Classical Numerical Analysis, Chapter 08

Abner J. Salgado and Steven M. Wise

asalgad1@utk.edu swise1@.utk.edu
University of Tennessee

Chapter 08, Part 1 of 2
Eigenvalue Problems

Computing Eigenvalues

The focus of this chapter will be the eigenvalue problem: given $A \in \mathbb{C}^{n \times n}$ we will be interested in finding pairs $(\lambda, \mathbf{x}) \in \mathbb{C} \times \mathbb{C}^n_\star$ such that

$$A\mathbf{x} = \lambda \mathbf{x}.$$

The possible applications of this problem are so vast that any attempt at listing them here will force us to misrepresent them.
One might argue that this chapter is in the *wrong* part of the book. Indeed, finding eigenvalues requires *nonlinear* methods, and it is a bit misleading to place it in the Numerical *linear* algebra part of our discussion. However, this is done for historical reasons.

Computing Eigenvalues

Why *nonlinear* rather than *linear*? Since, for a $\mathbf{x} \neq \mathbf{0}$, we have that $(\mathsf{A} - \lambda \mathsf{I})\mathbf{x} = \mathbf{0}$, and, therefore, that $\det(\mathsf{A} - \lambda \mathsf{I}) = 0$. Recall that the characteristic polynomial of the matrix A is defined by

$$\chi_{\mathsf{A}}(t) = \det(\mathsf{A} - t\mathsf{I}).$$

This shows that $\lambda \in \sigma(\mathsf{A})$ iff $\chi_{\mathsf{A}}(\lambda) = 0$. This suggests a naïve, approach to the problem at hand: to find the eigenvalues of a matrix, it is enough to find the roots of the characteristic polynomial, a nonlinear process. We now immediately see two issues. To find eigenvalues, following this approach we must:

1. Compute the characteristic polynomial χ_{A}.
2. Find its roots.

Example (Instability of the determinant)

Suppose that

$$A = \begin{bmatrix} 1 & 20 & & & \\ & 2 & 20 & & \\ & & \ddots & \ddots & \\ & & & 9 & 20 \\ & & & & 10 \end{bmatrix}.$$

There are zeros everywhere there is not a whole number above. It follows that $\sigma(A) = \{1, \dots, 10\}$. Now, for $0 < \varepsilon \ll 1$, consider also

$$A_\varepsilon = \begin{bmatrix} 1 & 20 & & & \\ & 2 & 20 & & \\ & & \ddots & \ddots & \\ & & & 9 & 20 \\ \varepsilon & & & & 10 \end{bmatrix}.$$

Example (Instability of the determinant, Cont.)

In other words, we have replaced one zero, in the bottom left corner, with a very small number. In any reasonable norm, $\|A_\varepsilon\|$ will be close to $\|A\|$. However, we can choose ε so that $0 \in \sigma(A_\varepsilon)$. To see this, consider

$$\begin{aligned}\det(A_\varepsilon) &= \det \begin{bmatrix} 2 & 20 & & \\ & \ddots & \ddots & \\ & & 9 & 20 \\ & & & 10 \end{bmatrix} - \varepsilon \det \begin{bmatrix} 20 & & & & \\ 2 & 20 & & & \\ & \ddots & \ddots & & \\ & & 9 & 20 & \\ & & & 10 & 20 \end{bmatrix} \\ &= 10! - 20^9 \varepsilon.\end{aligned}$$

Therefore, if

$$\varepsilon = \frac{10!}{20^9} \approx \frac{3 \times 10^6}{5 \times 10^{11}} \approx 7 \times 10^{-6}.$$

the matrix A_ε has zero as an eigenvalue.

Other Problems and the Solution

1. The determinant route is computationally expensive.
2. Small changes in the coefficients of the characteristic polynomial can lead to large changes in the roots.
3. The characteristic polynomials cannot be solved analytically.

Every practical eigenvalue approximation algorithm is iterative. We will discuss such approaches in this chapter.

Estimating Eigenvalues Using Gershgorin Discs

Definition (Gershgorin discs)

Let $n \geq 2$ and $\mathsf{A} \in \mathbb{C}^{n\times n}$. The **Gershgorin discs** D_i of A are

$$D_i = \{z \in \mathbb{C} \mid |z - a_{i,i}| \leq R_i\}, \quad R_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{i,j}|, \quad i = 1, \ldots, n.$$

Theorem (Gershgorin Circle Theorem)

Let $n \geq 2$ *and* $\mathsf{A} = [a_{i,j}] \in \mathbb{C}^{n \times n}$. *Then*

$$\sigma(\mathsf{A}) \subset \bigcup_{i=1}^{n} D_i .$$

Proof.

Suppose that $(\lambda, \mathbf{w})$ is an eigenpair of A. Then

$$\sum_{j=1}^{n} a_{i,j} w_j = \lambda w_i, \quad i = 1, 2, \ldots, n.$$

Suppose that

$$|w_k| = \|\mathbf{w}\|_\infty = \max_{i=1}^{n} |w_i|.$$

Observe that $w_k \neq 0$, since $\mathbf{w} \neq \mathbf{0}$.

Proof, Cont.

Then,

$$
\begin{aligned}
|\lambda - a_{k,k}| \cdot |w_k| &= |\lambda w_k - a_{k,k} w_k| \\
&= \left| \sum_{j=1}^{n} a_{k,j} w_j - a_{k,k} w_k \right| \\
&= \left| \sum_{\substack{j=1 \\ j \neq k}}^{n} a_{k,j} w_j \right| \\
&\leq \sum_{\substack{j=1 \\ j \neq k}}^{n} |a_{k,j} w_j| \\
&\leq \sum_{\substack{j=1 \\ j \neq k}}^{n} |a_{k,j}| \cdot |w_k| \\
&= R_k |w_k|.
\end{aligned}
$$

Thus $\lambda \in D_k$. $\square$

Theorem (Gershgorin second theorem)

Let $n \geq 2$ and $\mathsf{A} \in \mathbb{C}^{n \times n}$. Suppose that $1 \leq p \leq n-1$ and that the Gershgorin discs of the matrix A can be divided into disjoint subsets $D^{(p)}$ and $D^{(q)}$ containing p and $q = n - p$ discs, respectively. Then the union of the discs in $D^{(p)}$ contains p eigenvalues, and the union of the discs in $D^{(q)}$ contains q eigenvalues, counting multiplicities. In particular, if one disc is disjoint from all the others, it contains exactly one eigenvalue. And, if all of the discs are disjoint, then each contains exactly one eigenvalue.

Proof.

We will proceed with a technique that is commonly known as a *homotopy argument*. In particular, we construct a family of matrices $\mathsf{B}(\varepsilon)$, parameterized by $\varepsilon \in [0, 1]$, such that, we know the spectrum at $\varepsilon = 0$ completely. As ε increases from 0 to 1, we can follow the eigenvalue trajectories as they continuously deform. From this we will try to extract information for $\varepsilon = 1$. If we properly construct this family, then $\mathsf{B}(1) = \mathsf{A}$ and we will have proved the result.

Proof, Cont.

Define the matrix $\mathsf{B}(\varepsilon) = [b_{i,j}(\varepsilon)]_{i,j=1}^{n}$ as follows:

$$b_{i,j}(\varepsilon) = \begin{cases} a_{i,i}, & i = j, \\ \varepsilon a_{i,j}, & i \neq j. \end{cases}$$

Then $\mathsf{B}(1) = \mathsf{A}$ and $\mathsf{B}(0) = \mathbf{diag}(a_{1,1}, \ldots, a_{n,n})$. Each eigenvalue of $\mathsf{B}(0)$ is the center of one of the Gershgorin discs of A. Thus, exactly p of the eigenvalues of $\mathsf{B}(0)$ lie in the union of the discs in $D^{(p)}$.

The eigenvalues of the matrix $\mathsf{B}(\varepsilon)$ are the zeros of the characteristic polynomial of $\mathsf{B}(\varepsilon)$. The coefficients of this characteristic polynomial are continuous functions of the parameter ε, and, in turn, the zeros of the polynomial are continuous functions of ε. We will accept this rather deep fact without proof.

Proof, Cont.

As ε increases from 0 to 1, the eigenvalues of $\mathrm{B}(\varepsilon)$ move in continuous paths in the complex plane. Since the degree of the characteristic polynomial of $\mathrm{B}(\varepsilon)$ is always exactly n, none of the zeros of the characteristic polynomial diverge to infinity. At the same time, the radii of the Gershgorin discs increase from 0 to R_i, respectively. In particular, it is easy to see that they increase as

$$R_{\varepsilon,i} = \varepsilon R_i, \quad 0 \leq \varepsilon \leq 1, \quad i = 1, \ldots, n.$$

Since p of the eigenvalues of $\mathrm{B}(\varepsilon)$ lie in the union of the discs in $D^{(p)}$ when $\varepsilon = 0$, and these discs are disjoint from those in $D^{(q)}$, these p eigenvalues will stay within the union of the discs in $D^{(p)}$, for all values of ε, $0 \leq \varepsilon \leq 1$. □

Theorem (almost diagonal)

Let $n \geq 2$ and suppose that in the matrix $\mathsf{A} \in \mathbb{C}^{n\times n}$ all off–diagonal elements are smaller in modulus than $\varepsilon > 0$, i.e.,

$$|a_{i,j}| < \varepsilon, \quad 1 \leq i, j \leq n, \quad i \neq j.$$

Suppose that there is $\delta > 0$ such that, for some $r \in \{1, \ldots, n\}$, the diagonal element $a_{r,r}$ satisfies

$$|a_{r,r} - a_{i,i}| > \delta, \quad \forall\, i \neq r.$$

Then, provided

$$\varepsilon < \frac{\delta}{2(n-1)},$$

there is an eigenvalue $\lambda \in \sigma(\mathsf{A})$ such that

$$|\lambda - a_{r,r}| < \frac{2(n-1)}{\delta}\varepsilon^2 < \frac{\delta}{2(n-1)}.$$

Proof.

Let $\kappa > 0$. Define $\mathsf{K} = [k_{i,j}] \in \mathbb{C}^{n\times n}$ via

$$k_{i,j} = \begin{cases} \kappa, & i = j = r, \\ \delta_{i,j}, & \text{otherwise.} \end{cases}$$

Define the similar matrix $\mathsf{A}_\kappa = \mathsf{K}\mathsf{A}\mathsf{K}^{-1}$. For example, if $r = 3$, $n = 4$,

$$\mathsf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix} \in \mathbb{C}^{4\times 4},$$

then

$$\mathsf{A}_\kappa = \begin{bmatrix} a_{1,1} & a_{1,2} & \kappa^{-1} a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & \kappa^{-1} a_{2,3} & a_{2,4} \\ \kappa a_{3,1} & \kappa a_{3,2} & a_{3,3} & \kappa a_{3,4} \\ a_{4,1} & a_{4,2} & \kappa^{-1} a_{4,3} & a_{4,4} \end{bmatrix}.$$

Proof, Cont.

The Gershgorin disc of A_κ, with respect to row r, has its center at $a_{r,r}$, and its radius is

$$R_{\kappa,r} = \sum_{\substack{j=1 \\ j\neq r}}^{n} |\kappa a_{i,j}| = \kappa \sum_{\substack{j=1 \\ j\neq r}}^{n} |a_{i,j}| < \kappa \sum_{\substack{j=1 \\ j\neq r}}^{n} \varepsilon = \kappa(n-1)\varepsilon.$$

The disc of A_κ corresponding to row $i \neq r$ has $a_{i,i}$ as its center, and its radius is

$$R_{\kappa,i} = \sum_{\substack{j=1 \\ j\neq i \\ j\neq r}}^{n} |a_{i,j}| + \kappa^{-1}|a_{i,r}| < (n-2)\varepsilon + \frac{\varepsilon}{\kappa}.$$

Now, pick $\kappa = \frac{2\varepsilon}{\delta}$. Then,

$$R_{\kappa,r} < \frac{2\varepsilon^2(n-1)}{\delta}, \qquad R_{\kappa,i} < \frac{\delta}{2} + (n-2)\varepsilon, \quad i \neq r.$$

Proof, Cont.

Therefore, for $i \neq r$,

$$R_{\kappa,r} + R_{\kappa,i} < \frac{2\varepsilon^2(n-1)}{\delta} + \frac{\delta}{2} + (n-2)\varepsilon < \varepsilon + \frac{\delta}{2} + (n-2)\varepsilon < \delta,$$

on the assumption that $\varepsilon < \frac{\delta}{2(n-1)}$. Finally, we have

$$\delta < |a_{r,r} - a_{i,i}|,$$

but $R_{\kappa,r} + R_{\kappa,i} < \delta$. Therefore, the two discs $D_{\kappa,r}$ and $D_{\kappa,i}$ must be disjoint. Hence, there is an eigenvalue $\lambda \in \sigma(\mathsf{A}_\kappa) = \sigma(\mathsf{A})$, with $\lambda \in D_{\kappa,r}$. In other words,

$$|\lambda - a_{r,r}| \leq R_{\kappa,r} < \frac{2\varepsilon^2(n-1)}{\delta} < \frac{\delta}{2(n-1)}. \qquad \square$$

Stability

Theorem (Bauer–Fike)

Suppose that $\mathsf{A} \in \mathbb{C}^{n \times n}$ *is diagonalizable, i.e., there is an invertible matrix* $\mathsf{P} \in \mathbb{C}^{n \times n}$ *and a diagonal matrix* $\mathsf{D} = \mathbf{diag}\left(\lambda_1, \ldots, \lambda_n\right)$ *such that* $\mathsf{A} = \mathsf{P}\mathsf{D}\mathsf{P}^{-1}$*. Suppose that* $\lambda \notin \sigma(\mathsf{A}) = \{\lambda_i\}_{i=1}^n$ *is an eigenvalue of the perturbed matrix* $\mathsf{A} + \delta\mathsf{A}$*, with* $\delta\mathsf{A} \in \mathbb{C}^{n \times n}$*. Then, for any* $p \in [1, \infty]$*, we have*

$$\min_{i=1}^{n} |\lambda - \lambda_i| \leq \|\mathsf{P}\|_p \left\|\mathsf{P}^{-1}\right\|_p \|\delta\mathsf{A}\|_p = \kappa_p(\mathsf{P}) \|\delta\mathsf{A}\|_p .$$

Proof.

There is some $\mathbf{x} \in \mathbb{C}_\star^n$ such that

$$(\mathsf{A} + \delta\mathsf{A})\,\mathbf{x} = \lambda\mathbf{x}.$$

It follows that

$$(\lambda\mathsf{I}_n - \mathsf{A})\,\mathbf{x} = \delta\mathsf{A}\mathbf{x},$$

and

$$\mathsf{P}\left(\lambda\mathsf{I}_n - \mathsf{D}\right)\left(\mathsf{P}^{-1}\mathbf{x}\right) = \delta\mathsf{A}\mathsf{P}(\mathsf{P}^{-1}\mathbf{x}).$$

Proof, Cont.

Note that $(\lambda I_n - D)$ is invertible, since $\lambda \notin \sigma(A)$. Hence,

$$\left\|P^{-1}\mathbf{x}\right\|_p \le \left\|(\lambda I_n - D)^{-1}\right\|_p \left\|P^{-1}\delta A P\right\|_p \left\|P^{-1}\mathbf{x}\right\|_p .$$

Observe that, for any $p \in [1, \infty]$

$$\left\|(\lambda I_n - D)^{-1}\right\|_p = \max_{i=1}^{n} \frac{1}{|\lambda - \lambda_i|} = \frac{1}{\min_{i=1}^{n} |\lambda - \lambda_i|}.$$

Therefore,

$$\min_{i=1}^{n} |\lambda - \lambda_i| \le \|P\|_p \left\|P^{-1}\right\|_p \|\delta A\|_p ,$$

as desired. $\square$

Corollary (Hermitian matrix I)

If $\mathsf{A} \in \mathbb{C}^{n\times n}$ *is Hermitian, then*

$$\min_{i=1}^{n} |\lambda - \lambda_i| \leq \|\delta \mathsf{A}\|_2 \,.$$

Proof.

If A is Hermitian, it is unitarily diagonalizable. Since the 2–norm of a unitary matrix equals one, the result follows. □

Corollary (Hermitian matrix II)

Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ *is Hermitian,* $\lambda \in \mathbb{R}$*, and* λ *is closest to the eigenvalue* λ_r*, i.e.,*

$$r = \operatorname*{argmin}_{1\le i\le n} |\lambda_i - \lambda|.$$

Then, setting, $\lambda = \lambda_r + \delta\lambda_r$*,*

$$|\delta\lambda_r| \le \|\delta\mathsf{A}\|_2 .$$

Proof.

Exercise. □

Remark (stability)

In other words, for Hermitian matrices, the eigenvalue problem is stable to perturbations in the coefficient matrix A*. In the general case, the eigenvalue problem is stable, but a type of condition number,* $\|\mathsf{P}\|_p \left\|\mathsf{P}^{-1}\right\|_p$*, appears on the left hand side.*

Theorem (eigenvector perturbation)

Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ *is Hermitian and* $\sigma(\mathsf{A}) = \{\lambda_1, \ldots, \lambda_n\} \subset \mathbb{R}$. *Suppose that* $\mathbf{x} \in \mathbb{C}^n_\star$ *and* $\lambda \in \mathbb{C}$ *are given and define*

$$\mathbf{w} = \mathsf{A}\mathbf{x} - \lambda\mathbf{x}.$$

Then

$$\min_{1\le i\le n} |\lambda_i - \lambda| \le \frac{\|\mathbf{w}\|_2}{\|\mathbf{x}\|_2}.$$

Proof.

Exercise. □

The Rayleigh Quotient for Hermitian Matrices

Definition (Rayleigh quotient)

Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ is Hermitian. The **Rayleigh quotient** of $\mathbf{x} \in \mathbb{C}^n_\star$ is

$$R(\mathbf{x}) = \frac{(\mathsf{A}\mathbf{x}, \mathbf{x})_2}{(\mathbf{x}, \mathbf{x})_2}.$$

Proposition (properties of the Rayleigh quotient)

Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ *is Hermitian with spectrum* $\sigma(\mathsf{A}) = \{\lambda_i\}_{i=1}^n \subset \mathbb{R}$, *where the following ordering is imposed:*

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{n-1} \le \lambda_n.$$

Then

1. $R(\mathbf{x}) \in \mathbb{R}$, *for all* $\mathbf{x} \in \mathbb{C}^n_\star$.
2. *For all* $\mathbf{x} \in \mathbb{C}^n_\star$,
$$\min_{j=1}^{n} \lambda_j \le R(\mathbf{x}) \le \max_{j=1}^{n} \lambda_j.$$
3. $R(\mathbf{x}) = \lambda_k$ *iff* $\mathbf{x}$ *is an eigenvector associated to* λ_k.
4. *For a fixed* $\mathbf{x} \in \mathbb{C}^n_\star$, *the function*
$$Q(\alpha) = \|\mathsf{A}\mathbf{x} - \alpha\mathbf{x}\|_2^2, \quad \forall\, \alpha \in \mathbb{C},$$
has a unique global minimum, and, in fact, the Rayleigh quotient is the unique minimizer
$$R(\mathbf{x}) = \operatorname*{argmin}_{\alpha\in\mathbb{C}} \|\mathsf{A}\mathbf{x} - \alpha\mathbf{x}\|_2^2.$$

Theorem (eigenvalue estimate)

Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ *is Hermitian with spectrum* $\sigma(\mathsf{A}) = \{\lambda_i\}_{i=1}^n \subset \mathbb{R}$ *and an associated orthonormal basis of eigenvectors* $\{\mathbf{w}_i\}_{i=1}^n$*. Suppose that* $\mathbf{x} \in \mathbb{C}_\star^n$ *is a vector with the property that*

$$\|\mathbf{x} - \mathbf{w}_k\|_2 < \varepsilon, \quad \|\mathbf{x}\|_2 = 1,$$

for some positive number ε*. Then*

$$|R(\mathbf{x}) - \lambda_k| < 2\rho(\mathsf{A})\varepsilon^2.$$

Proof.

Suppose that $\mathbf{x} \neq \mathbf{w}_k$. There are unique constants $\alpha_i \in \mathbb{C}$, such that

$$\mathbf{x} = \sum_{j=1}^n \alpha_j \mathbf{w}_j, \qquad \|\mathbf{x}\|_2^2 = \sum_{j=1}^n |\alpha_j|^2 = 1.$$

Proof, Cont.

It follows from orthonormality that

$$0 < \|\mathbf{x} - \mathbf{w}_k\|_2^2 = \sum_{\substack{j=1 \\ j \neq k}}^{n} |\alpha_j|^2 + |\alpha_k - 1|^2.$$

Therefore, using our assumptions,

$$0 < \sum_{\substack{j=1 \\ j \neq k}}^{n} |\alpha_j|^2 + |\alpha_k - 1|^2 = \|\mathbf{x} - \mathbf{w}_k\|_2^2 < \varepsilon^2. \quad (1)$$

Since $\|\mathbf{x}\|_2^2 = 1$, it follows from (1) that

$$0 \leq \sum_{\substack{j=1 \\ j \neq k}}^{n} |\alpha_j|^2 = 1 - |\alpha_k|^2 < \varepsilon^2. \quad (2)$$

Proof, Cont.

Finally,

$$R(\mathbf{x}) = \frac{\sum_{j=1}^{n} \lambda_j |\alpha_j|^2}{\sum_{j=1}^{n} |\alpha_j|^2} = \lambda_k |\alpha_k|^2 + \sum_{\substack{j=1 \\ j \neq k}}^{n} \lambda_j |\alpha_j|^2,$$

so that, using (2),

$$\begin{aligned}
|R(\mathbf{x}) - \lambda_k| &= \left| \lambda_k \left(|\alpha_k|^2 - 1 \right) + \sum_{\substack{j=1 \\ j \neq k}}^{n} \lambda_j |\alpha_j|^2 \right| \\
&\leq |\lambda_k| \left(1 - |\alpha_k|^2 \right) + \sum_{\substack{j=1 \\ j \neq k}}^{n} |\lambda_j| \cdot |\alpha_j|^2 \\
&\leq \max_{j=1}^{n} |\lambda_j| \left(1 - |\alpha_k|^2 \right) + \max_{j=1}^{n} |\lambda_j| \sum_{\substack{j=1 \\ j \neq k}}^{n} |\alpha_j|^2 \\
&\leq \rho(\mathsf{A}) \varepsilon^2 + \rho(\mathsf{A}) \varepsilon^2 \\
&= 2\rho(\mathsf{A}) \varepsilon^2.
\end{aligned}$$