Numerical Methods for Solving Large Scale Eigenvalue Problems

Lecture 2, March 2, 2016: Numerical linear algebra basics

http://people.inf.ethz.ch/arbenz/ewp/

Peter Arbenz
Computer Science Department, ETH Zürich
E-mail: arbenz@inf.ethz.ch

1. Introduction
2. Numerical linear algebra basics
    - Definitions
    - Similarity transformations
    - Schur decompositions
    - SVD
3. Newtons method for linear and nonlinear eigenvalue problems
4. The QR Algorithm for dense eigenvalue problems
5. Vector iteration (power method) and subspace iterations
6. Krylov subspaces methods
    - Arnoldi and Lanczos algorithms
    - Krylov-Schur methods
7. Davidson/Jacobi-Davidson methods
8. Rayleigh quotient minimization for symmetric systems
9. Locally-optimal block preconditioned conjugate gradient (LOBPCG) method

- Basics
    - Notation
    - Statement of the problem
    - Similarity transformations
    - Schur decomposition
    - The real Schur decomposition
    - Hermitian matrices
    - Jordan normal form
    - Projections
    - The singular value decomposition (SVD)

Literature

- G. H. Golub and C. F. van Loan. *Matrix Computations*, 4th edition. Johns Hopkins University Press. Baltimore, 2012.
- R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
- Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, SIAM, Philadelphia, PA, 2011.
- E. Anderson et al. LAPACK Users Guide, 3rd edition. SIAM, Philadelphia, 1999. http://www.netlib.org/lapack/

Notations

$\mathbb{R}$: The field of real numbers
$\mathbb{C}$: The field of complex numbers
$\mathbb{R}^n$: The space of vectors of n *real* components
$\mathbb{C}^n$: The space of vectors of n *complex* components
Scalars : lowercase letters, a, b, c..., and $\alpha, \beta, \gamma \ldots$.
Vectors : boldface lowercase letters, **a**, **b**, **c**, ...

$$\mathbf{x} \in \mathbb{R}^n \iff \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad x_i \in \mathbb{R}.$$

We often make statements that hold for real or complex vectors.
$\longrightarrow \mathbf{x} \in \mathbb{F}^n$.

- The **inner product** of two n-vectors in $\mathbb{C}$:

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x_i \bar{y}_i = \mathbf{y}^* \mathbf{x},$$

- $\mathbf{y}^* = (\bar{y}_1,\ \bar{y}_2, \ldots, \bar{y}_n)$: conjugate transposition of complex vectors.
- $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal**, $\mathbf{x} \perp \mathbf{y}$, if $\mathbf{x}^* \mathbf{y} = 0$.
- **Norm** in $\mathbb{F}$, (Euclidean norm or 2-norm)

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})} = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}.$$

- $$A \in \mathbb{F}^{m\times n} \iff A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad a_{ij} \in \mathbb{F}.$$
  $$A^* \in \mathbb{F}^{n\times m} \iff A^* = \begin{pmatrix} \bar{a}_{11} & \bar{a}_{21} & \dots & \bar{a}_{m1} \\ \bar{a}_{12} & \bar{a}_{22} & \dots & \bar{a}_{m2} \\ \vdots & \vdots & & \vdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \dots & \bar{a}_{nm} \end{pmatrix}$$
  is the **Hermitian transpose** of A. For square matrices:
  - $A \in \mathbb{F}^{n\times n}$ is called **Hermitian** $\iff A^* = A$.
  - Real Hermitian matrix is called **symmetric**.
  - $U \in \mathbb{F}^{n\times n}$ is called **unitary** $\iff U^{-1} = U^*$.
  - Real unitary matrices are called **orthogonal**.
  - $A \in \mathbb{F}^{n\times n}$ is called **normal** $\iff A^*A = AA^*$.
    Both, Hermitian and unitary matrices are normal.

- **Norm** of a matrix (matrix norm induced by vector norm):

$$\|A\| := \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|.$$

- The condition number of a nonsingular matrix:

$$\kappa(A) = \|A\|\|A^{-1}\|.$$

U unitary $\implies$ $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ $\implies$ $\kappa(U) = 1$.

The **(standard) eigenvalue problem**:

Given a square matrix $A \in \mathbb{F}^{n \times n}$.
Find scalars $\lambda \in \mathbb{C}$ and vectors $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x} \neq \mathbf{0}$, such that

$$A\mathbf{x} = \lambda\mathbf{x}, \tag{1}$$

i.e., such that

$$(A - \lambda I)\mathbf{x} = \mathbf{0} \tag{2}$$

has a nontrivial (nonzero) solution.

We are looking for numbers λ such that $A - \lambda I$ is *singular*.
The pair $(\lambda, \mathbf{x})$ be a solution of (1) or (2).

- λ is called an **eigenvalue** of A,
- $\mathbf{x}$ is called an **eigenvector** corresponding to λ

- $(\lambda, \mathbf{x})$ is called **eigenpair** of A.
- The set $\sigma(A)$ of *all* eigenvalues of A is called **spectrum** of A.
- The set of all eigenvectors corresponding to an eigenvalue λ together with the vector $\mathbf{0}$ form a linear subspace of $\mathbb{C}^n$ called the **eigenspace** of λ.
- The eigenspace of λ is the null space of $\lambda I - A$: $\mathcal{N}(\lambda I - A)$.
- The dimension of $\mathcal{N}(\lambda I - A)$ is called **geometric multiplicity** $g(\lambda)$ of λ.
- An eigenvalue λ is a root of the **characteristic polynomial**

$$\chi(\lambda) := \det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0.$$

  The multiplicity of λ as a root of χ is called the **algebraic multiplicity** $m(\lambda)$ of λ.

$$1 \le g(\lambda) \le m(\lambda) \le n, \qquad \lambda \in \sigma(A), \quad A \in \mathbb{F}^{n\times n}.$$

- **y** is called **left eigenvector** corresponding to λ

$$\mathbf{y}^* A = \lambda \mathbf{y}^*$$

- Left eigenvector of A is a right eigenvector of A^*, corresponding to the eigenvalue $\bar{\lambda}$, $A^*\mathbf{y} = \bar{\lambda}\mathbf{y}$.
- A is an **upper triangular** matrix,

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ & a_{22} & \dots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{pmatrix}, \quad a_{ik} = 0 \text{ for } i > k.$$

$$\iff \det(\lambda I - A) = \prod_{i=1}^{n} (\lambda - a_{ii}).$$

(Generalized) eigenvalue problem

Given two square matrices $A, B \in \mathbb{F}^{n\times n}$.
Find scalars $\lambda \in \mathbb{C}$ and vectors $\mathbf{x} \in \mathbb{C}$, $\mathbf{x} \neq \mathbf{0}$, such that

$$A\mathbf{x} = \lambda B\mathbf{x}, \tag{3}$$

or, equivalently, such that

$$(A - \lambda B)\mathbf{x} = \mathbf{0} \tag{4}$$

has a nontrivial solution.

The pair $(\lambda, \mathbf{x})$ is a solution of (3) or (4).

- λ is called an **eigenvalue** of A **relative to** B,
- $\mathbf{x}$ is called an **eigenvector** of A **relative to** B corresponding to λ.
- $(\lambda, \mathbf{x})$ is called an **eigenpair** of A **relative to** B,
- The set $\sigma(A; B)$ of *all* eigenvalues of (3) is called the **spectrum** of A **relative to** B.

Similarity transformations I

Matrix A is similar to a matrix C, $A \sim C$, $\iff$ there is a nonsingular matrix S such that

$$S^{-1}AS = C. \tag{5}$$

The mapping $A \to S^{-1}AS$ is called a similarity transformation.

Theorem

Similar matrices have equal eigenvalues with equal multiplicities. If $(\lambda, \mathbf{x})$ is an eigenpair of A and $C = S^{-1}AS$ then $(\lambda, S^{-1}\mathbf{x})$ is an eigenpair of C.

Similarity transformations II

Proof:

$$A\mathbf{x} = \lambda\mathbf{x} \quad \text{and} \quad C = S^{-1}AS \Longrightarrow CS^{-1}\mathbf{x} = S^{-1}ASS^{-1}\mathbf{x} = \lambda S^{-1}\mathbf{x}$$

Hence A and C have equal eigenvalues and their geometric multiplicity is not changed by the similarity transformation.

$$\begin{aligned}\det(\lambda I - C) &= \det(\lambda S^{-1}S - S^{-1}AS) \\ &= \det(S^{-1}(\lambda I - A)S) \\ &= \det(S^{-1})\det(\lambda I - A)\det(S) \\ &= \det(\lambda I - A)\end{aligned}$$

the characteristic polynomials of A and C are equal and hence also the algebraic eigenvalue multiplicities are equal.

Unitary similarity transformations I

Two matrices A and C are called unitarily similar (orthogonally similar) if S ($C = S^{-1}AS = S^*AS$) is unitary (orthogonal).
Reasons for the importance of unitary similarity transformations:

1. U is unitary $\longrightarrow \|U\| = \|U^{-1}\| = 1 \longrightarrow \kappa(U) = 1$.
   Hence, if $C = U^{-1}AU \longrightarrow C = U^*AU$ and $\|C\| = \|A\|$.
   If A is disturbed by δA (roundoff errors introduced when storing the entries of A in finite-precision arithmetic)

   $$\longrightarrow U^*(A + \delta A)U = C + \delta C, \qquad \|\delta C\| = \|\delta A\|.$$

   Hence, errors (perturbations) in A are not amplified by a unitary similarity transformation. This is in contrast to arbitrary similarity transformations.

Unitary similarity transformations II

2. Preservation of symmetry: If A is symmetric
   $A = A^*, \quad U^{-1} = U^* : \quad C = U^{-1}AU = U^*AU = C^*$
3. For generalized eigenvalue problems, similarity transformations are not so crucial since we can operate with different matrices from both sides. If S and T are nonsingular

$$A\mathbf{x} = \lambda B\mathbf{x} \quad \Longleftrightarrow \quad TAS^{-1}S\mathbf{x} = \lambda TBS^{-1}S\mathbf{x}.$$

   This is called equivalence transformation of A, B.
   $\sigma(A; B) = \sigma(TAS^{-1}, TBS^{-1})$.
   Special Case: B is invertible & $B = LU$ is LU-factorization of B.
   $\longrightarrow$ Set $S = U$ and $T = L^{-1} \Rightarrow TBU^{-1} = L^{-1}LUU^{-1} = I$
   $\Rightarrow \sigma(A; B) = \sigma(L^{-1}AU^{-1}, I) = \sigma(L^{-1}AU^{-1})$.

Schur decomposition I

Theorem

If $A \in \mathbb{C}^{n\times n}$ then there is a unitary matrix $U \in \mathbb{C}^{n\times n}$ such that

$$U^* A U = T \tag{6}$$

is upper triangular. The diagonal elements of T are the eigenvalues of A.

Proof: By induction:

1. For $n = 1$, the theorem is obviously true.
2. Assume that the theorem holds for matrices of order $\leq n-1$.

Schur decomposition II

3. Let $(\lambda, \mathbf{x})$, $\|\mathbf{x}\| = 1$, be an eigenpair of A, $A\mathbf{x} = \lambda\mathbf{x}$. Construct a unitary matrix U_1 with first column $\mathbf{x}$ (e.g. the Householder reflector U_1 with $U_1\mathbf{x} = \mathbf{e}_1$). Partition $U_1 = [\mathbf{x}, \overline{U}]$. Then

$$U_1^* A U_1 = \begin{bmatrix} \mathbf{x}^* A\mathbf{x} & \mathbf{x}^* A\overline{U} \\ \overline{U}^* A\mathbf{x} & \overline{U}^* A\overline{U} \end{bmatrix} = \begin{bmatrix} \lambda & \times \cdots \times \\ \mathbf{0} & \hat{A} \end{bmatrix}$$

as $A\mathbf{x} = \lambda\mathbf{x}$ and $\overline{U}^*\mathbf{x} = \mathbf{0}$ by construction of U_1. By assumption, there exists a unitary matrix $\hat{U} \in \mathbb{C}^{(n-1)\times(n-1)}$ such that $\hat{U}^*\hat{A}\hat{U} = \hat{T}$ is upper triangular. Setting $U := U_1(1 \oplus \hat{U})$, we obtain (6).

Schur vectors I

$U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$
$U^*AU = T$ is a Schur decomposition of $A \Longleftrightarrow AU = UT$.
The k-th column of this equation is

$$A\mathbf{u}_k = \lambda \mathbf{u}_k + \sum_{i=1}^{k-1} t_{ik}\mathbf{u}_i, \qquad \lambda_k = t_{kk}. \tag{7}$$

$$\Longrightarrow A\mathbf{u}_k \in \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}, \quad \forall k.$$

The first k **Schur vectors** $\mathbf{u}_1, \ldots, \mathbf{u}_k$ form an invariant subspace for A. (A subspace $\mathcal{V} \subset \mathbb{F}^n$ is called invariant for A if $A\mathcal{V} \subset \mathcal{V}$.)

- From (7): the *first* Schur vector is an eigenvector of A.
- The other columns of U, are in general not eigenvectors of A.

The Schur decomposition is not unique. The eigenvalues can be arranged in any order in the diagonal of T.

The real Schur decomposition I

* Real matrices can have complex eigenvalues. If complex eigenvalues exist, then they occur in complex conjugate pairs! If λ is an eigenvalue of the real matrix $A \longrightarrow \bar{\lambda}$ is an eigenvalue of A.

Theorem

(Real Schur decomposition) *If $A \in \mathbb{R}^{n\times n}$ then there is an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ such that*

$$Q^T A Q = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ & R_{22} & \cdots & R_{2m} \\ & & \ddots & \vdots \\ & & & R_{mm} \end{bmatrix} \tag{8}$$

is upper quasi-triangular. The diagonal blocks R_{ii} are either 1×1 or 2×2 matrices. A 1×1 block corresponds to a real eigenvalue, a 2×2 block corresponds to a pair of complex conjugate eigenvalues.

The real Schur decomposition II

Remark: The matrix

$$\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}, \quad \alpha, \beta \in \mathbb{R},$$

has the eigenvalues $\alpha + i\beta$ and $\alpha - i\beta$.

Let $\lambda = \alpha + i\beta$, $\beta \neq 0$, be an eigenvalue of A with eigenvector $\mathbf{x} = \mathbf{u} + i\mathbf{v}$. Then $\bar{\lambda} = \alpha - i\beta$ is an eigenvalue corresponding to $\bar{\mathbf{x}} = \mathbf{u} - i\mathbf{v}$.

$$\begin{aligned} A\mathbf{x} &= A(\mathbf{u} + i\mathbf{v}) = A\mathbf{u} + iA\mathbf{v}, \\ \lambda\mathbf{x} &= (\alpha + i\beta)(\mathbf{u} + i\mathbf{v}) = (\alpha\mathbf{u} - \beta\mathbf{v}) + i(\beta\mathbf{u} + \alpha\mathbf{v}). \end{aligned}$$

$$\begin{aligned} \longrightarrow \quad A\bar{\mathbf{x}} &= A(\mathbf{u} - i\mathbf{v}) = A\mathbf{u} - iA\mathbf{v}, \\ &= (\alpha\mathbf{u} - \beta\mathbf{v}) - i(\beta\mathbf{u} + \alpha\mathbf{v}) \\ &= (\alpha - i\beta)\mathbf{u} - i(\alpha - i\beta)\mathbf{v} = (\alpha - i\beta)(\mathbf{u} - i\mathbf{v}) = \bar{\lambda}\bar{\mathbf{x}}. \end{aligned}$$

The real Schur decomposition III

k: the number of complex conjugate pairs.

Now, let's prove the theorem by induction on k.

Proof:

- First $k = 0$. In this case, A has real eigenvalues and eigenvectors. We can repeat the proof of the Schur decomposition Theorem in real arithmetic to get the decomposition ($U^*AU = T$) with $U \in \mathbb{R}^{n\times n}$ and $T \in \mathbb{R}^{n\times n}$. So, there are n diagonal blocks R_{jj} all of which are 1×1.

$$\begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ & R_{22} & \cdots & R_{2m} \\ & & \ddots & \vdots \\ & & & R_{mm} \end{bmatrix}$$

The real Schur decomposition IV

- Assume that the theorem is true for all matrices with fewer than k complex conjugate pairs. Then, with $\lambda = \alpha + i\beta$, $\beta \neq 0$ and $\mathbf{x} = \mathbf{u} + i\mathbf{v}$,

$$A[\mathbf{u}, \mathbf{v}] = [\mathbf{u}, \mathbf{v}] \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$

Let $\{\mathbf{x}_1, \mathbf{x}_2\}$ be an orthonormal basis of span$([\mathbf{u}, \mathbf{v}])$. Then, since $\mathbf{u}$ and $\mathbf{v}$ are linearly independent (If u and v were linearly dependent then it follows that β must be zero.), there is a nonsingular 2×2 real square matrix C with

$$[\mathbf{x}_1, \mathbf{x}_2] = [\mathbf{u}, \mathbf{v}]C.$$

The real Schur decomposition V

$$
\begin{aligned}
A[\mathbf{x}_1, \mathbf{x}_2] = A[\mathbf{u}, \mathbf{v}]C = [\mathbf{u}, \mathbf{v}] \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} C \\
= [\mathbf{x}_1, \mathbf{x}_2] C^{-1} \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} C =: [\mathbf{x}_1, \mathbf{x}_2] S.
\end{aligned}
$$

S and $\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$ are similar and therefore have equal eigenvalues. Now, construct an orthogonal matrix $[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n] =: [\mathbf{x}_1, \mathbf{x}_2, W]$.

$$
\begin{aligned}
\big[[\mathbf{x}_1, \mathbf{x}_2], W\big]^T A \big[[\mathbf{x}_1, \mathbf{x}_2], W\big] = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ W^T \end{bmatrix} \big[[\mathbf{x}_1, \mathbf{x}_2] S, AW\big] \\
= \begin{bmatrix} S & [\mathbf{x}_1, \mathbf{x}_2]^T AW \\ O & W^T AW \end{bmatrix}.
\end{aligned}
$$

The real Schur decomposition VI

The matrix $W^T AW$ has less than k complex-conjugate eigenvalue pairs. Therefore, by the induction assumption, there is an orthogonal $Q_2 \in \mathbb{R}^{(n-2)\times(n-2)}$ such that the matrix

$$Q_2^T(W^T AW)Q_2$$

is quasi-triangular. Thus, the orthogonal matrix

$$Q = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n] \begin{pmatrix} I_2 & O \\ O & Q_2 \end{pmatrix}$$

transforms A similarly to quasi-triangular form.

Hermitian matrices

Matrix $A \in \mathbb{F}^{n\times n}$ is Hermitian if $A = A^*$.

In the Schur decomposition $A = U\Lambda U^*$ for Hermitian matrices the upper triangular Λ is Hermitian and therefore diagonal.

$$\overline{\Lambda} = \Lambda^* = (U^*AU)^* = U^*A^*U = U^*AU = \Lambda,$$

each diagonal element λ_i of Λ satisfies $\overline{\lambda}_i = \lambda_i \Longrightarrow \Lambda$ must be real.

Hermitian/symmetric matrix is called positive definite (positive semi-definite) if all its eigenvalues are positive (nonnegative).

HPD or SPD $\Longrightarrow$ Cholesky factorization exists.

Spectral decomposition

Theorem

(Spectral theorem for Hermitian matrices) *Let A be Hermitian. Then there is a unitary matrix U and a real diagonal matrix Λ such that*

$$A = U\Lambda U^* = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^*. \tag{9}$$

The columns $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of U are eigenvectors corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$. They form an orthonormal basis for $\mathbb{F}^n$.

The decomposition (9) is called a spectral decomposition of A.
As the eigenvalues are real we can sort them with respect to their magnitude. We can, e.g., arrange them in ascending order such that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

- If $\lambda_i = \lambda_j$, then any nonzero linear combination of $\mathbf{u}_i$ and $\mathbf{u}_j$ is an eigenvector corresponding to λ_i,

$$A(\mathbf{u}_i\alpha + \mathbf{u}_j\beta) = \mathbf{u}_i\lambda_i\alpha + \mathbf{u}_j\lambda_j\beta = (\mathbf{u}_i\alpha + \mathbf{u}_j\beta)\lambda_i.$$

- Eigenvectors corresponding to different eigenvalues are orthogonal. $A\mathbf{u} = \mathbf{u}\lambda$ and $A\mathbf{v} = \mathbf{v}\mu$, $\lambda \neq \mu$.

$$\lambda\mathbf{u}^*\mathbf{v} = (\mathbf{u}^*A)\mathbf{v} = \mathbf{u}^*(A\mathbf{v}) = \mathbf{u}^*\mathbf{v}\mu,$$

and thus

$$(\lambda - \mu)\mathbf{u}^*\mathbf{v} = 0,$$

from which we deduce $\mathbf{u}^*\mathbf{v} = 0$ as $\lambda \neq \mu$.

Eigenspace

- The eigenvectors corresponding to a particular eigenvalue λ form a subspace, the eigenspace $\{\mathbf{x} \in \mathbb{F}^n, A\mathbf{x} = \lambda\mathbf{x}\} = \mathcal{N}(A - \lambda I)$.
- They are perpendicular to the eigenvectors corresponding to all the other eigenvalues.
- Therefore, the spectral decomposition is unique up to $\pm$ signs if all the eigenvalues of A are distinct.
- In case of multiple eigenvalues, we are free to choose any orthonormal basis for the corresponding eigenspace.

Remark: The notion of Hermitian or symmetric has a wider background. Let $\langle \mathbf{x}, \mathbf{y} \rangle$ be an inner product on $\mathbb{F}^n$. Then a matrix A is symmetric with respect to this inner product if $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A\mathbf{y} \rangle$ for all vectors $\mathbf{x}$ and $\mathbf{y}$. *All the properties of Hermitian matrices hold similarly for matrices symmetric with respect to a certain inner product.*

Matrix polynomials

$p(\lambda)$: polynomial of degree d,
$p(\lambda) = \alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \cdots + \alpha_d\lambda^d$.

$$A^j = (U\Lambda U^*)^j = U\Lambda^j U^*$$

Matrix polynomial:

$$p(A) = \sum_{j=0}^{d} \alpha_j A^j = \sum_{j=0}^{d} \alpha_j U\Lambda^j U^* = U\left(\sum_{j=0}^{d} \alpha_j \Lambda^j\right) U^*.$$

This equation shows that

- $p(A)$ has the same eigenvectors as the original matrix A.
- The eigenvalues are modified though, λ_k becomes $p(\lambda_k)$.
- More complicated functions of A can be computed if the function is defined on spectrum of A.

Theorem (Jordan normal form)

For every $A \in \mathbb{F}^{n\times n}$ *there is a nonsingular matrix* $X \in \mathbb{F}^{n\times n}$ *such that*

$$X^{-1}AX = J = \mathrm{diag}(J_1, J_2, \ldots, J_p),$$

where

$$J_k = J_{m_k}(\lambda_k) = \begin{bmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix} \in \mathbb{F}^{m_k\times m_k}$$

are called **Jordan blocks** *and* $m_1 + \cdots + m_p = n$. *The values* λ_k *need not be distinct. The Jordan matrix* J *is unique up to the ordering of the blocks. The transformation matrix* X *is not unique.*

Jordan normal form I

- Matrix diagonalizable $\Longleftrightarrow$ all Jordan blocks are 1×1 (trivial). In this case the columns of X are eigenvectors of A.
- One eigenvector associated with each Jordan block

$$J_2(\lambda)\mathbf{e}_1 = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \lambda \, \mathbf{e}_1.$$

- Nontrivial blocks give rise to generalized eigenvectors $\mathbf{e}_2, \ldots, \mathbf{e}_{m_k}$ since

$$(J_k(\lambda) - \lambda I)\mathbf{e}_{j+1} = \mathbf{e}_j, \quad j = 1, \ldots, m_k - 1.$$

- Computation of Jordan blocks is unstable.

Jordan normal form II

Let $Y := X^{-*}$ and let $X = [X_1, X_2, \ldots, X_p]$ and $Y = [Y_1, Y_2, \ldots, Y_p]$ be partitioned according to J. Then,

$$A = XJY^* = \sum_{k=1}^{p} X_k J_k Y_k^* = \sum_{k=1}^{p} (\lambda_k X_k Y_k^* + X_k N_k Y_k^*)$$

$$= \sum_{k=1}^{p} (\lambda_k P_k + D_k),$$

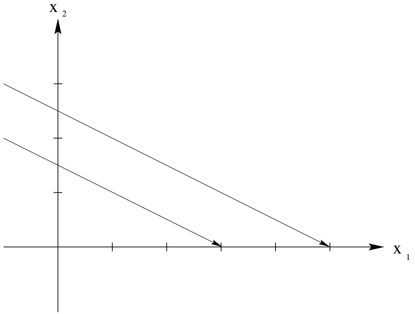
where $N_k = J_{m_k}(0)$, $P_k := X_k Y_k^*$, $D_k := X_k N_k Y_k^*$.

Since $P_k^2 = P_k$, P_k is a projector on $\mathcal{R}(P_k) = \mathcal{R}(X_k)$. It is called a spectral projector.

Projections I

A matrix P that satisfies $P^2 = P$ is called a projection.

A projection is a square matrix. If P is a projection then $P\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in the range $\mathcal{R}(P)$ of P. In fact, if $\mathbf{x} \in \mathcal{R}(P)$ then $\mathbf{x} = P\mathbf{y}$ for some $\mathbf{y} \in \mathbb{F}^n$ and $P\mathbf{x} = P(P\mathbf{y}) = P^2\mathbf{y} = P\mathbf{y} = \mathbf{x}$.



Projections II

Example: Let

$$P = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}.$$

The range of P is $\mathcal{R}(P) = \mathbb{F} \times \{\mathbf{0}\}$. The effect of P is depicted in Figure on previous page: All points $\mathbf{x}$ that lie on a line parallel to $\text{span}\{(2,-1)^*\}$ are mapped on the same point on the $\mathbf{x}_1$ axis. So, the projection is *along* $\text{span}\{(2,-1)^*\}$ which is the null space $\mathcal{N}(P)$ of P.

If P is a projection then $I - P$ is a projection.

If $P\mathbf{x} = \mathbf{0}$ then $(I - P)\mathbf{x} = \mathbf{x}$.
$\Longrightarrow$ range of $I - P$ equals null space of P: $\mathcal{R}(I - P) = \mathcal{N}(P)$.

It can be shown that $\mathcal{R}(P) = \mathcal{N}(P^*)^\perp$.

Projections III

Notice that $\mathcal{R}(P) \cap \mathcal{R}(I-P) = \mathcal{N}(I-P) \cap \mathcal{N}(P) = \{\mathbf{0}\}$.

So, any vector $\mathbf{x}$ can be uniquely decomposed into

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2, \qquad \mathbf{x}_1 \in \mathcal{R}(P), \quad \mathbf{x}_2 \in \mathcal{R}(I-P) = \mathcal{N}(P).$$

The most interesting situation occurs if the decomposition is orthogonal, i.e., if $\mathbf{x}_1^*\mathbf{x}_2 = 0$ for all $\mathbf{x}$.

A matrix P is called an orthogonal projection if

$$\begin{array}{ll} \text{(i)} & P^2 = P \\ \text{(ii)} & P^* = P. \end{array}$$

Projections IV

Example: Let $\mathbf{q}$ be an arbitrary vector of norm 1, $\|\mathbf{q}\| = \mathbf{q}^*\mathbf{q} = 1$. Then $P = \mathbf{q}\mathbf{q}^*$ is the orthogonal projection onto $\text{span}\{\mathbf{q}\}$.

Example: Let $Q \in \mathbb{F}^{n \times p}$ with $Q^*Q = I_p$. Then QQ^* is the orthogonal projector onto $\mathcal{R}(Q)$, which is the space spanned by the columns of Q.

Rayleigh quotient I

The Rayleigh quotient of A at $\mathbf{x}$ is defined as

$$\lambda = \rho(\mathbf{x}) := \frac{\mathbf{x}^* A\mathbf{x}}{\mathbf{x}^*\mathbf{x}}, \qquad \mathbf{x} \neq \mathbf{0}$$

If $\mathbf{x}$ is an approximate eigenvector, then $\rho(\mathbf{x})$ is a reasonable choice for the corresponding eigenvalue.
Using the spectral decomposition $A = U\Lambda U^*$,

$$\mathbf{x}^* A\mathbf{x} = \mathbf{x}^* U\Lambda U^*\mathbf{x} = \sum_{i=1}^{n} \lambda_i |\mathbf{u}_i^*\mathbf{x}|^2.$$

Similarly, $\mathbf{x}^*\mathbf{x} = \sum_{i=1}^{n} |\mathbf{u}_i^*\mathbf{x}|^2$.With $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, we have

$$\lambda_1 \sum_{i=1}^{n} |\mathbf{u}_i^*\mathbf{x}|^2 \leq \sum_{i=1}^{n} \lambda_i |\mathbf{u}_i^*\mathbf{x}|^2 \leq \lambda_n \sum_{i=1}^{n} |\mathbf{u}_i^*\mathbf{x}|^2.$$

Rayleigh quotient II

$$\Longrightarrow \lambda_1 \leq \rho(\mathbf{x}) \leq \lambda_n, \qquad \text{for all } \mathbf{x} \neq \mathbf{0}.$$

$$\rho(\mathbf{u}_k) = \lambda_k,$$

the extremal values λ_1 and λ_n are attained for $\mathbf{x} = \mathbf{u}_1$ and $\mathbf{x} = \mathbf{u}_n$.

Theorem

Let A be Hermitian. Then the Rayleigh quotient satisfies

$$\lambda_1 = \min \rho(\mathbf{x}), \qquad \lambda_n = \max \rho(\mathbf{x}). \tag{10}$$

As the Rayleigh quotient is a continuous function it attains *all* values in the closed interval $[\lambda_1, \lambda_n]$.

Theorem (Minimum-maximum principle)

Let A be Hermitian. Then

$$\lambda_p = \min_{X \in \mathbb{F}^{n\times p},\ \mathrm{rank}(X)=p} \max_{\mathbf{x}\neq \mathbf{0}} \rho(X\mathbf{x})$$

Proof: Let $U_{p-1} = [\mathbf{u}_1, \ldots, \mathbf{u}_{p-1}]$. For every $X \in \mathbb{F}^{n\times p}$ with full rank we can choose $\mathbf{x} \neq \mathbf{0}$ such that $U_{p-1}^* X\mathbf{x} = \mathbf{0}$. Then $\mathbf{0} \neq \mathbf{z} := X\mathbf{x} = \sum_{i=p}^{n} z_i \mathbf{u}_i$ and

$$\rho(\mathbf{z}) \geq \lambda_p.$$

For equality choose $X = [\mathbf{u}_1, \ldots, \mathbf{u}_p]$. □

Theorem (Monotonicity principle)

Let A be Hermitian and let $Q := [\mathbf{q}_1, \ldots, \mathbf{q}_p]$ with $Q^*Q = I_p$. Let $A' := Q^*AQ$ with eigenvalues $\lambda'_1 \le \cdots \le \lambda'_p$. Then

$$\lambda_k \le \lambda'_k, \qquad 1 \le k \le p.$$

Proof: Let $\mathbf{w}_1, \ldots, \mathbf{w}_p \in \mathbb{F}^p$, $\mathbf{w}_i^*\mathbf{w}_j = \delta_{ij}$, be the eigenvectors of A',

$$A'\mathbf{w}_i = \lambda'_i\mathbf{w}_i, \qquad 1 \le i \le p.$$

Vectors $Q\mathbf{w}_1, \ldots, Q\mathbf{w}_p$ are normalized and mutually orthogonal. Construct normalized vector $\mathbf{x}_0 = Q(a_1\mathbf{w}'_1 + \cdots + a_k\mathbf{w}'_k) \equiv Q\mathbf{a}$ that is orthogonal to the first $k-1$ eigenvectors of A, $\mathbf{x}_0^*\mathbf{u}_i = 0$, $1 \le i \le k-1$. Minimum-maximum principle:

$\Longrightarrow \quad \lambda_k \le R(\mathbf{x}_0) = \mathbf{a}^*Q^*AQ\mathbf{a} = \sum_{i=1}^k |a|_i^2 \lambda'_i \le \lambda'_k.$ □

Trace of a matrix

The trace of a matrix $A \in \mathbb{F}^{n\times n}$ is defined to be the sum of the diagonal elements of a matrix. Matrices that are similar have equal trace. Hence, by the spectral theorem,

$$\mathrm{trace}(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i .$$

Theorem

(Trace theorem)

$$\lambda_1 + \lambda_2 + \cdots + \lambda_p = \min_{X\in\mathbb{F}^{n\times p},\, X^*X=I_p} \mathrm{trace}(X^*AX)$$

The singular value decomposition (SVD) I

Theorem

(Singular value decomposition) *If $A \in \mathbb{C}^{m\times n}$ then there exist unitary matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ such that*

$$U^*AV = \Sigma = \begin{pmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_p) & 0 \\ 0 & 0 \end{pmatrix}, \qquad p = \min(m, n),$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$.

Relation of the singular value decomposition with the spectral decomposition of the Hermitian matrices A^*A and AA^*,

$$A = U\Sigma V^* \quad \Longrightarrow \quad A^*A = V\Sigma^2V^*, \qquad AA^* = U\Sigma^2U^*. \quad (11)$$

The singular value decomposition (SVD) II

The SVD of dense matrices is computed in a way that is very similar to the dense Hermitian eigenvalue problem. However, in the presence of roundoff error, it is not advisable to use the matrices A^*A or AA^*. Instead, let us consider the $(n+m)\times(n+m)$ Hermitian matrix

$$\begin{bmatrix} O & A \\ A^* & O \end{bmatrix}. \tag{12}$$

Using the SVD we get

$$\begin{bmatrix} O & A \\ A^* & O \end{bmatrix} = \begin{bmatrix} U & O \\ O & V \end{bmatrix} \begin{bmatrix} O & \Sigma \\ \Sigma^T & O \end{bmatrix} \begin{bmatrix} U^* & O \\ O & V^* \end{bmatrix}.$$

- Assume that $m \geq n$.

The singular value decomposition (SVD) III

- Then write $U = [U_1, U_2]$ where $U_1 \in \mathbb{F}^{m\times n}$ and $\Sigma = \begin{bmatrix} \Sigma_1 \\ O \end{bmatrix}$ with $\Sigma_1 \in \mathbb{R}^{n\times n}$.
- $$\begin{bmatrix} O & A \\ A^* & O \end{bmatrix} = \begin{bmatrix} U_1 & U_2 & O \\ O & O & V \end{bmatrix} \begin{bmatrix} O & O & \Sigma_1 \\ O & O & O \\ \Sigma_1 & O & O \end{bmatrix} \begin{bmatrix} U_1^* & O \\ U_2^* & O \\ O & V^* \end{bmatrix}$$
$$= \begin{bmatrix} U_1 & O & U_2 \\ O & V & O \end{bmatrix} \begin{bmatrix} O & \Sigma_1 & O \\ \Sigma_1 & O & O \\ O & O & O \end{bmatrix} \begin{bmatrix} U_1^* & O \\ O & V^* \\ U_2^* & O \end{bmatrix}.$$

  The first and third diagonal zero blocks have order n. The middle diagonal block has order $n - m$.

The singular value decomposition (SVD) IV

- Now, employ the fact that

$$\begin{bmatrix} 0 & \sigma \\ \sigma & 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma & 0 \\ 0 & -\sigma \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

to obtain

$$\begin{bmatrix} O & A \\ A^* & O \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} U_1 & \frac{1}{\sqrt{2}} U_1 & U_2 \\ \frac{1}{\sqrt{2}} V & -\frac{1}{\sqrt{2}} V & O \end{bmatrix} \begin{bmatrix} \Sigma_1 & O & O \\ O & -\Sigma_1 & O \\ O & O & O \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} U_1^* & \frac{1}{\sqrt{2}} V^* \\ \frac{1}{\sqrt{2}} U_1^* & -\frac{1}{\sqrt{2}} V^* \\ U_2^* & O \end{bmatrix}.$$

So there are 3 ways to treat the singular value decomposition as an eigenvalue problem. One of the two forms in (11) is used *implicitly* in the QR algorithm for dense matrices A, see Golub & van Loan or the LAPACK users guide.
The form (12) is suited if A is a sparse matrix.