, 1970.

On the Existence of Interpolating Projections Onto Spline Spaces*

by

Stephen Demko
School of Mathematics
Georgia Institute of Technology
Atlanta, Georgia 30332

AMS Subject Classification: 41A15

Key Words: Spline interpolation, equioscillation

*Supported in part by AFOSR Grant No. 82-0329.

TABLE OF CONTENTS

ABSTRACT

Sufficient conditions for the existence of a bounded interpolating projection onto subspaces of C[0,1] are found. For spaces of piecewise polynomial functions the projection can be bounded by the B-spline basis condition number. Infinite interpolation problems are also considered.

§1. Introduction

Let C[0,1] be the Banach space of continuous functions on the closed interval [0,1] and let S be a closed subspace. A linear operator $P: C[0,1] \to S$ is called an interpolating projection if there are points $\{t_i\}$ in [0,1] such that P has the definition: $Pf = s$ if and only if for all t_i $f(t_i) = s(t_i)$. We establish here some sufficient conditions on S for there to exist an interpolating projection onto S and give bounds on the norm of such a projection in terms of the geometry of S. The main corollary is that for spline spaces of fixed degree there is always an interpolating projection whose norm is less than or equal to B-spline basis condition number. This means that the norm of the projection does not depend on the number of knots or their distribution. The proof uses the facts that the B-spline collocation matrix is totally positive and that spline spaces are weak Chebyshev systems. These same properties were used by Goodman and Micchelli [7] recently to prove convergence of interpolating spline functions on a fixed periodic bi-infinite simple knot sequence as the degree of the splines goes to infinity. It seems that the existence of interpolating projections with norm depending on only the local degree of splines was known only for the cases of degree 1 (completely trivial), degree 2 (Marsden [11]), and degree 3 (de Boor [2]). Most approaches to spline interpolation have considered projection operators whose domains consisted of smooth functions or projections onto splines on uniform or quasi-uniform partitions or projections onto low degree spline spaces. The recent results of Jia [8] to the effect that, for high enough degree splines, interpolation at knot averages gives rise to projections whose norms grow with the number of mesh points (for a geometric mesh) make the results of this paper a bit more interesting than they would otherwise have been. The case of finite meshes is considered in Section 2 and that of infinite meshes in Section 3.

In what follows we assume that S has a (Schauder) basis $\{\phi_i\}$ with the following properties:

(1) there is a number m such that for all sequences $\{\alpha_i\}$

$$m \sup|\alpha_i| \leq \|\Sigma\alpha_i\phi_i\| \leq \sup|\alpha_i|$$

(2) for every choice of points $\{t_i\}$, the collocation matrices $(\phi_i(t_j))$ are totally non-negative, i.e., all minors are non-negative.

§2. Finite Dimensional Case

We first consider the case where the dimension of S is finite, say dim S = N. Since P is an interpolating projection with range S, the requirement $P\phi_k = \phi_k$ forces it to have the form

$$(Pf)(x) = \sum_{i=1}^{N} \left(\sum_{j=1}^{N} a_{ij} f(t_j) \right) \phi_i(x)$$

where $\{t_j\}$ are the interpolation points and the a_{ij}'s satisfy $(a_{ij})^{-1} = (\phi_j(t_i))$. By condition (1),

$$(3) \quad \|Pf\| \leq \max_i \left|\sum_j a_{ij} f(t_j)\right| \leq \|f\| \max_i \sum_j |a_{ij}|$$

and

$$\|Pf\| \geq m \cdot \max_i \left|\sum_j a_{ij} f(t_j)\right|.$$

If

$$\max_i \left|\sum a_{ij} f(t_j)\right| = \sum_j |a_{i_0 j}|,$$

then by choosing $\tilde{f}$ with $\|\tilde{f}\| = 1$ and $\tilde{f}(t_j) = \operatorname{sgn} a_{i_0 j}$ we get $\|P\| \geq \|Pf\| \geq m \cdot \max_i \sum_j |a_{ij}|$. Thus, the problem of bounding the norm of P is equivalent, modulo the quantity m, to bounding the ℓ_∞ operator norm of the inverse of the matrix whose $(i,j)^{th}$ entry is $\phi_j(t_i)$. With the ℓ_∞ vector norm $\|x\| := \max|x_i|$ and the associated matrix norm, we recall a result of de Boor [1].

Lemma 1. Let A be an $n \times n$ matrix whose $(n-1) \times (n-1)$ principal minors are all non-negative. Suppose there is a vector c such that $(Ac)(i)(-1)^i \geq \delta > 0$ for $1 \leq i \leq n$. Then, A is invertible and $\|A^{-1}\| \leq \|c\|/\delta$.

The key to our analysis is the following result which relies on a theorem of Jones and Karlovitz [9] which was used in a similar way in [7].

Lemma 2. Let S be an n-dimensional subspace C[0,1] having a basis $\{\phi_i\}_{i=1}^n$ satisfying (1) and (2). Then there is $s \in S$ and points $0 \leq t_1 < \ldots < t_n \leq 1$ such that $s(t_i) = (-1)^i = (-1)^i \|s\|$, $1 \leq i \leq n$.

Proof. By (2) $\{\phi_2,\ldots,\phi_n\}$ is a weak Chebyshev system. The Jones-Karlovitz result then says that there exist numbers $\alpha_2,\ldots,\alpha_n$ such that the function $\phi_1 - \sum_{i=2}^{n} \alpha_i\phi_i$ has points of equioscillation, $0 \leq t_1 < \ldots < t_n \leq 1$. Multiplying this function by an appropriate number will make it have norm 1.▲

Remark. This proof replaces an earlier version that (unnecessarily) invoked the Borsuk antipodal theorem. The author thanks Charles Micchelli for pointing out this simplification.

Theorem 1. Let S be an n-dimensional subspace of C[0,1] satisfying (1)-(2), then there are points $t_1 < \ldots < t_n$ such that the interpolating projection P: C[0,1] → S determined by these points has norm no greater than m^{-1} (from condition (1)).

Proof. Let $g \in S$ satisfy: $g(t_i) = (-1)^i = (-1)^i \|g\|$ for some $t_1 < \ldots < t_n$ as guaranteed by Lemma 2. Let $B = (b_{ij}) = (\phi_j(t_i))$ be the corresponding collocation matrix where $\{\phi_i\}$ is the basis for S satisfying (1)-(2). If $g = \Sigma c_j\phi_j$, then

$$1 = g(t_i)(-1)^i = \sum_j c_j\phi_j(t_i)(-1)^i = (-1)^i(Bc)(t_i).$$

Since $\|c\| \leq \frac{1}{m} \|g\| = \frac{1}{m}$, Lemma 1 gives $\|B^{-1}\| \leq m^{-1}$.

By (3), $||P|| \leq m^{-1}$. ∇

Since the condition number of the L_∞ normalized B-spline basis is $D_{k,\infty} \sim 2^k$, [3], we have

Corollary. Let S be a finite dimensional polynomial spline space in C[0,1], then there is an interpolating projection onto S with norm $\leq D_{k,\infty}$.

Remark 1. The use of Chebyshev's theorem makes the preceeding argument essentially univariate. The existence of nicely bounded interpolating projections -- or, more generally, projections determined by local, positive linear functionals -- in the case of multi-dimensional splines appears to be open.

2. We have not proven that the points coming out of Lemma 2 are unique. Nevertheless, Lemma 2 does give a verifiable condition that might lead to a Remez type algorithm for determining an equioscillating spline on a given mesh.

3. The fact that an interpolating projection onto a spline subspace could be bounded in terms of the (smallest) amplitudes of an oscillating spline function has been known for some time; it was made explicit by de Boor in [2].

4. The nature of the dependence of the points of interpolation on the given knot sequence and degree of the spline space is not revealed by the arguments in this paper. One natural question is whether or not one can choose good points of interpolation by using only local knot averages.

§3. Spline Interpolation on Bi-infinite Meshes

We consider the problem of interpolation of bounded data $\underline{y}: = \{y_i\}$ by functions of the form

$$\sum_{i=-\infty}^{\infty} \alpha_i N_{i,k}$$

where $\{N_{i,k}\}$ are B-splines of order k on some prescribed bi-infinite mesh. We assume condition (1) holds. In the spirit of the preceeding section we show only that it is possible to find points $\{t_i\}$ such that for any given $\underline{y}$ ℓ_∞ there is a unique $g = \Sigma\alpha_i N_{i,k}$ with $g(t_i) = y_i$ for all i and $||g|| \leq D_{k,\infty} ||y||$. Problems of infinite interpolation have been considered by several authors. In particular, both Micchelli's paper [12] and de Boor's [4] discuss their historical antecedants: the work of Schoenberg and Subbotin.

The results of §2 insure that for each $M \geq 1$, there are points $\{t_i^M: |i| \leq M\}$ such that

$$(4) \quad ||(N_{j,k}(t_i^M))^{-1}_{\substack{|i|\leq M\\|j|\leq M}}|| \leq D_{k,\infty}.$$

By [5], $t_i^M \in \{x: N_{i,k}(x) \geq D_{k,\infty}^{-1}\}$. By a diagonal argument we can find $\{t_i: -\infty < i < \infty\}$ and $\{M_\ell\}$ such that

$$\lim_{\ell\to\infty} t_i^{M_\ell} = t_i.$$

By continuity of matrix inversion,

$$\lim_{\ell\to\infty} ||(N_{j,k}(t_i))^{-1}_{\substack{|i|\leq M\\|j|\leq M}} - (N_{j,k}(t_i^{M_\ell}))^{-1}_{\substack{|i|\leq M\\|j|\leq M}}|| = 0$$

for any M. Thus, by (4) the finite sections of $(N_{j,k}(t_i))$ have inverses bounded by $D_{k,\infty}$. Let B_M denote the matrix

$$(N_{j,k}(t_i))_{\substack{|i|\le M\\ |j|\le M}} .$$

By Lemma 1 of [5],

$$\lim_{M\to\infty} B_M^{-1}(i,j) =: C(i,j)$$

exists for every i,j. Furthermore, the uniform (in M) exponential decay of $B_M^{-1}(i,j)$ as $|i-j| \to \infty$, [6], insures that $\sum_j |C(i,j)|$ converges; therefore, we can assert

$$\begin{aligned}\sum_j |C(i,j)| &= \sum_j \lim_M |B_M^{-1}(i,j)| \\ &= \lim_M \sum_j |B_M^{-1}(i,j)| \le D_{k,\infty}.\end{aligned}$$

Now, for any i,k $\sum_j C(i,j)N_j(t_k) = \sum_j \lim_M B_M^{-1}(i,j)N_j(t_k) =$ $\lim_M \sum_j B_M^{-1}(i,j)N_j(t_k) = \delta_{i,k}$. Similarly $(N_j(t_i))C = I$. So we have the following from which the interpolation results stated in the first paragraph of this section follow.

<u>Proposition</u>. Let $\{N_{j,k}\}$ be B-splines of order k on a given bi-infinite mesh. Then, there are points $\{t_i\}$ such that $\|(N_{j,k}(t_i))^{-1}\| \le D_{k,\infty}$.

Finally, it is clear that the arguments used here apply in the semi-infinite case.

REFERENCES

[1] C. de Boor, on the convergence of odd-degree spline interpolation, J. Approx. Th., 1 (1968), 452-468.

[2] ———, On bounding spline interpolation, J. Approx. Th., 14 (1975), 191-203.

[3] ———, A Practical Guide to Splines, Springer-Verlag, New York, 1978.

[4] ———, Odd-degree spline interpolation at a bi-infinite knot sequence, in Approximation Theory, R. Schaback and R. Scherer, eds., Lecture Notes in Mathematics, 556, Springer, Heidelberg, 1976, 30-53.

[5] ———, R. Jia, A. Pinkus, Structure of invertible (bi)infinite totally positive matrices, Lin. Alg. Appl., 47 (1982), 41-56.

[6] S. Demko, Inverses of band matrices and local convergence of spline projections, SIAM J. Numer. Anal., 14 (1977), 616-619.

[7] T. N. T. Goodman and C. A. Micchelli, Limits of spline functions with periodic knots, J. London Math. Soc. (2), 28 (1983), 103-112.

[8] R. Jia, Thesis, U. of Wisconsin, Madison, Wisconsin, 1983.

[9] R. C. Jones and L. A. Karlovitz, Equioscillation under nonuniqueness in approximation of continuous functions, J. Approx. Th., 3 (1970), 138-145.

[10] G. G. Lorentz, Approximation of Functions, Holt, Rinehart and Winston, New York, 1966.

[11] M. J. Marsden, Quadratic spline interpolation, BAMS, 80 (1974), 903-906.

[12] C. A. Micchelli, Infinite spline interpolation, in Approximation in Theorie und Praxis, G. Meindardus, ed., Bibliographisches Institut Mannheim, 1979, 209-238.