ISE-TR-82-29

# PERTURBATION THEOREMS FOR MATRIX EIGENVALUES

by

Yasuhiko Ikebe

Toshiyuki Inagaki

Sadaaki Miyamoto

September 11, 1982

INSTITUTE

OF

INFORMATION SCIENCES AND ELECTRONICS

UNIVERSITY OF TSUKUBA

# Perturbation Theorems for Matrix Eigenvalues

Yasuhiko Ikebe

Toshiyuki Inagaki

Sadaaki Miyamoto

Institute of Information Sciences and Electronics

University of Tsukuba, Tsukuba Science City

Ibaraki 305 JAPAN

ABSTRACT

This paper gives a unified derivation of a class of common perturbation theorems for matrix eigenvalues. We prove a fundamental inequality that appears not to have been reported despite its usefulness. By applying the fundamental inequality to special cases important in applications, we derive a class of inequalities that are useful in the localization of matrix eigenvalues.

## 1. INTRODUCTION

The matrix eigenvalue problem arises in a wide variety of areas in the physical and social sciences as well as in engineering, most typically, for example, in the stability analysis of physical systems that are modeled by linear systems of equations, differential equations, and so on.

Perturbation theorems on matrix eigenvalues are concerned with localization of eigenvalues, i.e., to produce regions in the complex plane in which eigenvalues of a given matrix lie. The theorems place bounds on the variation of the eigenvalues in terms of the variation of matrix elements. The information given by the theorems is useful in estimating true eigenvalues from computed or approximate eigenvalues, in analysing the stability of eigenvalues, and so on.

In this paper we are concerned with a unified derivation of a class of common perturbation theorems for matrix eigenvalues. To this end we prove first a basic inequality (see (2.1) below) which appears to be unreported in the literature. Some of the inequalities presented in this paper are well-known while others such as (2.1) and (3.9) appear to be less well-known despite their usefulness.

## 2. PRELIMINARIES

The vector and matrix norms considered in this paper are only the usual q-norm (or $\ell_q$-norm), $1 \leq q \leq \infty$ (see below for definition).

Some of the facts that follow may be generalized to a wider class of norms. However we will not discuss such generalization in this paper. Instead, we refer the interested reader to [2, Chapter 2].

We begin by reviewing basic facts on matrix norms. For each n = 1, 2,···, let $E^n$ denote a real or complex vector space of column vectors of dimension n, $x = (x_1,\ldots, x_n)^T$. The q-norm on $E^n$ is defined as follows:

$$(1.1)\qquad x = (x_1,\ldots, x_n)^T \in E^n \Rightarrow \begin{cases} \|x\|_q = \{|x_1|^q + \cdots + |x_n|^q\}^{1/q}, & 1 \le q < \infty, \\ \|x\|_\infty = \max_i |x_i|, & q = \infty. \end{cases}$$

Let B be any n x p real or complex matrix. Let $\|B\|_{q,q'}$ denote the norm of B as a linear transformation from $E^p$ to $E^n$, where $E^n$ is given the q-norm and $E^p$ is given the q'-norm, i.e.,

$$(1.2)\qquad \|B\|_{q,q'} = \max\{\|By\|_q / \|y\|_{q'} : y \neq 0,\ y \in E^p\}$$

$\|B\|_{q,q'}$ will be denoted by simply $\|B\|_q$ if q = q'. From the definition of the matrix norm $\|\cdot\|_{q,q'}$, we have

$$(1.3)\qquad \|By\|_q \le \|B\|_{q,q'} \cdot \|y\|_{q'},\quad y \in E^p.$$

<u>Example 1</u> [4, p.179]. Let B be any n x p matrix. Then

$$\|B\|_1 = \max_{1\le j\le p} \sum_{i=1}^{n} |b_{ij}|, \quad \text{\underline{matrix column-sum norm}};$$

$$\|B\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |b_{ij}|, \quad \text{\underline{matrix row-sum norm}};$$

$$\|B\|_{\infty,1} = \max_{\substack{1\le i\le n \\ 1\le j\le p}} |b_{ij}| .$$

Example 2 [3, p.21]. Let $\sigma_1^2(B) \geq \cdots \geq \sigma_p^2(B)$ be the eigenvalues of $B^H B$, a nonnegative-definite Hermitian p x p matrix, where B is n x p ($p \leq n$). The p numbers $\sigma_1(B), \ldots, \sigma_p(B) \geq 0$ are called the singular values of B. It may be proved that

$$\sigma_1(B) = \max_{y \neq 0} (\|By\|_2 / \|y\|_2) = \|B^H B\|_2^{1/2} = \|B\|_2 \text{, and}$$
$$\sigma_n(B) = \min_{y \neq 0} (\|By\|_2 / \|y\|_2) \text{ (always) } = \|(B^H B)^{-1}\|^{-1/2}$$

(provided $(B^H B)^{-1}$ exists).

Example 3 [3, p.52]. By diag $\{d_1, \ldots, d_n\}$ we denote the diagonal matrix with diagonal elements $d_1, \ldots, d_n$. Then

(1.4) $\|\text{diag}\{d_1, \ldots, d_n\}\|_{q,q'} = \max_i |d_i|$ , $1 \leq q' \leq q \leq \infty$.

In particular, $\|I\|_{q,q'} = 1$, $1 \leq q' \leq q \leq \infty$. Inequality (1.4) is generally false if $q' > q$. For example, $\|I\|_{1,\infty} = n$, where I is the n-th order identity matrix (cf. Example 1).

Remark. Let $\|\cdot\|$ be a norm on $E^n$. It is called an absolute norm if $\|x\| = \| \, |x| \, \|$ for every x in $E^n$, where $|x| = (|x_1|, \ldots, |x_n|)^T$ for $x = (x_1, \ldots, x_n)^T$. The norm $\|\cdot\|$ is called a monotone norm if $|x| \leq |y|$ (i.e., $|x_i| \leq |y_i|$ , $i = 1, \ldots, n$) implies $\|x\| \leq \|y\|$ .

The following three conditions are known to be equivalent [2, p.47]:

(i) $\|\cdot\|$ is an absolute norm;

(ii) $\|\cdot\|$ is a monotone norm;

(iii) $\|\cdot\|$ is such that for any diagonal matrix $D = \text{diag}\{d_1, \ldots, d_n\}$ ,

$$\max_{x \neq 0} (\|Dx\| / \|x\|) = \max_i |d_i|.$$

The q-norm on $E^n$ defined earlier is an example of absolute norm on $E^n$.

## 3. FUNDAMENTAL INEQUALITY

Let A, X and B be n x n, n x p and p x p matrices, respectively, where $p \leq n$. Let $\beta$ be an eigenvalue of B but not of A. Then for $1 \leq q, q' \leq \infty$,

$$\min_{y \neq 0}(\| Xy\|_q / \| y\|_{q'}) \leq \| (A-\beta I)^{-1}(AX-XB)\|_{q,q'} . \tag{2.1}$$

For proof, let $Bv = \beta v$, $v \neq 0$. Since $\beta$ is not an eigenvalue of A, $(A-\beta I)^{-1}$ exists and we compute

$$\begin{aligned} \| Xv\|_q &= \| (A-\beta I)^{-1}(A-\beta I)Xv\|_q = \| (A-\beta I)^{-1}(AX-XB)v\|_q \\ &\leq \| (A-\beta I)^{-1}(AX-XB)\|_{q,q'} \cdot \| v\|_{q'} \end{aligned}$$

by (1.3). From this (2.1) follows.

## 4. APPLICATIONS

We now give several applications of (2.1).

(1) Gerschgorin's Theorem [4, p.302]: Let $B = (b_{ij})$ be any n x n matrix and let $\beta$ be an eigenvalue such that $\beta \neq b_{ii}$, $i = 1, \ldots, n$. Let $A = \operatorname{diag}\{ b_{11}, \ldots, b_{nn}\}$ and let $X = I$. We apply (2.1) with A, X and B as indicated and $q = q' = \infty$. We obtain by Examples 1 and 3,

$$1 \leq \max_i \left( \sum_{\substack{j=1 \\ j \neq i}}^{n} |b_{ij}| / |b_{ii}-\beta| \right) ,$$

hence

$$|b_{ii}-\beta| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |b_{ij}| \quad \text{for some } i. \tag{3.1}$$

This inequality holds even when $\beta$ equals some diagonal element of B. Thus, every eigenvalue of B is contained in at least one of the <u>Gerschgorin disks</u> for B:

$$G_i = \{\lambda: |\lambda - b_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |b_{ij}|\} , \quad i = 1, \ldots, n.$$

(2) <u>Diagonalizable case</u>: Let $U^{-1}AU = \mathrm{diag}\{d_1, \ldots, d_n\} = D$ for some nonsingular matrix U. The $d_i$'s are the eigenvalues of A. Let X be any n x p matrix ($p \leq n$) whose columns are linearly independent. Let B be any p x p matrix. Let $\beta$ be an eigenvalue of B but not of A. Application of (2.1) to this case gives

$$1 \leq \|U\|_q \cdot \|(D-\beta I)^{-1}\|_q \cdot \|U^{-1}\|_q \|AX-XB\|_{q,q'} \;/\; \min_{y \neq 0} \left(\|Xy\|_q / \|y\|_{q'}\right)$$

where the minimum on the right-hand side is positive since X has linearly independent columns, and where

$$\|(D-\beta I)^{-1}\|_q = \max_i |d_i - \beta|^{-1} = \{\min_i |d_i - \beta|\}^{-1} \quad \text{(by Example 3)}.$$

Hence

$$\text{(3.2)} \quad \min_i |d_i - \beta| \leq \mathrm{cond}_q(U) \cdot \|AX-XB\|_{q,q'} \;/\; \min_{y \neq 0} \left(\|Xy\|_q / \|y\|_{q'}\right),$$

where $\mathrm{cond}_q(U) = \|U\|_q \cdot \|U^{-1}\|_q$, the q-<u>condition number</u> of U. Inequality (3.2) asserts that given any eigenvalue $\beta$ of B, there is an eigenvalue of A whose distance from $\beta$ does not exceed the number given by the right-hand side of (3.2).

(3) <u>Special case</u>: $q = q' = 2$. We obtain from (3.2)

$$\text{(3.3)} \quad \min_i |d_i - \beta| \leq \mathrm{cond}_2(U) \cdot \|AX-XB\|_2 \cdot \|(X^H X)^{-1}\|_2^{1/2} ,$$

where we used Example 2 and the fact that X has linearly independent columns so that $(X^H X)^{-1}$ exists.

(4) <u>Special case</u>: $q = q' = 2$ and A is normal ($A^H A = AA^H$). In this case a unitary matrix U ($U^H = U^{-1}$) exists such that $U^{-1}AU = \operatorname{diag}\{d_1, \ldots, d_n\}$. Since $\|U\|_2 = 1$ and $\|U^{-1}\|_2 = \|U^{-H}\|_2 = 1$ (by Example 2), (3.3) reduces to

$$(3.4) \quad \min_i |d_i - \beta| \leq \|AX-XB\|_2 \, \|(X^H X)^{-1}\|_2^{1/2} .$$

(5) <u>Special case</u>: $q = q'$, $n = p$ and $X = I$. We obtain from (3.2)

$$(3.5) \quad \min_i |d_i - \beta| \leq \operatorname{cond}_q(U) \cdot \|A-B\|_q .$$

This inequality is due to Bauer and Fike [1].

(6) <u>Special case</u> [3, p.53]: $p = 1$, $X = x \neq 0$ and $B = (\beta)$, where $\beta$ is a given number. Inequality (3.2) reduces to

$$(3.6) \quad \min_i |d_i - \beta| \leq \operatorname{cond}_q(U) \cdot \|Ax-\beta x\|_q \,/\, \|x\|_q .$$

Take the case $q = 2$. For a given A and an approximate eigenvector $x \neq 0$, the natural choice for $\beta$ is a well-known Rayleigh quotient $x^H Ax / x^H x$ which minimizes $\|Ax-\beta x\|_2$ as a function of $\beta$. This can be easily seen from the relation

$$\|Ax-\beta x\|_2^2 = \|Ax\|_2^2 - |x^H Ax|^2 + |z|^2$$

where $x^H x = 1$ and $\beta = x^H Ax + z$.

(7) Special case: $A = \begin{pmatrix} B & C \\ F & G \end{pmatrix}$, where A is n x n and B is p x p.

Let $U^{-1}AU = \mathrm{diag}\{d_1, \ldots, d_n\}$ for some U. We apply (3.2) with $X = (I_p, 0)^T$. We find

$$\| AX-XB \|_{q,q'} = \left\| \begin{pmatrix} B & C \\ F & G \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} - \begin{pmatrix} I \\ 0 \end{pmatrix} \cdot B \right\|_{q,q'} = \| F \|_{q,q'} .$$

Substitution into (3.2) gives

$$(3.7) \quad \min_i |d_i - \beta| \leq \mathrm{cond}_q(U) \cdot \| F \|_{q,q'} .$$

(8) Special case: $A = (a_{ij})$ is an n x n matrix, $B = (a_{kk})$ is a 1 x 1 matrix (p = 1) and $X = e_k$ = the k-th column of the identity matrix of order n. We still assume $U^{-1}AU = \mathrm{diag}\{d_1, \ldots, d_n\}$ for some U. Then (3.2) gives

$$(3.8) \quad \min_i |d_i - a_{kk}| \leq \mathrm{cond}_q(U) \cdot \{ \sum_{\substack{i=1 \\ i \neq k}}^{n} |a_{ik}|^q \}^{1/q} , \; k = 1, \ldots, n.$$

Taking the transpose we find

$$(3.8)' \quad \min_i |d_i - a_{kk}| \leq \mathrm{cond}_q(U) \cdot \{ \sum_{\substack{j=1 \\ j \neq k}}^{n} |a_{kj}|^q \}^{1/q} , \; k = 1, \ldots, n.$$

(9) Special case: $A = (a_{ij})$ is normal [2, Problem 6, p.86]. Let $U^{-1}AU = \mathrm{diag}\{\alpha_1, \ldots, \alpha_n\}$ with U unitary. Then $\mathrm{cond}_2(U) = 1$ as noted earlier. we take q = 2 in (3.8)' and obtain

$$(3.9) \quad \min_i |d_i - a_{kk}| \leq \{ \sum_{\substack{j=1 \\ j \neq k}}^{n} |a_{kj}|^2 \}^{1/2} , \; k = 1, \ldots, n.$$

Thus each disk

$$D_k = \{\lambda : |\lambda - a_{kk}| \leq \{\sum_{\substack{j=1 \\ j\neq k}}^{n} |a_{kj}|^2\}^{1/2}, \quad k = 1,\ldots, n$$

contains at least one eigenvalue of A _whether_ $D_k$ _overlaps with others or not_. The disks $D_k$ are smaller than the usual Gerschgorin disks for A (see (1) of this section). A stronger version of the Gerschgorin's theorem [4, p.303] states that if p Gerschgorin disks for A are disjoint from others, then the union of the p Gershcgorin disks contain exactly p eigenvalues of A counting multiple eigenvalues according to their multiplicities.

_Example 4_. Take a real symmetric (hence, normal) matrix

$$A = \begin{pmatrix} 1 & 10^{-2} & 10^{-4} \\ 10^{-2} & 10^{-4} & 10^{-6} \\ 10^{-4} & 10^{-6} & 10^{-8} \end{pmatrix}$$

Then

$$G_1 = \{\lambda : |\lambda-1| \leq 1.01 \times 10^{-2}\},$$
$$G_2 = \{\lambda : |\lambda-10^{-4}| \leq 1.0001 \times 10^{-2}\},$$
$$G_3 = \{\lambda : |\lambda-10^{-8}| \leq 1.01 \times 10^{-4}\}.$$

The disk $G_2$ contains $G_3$ but is disjoint from $G_1$. Hence the Gerschgorin's theorem asserts that there are two eigenvalues of A in $G_2$. On the other hand, (3.9) asserts that there is at least one eigenvalue of A in the disk

$$D_3 = \{\lambda : |\lambda-10^{-8}| \leq (1+10^{-8})^{1/2} \times 10^{-4}\},$$

which is properly contained in $G_3$.

## 5. CONCLUSION

We have provided a unified approach for deriving a class of perturbation theorems for matrix eigenvalues. In particular inequality (3.9) gives a useful complement to the well-known Gerschgorin's theorem in that the former requires no knowledge on the connectivity of Gerschgorin disks.

## REFERENCES


1. Bauer, F. L., and Fike, C. T. Norms and exclusion theorems, Numerische Mathematik, 2 (1960), 137-141.
2. Householder, A. S. The Theory of Matrices in Numerical Analysis, Blaisdell, New York (1964).
3. Ortega, J. M. Numerical Analysis: A Second Course, Academic Press, New York (1972).
4. Stewart, G. W. Introduction to Matrix Computations, Academic Press, New York (1973).

INSTITUTE OF INFORMATION SCIENCES AND ELECTRONICS
UNIVERSITY OF TSUKUBA
SAKURA-MURA, NIIHARI-GUN, IBARAKI 305 JAPAN

| REPORT DOCUMENTATION PAGE | REPORT NUMBER ISE-TR-82-29 |
|---|---|

TITLE

Pertubation Theorems for Matrix Eigenvalues

AUTHOR(S)

Yasuhiko Ikebe (Institute of Infromation Sciences and Electronics)

Toshiyuki Inagaki (Institute of Information Sciences and Electronics)

Sadaaki Miyamoto (Institute of Informaiton Sciences and Electronics)

| REPORT DATE | NUMBER OF PAGES |
|---|---|
| September 11, 1982 | 9 |
| MAIN CATEGORY | CR CATEGORIES |
| Numerical Analysis | 5.14, 5.11 |

KEY WORDS

Perturbation theorems, Matrix eigenvalues, Eigenvalue localization

ABSTRACT

This paper gives a unified derivation of a class of common perturbation theorems for matrix eigenvalues. We prove a fundamental inequality that appears not to have been reported despite its usefulness. By applying the fundamental inequality to special cases important in applications, we derive a class of inequalities that are useful in localization of matrix eigenvalues.

SUPPLEMENTARY NOTES