Lecture 8: Outline

- Chapter 6 + Appendix D: Location and perturbation of eigenvalues
- Some other results on perturbed eigenvalue problems
- Chapter 8: Nonnegative matrices

Eigenvector Perturbation Results, Motivation

We know from a previous lecture that $\rho(A) \leq ||A||$ for any matrix norm. That is, we know that all eigenvalues are in a circular disk with radius upper bounded by any matrix norm. Better results?

What can be said about the eigenvalues and eigenvectors of $A + \epsilon B$ when $\epsilon$ is small?

Geršgorin circles

Geršgorin’s Thm: Let $A = D + B$, where $D = \text{diag}(d_1, \ldots, d_n)$, and $B = [b_{ij}] \in M_n$ has zeros on the diagonal. Define

$$r_i'(B) = \sum_{j=1, j \neq i}^n |b_{ij}|$$

$$C_i(D, B) = \{ z \in \mathbb{C} : |z - d_i| \leq r_i'(B) \}$$

Then, all eigenvalues of $A$ are located in

$$\lambda_k(A) \in G(A) = \bigcup_{i=1}^n C_i(D, B) \quad \forall k$$

The $C_i(D, B)$ are called Geršgorin circles.

If $G(A)$ contains a region of $k$ circles that are disjoint from the rest, then there are $k$ eigenvalues in that region.

Geršgorin, Improvements

Since $A^T$ has the same eigenvalues as $A$, we can do the same but summing over columns instead of rows. We conclude that

$$\lambda_i(A) \in G(A) \cap G(A^T) \quad \forall i$$

Since $S^{-1}AS$ has the same eigenvalues as $A$, the above can be “improved” by

$$\lambda_i(A) \in G(S^{-1}AS) \cap G((S^{-1}AS)^T) \quad \forall i$$

for any choice of $S$. For it to be useful, $S$ should be “simple”, e.g., diagonal (see Corollary 6.1.6).
INVERTIBILITY AND STABILITY

If $A \in M_n$ is strictly diagonally dominant such that
\[ |a_{ii}| > \sum_{j=1, j \neq i}^{n} |a_{ij}| \quad \forall i \]
then
1. $A$ is invertible.
2. If all main diagonal elements are real and positive then all eigenvalues are in the right half plane.
3. If $A$ is Hermitian with all diagonal elements positive, then all eigenvalues are real and positive.

REDUCIBLE MATRICES

A matrix $A \in M_n$ is called reducible if
- $n = 1$ and $A = 0$ or
- $n \geq 2$ and there is a permutation matrix $P \in M_n$ such that
\[
P^TAP = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \begin{bmatrix} r \\ n-r \end{bmatrix} \]
for some integer $1 \leq r \leq n-1$.

A matrix $A \in M_n$ that is not reducible is called irreducible.

A matrix is irreducible iff it describes a strongly connected directed graph, "$A$ has the SC property".

IRREDUCIBLY DIAGONALLY DOMINANT

If $A \in M_n$ is called irreducibly diagonally dominant if
i) $A$ is irreducible (= $A$ has the SC property).
ii) $A$ is diagonally dominant,
\[ |a_{ii}| \geq \sum_{j=1, j \neq i}^{n} |a_{ij}| \quad \forall i \]
iii) For at least one row, $i$,
\[ |a_{ii}| > \sum_{j=1, j \neq i}^{n} |a_{ij}| \]

INVERTIBILITY AND STABILITY, STRONGER RESULT

If $A \in M_n$ is irreducibly diagonally dominant, then
1. $A$ is invertible.
2. If all main diagonal elements are real and positive then all eigenvalues are in the right half plane.
3. If $A$ is Hermitian with all diagonal elements positive, then all eigenvalues are real and positive.
**Perturbation theorems**

**Thm:** Let $A, E \in M_n$ and let $A$ be diagonalizable, $A = S\Lambda S^{-1}$.
Further, let $\hat{\lambda}$ be an eigenvalue of $A + E$. Then there is some eigenvalue $\lambda_i$ of $A$ such that
\[
|\hat{\lambda} - \lambda_i| \leq ||S|| ||S^{-1}|| ||E|| = \kappa(S)||E||
\]
for some particular matrix norms (e.g., $|| \cdot ||_1$, $|| \cdot ||_2$, $|| \cdot ||_\infty$).

**Cor:** If $A$ is a normal matrix, $S$ is unitary $\Rightarrow ||S||_2 = ||S^{-1}||_2 = 1$.
This gives
\[
|\hat{\lambda} - \lambda_i| \leq ||E||_2
\]
indicating that normal matrices are perfectly conditioned for eigenvalue computations.

**Perturbation cont’d**

If both $A$ and $E$ are Hermitian, we can use Weyl’s theorem (here we assume the eigenvalues are indexed in non-decreasing order):
\[
\lambda_1(E) \leq \lambda_k(A + E) - \lambda_k(A) \leq \lambda_n(E) \quad \forall k
\]
We also have for this case
\[
\sum_{k=1}^{n} |\lambda_k(A + E) - \lambda_k(A)|^2 \leq ||E||_2^{1/2}
\]
where $|| \cdot ||_2$ is the Frobenius norm.

**Perturbation of a simple eigenvalue**

Let $\lambda$ be a simple eigenvalue of $A \in M_n$ and let $y$ and $x$ be the corresponding left and right eigenvectors. Then $y^* x \neq 0$.

**Thm:** Let $A(t) \in M_n$ be differentiable at $t = 0$ and assume $\lambda$ is a simple eigenvalue of $A(0)$ with left and right eigenvectors $y$ and $x$. If $\lambda(t)$ is an eigenvalue of $A(t)$ for small $t$ such that $\lambda(0) = \lambda$ then
\[
\lambda'(0) = y^* A'(0) x / y^* x
\]

Example: $A(t) = A + tE$ gives $\lambda'(0) = y^* E x / y^* x$.

**Perturbation of eigenvalues cont’d**

Errors in eigenvalues may also be related to the residual $r = A\hat{x} - \hat{\lambda}\hat{x}$.
Assume for example that $A$ is diagonalizable $A = S\Lambda S^{-1}$ and let $\hat{x}$ and $\hat{\lambda}$ be a given complex vector and scalar, respectively. Then there is some eigenvalue of $A$ such that
\[
|\hat{\lambda} - \lambda| \leq \kappa(S) ||r|| / ||E||
\]
(for details and conditions see book).
We conclude that a small residual implies a good approximation of the eigenvalue.
LITERATURE WITH PERTURBATION RESULTS


Perturbation of eigenvectors with simple eigenvalues

**Thm:** Let $A(t) \in \mathbb{R}^{n \times n}$ be differentiable at $t = 0$ and assume $\lambda_0$ is a simple eigenvalue of $A(0)$ with left and right eigenvectors $y_0$ and $x_0$. If $\lambda(t)$ is an eigenvalue of $A(t)$, it has a right eigenvector $x(t)$ for small $t$ normalized such that $x_0 x(t) = 1$

with derivative

$$x'(0) = (\lambda_0 I - A(0))^{-1} \left( I - x_0 y_0^{T} \right) A'(0) x_0$$

$B^1$ denotes the Moore-Penrose pseudo inverse of a matrix $B$.


Perturbation of eigenvectors with simple eigenvalues: The real symmetric case

Assume that $A \in \mathbb{R}^{n \times n}$ is real symmetric matrix with normalized eigenvectors $x_i$ and eigenvalues $\lambda_i$. Further assume that $\lambda_1$ is a simple distinct eigenvalue. Let $\hat{A} = A + \epsilon B$ where $\epsilon$ is a small scalar, $B$ is real symmetric and let $\hat{x}_1$ be an eigenvector of $\hat{A}$ that approaches $x_1$ as $\epsilon \to 0$. Then a first order approximation (in $\epsilon$) is

$$\hat{x}_1 - x_1 = \epsilon \sum_{k=2}^{n} \frac{x_k^T B x_1}{\lambda_1 - \lambda_k} x_k$$

Warning: Non-unique derivative in the complex valued case!

Warning, Warning Warning: No extension to multiple eigenvalues!

Chapter 8: Nonnegative matrices

**Def:** A matrix $A = [a_{ij}]$ is nonnegative if $a_{ij} \geq 0$ for all $i, j$, and we write this as $A \geq 0$. (Note that this should not be confused with the matrix being nonnegative definite!) If $a_{ij} > 0$ for all $i, j$, we say that $A$ is positive and write this as $A > 0$. (We write $A > B$ to mean $A - B > 0$ etc.)

We also define $|A| = [|a_{ij}|]$.

Typical applications where nonnegative or positive matrices occur are problems in which we have matrices where the elements correspond to

- probabilities (e.g., Markov chains)
- power levels or power gain factors (e.g., in power control for wireless systems)
- any other application where only nonnegative quantities appear.
Nonnegative matrices: Some properties

Let \( A, B \in M_n \) and \( x \in \mathbb{C}^n \). Then

- \(|Ax| \leq |A||x|
- \(|AB| \leq |A||B|
- \) If \( A \geq 0 \), then \( A^m \geq 0 \); if \( A > 0 \), then \( A^m > 0 \).
- \) If \(|A| \leq |B|\), then \( \|A\| \leq \|B\| \), for any absolute norm \( \|\cdot\| \); that is, a norm for which \( \|A\| = \| |A| \| \).

Nonnegative matrices: Spectral radius

**Lemma:** If \( A \in M_n \), \( A \geq 0 \), and if the row sums of \( A \) are constant, then \( \rho(A) = \| |A| \|_\infty \). If the column sums are constant, then \( \rho(A) = \| |A| \|_1 \).

The following theorem can be used to give upper and lower bounds on the spectral radius of arbitrary matrices.

**Thm:** Let \( A, B \in M_n \). If \( |A| \leq B \), then \( \rho(A) \leq \rho(|A|) \leq \rho(B) \).

Positive matrices

For positive matrices we can say a little more.

**Perron’s theorem:** If \( A \in M_n \) and \( A > 0 \), then

1. \( \rho(A) > 0 \)
2. \( \rho(A) \) is an eigenvalue of \( A \)
3. There is an \( x \in \mathbb{R}^n \) with \( x > 0 \) such that \( Ax = \rho(A)x \)
4. \( \rho(A) \) is an algebraically (and geometrically) simple eigenvalue of \( A \)
5. \( |A| < \rho(A) \) for every eigenvalue \( \lambda \neq \rho(A) \) of \( A \)
6. \( |A| \rho(A)^m \rightarrow L \) as \( m \rightarrow \infty \), where \( L = xy^T \), \( Ax = \rho(A)x \), \( y^TA = \rho(A)y^T \), \( x > 0 \), \( y > 0 \), and \( x^Ty = 1 \).

The root \( \rho(A) \) is sometimes called a Perron root and the vector \( x = |x| \) a Perron vector if it is scaled such that \( \sum_{i=1}^n x_i = 1 \).
NONNEGATIVE MATRICES

Generalization of Perron’s theorem to general non-negative matrices?

**Thm:** If \( A \in M_n \) and \( A \geq 0 \), then
1. \( \rho(A) \) is an eigenvalue of \( A \)
2. There is a non-zero \( x \in \mathbb{R}^n \) with \( x \geq 0 \) such that \( Ax = \rho(A)x \)

For stronger results, we need a stronger assumption on \( A \).

IRREDUCIBLE MATRICES

**Remind:** A matrix \( A \in M_n \), \( n \geq 2 \) is called reducible if there is a permutation matrix \( P \in M_n \) such that
\[
P^TAP = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}
\]
for some integer \( 1 \leq r \leq n - 1 \).

A matrix \( A \in M_n \) that is not reducible is called irreducible.

**Thm:** A matrix \( A \in M_n \) with \( A \geq 0 \) is irreducible iff \( (I+A)^{n-1} > 0 \)

IRREDUCIBLE MATRICES

Frobenius’ theorem: If \( A \in M_n \), \( A \geq 0 \) is irreducible, then
1. \( \rho(A) > 0 \)
2. \( \rho(A) \) is an eigenvalue of \( A \)
3. There is an \( x \in \mathbb{R}^n \) with \( x > 0 \) such that \( Ax = \rho(A)x \)
4. \( \rho(A) \) is an algebraically (and geometrically) simple eigenvalue of \( A \)
5. If there are exactly \( k \) eigenvalues with \( |\lambda_p| = \rho(A) \), \( p = 1, \ldots, k \), then
   - \( \lambda_p = \rho(A)e^{i2\pi p/k} \), \( p = 0, 1, \ldots, k-1 \) (suitably ordered)
   - If \( \lambda \) is any eigenvalue of \( A \), then \( \lambda e^{i2\pi p/k} \) is also an eigenvalue of \( A \) for all \( p = 0, 1, \ldots, k-1 \)
   - \( \text{diag}[A^m] \equiv 0 \) for all \( m \) that are not multiples of \( k \) (e.g. \( m = 1 \)).

PRIMITIVE MATRICES

A matrix \( A \in M_n \), \( A \geq 0 \) is called primitive if
- \( A \) is irreducible
- \( \rho(A) \) is the only eigenvalue with \( |\lambda_p| = \rho(A) \).

**Thm:** If \( A \in M_n \), \( A \geq 0 \) is primitive, then
\[
\lim_{m \to \infty} [A/\rho(A)]^m = L
\]
where \( L = xy^T \), \( Ax = \rho(A)x \), \( y^TA = \rho(A)y^T \), \( x > 0 \), \( y > 0 \), and \( x^Ty = 1 \).

**Thm:** If \( A \in M_n \), \( A \geq 0 \), then it is primitive iff \( A^m > 0 \) for some \( m \geq 1 \).
Stochastic matrices

A nonnegative matrix with all its row sums equal to 1 is called a (row) stochastic matrix.

A column stochastic matrix is the transpose of a row stochastic matrix.

If a matrix is both row and column stochastic it is called doubly stochastic.

Stochastic matrices cont’d

The set of stochastic matrices in $M_n$ is a compact convex set.

Let $I = [1, 1, \ldots, 1]^T$. A matrix is stochastic if and only if $A I = 1 \implies 1$ is an eigenvector with eigenvalue $+1$ of all stochastic matrices.

An example of a doubly stochastic matrix is $A = [u_{ij}]$ where $U = |u_{ij}|$ is a unitary matrix. Also, notice that all permutation matrices are doubly stochastic.

Thm: A matrix is doubly stochastic if and only if it can be written as a convex combination of a finite number of permutation matrices.

Corr: The maximum of a convex function on the set of doubly stochastic matrices is attained at a permutation matrix.

Example, Markov processes

Consider a discrete stochastic process that at each time instant is in one of the states $S_1, \ldots, S_n$. Let $p_{ij}$ be the probability to change from state $S_i$ to state $S_j$, Note that the transition matrix $P = [p_{ij}]$, is a stochastic matrix.

Let $\mu_i(t)$ denote the probability of being in state $S_i$ at time $t$ and $\mu(t) = [\mu_1(t), \ldots, \mu_n(t)]$, then $\mu(t + 1) = \mu(t)P$ contains the corresponding probabilities for time $t + 1$. If $P$ is primitive (other terms are used in the statistics literature), then $\mu(t) \to \mu^\infty$ as $t \to \infty$ where $\mu^\infty = \mu^\infty P$, no matter what $\mu(0)$ is. $\mu^\infty$ is called the stationary distribution.


Further results

Other books contain more results.

In "Matrix Theory", vol. II by Gantmacher, for example, you can find results such as:

Thm: If $A \in M_n$, $A \geq 0$ is irreducible, then

$$(\alpha I - A)^{-1} > 0$$

for all $\alpha > \rho(A)$.

(Useful, for example, in connection with power control of wireless systems.)