Chapter 2

Basics

2.1 Notation

The fields of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. Elements in $\mathbb{R}$ and $\mathbb{C}$, *scalars*, are denoted by lowercase letters, a, b, $c, \ldots$, and α, β, $\gamma, \ldots$

Vectors are denoted by boldface lowercase letters, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}, \ldots$, and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, ... We denote the space of vectors of n *real* components by $\mathbb{R}^n$ and the space of vectors of n *complex* components by $\mathbb{C}^n$.

$$\mathbf{x} \in \mathbb{R}^n \Longleftrightarrow \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad x_i \in \mathbb{R}. \tag{2.1}$$

We often make statements that hold for real or complex vectors or matrices. Then we write, e.g., $\mathbf{x} \in \mathbb{F}^n$.

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

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x_i \bar{y}_i = \mathbf{y}^* \mathbf{x}, \tag{2.2}$$

that is, we require linearity in the first component and anti-linearity in the second.

$\mathbf{y}^* = (\bar{y}_1, \ \bar{y}_2, \ldots, \bar{y}_n)$ denotes conjugate transposition of complex vectors. To simplify notation we denote real transposition by an asterisk as well.

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are called **orthogonal**, $\mathbf{x} \perp \mathbf{y}$, if $\mathbf{x}^* \mathbf{y} = 0$.

The inner product (2.2) induces a **norm** in $\mathbb{F}$,

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})} = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}. \tag{2.3}$$

This norm is often called Euclidean norm or 2-norm.

The set of m-by-n **matrices** with components in the field $\mathbb{F}$ is denoted by $\mathbb{F}^{m \times n}$,

$$A \in \mathbb{F}^{m \times n} \Longleftrightarrow A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}, \quad a_{ij} \in \mathbb{F}. \tag{2.4}$$

The matrix $A^* \in \mathbb{F}^{n\times m}$,

$$A^* = \begin{pmatrix} \bar{a}_{11} & \bar{a}_{21} & \dots & \bar{a}_{m1} \\ \bar{a}_{12} & \bar{a}_{22} & \dots & \bar{a}_{m2} \\ \vdots & \vdots & & \vdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \dots & \bar{a}_{nm} \end{pmatrix} \tag{2.5}$$

is the **Hermitian transpose** of A. Notice, that with this notation n-vectors can be identified with n-by-1 matrices.

The following classes of square matrices are of particular importance:

- $A \in \mathbb{F}^{n\times n}$ is called **Hermitian** if and only if $A^* = A$.
- A *real* Hermitian matrix is called **symmetric**.
- $U \in \mathbb{F}^{n\times n}$ is called **unitary** if and only if $U^{-1} = U^*$.
- *Real* unitary matrices are called **orthogonal**.
- $A \in \mathbb{F}^{n\times n}$ is called **normal** if $A^*A = AA^*$. Both, Hermitian and unitary matrices are normal.

We define the norm of a matrix to be the norm induced by the vector norm (2.3),

$$\|A\| := \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|. \tag{2.6}$$

The condition number of a nonsingular matrix is defined as $\kappa(A) = \|A\|\|A^{-1}\|$. It is easy to show that if U is unitary then $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$. Thus the condition number of a unitary matrix is 1.

2.2 Statement of the problem

The **(standard) eigenvalue problem** is as follows.

> 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
>
> $$A\mathbf{x} = \lambda\mathbf{x}, \tag{2.7}$$
>
> i.e., such that
>
> $$(A - \lambda I)\mathbf{x} = \mathbf{0} \tag{2.8}$$
>
> has a nontrivial (nonzero) solution.

So, we are looking for numbers λ such that $A - \lambda I$ is *singular*.

Definition 2.1 Let the pair $(\lambda, \mathbf{x})$ be a solution of (2.7) or (2.8), respectively. Then

- 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. As the eigenspace of λ is the null space of $\lambda I - A$ we denote it by $\mathcal{N}(\lambda I - A)$.
- The dimension of $\mathcal{N}(\lambda I - A)$ is called **geometric multiplicity** $g(\lambda)$ of λ.
- An eigenvalue λ is a zero 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 zero of χ is called the **algebraic multiplicity** $m(\lambda)$ of λ. We will later see that

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

Remark 2.1. A nontrivial solution solution $\mathbf{y}$ of

$$\mathbf{y}^* A = \lambda \mathbf{y}^* \tag{2.9}$$

is called **left eigenvector** corresponding to λ. A left eigenvector of A is a right eigenvector of A^*, corresponding to the eigenvalue $\bar{\lambda}$, $A^*\mathbf{y} = \bar{\lambda}\mathbf{y}$. □

Problem 2.2 Let $\mathbf{x}$ be a (right) eigenvector of A corresponding to an eigenvalue λ and let $\mathbf{y}$ be a left eigenvector of A corresponding to a *different* eigenvalue $\mu \neq \lambda$. Show that $\mathbf{x}^*\mathbf{y} = 0$.

Remark 2.2. Let A be an **upper triangular** matrix,

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ & a_{22} & \dots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{pmatrix}, \qquad a_{ik} = 0 \text{ for } i > k. \tag{2.10}$$

Then we have

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

□

Problem 2.3 Let $\lambda = a_{ii}$, $1 \le i \le n$, be an eigenvalue of A in (2.10). Can you give a corresponding eigenvector? Can you detect a situation where $g(\lambda) < m(\lambda)$?

The **(generalized) eigenvalue problem** is as follows.

> 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
>
> $$A\mathbf{x} = \lambda B\mathbf{x}, \tag{2.11}$$
>
> or, equivalently, such that
>
> $$(A - \lambda B)\mathbf{x} = \mathbf{0} \tag{2.12}$$
>
> has a nontrivial solution.

Definition 2.4 Let the pair $(\lambda, \mathbf{x})$ be a solution of (2.11) or (2.12), respectively. Then

- 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 (2.11) is called the **spectrum** of A **relative to** B.

Let us look at some examples.

- Let B be nonsingular. Then

  $$A\mathbf{x} = \lambda B\mathbf{x} \Longleftrightarrow B^{-1}A\mathbf{x} = \lambda\mathbf{x} \tag{2.13}$$

- Let both A and B be Hermitian, $A = A^*$ and $B = B^*$. Let further be B positive definite and $B = LL^*$ be its Cholesky factorization. Then

  $$A\mathbf{x} = \lambda B\mathbf{x} \Longleftrightarrow L^{-1}AL^{-*}\mathbf{y} = \lambda\mathbf{y}, \quad \mathbf{y} = L^*\mathbf{x}. \tag{2.14}$$

- Let A be invertible. Then $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$. That is, $0 \notin \sigma(A; B)$. Therefore,

  $$A\mathbf{x} = \lambda B\mathbf{x} \Longleftrightarrow \mu\mathbf{x} = A^{-1}B\mathbf{x}, \quad \mu = \frac{1}{\lambda} \tag{2.15}$$

- Let $A = B \in \mathbb{R}^{n\times n}$ be invertible. Then

  $$A\mathbf{x} = \lambda B\mathbf{x} \Longleftrightarrow B^{-1}A\mathbf{x} = I\mathbf{x} = \mathbf{x}.$$

  Therefore, $\sigma(A; B) = \{1\}$. The associated eigenspace is $\mathbb{R}^n$. Every nonzero vector $\mathbf{x}$ is an eigenvector.

- *Difficult situation*: both A and B are singular.

  1. Let, e.g.,

     $$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

     Then,

     $$A\mathbf{e}_2 = \mathbf{0} = 0 \cdot B\mathbf{e}_2 = 0 \cdot \mathbf{e}_2,$$

     such that 0 is an eigenvalue of A relative to B. Since

     $$A\mathbf{e}_1 = \mathbf{e}_1 = \lambda B\mathbf{e}_1 = \lambda\mathbf{0}$$

     $\mathbf{e}_1$ cannot be an eigenvector of A relative to B.
     As in (2.15) we may swap the roles of A and B. Then

     $$B\mathbf{e}_1 = \mathbf{0} = \mu A\mathbf{e}_1 = 0\mathbf{e}_1.$$

     So, $\mu = 0$ is an eigenvalue of B relative to A. We therefore say, informally, that $\lambda = \infty$ is an eigenvalue of A relative to B. So, $\sigma(A; B) = \{0, \infty\}$.
  2. Let

     $$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = A.$$

     Then,

     $$\begin{aligned} A\mathbf{e}_1 &= 1 \cdot B\mathbf{e}_1, \\ A\mathbf{e}_2 &= \mathbf{0} = \lambda B\mathbf{e}_2 = \lambda\mathbf{0}, \ \text{for } \textbf{all } \lambda \in \mathbb{C}. \end{aligned}$$

     Therefore, in this case, $\sigma(A; B) = \mathbb{C}$.

2.3 Similarity transformations

Definition 2.5 A matrix $A \in \mathbb{F}^{n\times n}$ is **similar** to a matrix $C \in \mathbb{F}^{n\times n}$, $A \sim C$, if and only if there is a nonsingular matrix S such that

$$S^{-1}AS = C. \tag{2.16}$$

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

Theorem 2.6 *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.*

Proof. $A\mathbf{x} = \lambda\mathbf{x}$ and $C = S^{-1}AS$ imply that

$$CS^{-1}\mathbf{x} = S^{-1}ASS^{-1}\mathbf{x} = S^{-1}\lambda\mathbf{x}.$$

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

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

it follows that the characteristic polynomials of A and C are equal and hence also the algebraic eigenvalue multiplicities are equal. ■

Similarity transformations are used to transform matrices into similar matrices from which eigenvalues can be easily read. Diagonal matrices are the preferred matrix structure. However, not all matrices are diagonalizable. There is, e.g., no invertible matrix S that diagonalizes the matrix

$$\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}.$$

In the Jordan normal form introduced in section 2.8 the transformation is into a bidiagonal matrix. In the Schur normal form, see section 2.4 the transformation is into an upper tridiagonal matrix, but with an unitary S.

Definition 2.7 Two matrices A and B are called **unitarily similar** if S in (2.16) is unitary. If the matrices are real the term orthogonally similar is used.

Unitary similarity transformations are very important in numerical computations. Let U be unitary. Then $\|U\| = \|U^{-1}\| = 1$, the condition number of U is therefore $\kappa(U) = 1$. Hence, if $C = U^{-1}AU$ then $C = U^*AU$ and $\|C\| = \|A\|$. In particular, if A is disturbed by δA (e.g., roundoff errors introduced when storing the entries of A in finite-precision arithmetic) then

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

Hence, errors (perturbations) in A are not amplified by a unitary similarity transformation. This is in contrast to arbitrary similarity transformations. However, as we will see later, small eigenvalues may still suffer from large relative errors.

Another reason for the importance of unitary similarity transformations is the preservation of symmetry: If A is symmetric then $U^{-1}AU = U^*AU$ is symmetric as well.

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 sometimes called *equivalence transformation* of A, B. Thus, $\sigma(A; B) = \sigma(TAS^{-1}; TBS^{-1})$. Let us consider a special case: let B be invertible and let $B = LU$ be the LU-factorization of B. Then we set $S = U$ and $T = L^{-1}$ and obtain $TBU^{-1} = L^{-1}LUU^{-1} = I$. Thus, $\sigma(A; B) = \sigma(L^{-1}AU^{-1}; I) = \sigma(L^{-1}AU^{-1})$.

2.4 Schur decomposition

Theorem 2.8 ***(Schur decomposition)*** *If $A \in \mathbb{C}^{n\times n}$ then there is a unitary matrix $U \in \mathbb{C}^{n\times n}$ such that*

$$U^*AU = T \tag{2.17}$$

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

Proof. The proof is by induction. For $n = 1$, the theorem is obviously true.

Assume that the theorem holds for matrices of order $\le n-1$. Let $(\lambda, \mathbf{x})$, $\|\mathbf{x}\| = 1$, be an eigenpair of A, $A\mathbf{x} = \lambda\mathbf{x}$. We 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

$$U_1^*AU_1 = \begin{bmatrix} \mathbf{x}^*A\mathbf{x} & \mathbf{x}^*A\overline{U} \\ \overline{U}^*A\mathbf{x} & \overline{U}^*A\overline{U} \end{bmatrix} = \begin{bmatrix} \lambda & \times\cdots\times \\ \mathbf{0} & \hat{A} \end{bmatrix}$$

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 (2.17). ∎

Notice , that this proof is not constructive as we assume the knowledge of an eigenpair $(\lambda, \mathbf{x})$. So, we cannot employ it to actually compute the Schur form. The QR algorithm is used for this purpose. We will discuss this basic algorithm in Chapter 4.

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

$$A\mathbf{u}_k = \lambda\mathbf{u}_k + \sum_{i=1}^{k-1} t_{ik}\mathbf{u}_i, \qquad \lambda_k = t_{kk}. \tag{2.18}$$

This implies that

$$A\mathbf{u}_k \in \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}, \quad \forall k. \tag{2.19}$$

Thus, the first k **Schur vectors** $\mathbf{u}_1, \ldots, \mathbf{u}_k$ form an **invariant subspace**[1] for A. From (2.18) it is clear that the *first* Schur vector is an eigenvector of A. The other columns of U, however, are in general *not* eigenvectors of A. Notice, that the Schur decomposition is not unique. In the proof we have chosen *any* eigenvalue λ. This indicates that the eigenvalues can be arranged in any order in the diagonal of T. This also indicates that the order with which the eigenvalues appear on T's diagonal can be manipulated.

Problem 2.9 Let

$$A = \begin{bmatrix} \lambda_1 & \alpha \\ 0 & \lambda_2 \end{bmatrix}.$$

[1] A subspace $\mathcal{V} \subset \mathbb{F}^n$ is called invariant for A if $A\mathcal{V} \subset \mathcal{V}$.

Find an orthogonal 2×2 matrix Q such that

$$Q^* A Q = \begin{bmatrix} \lambda_2 & \beta \\ 0 & \lambda_1 \end{bmatrix}.$$

Hint: the first column of Q must be a normalized eigenvector of A corresponding to eigenvalue λ_2. Why?

2.5 The real Schur decomposition

Real matrices can have complex eigenvalues. If complex eigenvalues exist, then they occur in *complex conjugate pairs*! That is, if λ is an eigenvalue of the real matrix A, then also $\bar{\lambda}$ is an eigenvalue of A. The following theorem indicates that complex computation can be avoided.

Theorem 2.10 ***(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*

$$Q^T A Q = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ & R_{22} & \cdots & R_{2m} \\ & & \ddots & \vdots \\ & & & R_{mm} \end{bmatrix} \tag{2.20}$$

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.

Remark 2.3. The matrix

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

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

Proof. 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 $\bar{\mathbf{x}} = \mathbf{u} - i\mathbf{v}$. To see this we first observe that

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

Thus,

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

Now, the actual proof starts. Let k be the number of complex conjugate pairs. We prove the theorem by induction on k.

First we consider the case $k = 0$. In this case A has real eigenvalues and eigenvectors. It is clear that we can repeat the proof of the Schur decomposition of Theorem 2.8 in real arithmetic to get the decomposition (2.17) with $U \in \mathbb{R}^{n\times n}$ and $T \in \mathbb{R}^{n\times n}$. So, there are n diagonal blocks R_{jj} in (2.20) all of which are 1×1.

Let us now 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 $\mathbf{x} = \mathbf{u} + i\mathbf{v}$, as previously, we have

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

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[2], there is a nonsingular 2×2 real square matrix C with

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

Now,

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

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

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

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. ■

2.6 Normal matrices

Definition 2.11 A matrix $A \in \mathbb{F}^{n\times n}$ is called **normal** if

$$AA^* = A^*A. \tag{2.21}$$

Let $A = URU^*$ be the Schur decomposition of A. Then,

$$RR^* = U^*AUU^*A^*U = U^*AA^*U = U^*A^*AU = U^*A^*UU^*AU = R^*R.$$

Therefore, also the upper triangular R is normal. We look at the (1,1)-elements of RR^* and R^*R that evidently must be equal. On one hand we have

$$(R^*R)_{11} = \bar{r}_{11} r_{11} = |r_{11}|^2,$$

on the other hand

$$(RR^*)_{11} = \sum_{j=1}^{n} r_{1j}\bar{r}_{1j} = |r_{11}|^2 + \sum_{j=2}^{n} |r_{1j}|^2.$$

Therefore, the latter sum must vanish, i.e., $r_{1j} = 0$ for $j = 2, \ldots, n$. Comparing the (2,2)-elements, (3,3)-elements, etc., of RR^* and R^*R, we see that R is diagonal. In this way we arrive at

Theorem 2.12 *A matrix is normal if and only if it is diagonalizable by a unitary matrix.* ■

(Note that unitarily diagonalizable matrices are trivially normal.)

[2]If u and v were linearly dependent then it follows that β must be zero.

2.7 Hermitian matrices

Definition 2.13 A matrix $A \in \mathbb{F}^{n\times n}$ is **Hermitian** if

$$A = A^*. \tag{2.22}$$

The Schur decomposition for Hermitian matrices is particularly simple. We first note that A being Hermitian implies that the upper triangular Λ in the Schur decomposition $A = U\Lambda U^*$ is Hermitian and thus diagonal. In fact, because

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

each diagonal element λ_i of Λ satisfies $\overline{\lambda_i} = \lambda_i$. So, Λ has to be *real*. In summary have the following result.

Theorem 2.14 ***(Spectral theorem for Hermitian matrices)*** *Let A be Hermitian. Then there is a unitary matrix U and a real diagonal matrix Λ such that*

$$A = U\Lambda U^* = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^*. \tag{2.23}$$

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

The decomposition (2.23) 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 $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$.

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,

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

However, eigenvectors corresponding to *different* eigenvalues are orthogonal. Let $A\mathbf{u} = \mathbf{u}\lambda$ and $A\mathbf{v} = \mathbf{v}\mu$, $\lambda \ne \mu$. Then

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

and thus

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

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

In summary, 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 (2.23) 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 2.4. 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 $\mathbf{y}$. For the ordinary Euclidean inner product $(\mathbf{x}, \mathbf{y}) = \mathbf{x}^*\mathbf{y}$ we arrive at the element-wise Definition 2.7 if we set $\mathbf{x}$ and $\mathbf{y}$ equal to coordinate vectors.

It is important to note that all the properties of Hermitian matrices that we will derive subsequently hold similarly for matrices symmetric with respect to a certain inner product.
□

Example 2.15 We consider the one-dimensional Sturm-Liouville eigenvalue problem

$$-u''(x) = \lambda u(x), \quad 0 < x < \pi, \quad u(0) = u(\pi) = 0, \tag{2.24}$$

that models the vibration of a homogeneous string of length π that is *fixed* at both ends. The eigenvalues and eigenvectors or eigenfunctions of (2.24) are

$$\lambda_k = k^2, \quad u_k(x) = \sin kx, \qquad k \in \mathbb{N}.$$

Let $u_i^{(n)}$ denote the approximation of an (eigen)function u at the grid point x_i,

$$u_i \approx u(x_i), \quad x_i = ih, \quad 0 \le i \le n+1, \quad h = \frac{\pi}{n+1}.$$

We approximate the second derivative of u at the *interior* grid points by finite differences [3, 8]

$$\frac{1}{h^2}(-u_{i-1} + 2u_i - u_{i+1}) = \lambda u_i, \qquad 1 \le i \le n. \tag{2.25}$$

Collecting these equations and taking into account the boundary conditions, $u_0 = 0$ and $u_{n+1} = 0$, we get a (matrix) eigenvalue problem

$$T_n \mathbf{x} = \lambda \mathbf{x} \tag{2.26}$$

where

$$T_n := \frac{(n+1)^2}{\pi^2} \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

The matrix eigenvalue problem (2.26) can be solved explicitly [10, p.229]. Eigenvalues and eigenvectors are given by

$$\begin{aligned} \lambda_k^{(n)} &= \frac{(n+1)^2}{\pi^2}(2 - 2\cos\phi_k) = \frac{4(n+1)^2}{\pi^2}\sin^2\frac{k\pi}{2(n+1)}, \\ \mathbf{u}_k^{(n)} &= \left(\frac{2}{n+1}\right)^{1/2} [\sin\phi_k, \sin 2\phi_k, \ldots, \sin n\phi_k]^T, \qquad \phi_k = \frac{k\pi}{n+1}. \end{aligned} \tag{2.27}$$

Clearly, $\lambda_k^{(n)}$ converges to λ_k as $n \to \infty$. (Note that $\sin\xi \to \xi$ as $\xi \to 0$.) When we identify $\mathbf{u}_k^{(n)}$ with the piecewise linear function that takes on the values given by $\mathbf{u}_k^{(n)}$ at the grid points x_i then this function evidently converges to $\sin kx$.

Let $p(\lambda)$ be a polynomial of degree d, $p(\lambda) = \alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \cdots + \alpha_d\lambda^d$. As $A^j = (U\Lambda U^*)^j = U\Lambda^j U^*$ we can define a *matrix polynomial* as

$$p(A) = \sum_{j=0}^{d} \alpha_j A^j = \sum_{j=0}^{d} \alpha_j U\Lambda^j U^* = U\left(\sum_{j=0}^{d} \alpha_j \Lambda^j\right) U^*. \tag{2.28}$$

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

2.8 The Jordan normal form

Theorem 2.16 ***(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*

$$X^{-1}AX = J = \mathrm{diag}(J_1, J_2, \ldots, J_p), \tag{2.29}$$

where

$$J_k = J_{m_k}(\lambda_k) = \begin{bmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix} \in \mathbb{F}^{m_k\times m_k} \tag{2.30}$$

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.*

A matrix is diagonalizable if all Jordan blocks are 1×1, i.e., $m_k = 1$ for all k[3]. In this case the columns of X are eigenvectors of A.

More generally, there is one eigenvector associated with each Jordan block, e.g.,

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

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

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

This choice of generalized eigenvectors is not unique though, as $(J_k(\lambda)-\lambda I)(\mathbf{e}_2+\alpha\mathbf{e}_1) = \mathbf{e}_1$ for any α. This is one of the reasons for the non-uniqueness of the transformation matrix X in Theorem 2.16.

From the Jordan blocks we can read geometric and algebraic multiplicity of an eigenvalue: The number of Jordan blocks associated with a particular eigenvalue give the geometric multiplicity; the sum of its orders gives the algebraic multiplicity.

Numerically the size of the Jordan blocks cannot be determined stably as the following example shows. Let

$$\begin{bmatrix} \varepsilon & 1 \\ 0 & -\varepsilon \end{bmatrix} \approx \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = J_2(0)$$

be the approximation for $J_2(0)$ that some numerical algorithm has computed. This matrix has two distinct eigenvalues and thus two eigenvectors,

$$\begin{bmatrix} \varepsilon & 1 \\ 0 & -\varepsilon \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -2\varepsilon \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & -2\varepsilon \end{bmatrix} \begin{bmatrix} \varepsilon & 0 \\ 0 & -\varepsilon \end{bmatrix}.$$

For small ε the two eigenvectors are very close. They even collaps when $\varepsilon \to 0$. A numerical code cannot differ between the two cases ($\varepsilon = 0$, $\varepsilon \neq 0$) that have a completely different structure.

[3] 1×1 Jordan blocks are called trivial.

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 in (2.29), meaning that $X_j, Y_j \in \mathbb{F}^{n\times m_j}$. Then,

$$A = XJY^* = \sum_{k=1}^{p} X_k J_k Y_k^* = \sum_{k=1}^{p} (\lambda_k X_k Y_k^* + X_k N_k Y_k^*), \tag{2.31}$$

where $N_k = J_{m_k}(0)$. If $m_k = 1$ then N_k is zero. We define the matrices $P_k := X_k Y_k^*$ and $D_k := X_k N_k Y_k^*$. Then, 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**. From (2.31) we immediately obtain [9]

$$A = \sum_{k=1}^{p} (\lambda_k P_k + D_k). \tag{2.32}$$

Since $I_{m_k} N_k = N_k I_{m_k} = N_k$, we have

$$\begin{aligned} P_k D_\ell &= D_\ell P_k = \delta_{k\ell} D_\ell, \\ A P_k &= P_k A = P_k A P_k = \lambda_k P_k + D_k, \\ A^j P_k &= P_k A^j = P_k A^j P_k = P_k(\lambda_k I_n + D_k)^j = (\lambda_k I_n + D_k)^j P_k. \end{aligned}$$

The Jordan normal form can be computed from the Schur decomposition $A = U^*TU$, see, e.g., [2], although it is not recommended in general to do so.

1. Group equal eigenvalues on the diagonal of the triangular T. This is a generalization of the solution of Problem 2.4.

2. Let

$$T = \begin{bmatrix} T_1 & T_{12} & \cdots & T_{1s} \\ & T_2 & \cdots & T_{2s} \\ & & \ddots & \vdots \\ & & & T_s \end{bmatrix} \tag{2.33}$$

   where the s diagonal blocks T_k are related to the s distinct eigenvalues of T. The off-diagonal blocks $T_{j\ell}$ are zeroed one after the other. Each steps requires the solution of a **Sylvester equation** $T_{j\ell} = T_j Y - Y T_\ell$.

   ***Exercise*:** Consider the case of two (simple or multiple) eigenvalues,

$$T = \begin{bmatrix} T_1 & T_{12} \\ & T_2 \end{bmatrix}.$$

   Apply a similarity transformation with the matrix

$$X = \begin{bmatrix} I_1 & Y \\ & I_2 \end{bmatrix}.$$

   Determine Y? How can this be extended to the case (2.33) with s diagonal blocks?

3. The diagonal blocks $T_1, \ldots, T_s$ are brought to Jordan form.

The Jordan normal form can be nicely employed to define matrix functions, see [6].

2.9 Projections

Definition 2.17 A matrix P that satisfies

$$P^2 = P \tag{2.34}$$

is called a **projection**.

Obviously, 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 $P\mathbf{x} = P(P\mathbf{y}) = P^2\mathbf{y} = P\mathbf{y} = \mathbf{x}$.

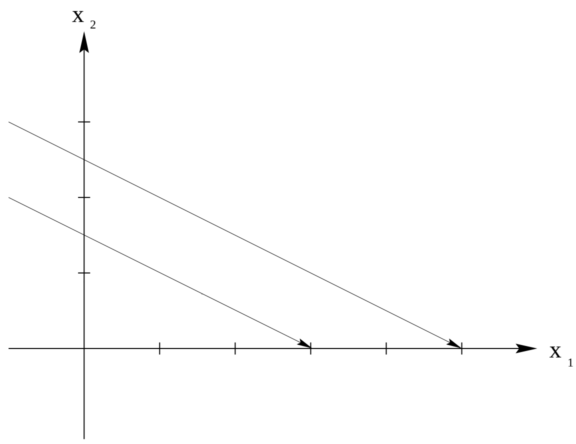

Figure 2.1: Oblique projection of example 2.9

Example 2.18 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 2.1: 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.

Example 2.19 Let $\mathbf{x}$ and $\mathbf{y}$ be arbitrary vectors such that $\mathbf{y}^*\mathbf{x} \neq 0$. Then

$$P = \frac{\mathbf{x}\mathbf{y}^*}{\mathbf{y}^*\mathbf{x}} \tag{2.35}$$

is a projection. Notice that the projector of the previous example can be expressed in the form (2.35).

Problem 2.20 Let $X, Y \in \mathbb{F}^{n \times p}$ such that Y^*X is nonsingular. Show that

$$P := X(Y^*X)^{-1}Y^*$$

is a projection.

Example 2.21 The spectral projectors $X_k Y_k^*$ introduced in (2.31) are projectors. Their range is the span of all eigenvectors and generalized eigenvectors associated with the eigenvalue λ_k.

If P is a projection then $I-P$ is a projection as well. In fact, $(I-P)^2 = I-2P+P^2 = I-2P+P = I-P$. If $P\mathbf{x} = \mathbf{0}$ then $(I-P)\mathbf{x} = \mathbf{x}$. Therefore, the range of $I-P$ coincides with the null space of P, $\mathcal{R}(I-P) = \mathcal{N}(P)$. It can be shown that $\mathcal{R}(P) = \mathcal{N}(P^*)^\perp$.

Notice that $\mathcal{R}(P) \cap \mathcal{R}(I-P) = \mathcal{N}(I-P) \cap \mathcal{N}(P) = \{\mathbf{0}\}$. For, if $P\mathbf{x} = \mathbf{0}$ then $(I-P)\mathbf{x} = \mathbf{x}$, which can only be zero if $\mathbf{x} = \mathbf{0}$. So, any vector $\mathbf{x}$ can be uniquely decomposed into

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2, \qquad \mathbf{x}_1 \in \mathcal{R}(P), \quad \mathbf{x}_2 \in \mathcal{R}(I-P) = \mathcal{N}(P). \tag{2.36}$$

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

Definition 2.22 A matrix P is called an **orthogonal projection** if

$$\begin{array}{rl} \text{(i)} & P^2 = P \\ \text{(ii)} & P^* = P. \end{array} \tag{2.37}$$

Proposition 2.23 *Let P be a projection. Then the following statements are equivalent.*
(i) $P^* = P$,
(ii) $\mathcal{R}(I-P) \perp \mathcal{R}(P)$, *i.e.* $(P\mathbf{x})^*(I-P)\mathbf{y} = 0$ *for all* $\mathbf{x}, \mathbf{y}$.

Proof. (ii) follows trivially from (i) and (2.34).

Now, let us assume that (ii) holds. Then

$$\begin{aligned} \mathbf{x}^*P^*\mathbf{y} &= (P\mathbf{x})^*\mathbf{y} = (P\mathbf{x})^*(P\mathbf{y} + (I-P)\mathbf{y}) \\ &= (P\mathbf{x})^*(P\mathbf{y}) \\ &= (P\mathbf{x} + (I-P)\mathbf{x})(P\mathbf{y}) = \mathbf{x}^*(P\mathbf{y}). \end{aligned}$$

This equality holds for any $\mathbf{x}$ and $\mathbf{y}$ and thus implies (i). ∎

Example 2.24 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 span$\{\mathbf{q}\}$.

Example 2.25 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.

Problem 2.26 Let $Q, Q_1 \in \mathbb{F}^{n\times p}$ with $Q^*Q = Q_1^*Q_1 = I_p$ such that $\mathcal{R}(Q) = \mathcal{R}(Q_1)$. This means that the columns of Q and Q_1, respectively, are orthonormal bases of the *same* subspace of $\mathbb{F}^n$. Show that the projector does not depend on the basis of the subspace, i.e., that $QQ^* = Q_1Q_1^*$.

Problem 2.27 Let $Q = [Q_1, Q_2]$, $Q_1 \in \mathbb{F}^{n\times p}$, $Q_2 \in \mathbb{F}^{n\times(n-p)}$ be a unitary matrix. Q_1 contains the first p columns of Q, Q_2 the last $n-p$. Show that $Q_1Q_1^* + Q_2Q_2^* = I$. Hint: Use $QQ^* = I$. Notice, that if $P = Q_1Q_1^*$ then $I - P = Q_2Q_2^*$.

Problem 2.28 What is the form of the orthogonal projection onto span$\{\mathbf{q}\}$ if the inner product is defined as $\langle \mathbf{x}, \mathbf{y}\rangle := \mathbf{y}^*M\mathbf{x}$ where M is a symmetric positive definite matrix?

2.10 The Rayleigh quotient

Definition 2.29 The quotient

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

is called the *Rayleigh quotient* of A at $\mathbf{x}$.

Notice, that $\rho(\mathbf{x}\alpha) = \rho(\mathbf{x})$, $\alpha \neq 0$. Hence, the properties of the Rayleigh quotient can be investigated by just considering its values on the unit sphere. Using the spectral decomposition $A = U\Lambda U^*$, we get

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

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

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

So,

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

As

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

the extremal values λ_1 and λ_n are actually attained for $\mathbf{x} = \mathbf{u}_1$ and $\mathbf{x} = \mathbf{u}_n$, respectively. Thus we have proved the following theorem.

Theorem 2.30 *Let A be Hermitian. Then the Rayleigh quotient satisfies*

$$\lambda_1 = \min_{\mathbf{x}\neq\mathbf{0}} \rho(\mathbf{x}), \qquad \lambda_n = \max_{\mathbf{x}\neq\mathbf{0}} \rho(\mathbf{x}). \tag{2.38}$$

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

The next theorem generalizes the above theorem to interior eigenvalues. It is attributed to Poincaré, Fischer and Pólya.

Theorem 2.31 ***(Minimum-maximum principle)*** *Let A be Hermitian. Then*

$$\lambda_p = \min_{X\in\mathbb{F}^{n\times p},\ \mathrm{rank}(X)=p} \ \max_{\mathbf{x}\neq\mathbf{0}} \rho(X\mathbf{x}) \tag{2.39}$$

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 $\mathbf{0} \neq \mathbf{z} := X\mathbf{x} = \sum_{i=p}^{n} z_i \mathbf{u}_i$. As in the proof of the previous theorem we obtain the inequality

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

To prove that equality holds in (2.39) we choose $X = [\mathbf{u}_1, \ldots, \mathbf{u}_p]$. Then

$$U_{p-1}^* X\mathbf{x} = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0}$$

implies that $\mathbf{x} = \mathbf{e}_p$, i.e., that $\mathbf{z} = X\mathbf{x} = \mathbf{u}_p$. So, $\rho(\mathbf{z}) = \lambda_p$. ∎

An important consequence of the minimum-maximum principle is the following

Theorem 2.32 ***(Monotonicity principle)*** *Let A be Hermitian and let $\mathbf{q}_1, \ldots, \mathbf{q}_p$ be normalized, mutually orthogonal vectors. Set $Q := [\mathbf{q}_1, \ldots, \mathbf{q}_p]$ and $A' := Q^*AQ \in \mathbb{F}^{p\times p}$. Then the p eigenvalues $\lambda'_1 \le \cdots \le \lambda'_p$ of A' satisfy*

$$\lambda_k \le \lambda'_k, \qquad 1 \le k \le p. \tag{2.40}$$

Proof. Let $\mathbf{w}_1, \ldots, \mathbf{w}_p \in \mathbb{F}^p$ be the eigenvectors of A',

$$A'\mathbf{w}_i = \lambda'_i \mathbf{w}_i, \qquad 1 \le i \le p, \tag{2.41}$$

with $\mathbf{w}_i^*\mathbf{w}_j = \delta_{ij}$. Then the vectors $Q\mathbf{w}_1, \ldots, Q\mathbf{w}_p$ are normalized and mutually orthogonal. Therefore, we can construct a normalized vector $\mathbf{x}_0$ with $\|\mathbf{x}_0\| = 1$,

$$\mathbf{x}_0 := a_1 Q\mathbf{w}_1 + \cdots + a_k Q\mathbf{w}_k = Q(a_1\mathbf{w}_1 + \cdots + a_k\mathbf{w}_k) = Q\mathbf{a},$$

that is orthogonal to the first $k-1$ eigenvectors of A,

$$\mathbf{x}_0^*\mathbf{u}_i = 0, \qquad 1 \le i \le k-1.$$

(Note, that $\|\mathbf{x}_0\| = 1$ implies $\|\mathbf{a}\| = 1$.) Then, with the minimum-maximum principle we get

$$\lambda_k = \min_{\substack{\mathbf{x}\ne\mathbf{0} \\ \mathbf{x}^*\mathbf{u}_1=\cdots=\mathbf{x}^*\mathbf{u}_{k-1}=0}} R(\mathbf{x}) \le R(\mathbf{x}_0) = \mathbf{x}_0^* A\mathbf{x}_0 = \mathbf{a}^*Q^*AQ\mathbf{a} = \sum_{i=1}^{k} |a|_i^2 \lambda'_i \le \lambda'_k.$$

■

***Exercise*:** It is possible to prove the inequalities (2.40) without assuming that the $\mathbf{q}_1, \ldots, \mathbf{q}_p$ are orthonormal. But then one has to use the eigenvalues λ'_k of

$$A'\mathbf{x} = \lambda' B\mathbf{x}, \quad B' = Q^*Q,$$

instead of (2.41). Prove this. □

Remark 2.5. Let $\mathbf{q}_i = \mathbf{e}_{j_i}$, $1 \le i \le k$. This means that we extract rows and columns $j_1, \ldots, j_k$ to construct A'. (The indices j_i are assumed to be distinct.) □

Remark 2.6. Let's remove a single row/column (with equal index) from A. Then $k = n-1$ in Remark 2.5 and the index set $j_1, \ldots, j_{n-1}$ contains all but one of the integers $1, \ldots, n$.

If we formulate a monotonicity principle based on the eigenvalues $\lambda_n, \lambda_{n-1}, \ldots$ as consecutive maxima of the Rayleigh quotient, then we arrive at the ***interlacing property***

$$\lambda_k \le \lambda'_k \le \lambda_{k+1}, \qquad 1 \le k < n. \tag{2.42}$$

This interlacing property can be generalized, see, e.g., [7, 5]. □

The ***trace*** of a matrix $A \in \mathbb{F}^{n\times n}$ is defined to be the sum of its diagonal elements. Matrices that are similar have equal trace [4, p.89]. Hence, by the spectral theorem,

$$\operatorname{trace}(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i. \tag{2.43}$$

The following theorem is a generalization of Theorem 2.30.

Theorem 2.33 ***(Trace theorem)***

$$\lambda_1 + \lambda_2 + \cdots + \lambda_p = \min_{X\in\mathbb{F}^{n\times p},\ X^*X=I_p} \operatorname{trace}(X^*AX) \tag{2.44}$$

Proof. Let $X \in \mathbb{F}^{n\times p}$ with $X^*X = I_p$ and let $\lambda'_1, \ldots, \lambda'_p$ be the eigenvalues of $A' = X^*AX$. Then the monotonicity principle applies, i.e., $\lambda_k \le \lambda'_k$ for $1 \le k \le p$. Thus,

$$\lambda_1 + \lambda_2 + \cdots + \lambda_p \le \lambda'_1 + \lambda'_2 + \cdots + \lambda'_p = \operatorname{trace}(X^*AX).$$

Equality holds if $X = [\mathbf{u}_1, \ldots, \mathbf{u}_p]$ is formed of the eigenvectors associated with the p smallest eigenvalues. ∎

Note that the choice of eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$ is not unique if one of the eigenvalues λ_1 to λ_p is multiple. Even if these eigenvalues are simple the minimizer X is not unique. In fact, any matrix $X = [\mathbf{u}_1, \ldots, \mathbf{u}_p]G$ with unitary $G \in \mathbb{F}^{p\times p}$ yields the same minimal trace.

2.11 Cholesky factorization

Definition 2.34 A Hermitian matrix is called **positive definite** (**positive semi-definite**) if all its eigenvalues are positive (nonnegative).

For a Hermitian positive definite matrix A, the LU decomposition can be written in a particular form reflecting the symmetry of A.

Theorem 2.35 ***(Cholesky factorization)*** *Let $A \in \mathbb{F}^{n\times n}$ be Hermitian positive definite. Then there is a* lower triangular *matrix L such that*

$$A = LL^*. \tag{2.45}$$

L is unique if we choose its diagonal elements to be positive.

Proof. We prove the theorem by giving an algorithm that computes the desired factorization.

Since A is positive definite, we have $a_{11} = \mathbf{e}_1^* A \mathbf{e}_1 > 0$. Therefore we can form the matrix

$$L_1 = \begin{bmatrix} l_{11}^{(1)} & & & \\ l_{21}^{(1)} & 1 & & \\ \vdots & & \ddots & \\ l_{n1}^{(1)} & & & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{a_{11}} & & & \\ \frac{a_{21}}{\sqrt{a_{1,1}}} & 1 & & \\ \vdots & & \ddots & \\ \frac{a_{n1}}{\sqrt{a_{1,1}}} & & & 1 \end{bmatrix}.$$

We now form the matrix

$$A_1 = L_1^{-1} A L_1^{-1*} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & a_{22} - \frac{a_{21}a_{12}}{a_{11}} & \ldots & a_{2n} - \frac{a_{21}a_{1n}}{a_{11}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} - \frac{a_{n1}a_{12}}{a_{11}} & \ldots & a_{nn} - \frac{a_{n1}a_{1n}}{a_{11}} \end{bmatrix}.$$

This is the first step of the algorithm. Since positive definiteness is preserved by a congruence transformation X^*AX (see also Theorem 2.38 below), A_1 is again positive definite. Hence, we can proceed in a similar fashion factorizing $A_1(2\!:\!n, 2\!:\!n)$, etc.

Collecting $L_1, L_2, \ldots,$ we obtain

$$I = L_n^{-1} \cdots L_2^{-1} L_1^{-1} A (L_1^*)^{-1} (L_2^*)^{-1} \cdots (L_n^*)^{-1}$$

or

$$(L_1 L_2 \cdots L_n)(L_n^* \cdots L_2^* L_1^*) = A.$$

which is the desired result. It is easy to see that $L_1L_2\cdots L_n$ is a lower triangular matrix and that

$$L_1L_2\cdots L_n = \begin{bmatrix} l_{11}^{(1)} & & & & \\ l_{21}^{(1)} & l_{22}^{(2)} & & & \\ l_{31}^{(1)} & l_{32}^{(2)} & l_{33}^{(3)} & & \\ \vdots & \vdots & \vdots & \ddots & \\ l_{n1}^{(1)} & l_{n2}^{(2)} & l_{n3}^{(3)} & \dots & l_{nn}^{(n)} \end{bmatrix}$$

■

Remark 2.7. To check if a (symmetric) matrix is positive definite, it is the best to try to compute the Cholesky factorization. The argorithm fails if one of the pivots becomes negative. □

Remark 2.8. When working with symmetric matrices, one often stores only half of the matrix, e.g. the lower triangle consisting of all elements including and below the diagonal. The L-factor of the Cholesky factorization can overwrite this information in-place to save memory. □

Definition 2.36 The **inertia** of a Hermitian matrix is the triple (ν, ζ, π) where ν, ζ, π is the number of negative, zero, and positive eigenvalues.

Definition 2.37 Two matrices A and B are called congruent if there is a nonsingular matrix X such that $B = X^*AX$. The mapping $A \mapsto X^*AX$ is called congruence transformation.

Theorem 2.38 ***(Sylvester's law of inertia)*** *The inertia is invariant under congruence transformations.*

Proof. The proof is given, for example, in [5]. ■

2.12 The singular value decomposition (SVD)

Theorem 2.39 ***(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*

$$U^*AV = \Sigma = \begin{pmatrix} \operatorname{diag}(\sigma_1, \dots, \sigma_p) & 0 \\ 0 & 0 \end{pmatrix}, \qquad p = \min(m, n), \tag{2.46}$$

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

Proof. If $A = O$, the theorem holds with $U = I_m, V = I_n$ and Σ equal to the $m \times n$ zero matrix.

We now assume that $A \neq O$. Let $\mathbf{x}$, $\|\mathbf{x}\| = 1$, be a vector that maximizes $\|A\mathbf{x}\|$ and let $A\mathbf{x} = \sigma\mathbf{y}$ where $\sigma = \|A\| = \|A\mathbf{x}\|$ and $\|\mathbf{y}\| = 1$. As $A \neq O$, $\sigma > 0$. Consider the scalar function

$$f(\alpha) := \frac{\|A(\mathbf{x}+\alpha\mathbf{y})\|^2}{\|\mathbf{x}+\alpha\mathbf{y}\|^2} = \frac{(\mathbf{x}+\alpha\mathbf{y})^*A^*A(\mathbf{x}+\alpha\mathbf{y})}{(\mathbf{x}+\alpha\mathbf{y})^*(\mathbf{x}+\alpha\mathbf{y})}$$

Because of the extremality of $A\mathbf{x}$, the derivative $f'(\alpha)$ of $f(\alpha)$ must vanish at $\alpha = 0$. This holds *for all* $\mathbf{y}$! The derivative $f'(\alpha)$ is given by

$$\frac{df}{d\alpha}(\alpha) = \frac{(\mathbf{x}^*A^*A\mathbf{y} + \bar{\alpha}\mathbf{y}^*A^*A\mathbf{y})\|\mathbf{x}+\alpha\mathbf{y}\|^2 - (\mathbf{x}^*\mathbf{y} + \bar{\alpha}\mathbf{y}^*\mathbf{y})\|A(\mathbf{x}+\alpha\mathbf{y})\|^2}{\|\mathbf{x}+\alpha\mathbf{y}\|^4}$$

Thus, we have for all $\mathbf{y}$,

$$\left.\frac{df}{d\alpha}(\alpha)\right|_{\alpha=0} = \frac{\mathbf{x}^*A^*A\mathbf{y}\|\mathbf{x}\|^2 - \mathbf{x}^*\mathbf{y}\|A\mathbf{x}\|^2}{\|\mathbf{x}\|^4} = 0.$$

As $\|\mathbf{x}\| = 1$ and $\|A\mathbf{x}\| = \sigma$, we get

$$(\mathbf{x}^*A^*A - \sigma^2\mathbf{x}^*)\mathbf{y} = (A^*A\mathbf{x} - \sigma^2\mathbf{x})^*\mathbf{y} = 0, \qquad \text{for all } \mathbf{y},$$

which implies

$$A^*A\mathbf{x} = \sigma^2\mathbf{x}.$$

Multiplying $A\mathbf{x} = \sigma\mathbf{y}$ from the left by A^* we get $A^*A\mathbf{x} = \sigma A^*\mathbf{y} = \sigma^2\mathbf{x}$ from which

$$A^*\mathbf{y} = \sigma\mathbf{x}$$

and $AA^*\mathbf{y} = \sigma A\mathbf{x} = \sigma^2\mathbf{y}$ follow. Therefore, $\mathbf{x}$ is an eigenvector of A^*A corresponding to the eigenvalue σ^2 and $\mathbf{y}$ is an eigenvector of AA^* corresponding to the same eigenvalue.

Now, we construct a unitary matrix U_1 with first column $\mathbf{y}$ and a unitary matrix V_1 with first column $\mathbf{x}$, $U_1 = [\mathbf{y}, \bar{U}]$ and $V_1 = [\mathbf{x}, \bar{V}]$. Then

$$U_1^*AV_1 = \begin{bmatrix} \mathbf{y}^*A\mathbf{x} & \mathbf{y}^*A\overline{V} \\ \overline{U}^*A\mathbf{x} & \overline{U}^*A\overline{V} \end{bmatrix} = \begin{bmatrix} \sigma & \sigma\mathbf{x}^*\overline{V} \\ \sigma\overline{U}^*\mathbf{y} & \overline{U}^*A\overline{V} \end{bmatrix} = \begin{bmatrix} \sigma & \mathbf{0}^* \\ \mathbf{0} & \hat{A} \end{bmatrix}$$

where $\hat{A} = \overline{U}^*A\overline{V}$. ■

The proof above is due to W. Gragg. It nicely shows the relation of the singular value decomposition with the spectral decomposition of the Hermitian matrices A^*A and AA^*,

$$A = U\Sigma V^* \quad \Longrightarrow \quad A^*A = V\Sigma^2V^*. \qquad AA^* = U\Sigma^2U^*, \tag{2.47}$$

Note that the proof given in [5] is shorter and maybe more elegant.

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 make use of the matrices A^*A and AA^*. Instead, let us consider the $(n+m)\times(n+m)$ Hermitian matrix

$$\begin{bmatrix} O & A \\ A^* & O \end{bmatrix}. \tag{2.48}$$

Making use of the SVD (2.46) we immediately get

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

Now, let us assume that $m \geq n$. Then we 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 $\Sigma_1 \in \mathbb{R}^{n\times n}$. Then

$$\begin{bmatrix} O & A \\ A^* & O \end{bmatrix} = \begin{bmatrix} U_1 & U_2 & O \\ O & O & V \end{bmatrix} \begin{bmatrix} O & O & \Sigma_1 \\ O & O & O \\ \Sigma_1 & O & O \end{bmatrix} \begin{bmatrix} U_1^* & O \\ U_2^* & O \\ O & V^* \end{bmatrix} = \begin{bmatrix} U_1 & O & U_2 \\ O & V & O \end{bmatrix} \begin{bmatrix} O & \Sigma_1 & O \\ \Sigma_1 & O & O \\ O & O & O \end{bmatrix} \begin{bmatrix} U_1^* & O \\ O & V^* \\ U_2^* & O \end{bmatrix}.$$

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

$$\begin{bmatrix} O & A \\ A^* & O \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}U_1 & \frac{1}{\sqrt{2}}U_1 & U_2 \\ \frac{1}{\sqrt{2}}V & -\frac{1}{\sqrt{2}}V & O \end{bmatrix} \begin{bmatrix} \Sigma_1 & O & O \\ O & -\Sigma_1 & O \\ O & O & O \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}}U_1^* & \frac{1}{\sqrt{2}}V^* \\ \frac{1}{\sqrt{2}}U_1^* & -\frac{1}{\sqrt{2}}V^* \\ U_2^* & O \end{bmatrix}. \tag{2.49}$$

Thus, there are three ways how to treat the computation of the singular value decomposition as an eigenvalue problem. One of the two forms in (2.47) is used *implicitly* in the QR algorithm for dense matrices A, see [5],[1]. The form (2.48) is suited if A is a sparse matrix.

Bibliography

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen, *LAPACK Users' Guide – Release 2.0*, SIAM, Philadelphia, PA, 1994. (Software and guide are available from Netlib at URL http://www.netlib.org/lapack/).

[2] R. Fletcher and D. C. Sorensen, *An algorithmic derivation of the Jordan canonical form*, Amer. Math. Monthly, 90 (1983), pp. 12–16.

[3] G. E. Forsythe and W. R. Wasow, *Finite-difference methods for partial differential equations*, Wiley, New York, 1960.

[4] F. R. Gantmacher, *The Theory of Matrices*, vol. 1, Chelsea, New York, 1959.

[5] G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 2nd ed., 1989.

[6] N. J. Higham, *Functions of matrices*, SIAM, Philadelphia, PA, 2008.

[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[8] R. J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations*, SIAM, Philadelphia, PA, 2007.

[9] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, SIAM, Philadelphia, PA, 2nd ed., 2011. (Classics in Applied Mathematics; 66).

[10] R. Zurmühl, *Matrizen und ihre technischen Anwendungen*, Springer, Berlin, 4th ed., 1964.