Numerical Matrix Analysis

Notes #18 — Eigenvalue Problems: Introduction

Peter Blomgren
⟨blomgren@sdsu.edu⟩

Department of Mathematics and Statistics
Dynamical Systems Group
Computational Sciences Research Center
San Diego State University
San Diego, CA 92182-7720

http://terminus.sdsu.edu/

Spring 2024
(Revised: March 26, 2024)

Outline

San Diego State University

Student Learning Targets, and Objectives

Target Eigenvalues — Introduction

- **Dictionary** Diagonalization, Unitary Diagonalization, Unitary Triangularization, Eigenvalue, Spectrum, Eigenspace, Invariant Subspace, Algebraic and Geometric Multiplicity, Characteristic Polynomial
- **Objective** Eigenvalue decomposition as a change of basis
- **Objective** Normality $\Rightarrow$ Unitary Diagonalizability
- **Objective** Schur Factorization
- **Objective** The Abel-Ruffini Theorem, and consequences

Eigenvalue Problems: Introduction

We (p)review the language and properties associated with eigenvalue problems, and describe three examples of matrix factorizations which reveal the eigenvalues of a matrix A:

- The **diagonalization** $A = X\Lambda X^{-1} \Leftrightarrow X^{-1}AX = \Lambda$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, and the columns of X contains the eigenvectors of A.
- The **unitary diagonalization** $A = Q\Lambda Q^* \Leftrightarrow Q^*AQ = \Lambda$.
- The **unitary triangularization** (a.k.a. **Schur factorization**) $A = QTQ^* \Leftrightarrow Q^*AQ = T$, where T is upper triangular, and the eigenvalues of A appear on the diagonal of T.

We discuss under what circumstances each of these factorizations exist.

Note: Fundamentals in [MATH 254], and deeper theory in [MATH 524].

Eigenvalues and Eigenvectors

Let $A \in \mathbb{C}^{m\times m}$ be a square matrix. A non-zero vector $\vec{x} \in \mathbb{C}^m$ is an **eigenvector** of A, and λ is the corresponding **eigenvalue** if

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

The set of all eigenvalues of a matrix A is the **spectrum** of A, commonly denoted by $\lambda(A)$, or $\Lambda(A)$.

The usefulness of eigenvalues and eigenvectors

Algorithmic Eigenvalue analysis can **simplify solutions** by reducing a coupled system to a collection of scalar problems.

Physical Eigenvalue analysis can give insight to the behavior of evolving systems governed by linear equations, *e.g.* the study of **resonance** and **stability** of physical systems.

Eigenvalue Decomposition

The eigenvalue decomposition of a square matrix is the factorization

$$A = X \Lambda X^{-1},$$

where X is non-singular and Λ diagonal. To make the connection between eigenvalues and eigenvectors clear, this decomposition can be rewritten

$$AX = X\Lambda.$$

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_m \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_m \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}$$

Showing that

$$A\vec{x}_j = \lambda_j \vec{x}_j, \quad j = 1, \ldots, m.$$

San Diego State University

Eigenvalue Decomposition: A Change of Basis

The eigenvalue decomposition can be viewed as a *change of basis* [NOTES#3.4 (MATH 254)] to **"eigenvector coordinates."**

In solving the linear system $A\vec{x} = \vec{b}$, with $A = X\Lambda X^{-1}$,

$$\Lambda(\underbrace{X^{-1}\vec{x}}_{\vec{y}}) = \underbrace{X^{-1}\vec{b}}_{\vec{c}}$$

we

- expand $\vec{x}$ (implicitly) and $\vec{b}$ (explicitly) in the basis $\mathfrak{X}$ given by the columns X;
- apply (solve with the diagonal) Λ; and
- interpret the result as a vector of coefficients $\vec{y} = [\vec{x}]_{\mathfrak{X}}$ of a linear combination of the columns of X, so that $\vec{x} = X\vec{y}$.

Eigenvalues: Geometric Multiplicity

The set of eigenvectors corresponding to a single eigenvalue, together with the zero-vector, form a subspace of $\mathbb{C}^m$ known as an **eigenspace**.

If $\lambda \in \Lambda(A)$, we denote the corresponding eigenspace by E_λ.

An eigenspace E_λ is an example of an **invariant subspace** of A, *i.e.* $AE_\lambda \subseteq E_\lambda$. — Shorthand for $\vec{x} \in E_\lambda \Rightarrow A\vec{x} \in E_\lambda$.

The dimension of E_λ can be interpreted as the maximum number of linearly independent eigenvectors that can be found corresponding to the eigenvalue λ. This is the **geometric multiplicity** [MATH 254] of λ, $\mathrm{gm}(\lambda)$.

We note that

$$\begin{aligned} E_\lambda &= \mathrm{null}(A - \lambda I_{m\times m}). \\ \mathrm{gm}(\lambda) &= \dim(E_\lambda) \end{aligned}$$

The Characteristic Polynomial $\rightsquigarrow$ Eigenvalues

The **characteristic polynomial** of $A \in \mathbb{C}^{m \times m}$, is the polynomial of degree m defined by

$$p_A(z) = \det(A - zI_{m\times m}).$$

The following theorem is (hopefully) well-known

Theorem (Eigenvalues are Roots of Characteristic Polynomial)

λ is an eigenvalue of A if and only if $p_A(\lambda) = 0$.

We note that even if A is real, the eigenvalues may be complex.

Further, we note that from previous discussion — recall Wilkinson's example in [NOTES#9] on the ill-conditioning of the root-finding problem. Looking for roots to the characteristic polynomial is **not** a stable way to identify eigenvalues!

Eigenvalues: Algebraic Multiplicity

By the **fundamental theorem of algebra**, $p_A(z)$ can be factored

$$p_A(z) = c(z-\lambda_1)^{m_1}(z-\lambda_2)^{m_2} \cdots (z-\lambda_r)^{m_r},$$

where

$$\sum_{k=1}^{r} m_k = m.$$

The integers $m_k \geq 1$ indicate the **algebraic multiplicity** of the eigenvalue $\lambda_k \in \mathbb{C}$.

The following is true

$$\textbf{Algebraic multiplicity}(\lambda_k) \geq \textbf{Geometric multiplicity}(\lambda_k)$$

This result comes from a discussion of **similarity transformations**.

Similarity Transformations

If $X \in \mathbb{C}^{m \times m}$ is non-singular, then the map

$$A \mapsto X^{-1}AX,$$

is called a **similarity transformation** of A. Two matrices A and B are **similar** if there exists a non-singular $X \in \mathbb{C}^{m \times m}$ such that $B = X^{-1}AX$.

We care about similarity transformations because:

Theorem

If $X \in \mathbb{C}^{m \times m}$ is non-singular, then A and $X^{-1}AX$ have the same characteristic polynomial, eigenvalues, and algebraic and geometric multiplicities.

Similarity Transformations...

The proof of the theorem is very straight-forward:

$$
\begin{aligned}
p_{X^{-1}AX}(z) &= \det(X^{-1}AX - zI) = \det(X^{-1}(A - zI)X) \\
&= \det(X^{-1}) \det(A - zI) \det(X) \\
&= \det(A - zI) = p_A(z).
\end{aligned}
$$

Since $p_{X^{-1}AX}(z) = p_A(z)$ the agreement on eigenvalues, and algebraic multiplicities follow. The agreement of geometric multipliers follows from the fact that if E_λ is an eigenspace for A, then $X^{-1}E_\lambda$ is an eigenspace for $X^{-1}AX$, and conversely. $\square$

With this result in our back-pocket we can show

Theorem

The algebraic multiplicity of an eigenvalue λ is at least as great as its geometric multiplicity.

Proof: Algebraic multiplicity $\geq$ Geometric multiplicity

Let n be the geometric multiplicity of λ for the matrix A. Form an $(m \times n)$ matrix $\widehat{V}$ whose n columns constitute an orthonormal basis of E_λ.

Proof: Algebraic multiplicity $\geq$ Geometric multiplicity

Let n be the geometric multiplicity of λ for the matrix A. Form an $(m \times n)$ matrix $\widehat{V}$ whose n columns constitute an orthonormal basis of E_λ. Let V be the square unitary matrix whose first n columns are given by $\widehat{V}$, and define B by

$$B = V^* A V = \begin{bmatrix} \lambda I_{n \times n} & C \\ 0 & D \end{bmatrix}, \quad C \in \mathbb{C}^{n \times (m-n)}, \; D \in \mathbb{C}^{(m-n) \times (m-n)}.$$

SAN DIEGO STATE UNIVERSITY

Proof: Algebraic multiplicity $\geq$ Geometric multiplicity

Let n be the geometric multiplicity of λ for the matrix A. Form an $(m \times n)$ matrix $\widehat{V}$ whose n columns constitute an orthonormal basis of E_λ. Let V be the square unitary matrix whose first n columns are given by $\widehat{V}$, and define B by

$$B = V^* A V = \begin{bmatrix} \lambda I_{n\times n} & C \\ 0 & D \end{bmatrix}, \quad C \in \mathbb{C}^{n\times(m-n)}, \; D \in \mathbb{C}^{(m-n)\times(m-n)}.$$

By the properties of the determinant,

$$\begin{aligned} \det(B - zI_{m\times m}) &= \det((\lambda - \mathbf{z})\mathbf{I}_{\mathbf{n}\times\mathbf{n}}) \det(D - zI_{(m-n)\times(m-n)}) \\ &= (\lambda - \mathbf{z})^{\mathbf{n}} \det(D - zI_{(m-n)\times(m-n)}). \end{aligned}$$

San Diego State University

Proof: Algebraic multiplicity $\geq$ Geometric multiplicity

Let n be the geometric multiplicity of λ for the matrix A. Form an $(m \times n)$ matrix $\widehat{V}$ whose n columns constitute an orthonormal basis of E_λ. Let V be the square unitary matrix whose first n columns are given by $\widehat{V}$, and define B by

$$B = V^* A V = \begin{bmatrix} \lambda I_{n \times n} & C \\ 0 & D \end{bmatrix}, \quad C \in \mathbb{C}^{n \times (m-n)}, \; D \in \mathbb{C}^{(m-n)\times(m-n)}.$$

By the properties of the determinant,

$$\begin{aligned} \det(B - zI_{m\times m}) &= \det((\lambda - \mathbf{z})\mathbf{I}_{\mathbf{n}\times\mathbf{n}}) \det(D - zI_{(m-n)\times(m-n)}) \\ &= (\lambda - \mathbf{z})^{\mathbf{n}} \det(D - zI_{(m-n)\times(m-n)}). \end{aligned}$$

Hence, the algebraic multiplicity of λ as an eigenvalue of B is at least n. Since similarity transformations preserve multiplicities, the same is true for A. $\square$

Defective (Non-Diagonalizable) Matrices

When **Algebraic multiplicity** $>$ **Geometric multiplicity**, the matrix is not diagonalizable. Consider

$$A = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 & \\ & 2 & 1 \\ & & 2 \end{bmatrix}.$$

Both A and B have $\lambda = 2$ with **algebraic multiplicity 3**. For A we can choose 3 linearly independent eigenvectors, but for B there is only one linearly independent eigenvector

$$\vec{x}_{A_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \vec{x}_{A_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \ \vec{x}_{A_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{x}_{B_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Geometric multiplicities of $\lambda = 2$ are **3 (for A) and 1 (for B)**.

Defective (Non-Diagonalizable) Matrices

An eigenvalue whose algebraic multiplicity exceeds its geometric multiplicity, is a **defective eigenvalue**. A matrix that has one or more defective eigenvalues is a **defective matrix**.

A non-defective matrix is **diagonalizable** —

Theorem

An $(m \times m)$ matrix A is non-defective if and only if it has an eigenvalue decomposition $A = X \Lambda X^{-1}$.

This result quantifies for what matrices the diagonalization is (theoretically) computable. — The matrix X may be highly ill-conditioned, which may prevent us from numerically performing the diagonalization.

Special Cases: Unitary Diagonalization

In rare circumstances, we come across a matrix $A \in \mathbb{C}^{m \times m}$ whose m eigenvectors not only are linearly independent, but also orthogonal.

In this case A is **unitarily diagonalizable**, *i.e.* there exists a unitary matrix Q such that

$$A = Q \Lambda Q^*.$$

Since $\|Q\|_2 = 1$, there is no ill-conditioning to worry about.

What kind of matrices are unitarily diagonalizable???

Unitarily Diagonalizable Matrices

Theorem ($A^* = A \Rightarrow$ Real Diagonalizable)

A Hermitian matrix is unitarily diagonalizable, and its eigenvalues are real.

Other example of unitarily diagonalizable matrices include

- **Skew-Hermitian** matrices, $S^* = -S$.
- **Unitary matrices**, $U^* = U^{-1}$, $U^*U = I$.
- **Circulant matrices**, C, whose rows are composed of cyclically shifted versions of a length-n list ℓ.
- Any of the above **plus** a multiple of the identity.

Unitarily Diagonalizable Matrices

Theorem ($A^* = A \Rightarrow$ Real Diagonalizable)

A Hermitian matrix is unitarily diagonalizable, and its eigenvalues are real.

Other example of unitarily diagonalizable matrices include

- **Skew-Hermitian** matrices, $S^* = -S$.
- **Unitary matrices**, $U^* = U^{-1}$, $U^*U = I$.
- **Circulant matrices**, C, whose rows are composed of cyclically shifted versions of a length-n list ℓ.
- Any of the above **plus** a multiple of the identity.

These types of matrices are all **normal**, *i.e.* $M^*M = MM^*$.

Unitarily Diagonalizable Matrices

Theorem ($A^* = A \Rightarrow$ Real Diagonalizable)

A Hermitian matrix is unitarily diagonalizable, and its eigenvalues are real.

Other example of unitarily diagonalizable matrices include

- **Skew-Hermitian** matrices, $S^* = -S$.
- **Unitary matrices**, $U^* = U^{-1}$, $U^*U = I$.
- **Circulant matrices**, C, whose rows are composed of cyclically shifted versions of a length-n list ℓ.
- Any of the above **plus** a multiple of the identity.

These types of matrices are all **normal**, *i.e.* $M^*M = MM^*$.

Theorem ($AA^* = A^*A \Rightarrow$ Complex Diagonalizable)

A matrix is unitarily complex diagonalizable if and only if it is normal.

[Complex/Real Spectral Theorem (Math 524 Notes#7.1)].

The Schur Factorization — Eigen**values** Only!

If we are interested in **numerically computing** the **eigenvalues** only, then the Schur factorization is the most useful approach.

The **Schur factorization** of a matrix A is a unitary factorization

$$A = QTQ^*,$$

where Q is unitary, and T is upper triangular. Since this is a similarity transform, the eigenvalues of A must appear on the diagonal of T.

The Schur Factorization — Eigen**values** Only!

The following theorem indicates why this is a useful approach —

Theorem

Every square matrix A has a Schur factorization.

Hence it should be possible to compute the eigenvalues for any matrix, without having to worry about ill-conditioning in the X (here Q) matrix which defines the similarity transformation.

The Schur Factorization: Existence Proof

The proof is by induction. The base case $m = 1$ is trivially true.

The Schur Factorization: Existence Proof

The proof is by induction. The base case $m = 1$ is trivially true.

Let $m \geq 2$ [**the inductive hypothesis says that there exists a Schur factorization of all $(m-1) \times (m-1)$ matrices**], and let $(\lambda, \vec{x})$ be any eigenvalue-eigenvector pair of A.

The Schur Factorization: Existence Proof

The proof is by induction. The base case $m = 1$ is trivially true.

Let $m \geq 2$ [**the inductive hypothesis says that there exists a Schur factorization of all $(m-1) \times (m-1)$ matrices**], and let $(\lambda, \vec{x})$ be any eigenvalue-eigenvector pair of A. Let $\vec{u}_1 = \vec{x}/\|\vec{x}\|_2$ be the first column of a unitary matrix U. Then by construction,

$$U^* A U = \begin{bmatrix} \lambda & B \\ 0 & C \end{bmatrix},$$

where $B \in \mathbb{C}^{1\times(m-1)}$, and $C \in \mathbb{C}^{(m-1)\times(m-1)}$.

San Diego State University

The Schur Factorization: Existence Proof

The proof is by induction. The base case $m = 1$ is trivially true.

Let $m \geq 2$ [**the inductive hypothesis says that there exists a Schur factorization of all $(m-1) \times (m-1)$ matrices**], and let $(\lambda, \vec{x})$ be any eigenvalue-eigenvector pair of A. Let $\vec{u}_1 = \vec{x} / \|\vec{x}\|_2$ be the first column of a unitary matrix U. Then by construction,

$$U^* A U = \begin{bmatrix} \lambda & B \\ 0 & C \end{bmatrix},$$

where $B \in \mathbb{C}^{1 \times (m-1)}$, and $C \in \mathbb{C}^{(m-1) \times (m-1)}$.

Now, by the induction hypothesis $C = VTV^*$ for some unitary $V \in \mathbb{C}^{(m-1) \times (m-1)}$, and upper-triangular $T \in \mathbb{C}^{(m-1) \times (m-1)}$. Therefore we can define

$$Q = U \begin{bmatrix} 1 & \\ & V \end{bmatrix}.$$

The Schur Factorization: Existence Proof

Q is unitary, and

$$
\begin{aligned}
Q^* A Q &= \begin{bmatrix} 1 & \\ & V^* \end{bmatrix} U^* A U \begin{bmatrix} 1 & \\ & V \end{bmatrix} \\
&= \begin{bmatrix} 1 & \\ & V^* \end{bmatrix} \begin{bmatrix} \lambda & B \\ 0 & C \end{bmatrix} \begin{bmatrix} 1 & \\ & V \end{bmatrix} \\
&= \begin{bmatrix} 1 & \\ & V^* \end{bmatrix} \begin{bmatrix} \lambda & BV \\ 0 & CV \end{bmatrix} \\
&= \begin{bmatrix} \lambda & BV \\ 0 & T \end{bmatrix}
\end{aligned}
$$

which is the Schur factorization we want. $\square$

Eigenvalue-Revealing Factorizations

We have described three eigenvalue-revealing factorizations

Type	Form	Restrictions on A	Vectors
Diagonalization	$A = X\Lambda X^{-1}$	Non-defective	$\surd$
Unitary Diagonalization	$A = Q\Lambda Q^*$	Normal, $A^*A = AA^*$	$\surd$
Schur Triangularization	$A = QTQ^*$	None	—

Note that the diagonalizations also give the eigenvectors, whereas the eigenvector information is lost in the Schur triangularization.

Factorizations based on unitary transformations tend to lead to algorithms that are numerically stable.

If A is normal, then the Schur form comes out diagonal; and if we know that A is Hermitian we can take advantage of the symmetry in order to save (approximately) half the work.

Computing Eigenvalues: Algorithms

Even though eigenvalues and eigenvectors have straight-forward definitions and clean characterizations, the best ways to compute them are not obvious.

Some of the most "obvious" ways of approaching the problem — *e.g.* by extracting the roots of the characteristic polynomial — are not stable.

The **power iteration** which generates the sequence

$$\frac{\vec{x}}{\|\vec{x}\|}, \quad \frac{A\vec{x}}{\|A\vec{x}\|}, \quad \frac{A^2\vec{x}}{\|A^2\vec{x}\|}, \quad \frac{A^3\vec{x}}{\|A^3\vec{x}\|}, \quad \ldots$$

will, under some weak conditions, converge to the eigenvector corresponding to the largest (in absolute value) eigenvalue. This approach is **slow** in general — and it only gives one eigenvector.

Computing Eigenvalues: Algorithms

General purpose eigenvalue algorithms are based on computing eigenvalue-**revealing** factorizations of A.

Depending on the properties of the matrix A, we can base our algorithms on diagonalization, unitary diagonalization, or unitary triangularization.

Clearly, we are going to pull out the tools we have already developed for generating algorithms that "put zeros into matrices."

Although, the flavor is related, eigenvalue computations are distinctly different (and fundamentally more difficult) than solutions of linear systems, or least squares problems.

The Difficulty of Eigenvalue Computations

We have seen that we can cast the eigenvalue problem as a root-finding problem (subject to potentially catastrophic ill-conditioning).

Conversely, any polynomial root-finding problem can be stated as an eigenvalue problem, *e.g.* given the polynomial

$$p(z) = z^m + a_{m-1}z^{m-1} + \cdots + a_1 z + a_0,$$

we can write $p(z) = (-1)^m \cdot \det(A - zI)$, where

$$A - zI = \begin{bmatrix} -z & & & & & -a_0 \\ 1 & -z & & & & -a_1 \\ & 1 & -z & & & -a_2 \\ & & 1 & \ddots & & \vdots \\ & & & \ddots & -z & -a_{m-2} \\ & & & & 1 & -(z + a_{m-1}) \end{bmatrix}.$$

The Difficulty of Eigenvalue Computations

Therefore the roots of $p(z)$ are the eigenvalues of the matrix

$$A = \begin{bmatrix} 0 & & & & & -a_0 \\ 1 & 0 & & & & -a_1 \\ & 1 & 0 & & & -a_2 \\ & & 1 & \ddots & & \vdots \\ & & & \ddots & 0 & -a_{m-2} \\ & & & & 1 & -a_{m-1} \end{bmatrix}$$

We are in quite a predicament! — It is well known that there is no formula for expressing the roots of an arbitrary polynomial given its coefficients.

The Difficulty of Eigenvalue Computations

Theorem (Abel-Ruffini Theorem)

For any $m \geq 5$, there is a polynomial $p(z)$ of degree m with rational coefficients that has a real root $p(r) = 0$ with the property that r **cannot** *be written using any expression involving rational numbers, addition, subtraction, multiplication, division, and kth roots.*

This theorem seems to spell out a lot of gloom-and-doom: even in exact arithmetic, there can be no computer program that produces the exact roots of an arbitrary polynomial in a finite number of steps.

The theorem is named after Paolo Ruffini, who provided an incomplete proof in 1799, and Niels Henrik Abel, who provided a proof in 1824. (Galois later proved more general statements, and provided a construction of a polynomial of degree 5 whose roots cannot be expressed in radicals from its coefficients.)

A Different Angle of Attack

The preceding discussion does not mean that we cannot generate a good eigenvalue solver. It does, however, indicate that we have to think "outside the box" (where the box is our present toolbox of algorithms).

Gaussian elimination and Householder reflections would solve linear systems of equations exactly in a finite number of steps if they could be implemented in exact arithmetic. However:

Fact

Any eigenvalue solver must be iterative.

We are going to generate sequences of numbers converging rapidly toward the eigenvalues. — The need for iterations may seem discouraging; however, in most cases we can define schemes that converge very rapidly — doubling or tripling the number of digits of accuracy in each iteration.