Schur Decomposition

Theorem 4. 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; complexity of computing the factorization is $\mathcal{O}(n^3)$)

- we will call the above decomposition the Schur decomposition or Schur (unitary) triangularization ($\mathbf{U}^H\mathbf{A}\mathbf{U} = \mathbf{T}$ 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 $\mathbf{A}$
- insight: eigenvalues of a triangular matrix are the diagonal entries of the matrix

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 cover, requires the Schur decomposition as the first key step

Implications of the Schur Decomposition

- proof of Theorem 3:
  - 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
  - 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
  - similar results apply to real symmetric $\mathbf{A}$, except that we use real $\mathbf{T}$, $\mathbf{U}$

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 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: $\mathbf{A}$ 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
- skew-symmetric matrices: $\mathbf{A}$ is said to be skew-symmetric if $\mathbf{A}^T = -\mathbf{A}$
  - example:
$$\mathbf{A} = \begin{bmatrix} 0 & 0 & -0.7 \\ 0 & 0 & 1 \\ 0.7 & -1 & 0 \end{bmatrix}$$
  - real skew-Hermitian is simply real skew-symmetric
- by the Schur decomposition, we can show that any skew-Hermitian (or real skew-symmetric) $\mathbf{A}$ admits an eigendecomposition with unitary $\mathbf{V}$ and the eigenvalues are (purely) imaginary (and possibly zero)

Eigenvalue-Revealing Factorizations

- eigenvalue-revealing factorizations of matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$
  - unitary diagonalization (eigendec.) $\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$ with unitary $\mathbf{V}$ (normal $\mathbf{A}$, i.e., $\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H$, including unitary, circ., Herm., and skew-Herm. matrices)
  - diagonalization (eigendec.) $\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}$ (nondefective $\mathbf{A}$)
  - 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

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

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

Real Schur Decomposition

- if $\mathbf{A}$ is real, a factorization with real matrices exists:

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

where $\mathbf{U}$ is orthogonal and $\mathbf{T}$ is block-triangular:

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \cdots & \mathbf{T}_{1k} \\ & \mathbf{T}_{22} & \cdots & \mathbf{T}_{2k} \\ & & \ddots & \vdots \\ & & & \mathbf{T}_{kk} \end{bmatrix}$$

with the diagonal blocks $\mathbf{T}_{ii}$ are 1×1 or 2×2

- the scalar diagonal blocks are real eigenvalues of $\mathbf{A}$

- the eigenvalues of the 2×2 diagonal blocks are complex eigenvalues (in conjugate pairs) of $\mathbf{A}$

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 $\mathbf{S}$.

- 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, then $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B})$.
  - If $\mathbf{A}$, $\mathbf{B}$ are similar, their characteristic polynomials are the same
  $$\det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{S}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{S}) = \det(\mathbf{B} - \lambda\mathbf{I}).$$
  - If $\mathbf{A}$, $\mathbf{B}$ are similar, then $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{B})$, $\det(\mathbf{A}) = \det(\mathbf{B})$.
  - If $\mathbf{A}$, $\mathbf{B}$ are similar, they have the same spectrum with the same algebraic multiplicity and geometric multiplicity
- if $\mathbf{S}$ is unitary, we say $\mathbf{B}$ is unitarily similar to $\mathbf{A}$ (cf. the cases of Schur decomp. and eigendecomp. for normal matrices)

Similarity Transformation

- we are more interested in whether a matrix can be similar to a diagonal matrix, i.e., diagonalizable by a similarity—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

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

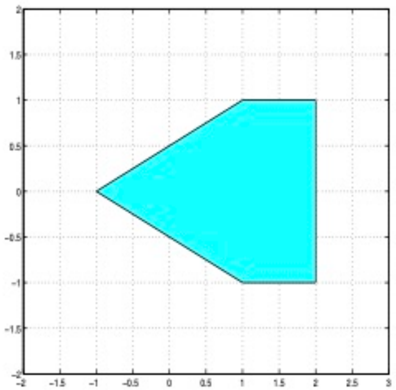
or equivalently,

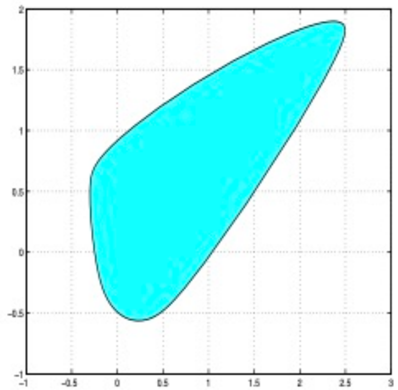
$$\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}.$$

- definition of "diagonalizable" based on similarity transformation
- the above equation can be equivalently rewritten as $\mathbf{AS} = \mathbf{SD}$ 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 $\mathbf{A}$.
- any square matrix is similar to a block diagonal (bidiag.) matrix (cf. Jordan dec.)
- if $\mathbf{S}$ is unitary, we say $\mathbf{A}$ is unitarily diagonalizable (i.e., normal matrices)

Spectrum and Numerical Range

- given $\mathbf{A} \in \mathbb{C}^{n\times n}$
  - spectrum of $\mathbf{A}$: $\sigma(\mathbf{A}) = \{\lambda \in \mathbb{C} \mid \det(\mathbf{A} - \lambda \mathbf{I}) = 0\}$
  - spectral radius of $\mathbf{A}$: $\rho(\mathbf{A}) = \max_{z\in\sigma(\mathbf{A})} |z|$
  - numerical range (field of values) of $\mathbf{A}$: $W(\mathbf{A}) = \{\mathbf{x}^H\mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^n,\ \|\mathbf{x}\|_2 = 1\}$
    * range of the Rayleigh quotient (to be defined later); obviously $\sigma(\mathbf{A}) \subseteq W(\mathbf{A})$
    * (Toeplitz-Hausdorff Theorem) $W(\mathbf{A})$ is convex and compact for any $\mathbf{A} \in \mathbb{C}^{n\times n}$





  - numerical radius of $\mathbf{A}$: $r(\mathbf{A}) = \max_{z\in W(\mathbf{A})} |z|$
    * $\rho(\mathbf{A}) \leq r(\mathbf{A})$

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, $W(\mathbf{A})$ is a line segment and the endpoints of $W(\mathbf{A})$ coincide with the smallest and the largest eigenvalues of $\mathbf{A}$ (will be proved later via Rayleigh-Ritz theorem)

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 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

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

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

Property 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 $\lambda_{\min}(\mathbf{H}) \leq \Re(\lambda_i(\mathbf{A})) \leq \lambda_{\max}(\mathbf{H})$ and $\lambda_{\min}(-\boldsymbol{j}\mathbf{S}) \leq \Im(\lambda_i(\mathbf{A})) \leq \lambda_{\max}(-\boldsymbol{j}\mathbf{S})$ 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 $\mathbf{H}$, $-\boldsymbol{j}\mathbf{S} \in \mathbb{H}^n$

Variational Characterizations: 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"

Variational Characterizations of Eigenvalues of Hermitian 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 (Rayleigh-Ritz Theorem)**.** Let $\mathbf{A} \in \mathbb{H}^n$. It holds that

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

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

- 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

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

  - we thus have $\max_{\|\mathbf{x}\|_2=1} \mathbf{x}^H\mathbf{A}\mathbf{x} \leq \lambda_1$
  - since $\mathbf{v}_1^H\mathbf{A}\mathbf{v}_1 = \lambda_1$, the above equality is attained
  - the results $\mathbf{x}^H\mathbf{A}\mathbf{x} \geq \lambda_n\|\mathbf{x}\|_2^2$ 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 6 (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 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 (like eigenvalue inequalities) 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 $\mathbf{u}_1,\dots\mathbf{u}_r$ $(r\leq n)$ to be orthogonal to each other
- it finds applications in matrix factorization and PCA (cf. SVD Topic)
- the Rayleigh quotients can be rewriten as

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

  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 7. 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 $1 \leq i_1 \leq i_2 \leq \cdots \leq i_k$,
$$\sum_{j=1}^{k} \lambda_{i_j}(\mathbf{A} + \mathbf{B}) \leq \sum_{j=1}^{k} \lambda_{i_j}(\mathbf{A}) + \sum_{j=1}^{k} \lambda_j(\mathbf{B}).$$