SI231 Matrix Analysis and Computations
Topic 5: Eigenvalues, Eigenvectors, and Eigendecomposition

Ziping Zhao

Spring Term 2020–2021

School of Information Science and Technology
ShanghaiTech University, Shanghai, China

Topic 5: Eigenvalues, Eigenvectors, and Eigendecomposition

- facts about eigenvalues and eigenvectors
- eigendecomposition, the case of Hermitian & real symmetric matrices
- Schur decomposition
- variational characterizations of eigenvalues of Hermitian & real symmetric matrices
- similarity transformation
- power iteration, inverse iteration, Rayleigh quotient iteration
- orthogonal iteration
- QR iteration
- PageRank: a case study
- generalized eigenvalue problem

Notation and Conventions

- a square matrix $\mathbf{A}$ is said to be symmetric if $a_{ij} = a_{ji}$ for all i, j with $i \neq j$, or equivalently, if $\mathbf{A}^T = \mathbf{A}$

  – example:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & -2 & 0.9 \\ 3 & 0.9 & 0 \end{bmatrix}$$

- a square matrix $\mathbf{A}$ is said to be Hermitian if $a_{ij} = a_{ji}^*$ for all i, j with $i \neq j$, or equivalently, if $\mathbf{A}^H = \mathbf{A}$

  – example:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 3 + \boldsymbol{j}0.7 \\ 0 & -2 & 0.9 - \boldsymbol{j} \\ 3 - \boldsymbol{j}0.7 & 0.9 + \boldsymbol{j} & 0 \end{bmatrix}$$

- we denote the set of all $n \times n$ real symmetric matrices by $\mathbb{S}^n$
- we denote the set of all $n \times n$ complex Hermitian matrices by $\mathbb{H}^n$

Notation and Conventions

- note the following subtleties:
  - by definition, a real symmetric matrix is also Hermitian
  - when we say that a matrix is Hermitian, we often imply that the matrix may be complex (at least for this course); a real Hermitian matrix is simply real symmetric
  - we can have a complex symmetric matrix, though we will not study it

Main Results

A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$) is said to admit an eigendecomposition if there exists a nonsingular $\mathbf{V} \in \mathbb{C}^{n\times n}$ and a collection of scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ such that

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1},$$

where $\mathbf{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$.

- the above $(\mathbf{V}, \mathbf{\Lambda})$ satisfies $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ for $i = 1, \ldots, n$, which are eigen-equations
- $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are required to be linearly independent
- eigendecomposition *does not* always exist

Main Results

A real symmetric matrix $\mathbf{A} \in \mathbb{S}^n$ always admits an eigendecomposition

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$$

where $\mathbf{V} \in \mathbb{R}^{n\times n}$ is orthogonal; $\mathbf{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_i \in \mathbb{R}$ for all i.

A Hermitian matrix $\mathbf{A} \in \mathbb{H}^n$ always admits an eigendecomposition

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$$

where $\mathbf{V} \in \mathbb{C}^{n\times n}$ is unitary; $\mathbf{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_i \in \mathbb{R}$ for all i.

- differences: a Hermitian or real symmetric matrix always has
  - an eigendecomposition
  - real λ_i's
  - a $\mathbf{V}$ that is not only nonsingular but also unitary

Eigenvalues and Eigenvectors

We start with the basic definition of eigenvalues and eigenvectors.

Problem: given a $\mathbf{A} \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$), find a vector $\mathbf{v} \in \mathbb{C}^n$ with $\mathbf{v} \neq \mathbf{0}$ such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \qquad \text{for some } \lambda \in \mathbb{C} \tag{$*$}$$

- $(*)$ is called an eigenvalue problem or eigen-equation
- let $(\mathbf{v}, \lambda)$ be a solution to $(*)$. We call
  - $(\mathbf{v}, \lambda)$ an eigen-pair of $\mathbf{A}$
  - λ an eigenvalue of $\mathbf{A}$; $\mathbf{v}$ an eigenvector of $\mathbf{A}$ associated with λ
- if $(\mathbf{v}, \lambda)$ is an eigen-pair of $\mathbf{A}$, $(\alpha\mathbf{v}, \lambda)$ is also an eigen-pair for any $\alpha \in \mathbb{C}, \alpha \neq 0$
- unless specified, we will assume $\|\mathbf{v}\|_2 = 1$ in the sequel

Eigenvalues and Eigenvectors

Problem: given a $\mathbf{A} \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$), find a vector $\mathbf{v} \in \mathbb{C}^n$ with $\mathbf{v} \neq \mathbf{0}$ such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \qquad \text{for some } \lambda \in \mathbb{C} \tag{$*$}$$

- from $(*)$, action of matrix $\mathbf{A}$ on a subspace $\mathcal{V}_\lambda \subseteq \mathbb{C}^n$ sometimes is equivalent to scalar multiplication

$$\mathcal{V}_\lambda = \{\mathbf{v} \in \mathbb{C}^n \mid \mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \text{ for some } \lambda \in \mathbb{C}\}$$

- the subspace $\mathcal{V}_\lambda$ satisfying $(*)$ is called the eigenspace of $\mathbf{A}$ associated with λ, and any $\mathbf{v} \in \mathcal{V}_\lambda$ with $\mathbf{v} \neq \mathbf{0}$ is an eigenvector
  - $\mathcal{V}_\lambda = \mathcal{N}(\mathbf{A} - \lambda\mathbf{I})$ since $\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \iff (\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$
  - $\mathcal{V}_\lambda$ is an invariant subspace of $\mathbf{A}$, i.e., $\mathbf{A}\mathcal{V}_\lambda \subset \mathcal{V}_\lambda$
- the set of all eigenvalues of $\mathbf{A}$, denoted by $\sigma(\mathbf{A}) \subseteq \mathbb{C}$, is called the spectrum of $\mathbf{A}$

Eigenvalues and Eigenvectors

Fact: Every $\mathbf{A} \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$) has n eigenvalues.

- from the eigenvalue problem we see that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \text{ for some } \mathbf{v} \neq \mathbf{0} \iff (\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0} \text{ for some } \mathbf{v} \neq \mathbf{0}$$
$$\iff \det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

- let $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$, called the characteristic polynomial of $\mathbf{A}$; eigenvalues are sometimes called characteristic values
- from the determinant def., it can be shown that $p(\lambda)$ is a polynomial of degree n, viz., $p(\lambda) = \alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \ldots + \alpha_n\lambda^n$ where α_i's depend on $\mathbf{A}$
  - specifically, $\alpha_0 = \det(\mathbf{A})$, $\ldots$, $\alpha_{n-1} = -(-1)^n\mathrm{tr}(\mathbf{A})$, $\alpha_n = (-1)^n$
- as $p(\lambda)$ is a polynomial of degree n, it can be factored as $p(\lambda) = \prod_{i=1}^n(\lambda_i - \lambda)$ where $\lambda_1, \ldots, \lambda_n$ are the roots of $p(\lambda)$
  - specifically, $\alpha_0 = \prod_{i=1}^n \lambda_i$, $\ldots$, $\alpha_{n-1} = (-1)^{n-1}\sum_{i=1}^n \lambda_i$, $\alpha_n = (-1)^n$
- we have $\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \iff \lambda \in \{\lambda_1, \ldots, \lambda_n\}$
- the spectrum is hence defined by $\sigma(\mathbf{A}) = \{\lambda \in \mathbb{C} \,|\, \det(\mathbf{A} - \lambda\mathbf{I}) = 0\}$

Eigenvalues and Eigenvectors

Let $\lambda_1, \ldots, \lambda_n$ denote the n eigenvalues of $\mathbf{A}$. We write

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \qquad i = 1, \ldots, n,$$

where $\mathbf{v}_i$ denotes an eigenvector of $\mathbf{A}$ associated with λ_i.

- we should be careful about the meaning of n eigenvalues: *they are defined as the n roots of the characteristic polynomial $p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$*
- example: consider

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

  - from the original definition $\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$, one can verify that $\lambda = 1$ is the only eigenvalue of $\mathbf{A}$
  - from the characteristic polynomial, which is $p(\lambda) = (1 - \lambda)^2$, we see two roots $\lambda_1 = \lambda_2 = 1$ as two eigenvalues
  - every matrix has at least one eigenvalue, and every eigenvalue appears at least once in $p(\lambda)$

Eigenvalues and Eigenvectors

Fact: an eigenvalue can be complex even if $\mathbf{A}$ is real.

- a polynomial $p(\lambda) = \alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \ldots + \alpha_n\lambda^n$ with real coefficients α_i's can have complex roots
- example: consider

$$\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

  – we have $p(\lambda) = \lambda^2 + 1$, so $\lambda_1 = \boldsymbol{j}$, $\lambda_2 = -\boldsymbol{j}$
- similarily, an eigenvalue can be real even if $\mathbf{A}$ is complex

Fact: if $\mathbf{A}$ is real and there exists a real eigenvalue λ of $\mathbf{A}$, the associated eigenvector $\mathbf{v}$ can be taken as real.

- obviously, when $\mathbf{A} - \lambda\mathbf{I}$ is real we can define $\mathcal{N}(\mathbf{A} - \lambda\mathbf{I})$ on $\mathbb{R}^n$
- or, if $\mathbf{v}$ is a complex eigenvector of a real $\mathbf{A}$ associated with a real λ, we can write $\mathbf{v} = \mathbf{v}_{\mathrm{R}} + \boldsymbol{j}\mathbf{v}_{\mathrm{I}}$, where $\mathbf{v}_{\mathrm{R}}, \mathbf{v}_{\mathrm{I}} \in \mathbb{R}^n$. It is easy to verify that $\mathbf{v}_{\mathrm{R}}$ and $\mathbf{v}_{\mathrm{I}}$ are eigenvectors associated with λ

Further Discussion: Repeated/Degenerate Eigenvalues

- w.l.o.g., order $\lambda_1, \ldots, \lambda_n$ such that $\{\lambda_1, \ldots \lambda_k\}$, $k \leq n$, is the set of all distinct eigenvalues of $\mathbf{A}$; i.e., $\lambda_i \neq \lambda_j$ for all $i, j \in \{1, \ldots, k\}$, $i \neq j$; $\lambda_i \in \{\lambda_1, \ldots \lambda_k\}$ for all $i \in \{1, \ldots, n\}$

- denote μ_i as the number of repeated eigenvalues of λ_i, $i = 1, \ldots, k$
  - i.e., μ_i is the multiplicity of λ_i as the root of $p(\lambda)$ (λ_i is simple if $\mu_i = 1$)
  - μ_i is called the algebraic multiplicity of the eigenvalue λ_i

- every λ_i can have more than one eigenvector (scaling not counted)
  - denote $\gamma_i = \dim \mathcal{N}(\mathbf{A} - \lambda_i \mathbf{I}) = \dim \mathcal{V}_{\lambda_i}$, $i = 1, \ldots, k$
  - i.e., we can find γ_i linearly independent $\mathbf{v}_i$'s in $\mathcal{V}_{\lambda_i}$
  - γ_i is called the geometric multiplicity of the eigenvalue λ_i

Property 5.1. We have $\mu_i \geq \gamma_i$ for all $i = 1, \ldots, k$ (not trivial, requires a proof)
  - **Implication:** no. of repeated eigenvalues $\geq$ no. of linearly indep. eigenvectors

Further Discussion: Repeated/Degenerate Eigenvalues

- eigenvalue λ_i is called defective if $\mu_i > \gamma_i$, i.e., its algebraic multiplicity exceeds its geometric multiplicity

- a matrix is called defective if it has one or more defective eigenvalues

  – example: consider

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

Eigenvalues and Eigenvectors

We have, or can easily prove, the following properties:

- $\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$
- $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$
- the eigenvalues of $\mathbf{A}^k$ are $\lambda_1^k, \ldots, \lambda_n^k$
- $\mathrm{rank}(\mathbf{A}) < n$ (i.e., rank-deficient) if and only if 0 is one eigenvalue of $\mathbf{A}$
- $\mathrm{rank}(\mathbf{A}) \geq$ number of nonzero eigenvalues of $\mathbf{A}$

Right and Left Eigenvectors

- right eigenvector or eigenvector associated with λ

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \quad \text{for} \quad \mathbf{v} \neq \mathbf{0}$$

- left eigenvector associated with λ

$$\mathbf{u}^H\mathbf{A} = \lambda\mathbf{u}^H \quad \text{for} \quad \mathbf{u} \neq \mathbf{0}$$

  (unless specified, eigenvectors are commonly referred to right eigenvectors)

  – left eigenvectors of $\mathbf{A}$ are nothing else but the (right) eigenvectors of $\mathbf{A}^T$

  – while the eigenvalues of $\mathbf{A}$ and $\mathbf{A}^T$ are the same, the sets of left- and right-eigenvectors may be different in general

Eigendecomposition

A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$) is said to be diagonalizable, or admit an eigendecomposition, if there exists a nonsingular $\mathbf{V} \in \mathbb{C}^{n\times n}$ such that

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1},$$

where $\mathbf{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$. (diagonalization of matrix $\mathbf{A}$: $\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{\Lambda}$)

- a.k.a. eigenvalue decomposition, spectral decomposition
- in defining diagonalizability, we didn't say that $(\mathbf{v}_i, \lambda_i)$ has to be an eigen-pair of $\mathbf{A}$. But

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} \iff \mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda},\ \mathbf{V} \text{ nonsingular}$$
$$\iff \mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i,\ i = 1, \ldots, n,\ \mathbf{V} \text{ nonsingular (hence, } \mathbf{v}_i \neq \mathbf{0})$$

  Also, $\lambda_1, \ldots, \lambda_n$ must be the n eigenvalues of $\mathbf{A}$; this can be seen from the characteristic polynomial $\det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{\Lambda} - \lambda\mathbf{I}) = \prod_{i=1}^{n}(\lambda_i - \lambda)$

Theorem 5.1. (the sufficient and necessary condition for the existence of eigendec.) For $\mathbf{A} \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$), it admits an eigendcomposition if and only if there exist n linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbf{A}$

Eigendecomposition

If $\mathbf{A}$ admits an eigendecomposition, the following properties can be shown (easily):

- $\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$
- $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$
- the eigenvalues of $\mathbf{A}^k$ are $\lambda_1^k, \ldots, \lambda_n^k$
- $\mathrm{rank}(\mathbf{A}) =$ number of nonzero eigenvalues of $\mathbf{A}$ (Quiz)
- suppose that $\mathbf{A}$ is also nonsingular. Then, $\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^{-1}$

Note: the first three properties can be shown to be valid for any $\mathbf{A}$ (do not depend on the existence of eigendec.); the fourth property may not be valid when $\mathbf{A}$ does not admit an eigendecomposition (recall the example next page); the third and fifth properties can be used for effcient computations of matrix powers and inversions

Eigendecomposition

Question: Does every $\mathbf{A} \in \mathbb{R}^{n \times n}$ (or $\mathbb{C}^{n \times n}$) admit an eigendecomposition?

- the answer is no.
- counter example: consider

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

  - the characteristic polynomial is $p(\lambda) = -\lambda^3$, so $\lambda_1 = \lambda_2 = \lambda_3 = 0$
  - it is easy to see that

$$\mathcal{V}_{\lambda_1} = \mathcal{N}(\mathbf{A} - \lambda_1 \mathbf{I}) = \mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)^\perp = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

  - any selection of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathcal{N}(\mathbf{A})$ is linearly dependent
- $\mathbf{A}$ does not admit an eigendecomposition if $\mu_i > \gamma_i$ for some $i \in \{1, \ldots, k\}$.

Eigendecomposition

Question: under which conditions of μ_i ans γ_i can a matrix admit an eigendec.?

- there exist matrix subclasses in which eigendecomposition is guaranteed to exist
  - one example is the circulant matrix subclass, as seen in Topic 3
  - another example is the Hermitian matrix subclass, as we will see
- there exist simple sufficient conditions under which eigendecomposition exists

Eigendecomposition

Property 5.2. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$), and suppose that λ_i's are ordered such that $\{\lambda_1, \ldots, \lambda_k\}$ is the set of all distinct eigenvalues of $\mathbf{A}$. Also, let $\mathbf{v}_i$ be *any* eigenvector associated with λ_i. Then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ must be linearly independent.

(requires a proof)

Implications:

- (a sufficient condition for existence of eigendec.) if all the eigenvalues of $\mathbf{A}$ are distinct, i.e.,

  $$\lambda_i \neq \lambda_j, \quad \text{for all } i, j \in \{1, \ldots, n\} \text{ with } i \neq j,$$

  then $\mathbf{A}$ admits an eigendecomposition
  - to have all the eigenvalues to be distinct is not that hard, as we will see later
- (the sufficient and necessary condition for existence of eigendec.) considering Property 5.1, $\mathbf{A}$ admits an eigendcomposition if and only if $\mu_i = \gamma_i$ for all i

Eigendecomposition for Hermitian & Real Symmetric Matrices

Consider the Hermitian matrix subclass.

Property 5.3. Let $\mathbf{A} \in \mathbb{H}^n$.

1. the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$ are real
2. suppose that λ_i's are ordered such that $\{\lambda_1, \ldots, \lambda_k\}$ is the set of all distinct eigenvalues of $\mathbf{A}$. Also, let $\mathbf{v}_i$ be *any* eigenvector associated with λ_i. Then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ must be orthogonal (or orthonormal).

(requires a proof)

- the above results apply to real symmetric matrices; recall $\mathbf{A} \in \mathbb{S}^n \Longrightarrow \mathbf{A} \in \mathbb{H}^n$
- Corollary: for a real symmetric matrix, all eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ can be chosen as real
- implication: for a Hermitian $\mathbf{A}$ with all its eigenvalues being distinct, then $\mathbf{A}$ admit an eigendecomposition with unitary $\mathbf{V}$.
- In fact, a Hermitian $\mathbf{A}$ always admits an eigendecomposition with unitary $\mathbf{V}$, i.e., is always unitarily diagonalizable (or orthogonally diagonalizable for the real case)!

Eigendecomposition for Hermitian & Real Symmetric Matrices

Theorem 5.2. Every $\mathbf{A} \in \mathbb{H}^n$ admits an eigendecomposition

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H,$$

where $\mathbf{V} \in \mathbb{C}^{n\times n}$ is unitary; $\mathbf{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_i \in \mathbb{R}$ for all i. Also, if $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{V}$ can be taken as real orthogonal.

- does not require the assumption of distinct eigenvalues
- Corollary: if $\mathbf{A}$ is Hermitian or real symmetric, $\mu_i = \gamma_i$ for all i (no. of repeated eigenvalues $=$ no. of linearly indep. eigenvectors)
- Proof? a consequence of a more powerful decomposition, namely, the Schur decomposition; we will go through it next

Schur Decomposition

Theorem 5.3. Let $\mathbf{A} \in \mathbb{C}^{n\times n}$, and let $\lambda_1, \ldots, \lambda_n$ be its eigenvalues. The matrix $\mathbf{A}$ admits a decomposition

$$\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H,$$

for some unitary $\mathbf{U} \in \mathbb{C}^{n\times n}$ and for some upper triangular $\mathbf{T} \in \mathbb{C}^{n\times n}$ with $t_{ii} = \lambda_i$ for all i. If $\mathbf{A}$ is real and $\lambda_1, \ldots, \lambda_n$ are all real, $\mathbf{U}$ and $\mathbf{T}$ can be taken as real.

(requires a proof)

- we will call the above decomposition the Schur decomposition or Schur unitary triangulation ($\mathbf{U}^H\mathbf{A}\mathbf{U} = \mathbf{T}$ called the Schur form of $\mathbf{A}$) in the sequel
- exists for any $\mathbf{A} \in \mathbb{C}^{n\times n}$ and can be viewed as a generalization of the eigendecomposition for $\mathbf{A}$ not diagonalizable
- some insight: Suppose $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$ for some unitary $\mathbf{U}$ and upper triangular $\mathbf{T}$, but it's not known if $t_{ii} = \lambda_i$. Then

  $$\det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{T} - \lambda\mathbf{I}) = \textstyle\prod_{i=1}^{n}(t_{ii} - \lambda)$$

  This implies that $t_{11}, \ldots, t_{nn}$ are the eigenvalues of $\mathbf{A}$

Schur Decomposition

- the Schur decomposition is a powerful tool
- e.g., we can use it to show that for *any* square $\mathbf{A}$ (with or without eigendec.),
  - $\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$
  - $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$
  - the eigenvalues of $\mathbf{A}^k$ are $\lambda_1^k, \ldots, \lambda_n^k$
- we may use it to prove the convergence of the power method (to be shown later) when eigendecomposition does not exist
- the Jordan canonical form, which we will not teach, requires the Schur decomposition as the first key step

Implications of the Schur Decomposition

- proof of Theorem 5.2:
  - let $\mathbf{A}$ be Hermitian, and let $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$ be its Schur decomposition. Observe

$$\mathbf{0} = \mathbf{A} - \mathbf{A}^H = \mathbf{U}\mathbf{T}\mathbf{U}^H - \mathbf{U}\mathbf{T}^H\mathbf{U}^H = \mathbf{U}(\mathbf{T} - \mathbf{T}^H)\mathbf{U}^H \quad \Longleftrightarrow \quad \mathbf{0} = \mathbf{T} - \mathbf{T}^H$$

  - since $\mathbf{T}$ is upper triangular and $\mathbf{T}^H$ is lower triangular, $\mathbf{T} = \mathbf{T}^H$ implies that $\mathbf{T}$ is diagonal; thus, the Schur decomposition is also the eigendecomposition
  - similar results apply to real symmetric $\mathbf{A}$, except that we use real $\mathbf{T}$, $\mathbf{U}$
  - note: $\mathbf{T} = \mathbf{T}^H$ also implies that t_{ii}'s are real; so the proof also confirms that λ_i's are real

Implications of the Schur Decomposition

- even though $\mathbf{A}$ does not admit an eigendecomposition, it is not hard to find an approximation of $\mathbf{A}$ which admits an eigendecomposition

  Proposition 5.1. Let $\mathbf{A} \in \mathbb{C}^{n\times n}$. For every $\varepsilon > 0$, there exists a matrix $\tilde{\mathbf{A}} \in \mathbb{C}^{n\times n}$ such that the n eigenvalues of $\tilde{\mathbf{A}}$ are distinct and

$$\|\mathbf{A} - \tilde{\mathbf{A}}\|_F \leq \varepsilon.$$

- Implication: for any square $\mathbf{A}$, we can always find an $\tilde{\mathbf{A}}$ that is arbitrarily close to $\mathbf{A}$ and admits an eigendecomposition

- proof:
  - let $\mathbf{D} = \mathrm{Diag}(d_1, \ldots, d_n)$ where $d_1, \ldots, d_n$ are chosen such that $|d_i| \leq \left(\frac{\varepsilon}{n}\right)^{1/2}$ for all i and such that $t_{11} + d_1, \ldots, t_{nn} + d_n$ are distinct
  - let $\mathbf{U}\mathbf{T}\mathbf{U}^H$ be the Schur decomposition of $\mathbf{A}$, and let $\tilde{\mathbf{A}} = \mathbf{U}(\mathbf{T} + \mathbf{D})\mathbf{U}^H$
  - we have $\|\mathbf{A} - \tilde{\mathbf{A}}\|_F^2 = \|\mathbf{D}\|_F^2 \leq \varepsilon$

Implications of the Schur Decomposition

- skew-Hermitian matrices: $\mathbf{A} \in \mathbb{C}^{n\times n}$ is said to be skew-Hermitian if $\mathbf{A}^H = -\mathbf{A}$
  - example:
$$\mathbf{A} = \begin{bmatrix} \boldsymbol{j}1 & 0 & -0.7+\boldsymbol{j}3 \\ 0 & -\boldsymbol{j}2 & 1+\boldsymbol{j}0.9 \\ 0.7+\boldsymbol{j}3 & -1+\boldsymbol{j}0.9 & 0 \end{bmatrix}$$
  - $\mathbf{A}$ is Hermitian if and only if $\boldsymbol{j}\mathbf{A}$ is skew-Hermitian
- real skew-symmetric matrices: $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be skew-symmetric if $\mathbf{A}^T = -\mathbf{A}$
  - example:
$$\mathbf{A} = \begin{bmatrix} 0 & 0 & -3 \\ 0 & 0 & 0.9 \\ 3 & -0.9 & 0 \end{bmatrix}$$
  - real skew-Hermitian is simply real skew-symmetric
- by the Schur decomposition, we can show that any skew-Hermitian $\mathbf{A}$ admits an eigendecomposition with unitary $\mathbf{V}$ and the eigenvalues are (purely) imaginary (i.e., $\Re\{\lambda\} = 0$ for $\lambda \in \mathbb{C}$)

Eigenvalue-Revealing Factorizations

- eigenvalue-revealing factorizations of matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$
  - diagonalization (eigendec.) $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ (nondefective $\mathbf{A}$, i.e., $\mu_i = \gamma_i$)
  - unitary diagonalization (eigendec.) $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{H}$ with unitary $\mathbf{V}$ (normal $\mathbf{A}$, including Hermitian and skew-Hermitian matrices)
  - unitary triangularization (Schur dec.) $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^{H}$ with unitary $\mathbf{U}$ (any $\mathbf{A}$)
  - Jordan canonical/normal form (Jordan dec.) $\mathbf{A} = \mathbf{S}\mathbf{J}\mathbf{S}^{-1}$ (any $\mathbf{A}$), where $\mathbf{J}$ is block diagonal as

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & & & \\ & \mathbf{J}_2 & & \\ & & \ddots & \\ & & & \mathbf{J}_k \end{bmatrix} \quad \text{with a square} \quad \mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

- In general, Schur factorization is used, because
  - unitary matrices are involved, so algorithm tends to be more stable

Highlights

- let $\mathbf{A} \in \mathbb{H}^n$, and let $\lambda_1(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})$ be the eigenvalues of $\mathbf{A}$ with ordering

$$\lambda_{\max}(\mathbf{A}) = \lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A}) = \lambda_{\min}(\mathbf{A})$$

  where $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote the min. and max. eigenvalues of $\mathbf{A}$, resp.

- variational characterizations of $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$:

$$\lambda_{\max}(\mathbf{A}) = \max_{\mathbf{x} \in \mathbb{C}^n, \|\mathbf{x}\|_2 = 1} \mathbf{x}^T \mathbf{A} \mathbf{x}, \qquad \lambda_{\min}(\mathbf{A}) = \min_{\mathbf{x} \in \mathbb{C}^n, \|\mathbf{x}\|_2 = 1} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

- (Courant-Fischer) for $k \in \{1, \ldots, n\}$,

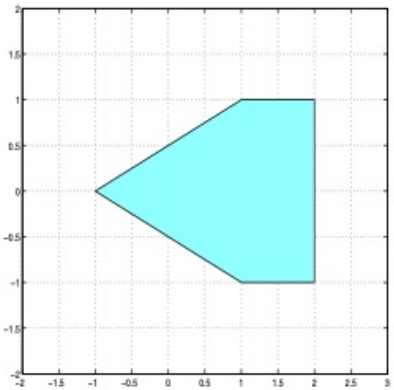
$$\lambda_k(\mathbf{A}) = \min_{\mathcal{S}_{n-k+1} \subseteq \mathbb{C}^n} \max_{\mathbf{x} \in \mathcal{S}_{n-k+1}, \|\mathbf{x}\|_2 = 1} \mathbf{x}^T \mathbf{A} \mathbf{x} = \max_{\mathcal{S}_k \subseteq \mathbb{C}^n} \min_{\mathbf{x} \in \mathcal{S}_k, \|\mathbf{x}\|_2 = 1} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

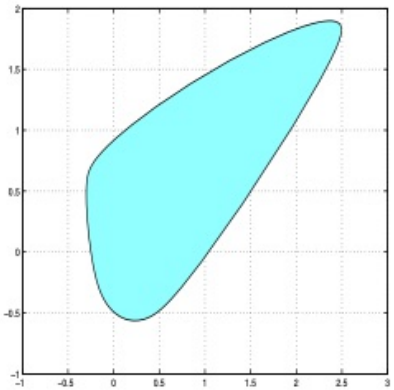
  where $\mathcal{S}_k$ denotes a subspace of dimension k

- real case: the same results apply; replace $\mathbb{C}$ by $\mathbb{R}$, $\mathbb{H}$ by $\mathbb{S}$, and "H" by "T"

Spectrum and Numerical Range

- given $\mathbf{A} \in \mathbb{C}^{n\times n}$ and its spectrum $\sigma(\mathbf{A}) = \{\lambda \in \mathbb{C} \,|\, \det(\mathbf{A} - \lambda\mathbf{I}) = 0\}$, we have
  - spectral radius of $\mathbf{A}$: $\rho(\mathbf{A}) = \max_{z\in\sigma(\mathbf{A})} |z|$
  - numerical range (field of values) of $\mathbf{A}$: $W(\mathbf{A}) = \{\mathbf{x}^H\mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^n,\ \|\mathbf{x}\|_2 = 1\}$
    - obviously $\sigma(\mathbf{A}) \subseteq W(\mathbf{A})$
    - (Toeplitz-Hausdorff Theorem) $W(\mathbf{A})$ is a convex set for any $\mathbf{A} \in \mathbb{C}^{n\times n}$
  - numerical radius of $\mathbf{A}$: $r(\mathbf{A}) = \max_{z\in W(\mathbf{A})} |z|$
    - $\rho(\mathbf{A}) \le r(\mathbf{A})$





Spectrum and Numerical Range

we can easily get the following properties on $W(\mathbf{A})$:

- for $\mathbf{A} \in \mathbb{C}^{n\times n}$ and a, $b \in \mathbb{C}$, $W(a\mathbf{A} + b\mathbf{I}) = aW(\mathbf{A}) + b$
- for $\mathbf{A} \in \mathbb{C}^{n\times n}$, $W(\mathbf{A}^T) = W(\mathbf{A})$ and $W(\mathbf{A}^H) = W(\mathbf{A}^*) = W(\mathbf{A})^*$
  - specially, if $\mathbf{A} \in \mathbb{R}^{n\times n}$, then $W(\mathbf{A})$ is symmetric with respect to the real axis
- for $\mathbf{A}$, $\mathbf{B} \in \mathbb{C}^{n\times n}$, $W(\mathbf{A} + \mathbf{B}) \subseteq W(\mathbf{A}) + W(\mathbf{B})$
- for $\mathbf{A} \in \mathbb{C}^{n\times n}$, $W(\mathbf{A}) \subset \mathbb{R}$ iff $\mathbf{A} \in \mathbb{H}^n$; in this case, the endpoints of $W(\mathbf{A})$ (a line segment) coincide with the smallest and the largest eigenvalues of $\mathbf{A}$ (will be proved later)

note: the first and second properties also apply to $\sigma(\mathbf{A})$; we also have if $\mathbf{A} \in \mathbb{R}^{n\times n}$, then $\sigma(\mathbf{A})$ is symmetric with respect to the real axis, i.e., complex eigenvalues of real matrices appear in conjugate pairs

Spectrum and Numerical Range

- for any $\mathbf{A} \in \mathbb{C}^{n\times n}$, it can be decomposed as

$$\mathbf{A} = \mathbf{H} + \mathbf{S}$$

  where $\mathbf{H} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^H)$ is Hermitian (with real eigenvalues) and $\mathbf{S} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^H)$ is skew-Hermitian (with purely imaginary eigenvalues)

  – when $n = 1$, it becomes $a = h + s$ with $h = \Re(a)$ and $s = \boldsymbol{j}\Im(a)$

Spectrum and Numerical Range

Property 5.4. If $\mathbf{H}$ and $\mathbf{S}$ are the Hermitian part and the skew-Hermitian part of $\mathbf{A} \in \mathbb{C}^{n\times n}$, respectively, then

$$\Re(W(\mathbf{A})) = W(\mathbf{H}) \quad \text{and} \quad \Im(W(\mathbf{A})) = -\boldsymbol{\jmath} W(\mathbf{S}) = W(-\boldsymbol{\jmath}\mathbf{S}).$$

($\Re(\cdot)$ and $\Im(\cdot)$ are used to denote the real and imaginary parts of a set, respectively.)
(proof is simple)

Property 5.5. Denote the spectrum of $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{S}$ as $\sigma(\mathbf{A})$, $\sigma(\mathbf{H})$, and $\sigma(\mathbf{S})$, and then we have $\lambda_{\min}(\mathbf{H}) \le \Re(\lambda_i(\mathbf{A})) \le \lambda_{\max}(\mathbf{H})$ and $\lambda_{\min}(-\boldsymbol{\jmath}\mathbf{S}) \le \Im(\lambda_i(\mathbf{A})) \le \lambda_{\max}(-\boldsymbol{\jmath}\mathbf{S})$ for all i.

- it can be hard to compute all the eigenvalues of a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, especially in the large-scale case
- Implications: we can estimate the geometrical locations or to find approximations of eigenvalues for any $\mathbf{A} \in \mathbb{C}^{n\times n}$ based on the extreme (i.e., largest and samllest) eigenvalues of $\mathbf{H}$, $-\boldsymbol{\jmath}\mathbf{S} \in \mathbb{H}^n$

Variational Characterizations of Eigenvalues of Hermitian & Real Symmetric Matrices

Notation and Conventions:

- $\lambda_1(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})$ denote the eigenvalues of a given $\mathbf{A} \in \mathbb{H}^n$ with ordering

$$\lambda_{\max}(\mathbf{A}) = \lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \ldots \geq \lambda_n(\mathbf{A}) = \lambda_{\min}(\mathbf{A}),$$

  where $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote the smallest and largest eigenvalues, resp.

- if not specified, $\lambda_1, \ldots, \lambda_n$ will be used to denote the eigenvalues of $\mathbf{A} \in \mathbb{H}^n$; they also follow the ordering

$$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n = \lambda_{\min}.$$

  Also, $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ will be used to denote the eigendecomposition of $\mathbf{A} \in \mathbb{H}^n$

Variational Characterizations of Eigenvalues

- let $\mathbf{A} \in \mathbb{H}^n$.
- for any $\mathbf{x} \in \mathbb{C}^n$ with $\mathbf{x} \neq \mathbf{0}$, the ratio

$$R(\mathbf{x}) = \frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}}$$

  is called the Rayleigh quotient.

- our interest: quadratic optimization such as

$$\max_{\mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}} = \max_{\mathbf{x} \in \mathbb{C}^n, \|\mathbf{x}\|_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x}$$

$$\min_{\mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}} = \min_{\mathbf{x} \in \mathbb{C}^n, \|\mathbf{x}\|_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x}$$

- Rayleigh quotient can be used for computing the eigenvalues of $\mathbf{A}$

Variational Characterizations of Eigenvalues: Rayleigh-Ritz

Theorem 5.4 (Rayleigh-Ritz)**.** Let $\mathbf{A} \in \mathbb{H}^n$. It holds that

$$\lambda_{\min}\|\mathbf{x}\|_2^2 \le \mathbf{x}^H\mathbf{A}\mathbf{x} \le \lambda_{\max}\|\mathbf{x}\|_2^2$$

$$\lambda_{\min} = \min_{\mathbf{x}\in\mathbb{C}^n, \|\mathbf{x}\|_2=1} \mathbf{x}^H\mathbf{A}\mathbf{x}, \qquad \lambda_{\max} = \max_{\mathbf{x}\in\mathbb{C}^n, \|\mathbf{x}\|_2=1} \mathbf{x}^H\mathbf{A}\mathbf{x}$$

- provides information about λ_1 and λ_n for $\mathbf{A}$
- proof:
  - by a change of variable $\mathbf{y} = \mathbf{V}^H\mathbf{x}$, we have

$$\mathbf{x}^H\mathbf{A}\mathbf{x} = \mathbf{y}^H\mathbf{\Lambda}\mathbf{y} = \sum_{i=1}^n \lambda_i|y_i|^2 \le \lambda_1 \sum_{i=1}^n |y_i|^2 = \lambda_1\|\mathbf{V}^H\mathbf{x}\|_2^2 = \lambda_1\|\mathbf{x}\|_2^2$$

  - we thus have $\max_{\|\mathbf{x}\|_2=1} \mathbf{x}^H\mathbf{A}\mathbf{x} \le \lambda_1$
  - since $\mathbf{v}_1^H\mathbf{A}\mathbf{v}_1 = \lambda_1$, the above equality is attained
  - the results $\mathbf{x}^H\mathbf{A}\mathbf{x} \ge \lambda_n\|\mathbf{x}\|_2^2$ and $\min_{\|\mathbf{x}\|_2=1} \mathbf{x}^H\mathbf{A}\mathbf{x} = \lambda_n$ are proven by the same way

Variational Characterizations of Eigenvalues: Courant-Fischer

Question: how about λ_k for any $k \in \{1, \ldots, n\}$? Do we have a similar variational characterization as that in the Rayleigh-Ritz theorem?

Theorem 5.5 (Courant-Fischer Minimax Theorem)**.** Let $\mathbf{A} \in \mathbb{H}^n$, and let $\mathcal{S}_k$ denote any subspace of $\mathbb{C}^n$ and of dimension k. For any $k \in \{1, \ldots, n\}$, it holds that

$$\begin{aligned} \lambda_k &= \min_{\mathcal{S}_{n-k+1} \subseteq \mathbb{C}^n} \max_{\mathbf{x} \in \mathcal{S}_{n-k+1}, \|\mathbf{x}\|_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x} \\ &= \max_{\mathcal{S}_k \subseteq \mathbb{C}^n} \min_{\mathbf{x} \in \mathcal{S}_k, \|\mathbf{x}\|_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x} \end{aligned}$$

(requires a proof)

- Rayleigh-Ritz Theorem 5.4 is a special case of the Courant-Fischer minimax theorem when $k = 1$ and $k = n$

Variational Characterizations of Eigenvalues: More Results

Some consequences and variants of the Courant-Fischer theorem: for any $\mathbf{A}, \mathbf{B} \in \mathbb{H}^n$, $\mathbf{z} \in \mathbb{C}^n$,

- (Weyl) $\lambda_k(\mathbf{A}) + \lambda_n(\mathbf{B}) \leq \lambda_k(\mathbf{A}+\mathbf{B}) \leq \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{B})$ for $k = 1, \ldots, n$
- (interlacing) $\lambda_{k+1}(\mathbf{A}) \leq \lambda_k(\mathbf{A} \pm \mathbf{z}\mathbf{z}^H)$ for $k = 1, \ldots, n-1$, and $\lambda_k(\mathbf{A} \pm \mathbf{z}\mathbf{z}^H) \leq \lambda_{k-1}(\mathbf{A})$ for $k = 2, \ldots, n$
- if $\mathrm{rank}(\mathbf{B}) \leq r$, then $\lambda_{k+r}(\mathbf{A}) \leq \lambda_k(\mathbf{A}+\mathbf{B})$ for $k = 1, \ldots, n-r$ and $\lambda_k(\mathbf{A}+\mathbf{B}) \leq \lambda_{k-r}(\mathbf{A})$ for $k = r+1, \ldots, n$
- (Weyl) $\lambda_{j+k-1}(\mathbf{A}+\mathbf{B}) \leq \lambda_j(\mathbf{A}) + \lambda_k(\mathbf{B})$ for $j, k \in \{1, \ldots, n\}$ with $j+k \leq n+1$
- for any $\mathcal{I} = \{i_1, \ldots, i_r\} \subseteq \{1, \ldots, n\}$, $\lambda_{k+n-r}(\mathbf{A}) \leq \lambda_k(\mathbf{A}_\mathcal{I}) \leq \lambda_k(\mathbf{A})$ for $k = 1, \ldots, r$
- for any semi-unitary $\mathbf{U} \in \mathbb{C}^{n \times r}$, $\lambda_{k+n-r}(\mathbf{A}) \leq \lambda_k(\mathbf{U}^H\mathbf{A}\mathbf{U}) \leq \lambda_k(\mathbf{A})$ for $k = 1, \ldots, r$
- many more...

Variational Characterizations of Eigenvalues: More Results

- we have considerd maximization or minimization of a Rayleigh quotient
- sometimes, we are interested in the problem of a sum of Rayleigh quotients:

$$\max_{\substack{\mathbf{U}\in\mathbb{C}^{n\times r}\\ \mathbf{u}_i\neq\mathbf{0}\ \forall i,\ \mathbf{u}_i^H\mathbf{u}_j=0\ \forall i\neq j}} \sum_{i=1}^{r}\frac{\mathbf{u}_i^H\mathbf{A}\mathbf{u}_i}{\mathbf{u}_i^H\mathbf{u}_i}$$

  where we want the vectors $\mathbf{u}_1,\ldots\mathbf{u}_r$ $(r\leq n)$ to be orthogonal to each other
- it finds applications in matrix factorization and PCA (cf. Topic 7)
- the Rayleigh quotients can be rewriten as

$$\max_{\substack{\mathbf{U}\in\mathbb{C}^{n\times r}\\ \mathbf{u}_i\neq\mathbf{0}\ \forall i,\ \mathbf{u}_i^H\mathbf{u}_j=0\ \forall i\neq j}} \sum_{i=1}^{r}\frac{\mathbf{u}_i^H\mathbf{A}\mathbf{u}_i}{\mathbf{u}_i^H\mathbf{u}_i} = \max_{\substack{\mathbf{U}\in\mathbb{C}^{n\times r}\\ \|\mathbf{u}_i\|_2=1\ \forall i,\ \mathbf{u}_i^H\mathbf{u}_j=0\ \forall i\neq j}} \sum_{i=1}^{r}\mathbf{u}_i^H\mathbf{A}\mathbf{u}_i$$

$$= \max_{\substack{\mathbf{U}\in\mathbb{C}^{n\times r}\\ \mathbf{U}^H\mathbf{U}=\mathbf{I}}} \operatorname{tr}(\mathbf{U}^H\mathbf{A}\mathbf{U}),$$

  where $\mathbf{U}$ is semi-unitary

Variational Characterizations of Eigenvalues: More Results

Then, we get an extension of the variational characterization to a sum of eigenvalues:

Theorem 5.6. Let $\mathbf{A} \in \mathbb{H}^n$. it holds that

$$\sum_{i=1}^{r} \lambda_i = \max_{\substack{\mathbf{U}\in\mathbb{C}^{n\times r} \\ \|\mathbf{u}_i\|_2=1 \ \forall i,\ \mathbf{u}_i^H\mathbf{u}_j=0 \ \forall i\neq j}} \sum_{i=1}^{r} \mathbf{u}_i^H \mathbf{A} \mathbf{u}_i = \max_{\substack{\mathbf{U}\in\mathbb{C}^{n\times r} \\ \mathbf{U}^H\mathbf{U}=\mathbf{I}}} \operatorname{tr}(\mathbf{U}^H\mathbf{A}\mathbf{U})$$

- can be proved by the eigenvalue inequality $\lambda_k(\mathbf{U}^H\mathbf{A}\mathbf{U}) \leq \lambda_k(\mathbf{A})$
- can also be proved by convex optimization

(requires a proof)

Variational Characterizations of Eigenvalues: More Results

Some more results (the proofs require more than just the Courant-Fischer theorem):

- (Cachy interlacing) Let

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{y} \\ \mathbf{y}^H & a \end{bmatrix} \in \mathbb{H}^n.$$

  Then, $\lambda_1(\mathbf{A}) \geq \lambda_1(\mathbf{B}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_{n-1}(\mathbf{B}) \geq \lambda_n(\mathbf{A})$.

- (von Neumann) Let $\mathbf{A}, \mathbf{B} \in \mathbb{H}^n$. It holds that

$$\sum_{i=1}^{n} \lambda_i(\mathbf{AB}) = \mathrm{tr}(\mathbf{AB}) \leq \sum_{i=1}^{n} \lambda_i(\mathbf{A})\lambda_i(\mathbf{B}).$$

- (Lidskii) Let $\mathbf{A}, \mathbf{B} \in \mathbb{H}^n$. For any $1 \leq i_1 \leq i_2 \leq \cdots \leq i_k$,

$$\sum_{j=1}^{k} \lambda_{i_j}(\mathbf{A}+\mathbf{B}) \leq \sum_{j=1}^{k} \lambda_{i_j}(\mathbf{A}) + \sum_{j=1}^{k} \lambda_j(\mathbf{B}).$$

Similarity Transformation

A matrix $\mathbf{B} \in \mathbb{C}^{n\times n}$ is said to be similar to another matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ if there exists a nonsingular $\mathbf{S} \in \mathbb{C}^{n\times n}$ such that

$$\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S},$$

and $\mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ called a similarity transformation of $\mathbf{A}$ via $\mathbf{S}$.

- Similar matrices are similar in the sense that their characteristic polynomials are the same. Specifically, if $\mathbf{A}$ is similar to $\mathbf{B}$ then we have

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{S}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{S}) = \det(\mathbf{B} - \lambda\mathbf{I}).$$

- It is easy to verify that similar matrices have the following properties:
  - If $\mathbf{B}$ is similar to $\mathbf{A}$, $\mathbf{A}$ is also similar to $\mathbf{B}$.
  - If $\mathbf{A}$, $\mathbf{B}$ are similar, they have the same set of eigenvalues (with the same algebraic multiplicity and geometric multiplicity)
  - If $\mathbf{A}$, $\mathbf{B}$ are similar, then $\det(\mathbf{A}) = \det(\mathbf{B})$.

  (proof is simple)
- if $\mathbf{S}$ is unitary, we say $\mathbf{B}$ is *unitarily similar* to $\mathbf{A}$ (recall Schur dec. and eigendec. for normal $\mathbf{A}$)

Similarity Transformation

- we are more interested in whether a matrix can be similar to a diagonal matrix—obviously because diagonal matrices are easy to deal with

A matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is said to be diagonalizable if it is similar to a diagonal matrix; i.e., there exists a nonsingular $\mathbf{S} \in \mathbb{C}^{n\times n}$ and a diagonal $\mathbf{D} \in \mathbb{C}^{n\times n}$ such that

$$\mathbf{D} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S},$$

or equivalently,

$$\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}.$$

- defination of "diagonalizable" based on similarity transformation
- the above equation can be equivalently rewritten as $\mathbf{A}\mathbf{S} = \mathbf{S}\mathbf{D}$ or

$$\mathbf{A}\mathbf{s}_i = d_i\mathbf{s}_i, \quad i = 1, \ldots, n,$$

where d_i denotes the (i,i)th entry of $\mathbf{D}$. Hence, every $(\mathbf{s}_i, d_i)$ must be an eigen-pair of $\mathbf{A}$.

Power Iteration

- power iteration/method: a method of numerically computing an eigenvector of a given matrix
- simple but provides the idea for a bunch of eigenvalue algorithms
- not the best in convergence speed
  - a comprehensive coverage of various computational methods for the eigenvalue problem can be found in the textbook **[Golub-Van Loan'13]**
- suitable for large-scale sparse problems, e.g., PageRank

Power Iteration

- assumptions:
  - $\mathbf{A}$ admits an eigendecomposition $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$
  - λ_i's are ordered such that $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$
  - $|\lambda_1| > |\lambda_2|$
  - we have an initial guess $\mathbf{x}$ that satisfies $[\mathbf{V}^{-1}\mathbf{x}]_1 \neq 0$ (random guess should do)
- consider $\mathbf{A}^k\mathbf{x}$. Let $\boldsymbol{\alpha} = \mathbf{V}^{-1}\mathbf{x}$, and observe

$$\mathbf{A}^k\mathbf{x} = \mathbf{V}\mathbf{\Lambda}^k\mathbf{V}^{-1}\mathbf{x} = \sum_{i=1}^{n} \alpha_i \lambda_i^k \mathbf{v}_i = \alpha_1\lambda_1^k\left(\mathbf{v}_1 + \underbrace{\sum_{i=2}^{n} \frac{\alpha_i}{\alpha_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k \mathbf{v}_i}_{=\mathbf{r}_k}\right)$$

  where $\mathbf{r}_k$ is a residual and has

$$\|\mathbf{r}_k\|_2 \leq \sum_{i=2}^{n} \left|\frac{\alpha_i}{\alpha_1}\right| \left|\frac{\lambda_i}{\lambda_1}\right|^k \|\mathbf{v}_i\|_2 \leq \left|\frac{\lambda_2}{\lambda_1}\right|^k \sum_{i=2}^{n} \left|\frac{\alpha_i}{\alpha_1}\right|$$

- convergence: let $c_k = \frac{|\alpha_1||\lambda_1|^k}{\alpha_1\lambda_1^k}$ (note $|c_k| = 1$). We have

$$\lim_{k\to\infty} c_k \frac{\mathbf{A}^k\mathbf{x}}{\|\mathbf{A}^k\mathbf{x}\|_2} = \mathbf{v}_1$$

Power Iteration

Algorithm: Power Iteration
input: $\mathbf{A} \in \mathbb{C}^{n\times n}$ and a starting vector $\mathbf{v}^{(0)} \in \mathbb{C}^n$
$k = 0$
repeat
$\tilde{\mathbf{v}}^{(k+1)} = \mathbf{A}\mathbf{v}^{(k)}$
$\mathbf{v}^{(k+1)} = \tilde{\mathbf{v}}^{(k+1)} / \|\tilde{\mathbf{v}}^{(k+1)}\|_2$ % normalization
$\lambda^{(k+1)} = R(\mathbf{v}^{(k+1)}) = (\mathbf{v}^{(k+1)})^H \mathbf{A}\mathbf{v}^{(k+1)}$
$k := k + 1$
until a stopping rule is satisfied
output: $\mathbf{v}^{(k)}$, $\lambda^{(k)}$

- it can be verified that $\mathbf{v}^{(k)} = \dfrac{\mathbf{A}^k\mathbf{v}^{(0)}}{\|\mathbf{A}^k\mathbf{v}^{(0)}\|_2}$
- it finds the dominant eigen-pair, i.e., dominant eigenvalue (largest eigenvalue in modulus λ_1) and dominant eigenvector $\mathbf{v}_1$ only, unless $\mathbf{v}_1$ is orthogonal to $\mathbf{v}^{(0)}$
- complexity per iteration: $\mathcal{O}(n^2)$, or $\mathcal{O}(\mathrm{nnz}(\mathbf{A}))$ for sparse $\mathbf{A}$

Power Iteration

- convergence rate depends on $\left|\frac{\lambda_2}{\lambda_1}\right|$
  - $\|\mathbf{v}^{(k)} - \mathbf{v}_1\|_2 = \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$ and $|\lambda^{(k)} - \lambda_1| = \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$
  - slower if $|\lambda_2|$ is closer to $|\lambda_1|$, i.e., $\left|\frac{\lambda_2}{\lambda_1}\right|$ is closer to 1
  - reduction per iteration is a constant, i.e., linear convergence
- now what if $|\lambda_1| = |\lambda_2| = \ldots = |\lambda_K| \geq \ldots \geq |\lambda_n|$ for some K?
  - by extending the convergence analysis, it can be shown $\mathbf{v}^{(k)}$ will converge to a vector in the subspace of $\mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K\}$
  - An important special case is when $\mathbf{A} \in \mathbb{R}^{2\times 2}$. If its complex eigenvalues come in conjugate pairs, then $\mathbf{v}^{(k)}$ will always be in the space spanned by the eigenvectors corresponding to the two eigenvalues.

Power Iteration with Deflation

- the power method finds the largest eigenvalue (in modulus) and the correponding eigenvector only
- how can we compute all the eigenvalues and eigenvectors?
- there are many ways and let's first consider a simple method called deflation
- consider a Hermitian $\mathbf{A}$ with $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n|$, and note the outer-product representation

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^H .$$

- Hotelling's deflation: use the power iteration to obtain $\mathbf{v}_1, \lambda_1$, do the subtraction

$$\mathbf{A} := \mathbf{A} - \lambda_1 \mathbf{v}_1 \mathbf{v}_1^H = \sum_{i=2}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^H ,$$

  and repeat until all the eigenvectors and eigenvalues are found
  – if we want the first k eigen-pairs only, deflation can also do that
- there are more deflation techniques which are not just for Hermitian matrices (learn by yourself)

Power Iteration With Shift

- $\mathbf{A} - \mu\mathbf{I}$ is with eigenvalue $(\lambda_i - \mu)$'s and the same eigenvector $\mathbf{v}_i$'s as $\mathbf{A}$
- if λ_1 is known (approximately), convergence can be made faster by applying power iteration for $\mathbf{A} - \mu\mathbf{I}$ s.t.
  - $(\lambda_1 - \mu)$ is the largest eigenvalue in modulus for $\mathbf{A} - \mu\mathbf{I}$
  - $\max_{i=2,\ldots,n} \left|\frac{\lambda_i - \mu}{\lambda_1 - \mu}\right|$ is as smaller as possible than $\left|\frac{\lambda_2}{\lambda_1}\right|$
- obviously, in practice hard to decide μ; the extent of acceleration is limited
- shift technique is commonly used together with inverse iteration and QR iteration

Inverse Iteration

- $(\mathbf{A}-\mu\mathbf{I})^{-1}$ is with eigenvalue $(\lambda_i-\mu)^{-1}$'s and the same eigenvector $\mathbf{v}_i$'s as $\mathbf{A}$
- inverse (power) iteration with shift: apply power iteration on $(\mathbf{A}-\mu\mathbf{I})^{-1}$
- if $\mu \approx \lambda_J$ for some J, $|(\lambda_J-\mu)^{-1}|$ may be far larger than $|(\lambda_i-\mu)^{-1}|$ for $i \neq J$, so power iteration can converge rapidly

Algorithm: Inverse Iteration
input: $\mathbf{A} \in \mathbb{C}^{n\times n}$ and a starting vector $\mathbf{v}^{(0)} \in \mathbb{C}^n$
$k = 0$
repeat
$\tilde{\mathbf{v}}^{(k+1)} = (\mathbf{A}-\mu\mathbf{I})^{-1}\mathbf{v}^{(k)}$ % solve $(\mathbf{A}-\mu\mathbf{I})\tilde{\mathbf{v}}^{(k+1)} = \mathbf{v}^{(k)}$
$\mathbf{v}^{(k+1)} = \tilde{\mathbf{v}}^{(k+1)}/\|\tilde{\mathbf{v}}^{(k+1)}\|_2$ % normalization
$\lambda^{(k+1)} = R(\mathbf{v}^{(k+1)}) = (\mathbf{v}^{(k+1)})^H\mathbf{A}\mathbf{v}^{(k+1)}$
$k := k+1$
until a stopping rule is satisfied
output: $\mathbf{v}^{(k)}$, $\lambda^{(k)}$

- converges to eigenvector $\mathbf{v}_J$ if parameter μ is close to λ_J
- complexity per iteration: $\mathcal{O}(n^2)$ (matrix $(\mathbf{A}-\mu\mathbf{I})$ is processed in advance)

Inverse Iteration

- convergence rate with
  - $\|\mathbf{v}^{(k)} - \mathbf{v}_J\|_2 = \mathcal{O}\left(\left(\max_{i=1,\ldots,n}\left|\frac{\lambda_J - \mu}{\lambda_i - \mu}\right|\right)^k\right)$
  - $|\lambda^{(k)} - \lambda_J| = \mathcal{O}\left(\left(\max_{i=1,\ldots,n}\left|\frac{\lambda_J - \mu}{\lambda_i - \mu}\right|\right)^k\right)$

    where λ_J is the closest eigenvalue to μ
  - reduction per iteration is a constant, i.e., linear convergence

- standard method for determining any eigenvector given an eigenvalue

- a linear system needs to be solved; similar to power iteration, can only compute one eigenpair

- inverse iteration without shift: taking $\mu = 0$ the algorithm converges to the eigenvector corresponding to the smallest eigenvalue of $\mathbf{A}$ (in modulus)

Rayleigh Quotient Iteration

- parameter μ is constant in inverse iteration, but convergence is better for μ close to the eigenvalue
- improvement: setting μ as the last computed Rayleigh quotient at each iteration

Algorithm: Rayleigh Quotient Iteration
input: $\mathbf{A} \in \mathbb{C}^{n \times n}$ and a starting vector $\mathbf{v}^{(0)} \in \mathbb{C}^n$
$k = 0$
$\mu^{(k)} = R(\mathbf{v}^{(k)})$
repeat
$\tilde{\mathbf{v}}^{(k+1)} = (\mathbf{A} - \mu^{(k)}\mathbf{I})^{-1}\mathbf{v}^{(k)}$ % solve $(\mathbf{A} - \mu^{(k)}\mathbf{I})\tilde{\mathbf{v}}^{(k+1)} = \mathbf{v}^{(k)}$
$\mathbf{v}^{(k+1)} = \tilde{\mathbf{v}}^{(k+1)} / \|\tilde{\mathbf{v}}^{(k+1)}\|_2$ % normalization
$\mu^{(k+1)} = \lambda^{(k+1)} = R(\mathbf{v}^{(k+1)}) = (\mathbf{v}^{(k+1)})^H \mathbf{A}\mathbf{v}^{(k+1)}$
$k := k + 1$
until a stopping rule is satisfied
output: $\mathbf{v}^{(k)}$, $\lambda^{(k)}$

- at least quadratic convergence, but uncertain to which eigenvalue it will converge
- complexity per iteration: $\mathcal{O}(n^3)$ (solving a different linear system each iteration)

Orthogonal Iteration

- for the previous methods, only find one eigenpair each time
  - what if we want more eigenvalues rather than λ_1
  - what if λ_1 and λ_2 are close or equal and we cannot decide shift μ; in this case, we might want to look for an invariant subspace associated with λ_1 and λ_2

- subspace iteration: starting with a set of linearly independent vectors or a subspace $\mathcal{V}^{(0)} = \text{span}\{\mathbf{v}_1^{(0)}, \mathbf{v}_2^{(0)}, \cdots, \mathbf{v}_r^{(0)}\}$, $\mathcal{V}^{(k)} = \mathbf{A}^k\mathcal{V}^{(0)}$ will converge (under suitable assumptions) to a subspace spanned by eigenvectors associated with the r largest eigenvalues in magnititude, i.e., dominant invariant subspace
  - in contrast, the power iteration is sometimes called vector iteration
  - use thin QR to get the bases $\mathbf{V}^{(k)}$ as $\mathbf{V}^{(k)}\mathbf{R}^{(k)} = \mathbf{A}^k\begin{bmatrix}\mathbf{v}_1^{(0)} & \mathbf{v}_2^{(0)} & \cdots & \mathbf{v}_r^{(0)}\end{bmatrix}$

- the above subspace iteration is an unnormalized simultaneous (power) iteration; since all of $\{\mathbf{A}^k\mathbf{v}_1^{(0)}, \mathbf{A}^k\mathbf{v}_2^{(0)}, \cdots, \mathbf{A}^k\mathbf{v}_r^{(0)}\}$ will converge to a multiple of $\mathbf{v}_1$, columns of $\mathbf{V}^{(k)}$ will form an extremely ill-conditioned basis for $\mathcal{V}^{(k)}$

- in practice, we use the orthogonal (simultaneous power) iteration

Orthogonal Iteration

- suppose there is a gap between the r $(1 \leq r \leq n)$ largest eigenvalues in magnititude and λ_{r+1}, i.e, $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_r| > |\lambda_{r+1}|$

Algorithm: Orthogonal Iteration
input: $\mathbf{A} \in \mathbb{C}^{n \times n}$ and a starting semi-unitary matrix $\mathbf{V}^{(0)} \in \mathbb{C}^{n \times r}$
$k = 0$
repeat
$\tilde{\mathbf{V}}^{(k+1)} = \mathbf{A}\mathbf{V}^{(k)}$
$\mathbf{V}^{(k+1)}\mathbf{R}^{(k+1)} = \tilde{\mathbf{V}}^{(k+1)}$ % normalization; perform thin QR for $\tilde{\mathbf{V}}^{(k+1)}$
$\left\{\lambda_1^{(k+1)}, \lambda_2^{(k+1)}, \ldots, \lambda_r^{(k+1)}\right\} = \sigma\left((\mathbf{V}^{(k+1)})^H \mathbf{A}\mathbf{V}^{(k+1)}\right)$
$k := k + 1$
until a stopping rule is satisfied
output: $\mathbf{V}^{(k)}$, $\left\{\lambda_1^{(k+1)}, \lambda_2^{(k+1)}, \ldots, \lambda_r^{(k+1)}\right\}$

- it can be verified that $\mathcal{R}\left(\mathbf{V}^{(k)}\right) = \mathcal{R}\left(\tilde{\mathbf{V}}^{(k)}\right) = \mathcal{R}\left(\mathbf{A}\mathbf{V}^{(k-1)}\right)$ (recall Topic 4)
- then $\mathcal{R}\left(\mathbf{V}^{(k)}\right) = \mathcal{R}\left(\mathbf{A}^k\mathbf{V}^{(0)}\right)$ (verify by yourself)

Orthogonal Iteration

- denote the Schur decomposition of $\mathbf{A}$ by $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$, s.t. $|t_{11}| \geq |t_{22}| \geq \cdots \geq |t_{rr}| > |t_{r+1,r+1}| \geq \cdots \geq |t_{nn}|$

- $\mathbf{V}^{(k)}$ converges linearly to an orthonormal basis for the dominant invariant subspace associated with the r largest eigenvalues in magnitude $\mathcal{R}(\mathbf{U}(:, 1:r))$

- $\left[\lambda_1^{(k)} \ \lambda_2^{(k)} \ \cdots \ \lambda_r^{(k)}\right] = \mathrm{diag}\left(\left(\mathbf{V}^{(k)}\right)^H \mathbf{A}\mathbf{V}^{(k)}\right) \rightarrow [\lambda_1 \ \lambda_2 \ \cdots \ \lambda_r]$

- $\left|\lambda_i^{(k)} - \lambda_i\right| = \mathcal{O}\left(\left(\mathrm{max}_{i=1,\ldots,r}\left|\frac{\lambda_{i+1}}{\lambda_i}\right|\right)^k\right), \ i = 1, \ 2, \ \cdots, \ r$

Orthogonal Iteration

- let's take a look at the span of the columns in $\mathbf{V}^{(k)}$
- given $\mathbf{V}^{(k+1)}\mathbf{R}^{(k+1)} = \mathbf{A}\mathbf{V}^{(k)}$, notice that the first p $(p = 1, \ldots, r)$ columns $\mathbf{V}^{(k+1)}$ satisfies the recurrence

$$\mathbf{V}^{(k+1)}\mathbf{R}^{(k+1)}(:, 1:p) = \mathbf{V}^{(k+1)}(:, 1:p)\mathbf{R}^{(k+1)}(1:p, 1:p) = \mathbf{A}\mathbf{V}^{(k)}(:, 1:p)$$

  then $\mathcal{R}\big(\mathbf{V}^{(k)}(:, 1:p)\big) = \mathcal{R}\big(\mathbf{A}^k\mathbf{V}^{(0)}(:, 1:p)\big)$

- over iterations, the first p columns of $\mathbf{V}^{(k)}$ converge to a basis for the dominant p-dimensional invariant subspace
- setting the initial $\mathbf{V}^{(0)} \in \mathbb{C}^{n\times n}$, directly get a n-dimensional invariant subspaces
  - when $r = n$, orthogonal iteration resembles the QR iteration algorithm

QR Iteration

Algorithm: QR Iteration
input: $\mathbf{A} \in \mathbb{C}^{n \times n}$
$\mathbf{A}^{(0)} = \mathbf{A}$
$k = 0$
repeat
  $\mathbf{Q}^{(k+1)}\mathbf{R}^{(k+1)} = \mathbf{A}^{(k)}$ % perform QR for $\mathbf{A}^{(k)}$
  $\mathbf{A}^{(k+1)} = \mathbf{R}^{(k+1)}\mathbf{Q}^{(k+1)}$
  $k := k + 1$
until a stopping rule is satisfied
output: $\mathbf{A}^{(k)}$

- $\mathbf{A}^{(k)}$ is unitarily similar to $\mathbf{A}^{(k+1)}$ in that $\mathbf{A}^{(k+1)} = \left(\mathbf{Q}^{(k+1)}\right)^H \mathbf{A}^{(k)} \mathbf{Q}^{(k+1)}$
  - and hence to $\mathbf{A}^{(0)}$ since $\mathbf{A}^{(0)} = \left(\mathbf{Q}^{(1)} \cdots \mathbf{Q}^{(k)}\right) \mathbf{A}^{(k)} \left(\mathbf{Q}^{(1)} \cdots \mathbf{Q}^{(k)}\right)^H$
- denote the Schur decomposition of $\mathbf{A}$ by $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$
- under some mild assumptions, $\mathbf{A}^{(k)}$ converges linearly to $\mathbf{T}$
  - if our problem is to compute all the eigenvalues of $\mathbf{A}$, picking the diagonal elements of $\mathbf{A}^{(k)}$ for a sufficiently large k would do

QR Iteration

- QR iteration is equivalent to orthogonal iteration with $\mathbf{V}^{(0)} = \mathbf{I}$
- the most popular method for computing all the eigenvalues of a general $\mathbf{A}$
  - the practical QR algorithm used in modern software is more sophisticated
- how to find the eigenvectors?
  - for λ_i, solve the eigen-equation $(\mathbf{T} - \lambda_i \mathbf{I})\mathbf{v} = \mathbf{0}$, which is an upper-triangular linear system
- as a counterpart to power iteration, to accelerate the convergence speed of orthogonal iteration there exist orthogonal simultaneous inverse iteration, orthogonal simultaneous iteration with shift, etc.
  - in the next we investigate QR iteration with shift

QR Iteration With Shift

Example: consider matrix $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \mathbf{A}^{(0)}$

$$\underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{\mathbf{Q}^{(0)}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{R}^{(0)}} = \mathbf{A}^{(0)}$$

$$\mathbf{A}^{(1)} = \mathbf{R}^{(0)}\mathbf{Q}^{(0)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \mathbf{A}^{(0)}$$

no convergence of $\mathbf{A}^{(k)}$ observed

- shift can also help to make QR iteration converge, i.e., $\mathbf{A}^{(k)}$ converge to a upper triangular matrix

QR Iteration with Shift

Algorithm: QR Iteration With Shift
input: $\mathbf{A} \in \mathbb{C}^{n \times n}$
$\mathbf{A}^{(0)} = \mathbf{A}$
$k = 0$
repeat
  choose a shift $\mu^{(k)}$
  $\mathbf{Q}^{(k+1)}\mathbf{R}^{(k+1)} = \mathbf{A}^{(k)} - \mu^{(k)}\mathbf{I}$ % perform QR for $\mathbf{A}^{(k)} - \mu^{(k)}\mathbf{I}$
  $\mathbf{A}^{(k+1)} = \mathbf{R}^{(k+1)}\mathbf{Q}^{(k+1)} + \mu^{(k)}\mathbf{I}$
  $k := k + 1$
until a stopping rule is satisfied
output: $\mathbf{A}^{(k)}$

- similar to QR iteration, $\mathbf{A}^{(k)}$ is unitarily similar to $\mathbf{A}^{(k+1)}$
  - $\mathbf{A}^{(k+1)} = \mathbf{R}^{(k+1)}\mathbf{Q}^{(k+1)} + \mu^{(k)}\mathbf{I} = \left(\mathbf{Q}^{(k+1)}\right)^H \left(\mathbf{A}^{(k)} - \mu^{(k)}\mathbf{I}\right)\mathbf{Q}^{(k+1)} + \mu^{(k)}\mathbf{I} = \left(\mathbf{Q}^{(k+1)}\right)^H \mathbf{A}^{(k)}\mathbf{Q}^{(k+1)}$
- shift $\mu^{(k)}$ may differ from iteration to iteration

QR Iteration with Shift

- Raleigh quotient shift
  - $\mu^{(k)} = \mathbf{A}^{(k)}(n,n)$ which will converge to the smallest eigenvalue in modulus
  - no guarantee on convergence
  - if converged, at least quadratic convergence
- Wilkinson shift
  - denote the lower-rightmost 2×2 matrix of $\mathbf{A}^{(k)}$ by

$$\bar{\mathbf{A}}^{(k)} = \begin{bmatrix} \mathbf{A}^{(k)}(n-1,n-1) & \mathbf{A}^{(k)}(n-1,n) \\ \mathbf{A}^{(k)}(n,n-1) & \mathbf{A}^{(k)}(n,n) \end{bmatrix}$$

  - chose the eigenvalue of $\bar{\mathbf{A}}^{(k)}$ closer to $\mathbf{A}^{(k)}(n,n)$
  - always converge with at least linear convergence **[Wilkinson'68]**

QR Iteration

- for $\mathbf{A} \in \mathbb{C}^{n\times n}$, each iteration requires $\mathcal{O}(n^3)$ to compute the QR factorization and the matrix multiplication; too computational expensive!

Question: can we directly transform $\mathbf{A}$ into an upper triangular matrix (i.e., introducing zeros below the diagonal) based on unitary similarity transformations?

- a naive try via the Householder reflections: let $\mathbf{Q}_1$ be the Householder reflection matrix that reflects $\mathbf{a}_1$ to $-\text{sign}(\mathbf{a}_1(1))\|\mathbf{a}_1\|_2\mathbf{e}_1$

$$\mathbf{A} = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix} \rightarrow \underbrace{\begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & \times & \times & \times \\ & \times & \times & \times \end{bmatrix}}_{\mathbf{Q}_1^H\mathbf{A}} \rightarrow \underbrace{\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix}}_{\mathbf{Q}_1^H\mathbf{A}\mathbf{Q}_1}$$

Property 5.6. Any $\mathbf{A} \in \mathbb{C}^{n\times n}$ is unitarily similar to an upper Hessenberg matrix $\mathbf{H}$ (i.e., introducing zeros below the first subdiagonal), i.e., $\mathbf{Q}^H\mathbf{A}\mathbf{Q} = \mathbf{H}$.

Hessenberg Reduction

- a upper Hessenberg matrix is given as

$$\mathbf{H} = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \end{bmatrix}$$

- Hessenberg reduction via Householder reflections: let $\tilde{\mathbf{a}}_1 = \mathbf{A}(2:n,1)$ and $\mathbf{Q}_1$ be the Householder reflection matrix that reflects $\tilde{\mathbf{a}}_1$ to $-\text{sign}(\tilde{\mathbf{a}}_1(1))\|\tilde{\mathbf{a}}_1\|_2\mathbf{e}_1$

$$\mathbf{A} = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix} \rightarrow \underbrace{\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \\ & \times & \times & \times \end{bmatrix}}_{\mathbf{Q}_1^H\mathbf{A}} \rightarrow \underbrace{\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \\ & \times & \times & \times \end{bmatrix}}_{\mathbf{Q}_1^H\mathbf{A}\mathbf{Q}_1}$$

- repeat the above procedure: $\underbrace{(\mathbf{Q}_1 \cdots \mathbf{Q}_{n-2})^H}_{\mathbf{Q}^H} \mathbf{A} \underbrace{\mathbf{Q}_1 \cdots \mathbf{Q}_{n-2}}_{\mathbf{Q}} = \mathbf{H}$

- compexity: $\frac{10}{3}n^3 + \mathcal{O}(n^2)$

Hessenberg Reduction

- for any $\mathbf{A} \in \mathbb{C}^{n\times n}$, the following algorithm reduces $\mathbf{A}$ to be upper Hessenberg

Algorithm: Householder Reduction to Upper Hessenberg Form
input: $\mathbf{A} \in \mathbb{C}^{n\times n}$
for $k = 1 : n-2$
$\mathbf{x} = \mathbf{A}(k+1:n, k)$
$\mathbf{v}_k = \text{sign}(\mathbf{x}(1))\|\mathbf{x}\|_2\mathbf{e}_1 + \mathbf{x}$
$\mathbf{v}_k = \mathbf{v}_k/\|\mathbf{v}_k\|_2$
$\mathbf{A}(k+1:n, k:n) = \mathbf{A}(k+1:n, k:n) - 2\mathbf{v}_k(\mathbf{v}_k^H\mathbf{A}(k+1:n, k:n))$
$\mathbf{A}(1:n, k+1:n) = \mathbf{A}(1:n, k+1:n) - 2(\mathbf{A}(1:n, k+1:n)\mathbf{v}_k)\mathbf{v}_k^H$
end
output: $\mathbf{A}$

Property 5.7. Any $\mathbf{A} \in \mathbb{H}^n$ is unitarily similar to a tridiagonal matrix.

- for $\mathbf{A} \in \mathbb{H}^n$, the above algorithm reduces $\mathbf{A}$ to be tridiagonal (Householder Reduction to Tridiagonal Form)

Hessenberg QR Iteration

Algorithm: Hessenberg QR Iteration
input: $\mathbf{A} \in \mathbb{C}^{n \times n}$
$\mathbf{H} = \mathbf{Q}^H \mathbf{A} \mathbf{Q}$, $\mathbf{A}^{(0)} = \mathbf{H}$ % Hessenberg reduction for $\mathbf{A}$
$k = 0$
repeat
$\mathbf{Q}^{(k+1)} \mathbf{R}^{(k+1)} = \mathbf{A}^{(k)}$ % perform QR for $\mathbf{A}^{(k)}$
$\mathbf{A}^{(k+1)} = \mathbf{R}^{(k+1)} \mathbf{Q}^{(k+1)}$
$k := k + 1$
until a stopping rule is satisfied
output: $\mathbf{A}^{(k)}$

- **Fact:** $\mathbf{A}^{(k)}$ preserves the upper Hessenberg property over iterations
  - $\mathbf{A}^{(k-1)}$ is upper Hessenberg and $\mathbf{R}^{(k)}$ is upper triangular, then $\mathbf{Q}^{(k)}$ is upper Hessenberg; since $\mathbf{R}^{(k)}$ is upper triangular, $\mathbf{A}^{(k)}$ is upper Hessenberge

Hessenberg QR Iteration

- QR factorization step for $\mathbf{A}^{(k)}$ and the matrix multiplication step requires $\mathcal{O}(n^2)$
  - using Givens rotations to compute $\mathbf{A}^{(k+1)} = \left(\mathbf{Q}^{(k+1)}\right)^H \mathbf{A}^{(k)} \mathbf{Q}^{(k+1)}$

$$\mathbf{A}^{(k)} = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \end{bmatrix} \rightarrow \underbrace{\begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix}}_{\mathbf{G}_1^H \mathbf{A}} \rightarrow \underbrace{\begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix}}_{\mathbf{G}_2^H \mathbf{G}_1^H \mathbf{A}} \rightarrow \underbrace{\begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix}}_{\mathbf{G}_3^H \mathbf{G}_2^H \mathbf{G}_1^H \mathbf{A} = \mathbf{R}^{(k+1)}}$$

$$\mathbf{R}^{(k+1)} = \begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix} \rightarrow \underbrace{\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix}}_{\mathbf{R}^{(k+1)} \mathbf{G}_1} \rightarrow \underbrace{\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \\ & & & \times \end{bmatrix}}_{\mathbf{R}^{(k+1)} \mathbf{G}_1 \mathbf{G}_2} \rightarrow \underbrace{\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \end{bmatrix}}_{\mathbf{R}^{(k+1)} \mathbf{G}_1 \mathbf{G}_2 \mathbf{G}_3 = \mathbf{A}^{(k+1)}}$$

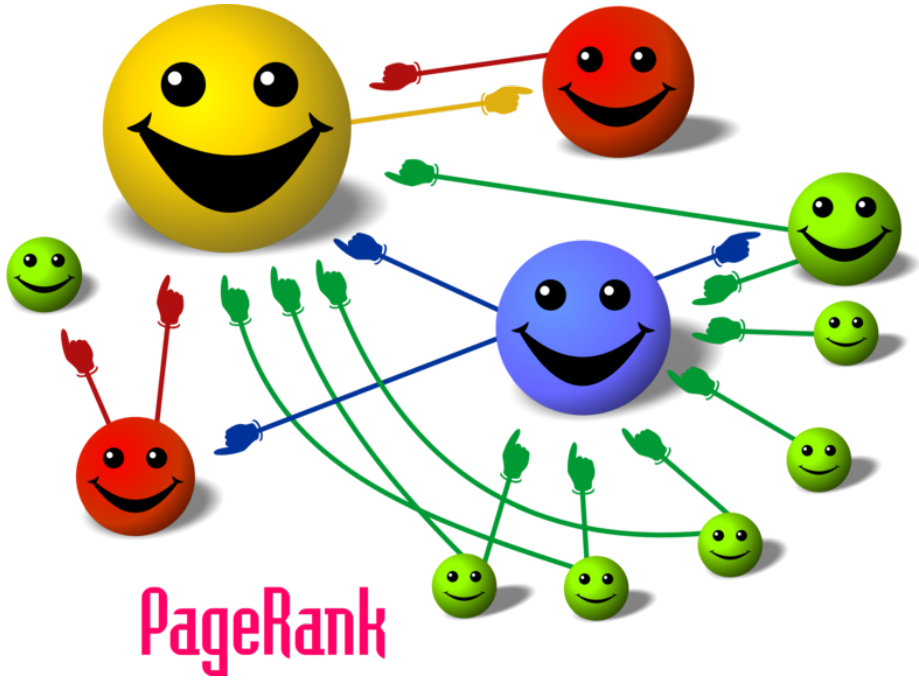
- for $\mathbf{A} \in \mathbb{H}^n$, QR factorization of a tridiagonal matrix requires only $\mathcal{O}(n)$ flops

Hessenberg QR Iteration with Shift

Algorithm: Hessenberg QR Iteration With Shift
input: $\mathbf{A} \in \mathbb{C}^{n \times n}$
$\mathbf{H} = \mathbf{Q}^H \mathbf{A} \mathbf{Q}$, $\mathbf{A}^{(0)} = \mathbf{H}$ % Hessenberg reduction for $\mathbf{A}$
$k = 0$
repeat
choose a shift $\mu^{(k)}$
$\mathbf{Q}^{(k+1)} \mathbf{R}^{(k+1)} = \mathbf{A}^{(k)} - \mu^{(k)} \mathbf{I}$ % perform QR for $\mathbf{A}^{(k)} - \mu^{(k)} \mathbf{I}$
$\mathbf{A}^{(k+1)} = \mathbf{R}^{(k+1)} \mathbf{Q}^{(k+1)} + \mu^{(k)} \mathbf{I}$
$k := k + 1$
until a stopping rule is satisfied
output: $\mathbf{A}^{(k)}$

PageRank: A Case Study

- PageRank is an algorithm used by Google to rank the pages of a search result.
- the idea is to use counts of links of various pages to determine pages' importance.



Source: Wiki.

- further reading: **[Bryan-Tanya2006]**

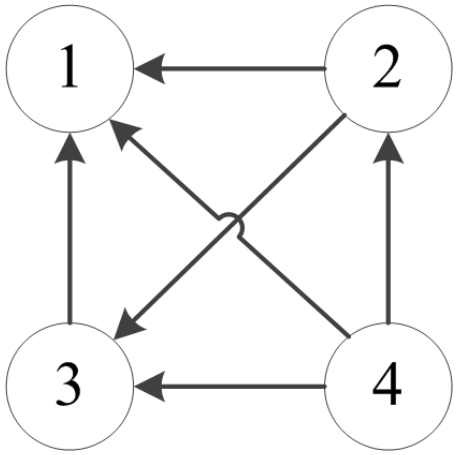
PageRank Model

- Model:

$$\sum_{j \in \mathcal{L}_i} \frac{v_j}{c_j} = v_i, \quad i = 1, \ldots, n,$$

where c_j is the number of outgoing links from page j; $\mathcal{L}_i$ is the set of pages with a link to page i; v_i is the importance score of page i.

- example:



$$\overbrace{\begin{bmatrix} 0 & \frac{1}{2} & 1 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}}^{\mathbf{A}} \overbrace{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}}^{\mathbf{v}} = \overbrace{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}}^{\mathbf{v}}.$$

PageRank Problem

- let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a matrix such that $a_{ij} = 1/c_j$ if $j \in \mathcal{L}_i$ and $a_{ij} = 0$ if $j \notin \mathcal{L}_i$
- Problem: find a non-negative $\mathbf{v}$ such that $\mathbf{A}\mathbf{v} = \mathbf{v}$
  - $\mathbf{A}$ is extremely large and sparse, and we want to use the power method
- Questions:
  - does a solution to $\mathbf{A}\mathbf{v} = \mathbf{v}$ exist? Or, is $\lambda = 1$ an eigenvalue of $\mathbf{A}$?
  - does $\mathbf{A}\mathbf{v} = \mathbf{v}$ have a non-negative solution? Or, does a non-negative eigenvector associated with $\lambda = 1$ exist?
  - is the solution to $\mathbf{A}\mathbf{v} = \mathbf{v}$ unique? Or, would there exist more than one eigenvector associated with $\lambda = 1$?
    - a unique solution is desired for this problem
  - is $\lambda = 1$ the only eigenvalue that is the largest in modulus?
    - this is required for the power method

Some Notation and Conventions

- notation:
  - $\mathbf{x} \geq \mathbf{y}$ means that $x_i \geq y_i$ for all i
  - $\mathbf{x} > \mathbf{y}$ means that $x_i > y_i$ for all i
  - $\mathbf{x} \ngeq \mathbf{y}$ means that $\mathbf{x} \geq \mathbf{y}$ does not hold
  - the same notations apply to matrices

- conventions:
  - $\mathbf{x}$ is said to be non-negative if $\mathbf{x} \geq \mathbf{0}$, and non-positive if $-\mathbf{x} \geq \mathbf{0}$
  - $\mathbf{x}$ is said to be positive if $\mathbf{x} > \mathbf{0}$, and negative if $-\mathbf{x} > \mathbf{0}$
  - the same conventions apply to matrices
  - a square $\mathbf{A}$ is said to be column-stochastic if $\mathbf{A} \geq \mathbf{0}$ and $\mathbf{A}^T\mathbf{1} = \mathbf{1}$
    - a column-stochastic $\mathbf{A}$ has every column $\mathbf{a}_i$ satisfying $\mathbf{a}_i^T\mathbf{1} = \sum_{j=1}^{n} a_{ji} = 1$

PageRank Matrix Properties

- in PageRank, $\mathbf{A}$ is column-stochastic if all pages have outgoing links
  - see the literature to see how to deal with cases where some pages do not have outgoing links (dangling nodes)

  Property 5.8. Let $\mathbf{A}$ be column-stochastic. Then,

  1. $\lambda = 1$ is an eigenvalue of $\mathbf{A}$
  2. $|\lambda| \leq 1$ for any eigenvalue λ of $\mathbf{A}$

- Implications:
  - a solution to $\mathbf{Av} = \mathbf{v}$ does exist, though it doesn't say if $\mathbf{v} \geq \mathbf{0}$ or not
  - $\lambda = 1$ is an eigenvalue that has the largest modulus, but we don't know if it is the *only* eigenvalue that has the largest modulus

- we resort to non-negative matrix theory to answer the rest of the questions

Non-Negative Matrix Theory

Theorem 5.7 (Perron-Frobenius)**.** Let $\mathbf{A}$ be square positive. There exists an eigenvalue ρ of $\mathbf{A}$ such that

1. ρ is real and $\rho > 0$
2. $|\lambda| < \rho$ for any eigenvalue λ of $\mathbf{A}$ with $\lambda \neq \rho$
3. there exists a positive eigenvector associated with ρ
4. the algebraic multiplicity of ρ is 1 (so the geometric multiplicity of ρ is also 1)

A weaker result for general non-negative matrices:

Theorem 5.8. Let $\mathbf{A}$ be square non-negative. There exists an eigenvalue ρ of $\mathbf{A}$ such that

1. ρ is real and $\rho \geq 0$
2. $|\lambda| \leq \rho$ for any eigenvalue λ of $\mathbf{A}$
3. there exists a non-negative eigenvector associated with ρ

PageRank Matrix Properties

- further implication by Theorem 5.8:
  - a non-negative solution to $\mathbf{A}\mathbf{v} = \mathbf{v}$ exists, though it doesn't say if there exists another solution
  - even worse, it is not known if there exists another solution $\mathbf{v}$ such that $\mathbf{v} \not\geq \mathbf{0}$

PageRank Matrix Properties

- PageRank actually considers a modified version of $\mathbf{A}$

$$\tilde{\mathbf{A}} = (1-\beta)\mathbf{A} + \beta \begin{bmatrix} 1/n & \dots & 1/n \\ \vdots & & \vdots \\ 1/n & \dots & 1/n \end{bmatrix}$$

  where $0 < \beta < 1$ (typical value is $\beta = 0.15$)

- $\tilde{\mathbf{A}}$ is positive
- further implications by Theorem 5.7:
  - $\lambda = 1$ is the *only* eigenvalue that has the largest modulus
  - there exists *only* one eigenvector associated with $\lambda = 1$; that eigenvector is either positive or negative
  - so the power method should work

References

[Golub-Van Loan'13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th edition, JHU Press, 2013.

[Bryan-Tanya2006] K. Bryan and L. Tanya, "The $25,000,000,000$ eigenvector: The linear algebra behind Google," *SIAM Review*, vol. 48, no. 3, pp. 569–581, 2006.

[Wilkinson'68] J. H. Wilkinson, "Global convergence of tridiagonal QR algorithm with origin shifts," *Linear Algebra and its Applications*, vol. 1, no. 3, pp. 409–420, 1968.