Linear Algebra II
Lectures Notes

MATH 425 Linear Algebra II, Spring 2015
LCD-undergrad 24908; LCD-grad 24909,
MWF 10:00-10:50, Taft Hall 308

Instructor: Shmuel Friedland
Office: 715 SEO, phone: 413-2176, *e:mail*: friedlan@uic.edu,
web: http://www.math.uic.edu/∼friedlan

Last update April 30, 2015

Remarks on the notes: These notes are updated during the course. Right now most of the notes are from Spring 2013. I will try to date the new material.

Contents

1 Basic facts on vector spaces, matrices and linear transformations 2
1.1 Fields and vector spaces . . . 2
1.1.1 Problems . . . 4
1.2 Dimension and basis . . . 5
1.2.1 More details on vector space . . . 5
1.2.2 Dimension . . . 7
1.2.3 Problems . . . 7
1.3 Matrices . . . 8
1.3.1 Basic properties of matrices . . . 8
1.3.2 Elementary row operation . . . 8
1.3.3 Invertible matrices . . . 9
1.3.4 Row and column spaces of A . . . 10
1.3.5 Special types of matrices . . . 11
1.3.6 Problems . . . 12
1.4 Determinants . . . 12
1.4.1 The group of bijections on a set $\mathcal{X}$ . . . 12
1.4.2 The permutation group . . . 13
1.5 Linear, multilinear maps and functions . . . 14
1.6 Definition and properties of determinants . . . 15
1.7 Polynomials . . . 19
1.7.1 Finite extension of fields . . . 20
1.8 Complex numbers . . . 21
1.9 Linear transformations . . . 22

2 Jordan canonical form **24**
2.1 Eigenvalues and eigenvectors 24
2.2 Matrix polynomials 28
2.3 Minimal polynomial and decomposition to invariant subspaces . . . 30
2.4 Quotient of vectors space and induced linear operators 35
2.5 Existence and uniqueness of the Jordan canonical form 35
2.6 Cyclic subspaces 39

3 Applications of Jordan Canonical form **45**
3.1 Functions of Matrices 45
3.2 Power stability, convergence and boundedness of matrices 50
3.3 e^{At} and stability of certain systems of ODE 53

4 Inner product spaces **56**
4.1 Inner product 56
4.2 Explanation of G-S process in standard Euclidean space 58
4.3 An example of G-S process 58
4.4 QR Factorization 58
4.5 An example of QR algorithm 59
4.6 Geometric interpretation of the determinant 62
4.7 Special transformations in IPS 64
4.8 Symmetric bilinear and hermitian forms 70
4.9 Max-min characterizations of eigenvalues 72
4.10 Positive definite operators and matrices 77
4.11 Inequalities for traces 81
4.12 Singular Value Decomposition 83
4.13 Characterizations of singular values 87
4.14 Moore-Penrose generalized inverse 93

5 Perron-Frobenius theorem **95**

1 Basic facts on vector spaces, matrices and linear transformations

1.1 Fields and vector spaces

A commutative group, also called an abelian group, denoted by $\mathbf{G}$, is a set of elements with a binary operation $\oplus$, i.e. $a \oplus b \in \mathbf{G}$ for each $a, b \in \mathbf{G}$. This operation is

1. associative: $(a \oplus b) \oplus c = a \oplus (b \oplus c)$;
2. commutative: $a \oplus b = b \oplus a$;
3. there exist a *neutral* element $\bigcirc$ such that $a + \bigcirc = a$ for each $a \in \mathbf{G}$;
4. for each $a \in \mathbf{G}$ there exists a unique $\ominus a$ such that $a + (\ominus a) = \bigcirc$.

Examples of commutative group:

1. The following subset of complex numbers where $\oplus$ is the standard addition $+$, $\ominus$ is $-$ and the neutral element is 0.
   (a) The set of integers $\mathbb{Z}$.
   (b) The set of rational numbers $\mathbb{Q}$.
   (c) The set of real numbers $\mathbb{R}$.
   (d) The set of complex numbers $\mathbb{C}$
2. The following subsets of $\mathbb{C}^* := \mathbb{C}\backslash\{0\}$, i.e. all nonzero complex numbers, where the operation $\oplus$ is the *product*, the neutral element is 1, and $\ominus a$ is a^{-1}.
   (a) $\mathbb{Q}^* := \mathbb{Q}\backslash\{0\}$.
   (b) $\mathbb{R}^* := \mathbb{R}\backslash\{0\}$.
   (c) $\mathbb{C}^* := \mathbb{C}\backslash\{0\}$.

From now on, we will denote the operation $\oplus$ by $+$. An abelian group $\mathbf{R}$ is called a *ring*, if there exist a second operation, called *product*, denoted by ab, which satisfies:

1. associativity: $(ab)c = a(bc)$;
2. distributibity: $a(b+c) = ab + ac,\ (b+c)a = ba + ca$;
3. existence of identity $1 \in \mathbf{R}$: $1a = a1 = 1$ for all $a \in \mathbf{R}$.

$\mathbf{R}$ is called a commutative ring if $ab = ba$ for all $a, b \in \mathbf{R}$.

Examples of rings:

1. The following subsets of $n \times n, n > 1$ complex valued matrices, denoted by $\mathbb{C}^{n\times n}$ with the addition $A + B$, multiplication AB, with the identity I_n, $n \times n$ identity matrix.
   (a) $\mathbb{Z}^{n\times n}$, the ring of $n \times n$ matrices with integer entries. (Noncommutative ring.)
   (b) $\mathbb{Q}^{n\times n}$, the ring of $n \times n$ matrices with rational entries. (Noncommutative ring.)
   (c) $\mathbb{R}^{n\times n}$, the ring of $n \times n$ matrices with real entries. (Noncommutative ring.)
   (d) $\mathbb{C}^{n\times n}$. (Noncommutative ring.)
   (e) $\mathbf{D}(n, S)$, the set of $n \times n$ diagonal matrices with entries in $S = \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$. (Commutative ring.)
2. $S[z]$, the commutative ring of polynomials in one variable z with coefficients in $S = \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$.
3. $\mathbb{Z}_m = \mathbb{Z}/(m\mathbb{Z})$, all integers *modulo* a positive integer m, with the addition and multiplication modulo m. $\#\mathbb{Z}_m$, the number of elements in $\mathbb{Z}_m$, is m. $\mathbb{Z}_m$ can be identified with $\{0, 1, \ldots, m-1\}$.

A commutative ring $\mathbf{R}$ is called a *field* if each nonzero element $a \in \mathbf{R}$ has a unique inverse, denoted by a^{-1} such that $aa^{-1} = 1$. A field is denoted by $\mathbb{F}$.

Examples of fields:

1. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$.
2. $\mathbb{Z}_m$, where m is a prime integer.

An abelian group $\mathbf{V}$ is called a vector space over a field $\mathbb{F}$, if for any $a \in \mathbb{F}, \mathbf{v} \in \mathbf{V}$ the product $a\mathbf{v}$ is an element in $\mathbf{V}$. This operation satisfies the following properties:

$$a(\mathbf{u}+\mathbf{v}) = a\mathbf{u} + a\mathbf{v}, (a+b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}, (ab)(\mathbf{u}+\mathbf{v}) = a(b(\mathbf{u}+\mathbf{v})), 1\mathbf{v} = \mathbf{v},$$

for all $a, b \in \mathbb{F}, \mathbf{u}, \mathbf{v} \in \mathbf{V}$.

A ring $\mathbf{R}$ is called an algebra over $\mathbb{F}$ if $\mathbf{R}$ is a vector space over $\mathbb{F}$ and $a(\mathbf{xy}) = (a\mathbf{x})\mathbf{y} = \mathbf{x}(a\mathbf{y})$ for any $a \in \mathbb{F}, \mathbf{x}, \mathbf{y} \in \mathbf{R}$. A standard example of an algebra over $\mathbb{F}$ is the set of $n \times n$ matrices with entries in $\mathbb{F}$, denoted by $\mathbb{F}^{n\times n}$. Another example is $\mathbb{F}[z]$, the ring of polynomials in one variable $\mathbf{z}$ with coefficients in $\mathbb{F}$.

A set $\mathbf{H}$ is called a group, if there exists a binary operation $ab \in \mathbf{H}$ for each $a, b \in \mathbf{H}$ such that

1. associative: $(ab)c = a(bc)$;
2. identity: $1a = a1 = a$;
3. inverse: $a^{-1}a = aa^{-1} = 1$ for each $a \in \mathbf{H}$.

$\mathbf{H}$ is abelian if $ab = ba$ for each $a, b \in \mathbf{H}$. This is the case discussed in the beginning of this section for $\mathbf{G}$, where the product operation is replaced by $\oplus$ and the identity by the neutral element.

A standard example of noncommutative group is $\mathrm{GL}(n, \mathbb{C}) \subset \mathbb{C}^{n\times n}, n > 1$, the group of $n \times n$ complex valued invertible matrices.

1.1.1 Problems

1. $\zeta \in \mathbb{C}$ is called a primitive unity root of order $m > 1$ if $\zeta^m = 1$ and $\zeta^k \neq 1$ for $k \in [m-1]$. Let $\mathbb{Q}[\zeta] := \{a_0 + a_1\zeta + \ldots + a_{m-1}\zeta^{m-1},\ a_0, \ldots, a_{m-1} \in \mathbb{Q}\}$. Show that $\mathbb{Q}[\zeta]$ is a field. It is a finite extension of $\mathbb{Q}$. More precisely, $\mathbb{Q}[\zeta]$ can be viewed as a vector space over $\mathbb{Q}$ of dimension m.

2. *Quaternions* $\mathbf{Q}$ is a four dimensional noncommutative division algebra that generalizes complex numbers. Any quaternion is written a vector $\mathbf{q} = a_0 + a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$. The product of two quaternions is determined by using the following equalities.

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1,\ \mathbf{ij} = -\mathbf{ji} = \mathbf{k},\ \mathbf{jk} = -\mathbf{kj} = \mathbf{i},\ \mathbf{ki} = -\mathbf{ik} = \mathbf{j}.$$

   Let $\bar{\mathbf{q}} = a_0 - a_1\mathbf{i} - a_2\mathbf{j} - a_3\mathbf{k}$. Show

   (a) The map $T(\mathbf{q}) = \bar{\mathbf{q}}$ is an *involution*. (T is an isomorphism preserving addition and multiplication and $T^2 = Id$.)

   (b) $\mathbf{Q}$ is a noncommutiative division algebra. I.e. if $\mathbf{q} \neq \mathbf{0}$ that there exists a unique $\mathbf{r}$ such that $\mathbf{rq} = \mathbf{qr} = 1$. (You can try to use the identity $\mathbf{q}\bar{\mathbf{q}} = \bar{\mathbf{q}}\mathbf{q} = |\mathbf{q}|^2 = \sum_{i=0}^{3} a_i^2$.)

1.2 Dimension and basis

Let $\mathbf{V}$ be a vector space over a field $\mathbb{F}$. (For simplicity of the exposition you may assume that $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$.) $\mathbf{U} \subset \mathbf{V}$ is called a subspace of $\mathbf{V}$, if $a_1\mathbf{u}_1 + a_2\mathbf{u}_2 \in \mathbf{U}$ for each $\mathbf{u}_1, \mathbf{u}_2 \in \mathbf{U}, a_1, a_2 \in \mathbb{F}$. The subspace $\mathbf{U} := \{\mathbf{0}\}$ is called the zero, or trivial, subspace of $\mathbf{V}$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbf{V}$ be n vectors. Then for any $a_1, \ldots, a_n \in \mathbb{F}$, called scalars, $a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n$ is called a linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_n$. $\mathbf{0} = \sum_{i=1}^n 0\mathbf{x}_i$ is the trivial combination. $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are *linearly independent*, abbreviated to l.i. or lin. ind., if the equality $\mathbf{0} = \sum_{i=1}^n a_i\mathbf{x}_i$ implies that $a_1 = \ldots = a_n = 0$. Otherwise $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are *linearly dependent*, abbreviated to l.d. or lin. dep.. The set of all linear combination of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is called the span of $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and denoted by $\operatorname{span}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. $\operatorname{span}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is a subspace of $\mathbf{V}$.

1.2.1 More details on vector space

$\mathbf{V}$ is called *finitely generated* if $\mathbf{V} = \operatorname{span}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ for some vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$. $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a *spanning set* of $\mathbf{V}$. (In these course we consider only finite generated vector spaces. We will show that every finitely generated vector space is finite dimensional). We will use the following notation for a positive integer n:

$$[n] := \{1, 2, \ldots, n-1, n\} \tag{1.1}$$

Observe the following "obvious" relation for any $n \in \mathbb{N}$. (Here $\mathbb{N}$ is the set of all positive integers $1, 2, \ldots,$.)

$$\operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_n) \subseteq \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n), \text{ for each } i \in [n]. \tag{1.2}$$

It is enough to consider the case $i = n$. (We assume here that for $n = 1$ the span of empty set is the zero subspace $\mathbf{U} = \{\mathbf{0}\}$.) So for $n = 1$ spaninc holds. Suppose that $n > 1$. Clearly, any linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$, which is $\mathbf{u} = \sum_{i=1}^{n-1} b_i\mathbf{u}_i$ is a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_n$: $\mathbf{u} = 0\mathbf{u}_n + \sum_{i=1}^{n-1} b_i\mathbf{u}_i$. Hence (1.2) holds.

Theorem 1.1 *Suppose that $n > 1$. Then there exists $i \in [n]$ such that*

$$\operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n) = \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_n) \tag{1.3}$$

if and only if $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly dependent vectors.

Proof. Suppose that (1.3) holds. So $\mathbf{u}_i = b_1\mathbf{u}_1 + \ldots b_{i-1}\mathbf{u}_{i-1} + b_{i+1}\mathbf{u}_{i+1} + \ldots b_n\mathbf{u}_n$. Hence $\sum_{i=1}^n a_i\mathbf{u}_i = \mathbf{0}$, where $a_j = b_j$ for $j \neq i$ and $a_i = -1$. Hence $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly dependent. Assume that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly dependent. Then there exist scalars $a_1, \ldots, a_n \in \mathbb{F}$ not all of them are equal to zero so that $a_1\mathbf{u}_1 + \ldots + a_n\mathbf{u}_n = 0$. Assume that $a_i \neq 0$. For simplicity of notation, (or by renaming indices in $[n]$), we may assume that $i = n$. So $\mathbf{u}_n = -\frac{1}{a_n}\sum_{i=1}^{n-1} a_i\mathbf{u}_i$. Let $\mathbf{u} \in \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$. So

$$\mathbf{u} = \sum_{i=1}^{n} b_i\mathbf{u}_i = b_n\mathbf{u}_n + \sum_{i=1}^{n-1} b_i\mathbf{u}_i = -b_n\sum_{i=1}^{n-1}\frac{a_i}{a_n}\mathbf{u}_i + \sum_{i=1}^{n-1} b_i\mathbf{u}_i = \sum_{i=1}^{n-1}\frac{a_nb_i - a_ib_n}{a_n}\mathbf{u}_i.$$

That is $\mathbf{u} \in \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_{n-1})$. This proves the theorem. □

Corollary 1.2 *Let $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbf{V}$. Assume that not all $\mathbf{u}_i$ are zero vectors. Then there exists $d \geq 1$ integers $1 \leq i_1 < i_2 < \ldots < i_d \leq n$ such that $\mathbf{u}_{i_1}, \ldots, \mathbf{u}_{i_d}$ are linearly independent and $\operatorname{span}(\mathbf{u}_{i_1}, \ldots, \mathbf{u}_{i_d}) = \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$.*

Proof. Suppose that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly independent. Then $d = n$ and $i_k = k$ for $k = 1, \ldots, n$ and we are done.

Assume that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly dependent. Apply Theorem 1.1. So consider now $n-1$ vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_n$ as given by Theorem 1.1. Note that it is not possible that all vectors in $\{\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_n\}$ are zero since this will imply that $\mathbf{u}_i = \mathbf{0}$. This will contradict the assumption that not all $\mathbf{u}_i$ are zero. Apply the previous arguments to $\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_n$ and continue in this fashion until one gets d linearly independent vectors $\mathbf{u}_{i_1}, \ldots, \mathbf{u}_{i_d}$. □

Corollary 1.3 *Let $\mathbf{V}$ be a finitely generated nontrivial vectors space, i.e. contains more than one element. Then there exist $n \in \mathbb{N}$ and n linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ such that $\mathbf{V} = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.*

In the following lemma we use the fact that any m homogeneous linear equations with n variables has a nontrivial solution if $m < n$. (This will be proved later using REF of matrices.)

Lemma 1.4 *Let $n > m \geq 1$ be integers. Then any $\mathbf{w}_1, \ldots, \mathbf{w}_n \in \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_m)$ are linearly dependent*

Proof. Observe that $\mathbf{w}_j = \sum_{i=1}^m a_{ij}\mathbf{u}_j$. So

$$\sum_{j=1}^n x_j \mathbf{w}_j = \sum_{j=1}^n x_j \sum_{i=1}^m a_{ij}\mathbf{u}_i = \sum_{i=1}^m (\sum_{j=1}^n a_{ij}x_j)\mathbf{u}_i.$$

Consider the following m homogeneous equations in n unknowns: $\sum_{j=1}^n a_{ij}x_j = 0$ for $i = 1, \ldots, m$. Since $n > m$ we have a nontrivial solution $(x_1, \ldots, x_n)^\top$. So $\sum_{j=1}^n x_j \mathbf{w}_j = \mathbf{0}$. □

Theorem 1.5 *Let $\mathbf{V} = \operatorname{span}(\mathbf{v}_1, \ldots \mathbf{v}_n)$ and assume that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent. Then the following holds.*

1. *Any vector $\mathbf{u}$ can be expressed as a unique linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.*
2. *For an integer $N > n$ any N vectors in $\mathbf{V}$ are linearly dependent.*
3. *Assume that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly independent. Then $\mathbf{V} = \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$.*

Proof. Assume that $\sum_{i=1}^n x_i \mathbf{v}_i = \sum_{i=1}^n y_i \mathbf{v}_i$. So $\sum_{i=1}^n (x_i - y_i)\mathbf{v}_i = \mathbf{0}$. As $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent it follows that $x_i - y_i = 0$ for $i = 1, \ldots, n$. Hence 1 holds.

Lemma 1.4 implies 2.

Suppose that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linearly independent. Let $\mathbf{v} \in \mathbf{V}$ and consider $n+1$ vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{v}$. 2 implies that $\mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{v}$ are linearly dependent. Thus there exists $n+1$ scalars $a_1, \ldots, a_{n+1}$, not all of them zero, such that $a_{n+1}\mathbf{v} + \sum_{i=1}^n a_i \mathbf{v}_i =$

0. Assume first that $a_{n+1} = 0$. Then $\sum_{i=1}^{n} a_i \mathbf{u}_i = \mathbf{0}$. Since not all a_i zero it follows that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ linearly dependent, which is a contradiction to our assumption. Hence $a_{n+1} \neq 0$ and $\mathbf{v} = \sum_{i=1}^{n} \frac{-a_i}{a_{n+1}} \mathbf{u}_i$. □

Definition 1.6 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *is called a* spanning set *of a linear space* $\mathbf{V}$ *if* $\mathbf{V} = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

1.2.2 Dimension

The dimension of a trivial vector space, consisting of the zero vector $\mathbf{V} = \{\mathbf{0}\}$ is zero. The dimension of a finite dimensional nonzero vector space $\mathbf{V}$ is the number of vectors in any spanning linearly independent set. The dimension of $\mathbf{V}$ is denoted by $\dim \mathbf{V}$. Assume that $\dim \mathbf{V} = n$. Suppose that $\mathbf{V} = \operatorname{span}(\mathbf{x}_1, \ldots, x_n)$. Then $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a *basis* of $\mathbf{V}$. I.e., each vector $\mathbf{x}$ can be uniquely expressed as $\sum_{i=1}^{n} a_i \mathbf{x}_i$. Thus to each $\mathbf{x} \in \mathbf{V}$ one corresponds a unique column vector $\mathbf{a} = (a_1, \ldots, a_n)^\top \in \mathbb{F}^n$, where $\mathbb{F}^n$ is the vector space on column vectors with n coordinates in $\mathbb{F}$. The correspondence $\mathbf{x} \mapsto \mathbf{a}$ is an isomorphism of $\iota : \mathbf{V} \to \mathbb{F}^n$. It will be convenient to denote $\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{x}_i$ be the equality $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]\mathbf{a}$ $(= \sum_{i=1} \mathbf{x}_i a_i)$. (Note that we use the standard way to multiply row by column.)

Assume now that $\mathbf{y}_1, \ldots, \mathbf{y}_n$ be n vectors in $\mathbf{V}$. Then

$$\mathbf{y}_i = \sum_{j=1}^{n} y_{ji} \mathbf{x}_j, \quad i = 1, \ldots, n.$$

Denote by $Y = [y_{ji}]_{i=j=1}^{n}$ the $n \times n$ matrix with the element y_{ji} in j-th row and i-th column. The above equalities are equivalent to the identity

$$[\mathbf{y}_1, \ldots, \mathbf{y}_n] = [\mathbf{x}_1, \ldots, \mathbf{x}_n]Y. \tag{1.4}$$

Then $\mathbf{y}_1, \ldots, \mathbf{y}_n$ is a basis if and only if Y is invertible matrix. (See for details next subsection.) Y is called the matrix of the change of basis from $[\mathbf{y}_1, \ldots, \mathbf{y}_n]$ to $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$. The matrix of the change of basis from $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ to $[\mathbf{y}_1, \ldots, \mathbf{y}_n]$ is given by Y^{-1}. (Just multiply the identity (1.4) by Y^{-1} from the right.) Suppose also that $[\mathbf{z}_1, \ldots, \mathbf{z}_n]$ is a basis in $\mathbf{V}$. Let $[\mathbf{z}_1, \ldots, \mathbf{z}_n] = [\mathbf{y}_1, \ldots, \mathbf{y}_n]Z$. Then $[\mathbf{z}_1, \ldots, \mathbf{z}_n] = [\mathbf{x}_1, \ldots, \mathbf{x}_n](YZ)$. (Use (1.4).)

1.2.3 Problems

Let $\mathcal{P}_m$ be subset of all polynomials of degree m at most over the field $\mathbb{F}$.

1. Show that the polynomials $1, x, \ldots, x^m$ form a basis in $\mathcal{P}_m$.
2. Let f_j be a nonzero polynomial in $\mathcal{P}_m$ of degree j for $j = 0, \ldots, m$. Show that $f_0, \ldots, f_m$ form a basis in $\mathcal{P}_m$.
3. Find a basis in $\mathcal{P}_4$ such that each polynomial is of degree 4.
4. Does there exists a basis in $\mathcal{P}_4$ such that each polynomial is of degree 3?
5. Write down the set of all vectors in $\mathbb{R}^4$ whose coordinates are two 0's and two 1's. (There are six such vectors). Show that these six vectors span $\mathbb{R}^4$. Find all the collections of four vectors out of these six which form a basis in $\mathbb{R}^4$.

1.3 Matrices

1.3.1 Basic properties of matrices

The set of $m \times n$ matrices with entries in $\mathbb{F}$ is denoted by $\mathbb{F}^{m\times n}$. $\mathbb{F}^{m\times n}$ is a vector space over $\mathbb{F}$ of dimension mn. A standard basis of $\mathbb{F}^{m\times n}$ is given by $E_{ij}, i = 1,\dots,m, j = 1,\dots,n$ where E_{ij} is a matrix with 1 in the entry (i,j) and 0 in all other entries. $\mathbb{F}^{m\times 1} = \mathbb{F}^m$ the set of column vectors with m coordinates. A standard basis of $\mathbb{F}^m$ is $\mathbf{e}_i = (\delta_{1i},\dots,\delta_{mi})^\top, i = 1,\dots,m$.

For $A \in \mathbb{F}^{m\times n}, B \in \mathbb{F}^{p\times q}$ the product AB is defined if and only if the number of columns in A is equal to the number of columns in B, i.e. $n = p$. In that case the resulting matrix $C = [c_{ij}]$ is $m \times q$. The entry c_{ij} is obtained by multiplying row i of A by the column j of B. Recall that for $\mathbf{x} = (x_1,\dots,x_n)^\top, \mathbf{y} = (y_1,\dots,y_n)^\top \in \mathbb{F}^n$ $\mathbf{x}^\top\mathbf{y} = \sum_{i=1}^n x_i y_i$. (This product can be regarded as a product of $1 \times n$ matrix $\mathbf{x}^\top$ with $n \times 1$ product matrix $\mathbf{y}$.) The product of corresponding sizes of matrices satisfies the properties:

1. associativity $(AB)C = A(BC)$;
2. distributivity $(A_1 + A_2)B = A_1B + A_2B, A(B_1 + B_2)$;
3. $a(AB) = (aA)B = A(aB)$ for each $a \in \mathbb{F}$;
4. identities $I_mA = AI_n$, where $A \in \mathbb{F}^{m\times n}$ and $I_m = [\delta_{ij}]_{i=j=1}^m \in \mathbb{F}^{m\times m}$ is the identity matrix.

For $A = [a_{ij}] \in \mathbb{F}^{m\times n}$ denote by $A^\top \in \mathbb{F}^{n\times m}$ the transposed matrix of A. So the (i,j) entry of $A^\top$ is the (j,i) entry of A. The following properties hold

$$(aA)^\top = aA,\ (A+B)^\top = A^\top + B^\top,\ (AC)^\top = C^\top A^\top.$$

$A = [a_{ij}] \in \mathbb{F}^{m\times n}$ is called diagonal if $a_{ij} = 0$ for $i \neq j$. Denote by $\mathbf{D}(m,n) \subset \mathbb{F}^{m\times n}$ the vector subspace of diagonal matrices. A square diagonal matrix with the diagonal entries $d_1,\dots,d_n$ is denoted by $\operatorname{diag}(d_1,\dots,d_n) \in \mathbb{F}^{n\times n}$. A is called a block matrix if $A = [A_{ij}]_{i=j=1}^{p,q}$ where each entry A_{ij} is an $m_i \times m_j$ matrix. So $A \in \mathbb{F}^{m\times n}$ where $m = \sum_{i=1}^p m_i, n = \sum_{j=1}^q n_j$. A block matrix A is called block diagonal if $A_{ij} = 0_{m_i\times m_j}$ for $i \neq j$. A block diagonal matrix with $p = q$ is denoted by $\operatorname{diag}(A_{11},\dots,A_{pp})$. A different notation is $\oplus_{l=1}^p A_{ll} := \operatorname{diag}(A_{11},\dots,A_{pp})$.

1.3.2 Elementary row operation

Elementary operations on the rows of $A \in \mathbb{F}^{m\times n}$ are defined as follows:

1. Multiply row i by a nonzero $\mathbf{a} \in \mathbb{F}$.
2. Interchange two distinct rows i and j in A.
3. Add to row j row i multiplied by a, where $i \neq j$.

$A \in \mathbb{F}^{m\times n}$ can be brought to a unique canonical form, called the *reduced row echelon form*, abbreviated as RREF by the elementary row operation. The RREF of zero matrix $0_{m\times n}$ is $0_{m\times n}$. For $A \neq 0_{m\times n}$, RREF of A given by B is of the following form. the first k rows of B are nonzero rows. The last $m - k$ rows are zero rows.

The number of nonzero rows is called the rank of A, and is denoted by rank A. The first nonzero element in each i-th row, $i \leq \text{rank } A$, located at the column j_i is 1. It is called a *pivot*. All other elements of B, in the same columns as the pivot, are zero elements. Futhermore the following condition hold $1 \leq j_1 < j_2 < \ldots < j_k \leq n$.

1.3.3 Invertible matrices

$A \in \mathbb{F}^{m\times m}$ is called *invertible* if there exists $B \in \mathbb{F}^{m\times m}$ such that $AB = BA = I_m$. Note that B is unique. If $AC = CA = I_m$ then $B = BI_m = B(AC) = (BA)C = I_mC = C$. So we denote B, the inverse of A by A^{-1}. Denote by $\mathrm{GL}(m, \mathbb{F}) \subset \mathbb{F}^{m\times m}$ the set of all invertible matrices. Note that $\mathrm{GL}(m, \mathbb{F})$ is a group under the multiplication, with the unit I_n. (Observe that for $A, B \in \mathrm{GL}(m, \mathbb{F})$ $(AB)^{-1} = B^{-1}A^{-1}$.)

A matrix $E \in \mathbb{F}^{m\times m}$ is called *elementary* if it is obtained from I_m by applying an appropriate elementary row operation. Note that applying an elementary row operation on A is equivalent to EA, where E is the corresponding elementary row operation. By reversing the corresponding elementary operation we see that E is invertible, and E^{-1} is also an elementary matrix.

Theorem 1.7 *Let $A \in \mathbb{F}^{m\times m}$. The following are equivalent.*

1. *A is invertible.*
2. *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
3. *The RREF of A is I_m.*
4. *A is a product of elementary matrices. A matrix $A \in \mathrm{GL}(m, \mathbb{F})$ if and only if I_m is its reduced row*

Proof. $1 \Rightarrow 2$. $A\mathbf{x} = \mathbf{0}$ implies that $A^{-1}(A\mathbf{x}) = I_m\mathbf{x} = \mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}$.
$2 \Rightarrow 3$. The system $A\mathbf{x} = \mathbf{0}$ does free variables, hence the number of pivots is the number of row, i.e. RREF of A is n.
$3 \Rightarrow 4$. There exists a sequence of elementary matrices so that $E_pE_{p-1}\cdots E_1A = I_m$. Hence $A = E_1^{-1}E_2^{-1}\cdots E_p^{-1}$, and each E_j^{-1} is also elementary.
$4 \Rightarrow 1$. As each elementary matrix is invertible so is their product, which is equal to A. □

Theorem 1.8 *Let $A \in \mathbb{F}^{n\times n}$. Define the matrix $B = [A\ I_n] \in F^{n\times(2n)}$. Let $C = [C_1\ C_2], C_1, C_2 \in F^{n\times n}$ be the RREF of B. Then A is invertible if and only if $C_1 = I_n$. Furthermore, if $C_1 = I_n$ then $A^{-1} = C_2$.*

Proof. The fact that A is invertible if and only if $C_1 = I_n$ follows straightforward from Theorem 1.7. Note that for any matrix $F = [F_1F_2], F_1, F_2 \in \mathbb{F}^{n\times p}, G \in \mathbb{F}^{l\times m}$ we have $GF = [(GF_1)(GF_2)]$. Let H be a product of elementary matrices such that $HB = [IC_2]$. So $HA = I_n, HI_n = C_1$. Hence $H = A^{-1}$ and $C_1 = A^{-1}$. □

1.3.4 Row and column spaces of A

$A, B \in \mathbb{F}^{m\times n}$ are called *left equivalent* if A and B have the same reduced row echelon form, or B can be obtained from A by applying elementary row operations. Equivalently, $B = UA$ for some $U \in \mathrm{GL}(m, F)$. This is denoted as $A \sim_l B$. It is straightforward to show that $\sim_l$ is an equivalence relation.

Let $A \in \mathbb{F}^{m\times n}$. Denote by $\mathbf{c}_1, \ldots, \mathbf{c}_n$ the n column of A. We write $A = [\mathbf{c}_1 \ \mathbf{c}_2 \ \ldots \ \mathbf{c}_n]$. Then $\mathrm{span}(\mathbf{c}_1, \ldots, \mathbf{c}_n)$ is called the column space of A. Its dimension is rank A. Similarly, the dimension of the columns space of $A^\top$, which is equal to the dimension of the row space of A, is rank A. I.e. rank $A^\top$ = rank A. Let $\mathbf{x} = (x_1, \ldots, x_n)^\top$. Then $A\mathbf{x} = \sum_{i=1}^n x_i\mathbf{c}_i$. Hence the system $A\mathbf{x} = \mathbf{b}$ is solvable if and only if $\mathbf{b}$ is in the column space of A. The set of all $\mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \mathbf{0}$ is called the *null space* of A. It is a subspace of dimension of $n - \mathrm{rank}\, A$, and is denoted as $\mathrm{N}(A) \subset \mathbb{F}^m$. $\dim \mathrm{N}(A)$ is called the *nullity* of A, and denoted by nul A. If $A \in \mathrm{GL}(n, \mathbb{F})$ then $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Theorem 1.9 *Let $A \in \mathbb{F}^{m\times n}$. Assume that $B \in \mathbb{F}^{m\times n}$ is a row echelon form of A. Let $k = \mathrm{rank}\, A$. Then*

1. *The nonzero rows of B form a basis in the row space of A.*
2. *Assume that the pivots of B are the columns $1 \le j_1 < \ldots < j_k \le n$, i.e. $x_{j_1}, \ldots, x_{j_k}$ are the lead variables in the system $A\mathbf{x} = \mathbf{0}$. Then the columns $j_1, \ldots, j_k$ of A form a basis of the column space of A.*
3. *If $n = rankA$ then the null space of A consists only of $\mathbf{0}$. Otherwise the null space of A has the following basis. For each free variable x_p let $\mathbf{x}_p$ be the unique solution of $A\mathbf{x} = \mathbf{0}$, where $x_p = 1$ and all other free variables are zero. Then this $n - \mathrm{rank}\, A$ vectors form a basis in* nul A.

Proof. 1. Note that if $E \in \mathbb{F}^{m\times m}$ then the row space of EA is contained in A, since any row in EA is a linear combination of the rows of A. Since $A = E^{-1}(EA)$ it follows that the row space of A is contained in the row space of EA. Hence the row space of A and EA are the same. Therefore the row space of A and B are the same. Since the last $m - k$ rows of B are zero rows, it follows that the row space of B spanned by the first k rows $\mathbf{b}_1^\top, \ldots, \mathbf{b}_k^\top$. We claim that $\mathbf{b}_1^\top, \ldots, \mathbf{b}_k^\top$ are linearly independent. Indeed, consider $x\mathbf{b}_1^\top + \ldots + \mathbf{x}_k\mathbf{b}_k = \mathbf{0}^\top$. Since all the entries below the first pivot in B are zero we deduce that $x_1 = 0$. Continue in the same manner to obtain that $x_2 = 0, \ldots, x_k = 0$. So $\mathbf{b}_1^\top, \ldots, \mathbf{b}_k^\top$ form a basis in the row space of B and A.

2. We first show that $\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_k}$ linearly independent. Consider the equality $\sum_{j=1}^k x_{i_j}\mathbf{c}_{i_j} = \mathbf{0}$. This is equivalent to the system $A\mathbf{x} = \mathbf{0}$, where $\mathbf{x} = (x_1, \ldots, x_n)^\top$ and $x_p = 0$ if x_p is a free variable. Since all free variables are zero, all lead variables are zero, i.e. $x_{j_i} = 0$ for $i = 1, \ldots, k$. So $\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_k}$ linearly independent. It is left to show that $\mathbf{c}_p$, where x_p is a free variable is a linear combination of $\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_k}$. Again, consider $A\mathbf{x} = \mathbf{0}$, where each $x_p = 1$ and all other free variables are zero. We have a unique solution to $A\mathbf{x} = \mathbf{0}$, which states that $\mathbf{0} = \mathbf{c}_p + \sum_{i=1}^k x_{j_i}\mathbf{c}_{j_i}$, i.e. $\mathbf{c}_p = \sum_{i=1}^k -x_{j_i}\mathbf{c}_{j_i}$.

3. This follows for the way we write down the general solution of the system $A\mathbf{x} = \mathbf{0}$ in terms of free variables. □

Remark: Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{F}^m$. To find a basis in $\mathbf{U} := \text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ consisting of k vectors in $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ apply Theorem 1.9 to the column space of the matrix $A = [\mathbf{x}_1\ \mathbf{x}_2 \ldots \mathbf{x}_n]$. To find a nice basis in $\mathbf{U}$ apply Theorem 1.9 to the row space of $A^\top$.

$A, B \in \mathbb{F}^{m\times n}$ are called *right equivalent* if $B = AV$ for some $V \in \text{GL}(n, \mathbb{F})$. It is denoted as $A \sim_r B$. It is straightforward to show that $\sim_r$ is an equivalence relation. Note that $A \sim_r B$ if and only if $A^\top \sim_r B^\top$.

$A, B \in \mathbb{F}^{m\times n}$ are called *equivalent* if $B = UAV$ for some $U \in \text{GL}(m, \mathbb{F}), V \in \text{GL}(n, \mathbb{F})$. Equivalently, B is equivalent to A, if B can be obtained from A by applying elementary row and column operation. This is denoted as $A \sim B$. It is straightforward to show that $\sim$ is an equivalence relation.

Lemma 1.10 *Let $A \in \mathbb{F}^{m\times n}$. Then A is equivalent to a block diagonal matrix $I_{\text{rank } A} \oplus 0$.*

Proof. First bring A to REF matrix B with all pivots equal to 1. Now perform the elementary column operations corresponding to the elementary row operations on $B^\top$. This will give a matrix C with r pivots equal to 1 and all other elements zero. Now interchange the corresponding columns to obtain $I_r \oplus 0$. It is left to how that $r = rankA$. First recall that if $B = UA$ then the row space of A and B are the same. So rank A = rank B, which is the dimension of the row space. Next, if $C = BV$ then C and B has the same column space. So rank B = rank C, which is the dimension of the column space. So r = rank A. □

1.3.5 Special types of matrices

The following are some special subsets of $\mathbb{F}^{n\times n}$.

1. A is symmetric if $A^\top = A$. Denote by $\text{S}(n, \mathbb{F})$ the subspace of $n \times n$ symmetric polynomials.

2. A is skew-symmetric, or anti-symmetric, if $A^\top = -A$. Denote by $\mathcal{A}(n, \mathbb{F})$ the subspace of the skew-symmetric matrices.

3. $A = [a_{ij}]$ is called upper triangular if $a_{ij} = 0$ for each $j < i$. Denote by $\text{U}(n, \mathbb{F})$ the subspace of upper triangular matrices.

4. $A = [a_{ij}]$ is called lower triangular if $a_{ij} = 0$ for each $j > i$. Denote by $\text{L}(n, \mathbb{F})$ the subspace of lower triangular matrices.

   Denote $[m] := \{1, , \ldots, m\}$ the set of consecutive integers from 1 to m. (Sometimes in the notes I use $\langle m\rangle$ for $[m]$.) Let $\boldsymbol{\alpha} = \{\alpha_1 < \alpha_2 < \ldots < \alpha_p\} \subset [m], \boldsymbol{\beta} = \{\beta_1 < \beta_2 < \ldots < \beta_p\} \subset [n]$ Then $A[\boldsymbol{\alpha}, \boldsymbol{\beta}]$ is an $p \times q$ submatrix of A, which is obtained by erasing in A all rows and columns which are not in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively. Assume that $p < m, q < n$. Denote by $A(\boldsymbol{\alpha}, \boldsymbol{\beta})$ the $(m-p) \times (n-q)$ submatrix of A, which is obtained by erasing in A all rows and columns which are in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively. For $i \in [m], j \in [n]$ denote

by $A(i,j) := A(\{i\},\{j\})$. Assume that $A \in \mathbb{F}^{n\times n}$. The $A[\boldsymbol{\alpha},\boldsymbol{\beta}]$ is called a *principal* if and only if $\boldsymbol{\alpha} = \boldsymbol{\beta}$.

1.3.6 Problems

1. Let $A \in \mathbb{F}^{n\times n}$ be an invertible matrix. Show
   (a) $(A^{-1})^\top = (A^\top)^{-1}$.
   (b) A is symmetric if and only if A^{-1} is symmetric.
   (c) A is skew symmetric if and only if A^{-1} is skew symmetric.
   (d) A is lower triangular if and only if A^{-1} is lower triangular.
   (e) A is upper triangular if and only if A^{-1} is upper triangular.
2. Let $A = [a_{ij}] \in \mathbb{F}^{n\times n}$ with the following property. Its row echelon form (REF) is an upper triangular matrix U with 1's on the main diagonals and there is no pivoting. That is, first $a_{11} \neq 0$. Apply the Gauss elimination to obtain a matrix $A_1 = [a^{(1)}_{ij}]$ where the first column of A_1 is $\mathbf{e}_1$ the first column of the identity. Let $B_1 = [a^{(1)}_{ij}]^n_{i=j=2}$ be the matrix obtained by deleting the first row and column of A_1. Then B_1 satisfies the same assumptions as A, i.e. $a^{(1)}_{22} \neq 0$. Now continue as above to obtain U. Show that $A = LU$ where L is a corresponding nonsingular lower triangular matrix.
3. Let $A \in \mathbb{F}^{l\times m}, B \in \mathbb{F}^{m\times n}$. Show rank $(AB) \leq \min(\text{rank } A, \text{rank } B)$. Furthermore, if A or B are invertible then equality holds. Give an example where inequality holds.

1.4 Determinants

1.4.1 The group of bijections on a set $\mathcal{X}$

Let $\mathcal{X},\mathcal{Y}$ be two sets of objects. Then $\phi : \mathcal{X} \to \mathcal{Y}$ is called a mapping. I.e. for each $x \in \mathcal{X}$ $\phi(x)$ is an element in $\mathcal{Y}$. $\phi : \mathcal{X} \to \mathcal{X}$ is called the identity map if $\phi(x) = x$ for each $x \in X$. The identity map is denoted as id or $\text{id}_\mathcal{X}$. Let $\psi : \mathcal{Y} \to \mathcal{Z}$. Then one defines the composition map $\psi\circ\phi : \mathcal{X} \to \mathcal{Z}$ as follows $(\psi\circ\phi)(x) = \psi(\phi(x))$. $\phi : \mathcal{X} \to \mathcal{Y}$ is called bijection if there exists $\psi : \mathcal{Y} \to \mathcal{X}$ such that $\psi \circ \phi = \text{id}_\mathcal{X}, \phi \circ \psi = \text{id}_\mathcal{Y}$. ψ is denoted as ϕ^{-1}. Denote by $\text{Bi}(\mathcal{X})$ the set of all bijections of $\mathcal{X}$ onto itself. It is easy to show that $\text{Bi}(\mathcal{X})$ forms a group under the composition, with the identity element $\text{id}_\mathcal{X}$. See the below problem.

Problems

1. Let $\mathcal{X}$ be any nonempty set. Show that $\text{Bi}(\mathcal{X})$ is a group.
2. Show that if $\mathcal{X}$ has one or two elements then $\text{Bi}(\mathcal{X})$ is a commutative group.
3. Show that if $\mathcal{X}$ has at least three elements then $\text{Bi}(\mathcal{X})$ is not a commutative group.

1.4.2 The permutation group

Denote by $\mathcal{S}_n$ is the group of the bijections of the set $[n] := \{1, \ldots, n\}$ onto itself. Introduce the following polynomials

$$P_\omega(\mathbf{x}) := \prod_{1 \le i < j \le n} (x_{\omega(i)} - x_{\omega(j)}), \mathbf{x} = (x_1, \ldots, x_n), \text{ for each } \omega \in \mathcal{S}_n. \tag{1.5}$$

Define

$$\operatorname{sign}(\omega) := \frac{P_\omega(\mathbf{x})}{P_{\mathrm{id}}(\mathbf{x})}. \tag{1.6}$$

Theorem 1.11 *For each $\omega \in \mathcal{S}_n$ $\operatorname{sign}(\omega) \in \{1, -1\}$. The map $\operatorname{sign} : \mathcal{S}_n \to \{1, -1\}$ is a group homomorphism, i.e.*

$$\operatorname{sign}(\omega \circ \sigma) = \operatorname{sign}(\omega)\operatorname{sign}(\sigma) \text{ for each pair } \omega, \sigma \in \mathcal{S}_n. \tag{1.7}$$

Proof. Clearly, if $\omega(i) < \omega(j)$ then the factor $(x_{\omega(i)} - x_{\omega(j)})$ appear in $P_{\mathrm{id}}(\mathbf{x})$. If $\omega(i) > \omega(j)$ then the factor $-(x_{\omega(i)} - x_{\omega(j)})$ appear in $P_{\mathrm{id}}(\mathbf{x})$. Hence $\operatorname{sign}(\omega) \in \{1, -1\}$. Observe next that

$$\frac{P_{\omega\circ\sigma}(\mathbf{x})}{P_{\mathrm{id}}(\mathbf{x})} = \frac{P_{\omega\circ\sigma}(\mathbf{x})}{P_\sigma(\mathbf{x})} \frac{P_\sigma(\mathbf{x})}{P_{\mathrm{id}}(\mathbf{x})} = \operatorname{sign}(\omega)\operatorname{sign}(\sigma).$$

(To show the equality $\frac{P_{\omega\circ\sigma}(\mathbf{x})}{P_\sigma(\mathbf{x})} = \operatorname{sign}(\omega)$ introduce new variables $y_i = x_{\sigma(i)}$ for $i \in [n]$.)

□

$\tau \in \mathcal{S}_n$ is a transposition if there exists a pair of integers $1 \le i < j \le n$ such that $\tau(i) = j, \tau(j) = i$. Furthermore $\tau(k) = k$ for $k \ne i, j$. Note that $\tau \circ \tau =$id, i.e. $\tau^{-1} = \tau$.

Theorem 1.12 *For an integer $n \ge 2$ each $\omega \in \mathcal{S}_n$ is a product of transpositions*

$$\omega = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_m. \tag{1.8}$$

The product decomposition is not unique. However the parity of m is unique. More precisely $\operatorname{sign}(\omega) = (-1)^m$.

Proof. We agree that in (1.8) $m = 0$ iff $\omega =$ id. (This is true for any $n \in \mathbb{N}$.) We prove the theorem by induction on n. For $n = 2$ $\mathcal{S}_2$ consists of two elements id and a unique permutation τ, which satisfies $\tau(1) = 2, \tau(2) = 1$.) So $\tau \circ \tau =$id. In this case the lemma follows straightforward.

Suppose that theorem holds for $n = N \ge 2$ and assume that $n = N + 1$. Let $\omega \in \mathcal{S}_n$. Suppose first that $\omega(n) = n$. So ω can be identified with the bijection $\omega' \in \mathcal{S}_{n-1}$, where $\omega'(i) = \omega(i)$ for $i = 1, \ldots, n-1$. Use the induction hypothesis to express ω' as a product of m transposition $\tau'_1 \circ \tau'_2 \circ \cdots \circ \tau'_m$ in cS_{n-1}. Clearly each τ'_i extends to a transposition τ_i in $\mathcal{S}_n$. Hence (1.8) holds.

Suppose now that $\omega(n) = i < n$. Let τ be the transposition that interchange i and n. Let $\omega' = \tau \circ \omega$. Then $\omega'(n) = \tau(\omega(n)) = n$. The previous arguments show that $\omega' = \tau_1 \circ \ldots \circ \tau_l$. So $\omega = \tau \circ \omega' = \tau \circ \tau_1 \circ \ldots \circ \tau_l$.

It is left to show that the parity of m is unique. First observe that $\operatorname{sign}(\tau) = -1$. See problem 4. Theorem 1.11 yields that $\operatorname{sign}(\omega) = (-1)^m$. Hence the parity of m is fixed. □

Problems Assume that $\omega, \tau \in \mathrm{S}_n$. Show

1. $\operatorname{sign}(\mathrm{id}) = 1$.
2. $\operatorname{sign}(\omega^{-1}) = \operatorname{sign}(\omega)$.
3. Assume that $1 \le i < j \le n$. We say that ω changes the order of the pair (i, j) if $\omega(i) > \omega(j)$. Let $N(\omega)$ be number of pairs (i, j) such that ω changes their order. Then $\operatorname{sign}(\omega) = (-1)^{N(\omega)}$.
4. Assume that $1 \le i < j \le n$. Let τ be the transposition $\tau(i) = j, \tau(j) = i$. Then $N(\tau) = 2(j - i) - 1$. Hence $\operatorname{sign}(\tau) = -1$ for any transposition τ.

1.5 Linear, multilinear maps and functions

Let $\mathbf{U}, \mathbf{V}$ be two vector spaces over a field $\mathbb{F}$. A map $T : \mathbf{U} \to \mathbf{V}$ is called linear if $T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$ for all $a, b \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in \mathbf{V}$. Let $\mathbf{U}_1, \dots, \mathbf{U}_k$ be $k \ge 2$ vectors spaces. Recall that $\times_{i=1}^k \mathbf{U}_i := \mathbf{U}_1 \times \dots \times \mathbf{U}_k$ is the set of all tuples $(\mathbf{u}_1, \dots, \mathbf{u}_k)$ where $\mathbf{u}_i \in \mathbf{U}_i$ for $i \in [k]$. A map $T : \times_{i=1}^k \mathbf{U}_i \to \mathbf{V}$ is called a *multilinear* if it is linear with respect to each variable $\mathbf{u}_i$, while all other variable are fixed. So

$$T(\mathbf{u}_1, \dots, \mathbf{u}_{i-1}, a\mathbf{v}_i + b\mathbf{w}_i, \mathbf{u}_{i+1}, \dots, \mathbf{u}_k) =$$
$$aT(\mathbf{u}_1, \dots, \mathbf{u}_{i-1}, \mathbf{v}_i, \mathbf{u}_{i+1}, \dots, \mathbf{u}_k) + bT(\mathbf{u}_1, \dots, \mathbf{u}_{i-1}, \mathbf{w}_i, \mathbf{u}_{i+1}, \dots, \mathbf{u}_k),$$

for all $a, b \in \mathbb{F}$, $\mathbf{u}_j \in \mathbf{U}_j, j \in [k] \setminus \{i\}, \mathbf{v}_i, \mathbf{w}_i \in \mathbf{U}_i$ and $i \in [k]$.

Suppose that $\mathbf{U}_1 = \dots = \mathbf{U}_k = \mathbf{U}$. Denote $\times^k \mathbf{U} := \underbrace{\mathbf{U} \times \dots \times \mathbf{U}}_{k}$. A map $T : \times^k \mathbf{U} \to \mathbf{V}$ is called a *symmetric* map if

$$T(\mathbf{u}_{\sigma(1)}, \dots, \mathbf{u}_{\sigma(k)}) = T(\mathbf{u}_1, \dots, \mathbf{u}_k) \text{ for each permutation } \sigma \in \mathcal{S}_k$$

and each $\mathbf{u}_1, \dots, \mathbf{u}_k$. A map $T : \times^k \mathbf{U} \to \mathbf{V}$ is called a *skew symmetric* map if

$$T(\mathbf{u}_{\sigma(1)}, \dots, \mathbf{u}_{\sigma(k)}) = \operatorname{sign}(\sigma) T(\mathbf{u}_1, \dots, \mathbf{u}_k) \text{ for each permutation } \sigma \in \mathcal{S}_k.$$

Since each permutation is a product of transpositions, and $\operatorname{sign}(\tau) = -1$ for any transposition a map T is skew symmetric if and only if

$$T(\mathbf{u}_{\tau(1)}, \dots, \mathbf{u}_{\tau(k)}) = -T(\mathbf{u}_1, \dots, \mathbf{u}_k) \text{ for each transposition } \tau \in \mathcal{S}_k. \tag{1.9}$$

In this course most of the maps are linear or multilinear. If $\mathbf{V} = \mathbb{F}$ then the map T is called linear, multilinear, symmetric and skew symmetric functions respectively.

Example 1: $T : \mathbb{F}^m \times \mathbb{F}^n \to \mathbb{F}^{m \times n}$ given by $T(\mathbf{x}, \mathbf{y}) = \mathbf{x}\mathbf{y}^\top$. A multilinear map. $T(\mathbb{F}^m \times \mathbb{F}^n)$ is the set of matrices of rank one and zero.

Example 2: $T, Q : \mathbb{F}^m \times \mathbb{F}^m \to \mathrm{S}(n, \mathbb{F})$ given by $T(\mathbf{x}, \mathbf{y}) = \mathbf{x}\mathbf{y}^\top + \mathbf{y}\mathbf{x}^\top$, $Q(\mathbf{x}, \mathbf{y}) = \mathbf{x}\mathbf{y}^\top - \mathbf{y}\mathbf{x}^\top$. Then T, Q are multilinear symmetric and skew map respectively. $T(\mathbb{F}^m \times \mathbb{F}^m)$ is the set of matrices symmetric matrices of rank two, one and zero.

Example 3: Let $m_1, \dots, m_k$ are positive integers. Denote $\mathbb{F}^{m_1 \times \dots m_k}$ the k-multidimensional arrays. so $C \in \mathbb{F}^{m_1 \times \dots m_k}$ is an k-array $[c_{i_1,\dots,i_k}]$ where $i_j \in [m_j]$ for $j \in k$. C is also called k-mode tensors. So $\mathbb{F}^{m_1 \times \dots m_k}$ is a vector space where $aC = [ac_{i_1,\dots,i_k}]$ and the addition is carried out entrywise. Then $T : \times_{i \in [k]} \mathbb{F}^{m_i} \to \mathbb{F}^{m_1 \times \dots m_k}$ is given by the tensor product $T(\mathbf{x}_1, \dots, \mathbf{x}_k) = \otimes_{i \in [k]} \mathbf{x}_i$

$$\otimes_{i \in [k]} \mathbf{x}_i := [x_{i_1,1} \dots x_{i_k,k}],\ i_j \in [m_j],\ \mathbf{x}_j := (x_{1,j}, \dots, x_{m_j,j})^\top, j \in [k]. \tag{1.10}$$

So T is a multilinear map. $\mathbb{F}^{m_1 \times \dots m_k}$ is also denoted as $\otimes_{j \in [k]} \mathbb{F}^{m_j}$. Note that $\mathrm{span}(T(\times_{j \in [k]} \mathbf{U}_j) = \otimes_{j \in [k]} \mathbb{F}^{m_j}$.

Suppose furthermore that $m_j = m$ for $j \in k$. Then $\otimes_{j \in [k]} \mathbb{F}^{m_j}$ is denoted by $\otimes^k \mathbb{F}^m$. Let $T : \times^k \mathbb{F}^m \to \otimes^k \mathbb{F}^m$ be the following maps

$$T(\mathbf{x}_1, \dots, \mathbf{x}_k) := \sum_{\sigma \in \mathcal{S}_k} \otimes_{i \in [k]} \mathbf{x}_{\sigma(i)}, \tag{1.11}$$

$$Q(\mathbf{x}_1, \dots, \mathbf{x}_k) := \sum_{\sigma \in \mathcal{S}_k} \mathrm{sign}(\sigma) \otimes_{i \in [k]} \mathbf{x}_{\sigma(i)}. \tag{1.12}$$

Then T ad Q are multilinear, symmetric and skew symmetric maps respectively.

Example 4: $\mathbf{U} = \mathbf{V} = \mathbb{R}^3$. Let $T(\mathbf{x}, \mathbf{y}) = \mathbf{x} \times \mathbf{y}$ be the cross product in $\mathbb{R}^3$. Then $T : \mathbb{R}^3 \times \mathbb{R}^3$ is a multilinear skew symmetric map.

1.6 Definition and properties of determinants

For $A \in \mathbb{F}^{n \times n}$ we define the determinant of A as follows, View row i of A as a vector $\mathbf{r}_i$. Then the determinant of A is a skew symmetric function $\det : \times^n \mathbb{F}^n \to \mathbb{F}$ normalized by the condition

$$\det I_n = 1. \tag{1.13}$$

Theorem 1.13 *Assume that the characteristic of $\mathbb{F}$ is different from 2, i.e. $2 \neq 0$ in $\mathbb{F}$. Then determinant function $\det : \mathbb{F}^{n \times n} \to \mathbb{F}$ exists and uniquely defined by the formula*

$$\det A = \sum_{\omega \in \mathcal{S}_n} \mathrm{sign}(\omega) a_{1\omega(1)} a_{2\omega(2)} \dots a_{n\omega(n)}, \quad A = [a_{ij}]. \tag{1.14}$$

(It is agreed that even in the case that the characterstic of $\mathbb{F}$ is 2 the above formula is the definition of the determinant.) Furthermore it satisfies the following properties.

1. $\det A = \det A^\top$.
2. $\det A$ *is a multilinear function in rows or columns of A.*
3. *The determinant of lower triangular or upper triangular matrix is a product of the diagonal entries of A.*

4. *Let B obtained from A by permuting two rows or columns of A. Then* $\det B = -\det A$.

5. *If A has two equal rows or columns then* $\det A = 0$.

6. *A is invertible iff* $\det A \neq 0$.

7. *Let $A, B \in \mathbb{F}^{n\times n}$. Then* $\det AB = \det A \det B$.

8. *(*Laplace row and column expansion for determinants*) For $i, j \in [n]$ denote by $A(i,j) \in \mathbb{F}^{(n-1)\times(n-1)}$ the matrix obtained from A by deleting its i-th row and j-th column. Then*

$$\det A = \sum_{j=1}^{n} a_{ij}(-1)^{i+j} \det A(i,j) = \sum_{i=1}^{n} a_{ji}(-1)^{i+j} \det A(i,j) \tag{1.15}$$

for $i = 1, \ldots, n$.

Proof. First observe that if the determinant function exists then it must be defined by (1.14). Indeed,

$$\mathbf{r}_i = \sum_{j=1}^{n} a_{ij}\mathbf{e}_j^\top, \quad \mathbf{e}_j = (\delta_{j1}, \ldots, \delta_{jn})^\top.$$

Let $T : \times^n \mathbb{F}^n \to \mathbb{F}$ be a multilinear skew symmetric function. Use multilinearity of T to deduce

$$T(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \sum_{i_1,\ldots,i_n \in [n]} a_{1i_1} \ldots a_{ni_n} T(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}).$$

Since T is skew symmetric $T(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}) = 0$ if $i_p = i_q$ for some $1 \leq< p < q \leq n$. Indeed if we interchange $\mathbf{e}_{i_p} = \mathbf{e}_{i_q}$ then one one hand we do not change the value of the T. On the other hand since T is is skew symmetric then the value of T is equal to $-T$ when we interchange $\mathbf{e}_{i_p} = \mathbf{e}_{i_q}$. As the characterstic of $\mathbb{F}$ is not 2 we deduce that $T(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}) = 0$ if $i_p = i_q$. Hence in the above expansion of $T(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ the only nonzero terms are where $\{i_1, \ldots, i_n\} = [n]$. Each such set of n indices $\{i_1, \ldots, i_n\}$ corresponds uniquely to a permutation $\omega \in \mathcal{S}_n$, where $\omega(j) = i_j$ for $j \in [n]$. Now $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ can be brought to $\{\mathbf{e}_{\sigma(1)}, \ldots, \mathbf{e}_{\sigma(n)}\}$ by using transpositions. The composition of this transpositions yield σ. Since T is skew symmetric, it follows that $T(\mathbf{e}_{\sigma(1)}, \ldots, \mathbf{e}_{\sigma(n)}) = \operatorname{sign}(\omega) T(\mathbf{e}_1, \ldots, \mathbf{e}_n)$. As $T(I_n) = 1$ we deduce that $T(\mathbf{e}_{\sigma(1)}, \ldots, \mathbf{e}_{\sigma(n)}) = \operatorname{sign}(\omega)$.

1. In (1.14) note that if $j = \omega(i)$ then $i = \omega^{-1}(j)$. sice ω is a bijection,we deduce that when $i = 1, \ldots, n$ then j takes each value in $[n]$. Since $\operatorname{sign}(\omega) = \operatorname{sign}(\omega^{-1})$ we obtain Hence

$$\det A = \sum_{\omega \in \mathcal{S}_n} \operatorname{sign}(\omega) a_{\omega^{-1}(1)1} a_{\omega^{-1}(2)2} \ldots a_{\omega^{-1}(n)n} =$$
$$\sum_{\omega \in \mathcal{S}_n} \operatorname{sign}(\omega^{-1}) a_{\omega^{-1}(1)1} a_{\omega^{-1}(2)2} \ldots a_{\omega^{-1}(n)n} = \det A^\top.$$

(Note since $\mathcal{S}_n$ is a group, when ω varies over $\mathcal{S}_n$ so is ω^{-1}.)
2. Fix all rows of A except the row i. From (1.14) it follows that $\det A$ is a linear function in the row i of A. I.e. $\det A$ is a multilinear function in the rows of A. In view of the identity $\det A = \det A^\top$ we deduce that $\det A$ is a multilinear function of the columns of A.
3. Assume that A is upper triangular. Then $a_{n\omega(n)} = 0$ if $\omega(n) \neq n$. So all nonzero terms in (1.14) are zero unless $\omega(n) = n$. So assume that $\omega(n) = n$. Then $a_{(n-1)\omega(n-1)} = 0$ unless $\omega(n-1) = n-1$. So all nonzero terms in (1.14) must come from all ω satisfying $\omega(n) = n, \omega(n-1) = n-1$. Continuing in the same manner we deduce that the only nonzero term in (1.14) comes from $\omega =$id. As sign(id) $= 1$ it follows that $\det A = \prod_{i=1}^n a_{ii}$. Since $\det A = \det A^\top$ we deduce the claim for lower triangular matrix.
4. Note that a permutation of two elements $1 \leq i < j \leq n$ in $[n]$ is achieved by a transposition $\tau \in \mathcal{S}_n$. So $B = [b_{pq}]$ and $b_{pq} = a_{\tau(p)q}$. As in the proof of 1 we let $\tau(i) = j$, then $j = \tau^{-1}(i) = \tau(i)$. Hence

$$\det B = \sum_{\omega \in \mathcal{S}_n} \operatorname{sign}(\omega) a_{\tau(1)\omega(1)} a_{\tau(2)\omega(2)} \dots a_{\tau(n)\omega(n)} =$$
$$\sum_{\omega \in \mathcal{S}_n} \operatorname{sign}(\omega) a_{1\omega(\tau(1))} a_{2\omega(\tau(2))} \dots a_{n\omega(\tau(n))} =$$
$$\sum_{\omega \in \mathcal{S}_n} -\operatorname{sign}(\omega \circ \tau) a_{1(\omega\circ\tau)(1)} a_{2(\omega\circ\tau)(2)} \dots a_{n(\omega\circ\tau)(n)} = -\det A.$$

5. Suppose first that we are not in characteristic 2, i.e. $0 \neq 2$. Suppose A has two identical rows. Interchange these two rows to obtain $B = A$. So $\det A = \det B = -\det A$, where the last equality is established in 4. So $2 \det A = 0$. Hence $\det A = 0$.

Suppose now that $2 = 0$, i.e. $1 = -1$. In this case $\det A$ is equal to the *permanent* of A:

$$\operatorname{perm} A = \sum_{\omega \in \mathcal{S}_n} a_{1\omega(1)} a_{2\omega(2)} \dots a_{n\omega(n)}. \tag{1.16}$$

Since A has two identical rows each term appears twice. (For example if row one is equal to two two them $a_{1i}a_{2j} = a_{1j}a_{2i}$. So we have only $\frac{n!}{2}$terms and each term is multiplied by $2 = 0$. Hence $\det A = 0$. Use $\det A = \det A^\top$ to deduce that $\det A = 0$ if A has two identical columns.
6. Use 2, 3 and 5 to deduce that $\det EA = \det E \det A$ if E is an elementary matrix. (Note that $\det E \neq 0$.) Hence if $E_1, \dots, E_k$ are elementary matrices we deduce that $\det(E_k \dots E_1) = \prod_{i=1}^k \det E_i$. Let B be the reduced row echelon for of A. So $B = E_k E_{k-1} \dots E_1 A$. Hence $\det B = (\prod_{i=1}^n \det E_i) \det A$. If $I_n \neq B$, then the last row of B is zero, so $\det B = 0$ which implies that $\det A = 0$. If $B = I_n$ then $\det A = \prod_{i=1}^n (\det E_i)^{-1}$.
7. Assume that either $\det A = 0$ or $\det B = 0$. We claim that $(AB)\mathbf{x} = \mathbf{0}$ has a nontrivial solution. Suppose $\det B = 0$. 6 and Theorem 1.7 yields that $B\mathbf{x} = 0$ has a nontrivial solution which satisfies $AB\mathbf{x} = \mathbf{0}$. Suppose that B is invertible and $A\mathbf{y} = \mathbf{0}$ for some $\mathbf{y} \neq \mathbf{0}$. Then $AB(B^{-1}\mathbf{y}) = \mathbf{0}$ which implies that $\det AB = 0$. Hence in these cases $\det AB = 0 = \det A \det B$. Suppose that A, B are invertible. Then each of them is a product of elementary matrices. Use the arguments in the proof of 6 to deduce that $\det AB = \det A \det B$.

8. First we prove the first part of the formula 1.15 for $i = n$. Clearly, (1.14) yield the equality

$$\det A = \sum_{j=1}^{n} a_{nj} \sum_{\omega \in \mathcal{S}_n, \omega(n)=j} \operatorname{sign}(\omega) a_{1\omega(1)} \dots a_{(n-1)\omega(n-1)}. \tag{1.17}$$

In the above sum let $j = n$. So $\omega(n) = n$. So ω can be viewed as $\omega' \in \mathcal{S}_{n-1}$. Also $\operatorname{sign}(\omega) = \operatorname{sign}(\omega')$. Hence $\sum_{\omega' \in \mathcal{S}_{n-1}} \operatorname{sign}(\omega) a_{1\omega'(1)} \dots a_{(n-1)\omega'(n-1)} = \det A(n, n)$. Note that $(-1)^{n+n} = 1$. This justify the form of the last term of the expansion 1.15 for $i = n$. To justify the sign of any term in 1.15 for $i = n$ we take the column j and interchange it first with column $j + 1$, then with column $j + 2$, and at last with the column n. The sign of $\det A(n, j)$ is $(-1)^{n+j}$. This proves the case $i = n$.

By interchanging any row $i < n$ with row $i + 1$, row $i + 2$, and finally with row n we deduce the first part of the formula 1.15 for any i. By considering $\det A^\top$ we deduce the second part of the formula 1.15.

□

Observe next that $\det I_n = 1$. Hence

$$1 = \det I_n = \det(AA^{-1}) = \det A \det A^{-1} \Rightarrow \det(A^{-1}) = \frac{1}{\det A}.$$

For $A = [a_{ij}] \in \mathbb{F}^{n \times n}$ denote by A_{ij} the determinant of the matrix obtained from A by deleting i-th row and j-th column and multiplied by $(-1)^{i+j}$. (This is called the (i, j) *cofactor* of A.)

$$A_{ij} := (-1)^{i+j} \det A(i, j). \tag{1.18}$$

Then the expansion of $\det A$ by the row i and the column j respectively, i.e. (1.17), is given by the equalities

$$\det A = \sum_{j=1}^{n} a_{ij} A_{ij} = \sum_{i=1}^{n} a_{ij} A_{ij}. \tag{1.19}$$

Then the *adjoint* matrix of A is defined as follows:

$$\operatorname{adj} A := \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}. \tag{1.20}$$

Proposition 1.14 *Let $A \in \mathbb{F}^{n \times n}$. Then*

$$A(\operatorname{adj} A) = (\operatorname{adj} A)A = (\det A) I_n. \tag{1.21}$$

Hence A is invertible if and only if $\det A \neq 0$. Furthermore, $A^{-1} = (\det A)^{-1} \operatorname{adj} A$.

Proof. Consider an (i, k) entry of $A(\operatorname{adj} A)$. It is given as $\sum_{j=1}^{n} a_{ij} A_{kj}$. For $i = k$ (1.19)yields that $\sum_{j=1}^{n} a_{ij} A_{ij} = \det A$. Suppose that $i \neq k$. Let B_k be the matrix obtained from A be replacing the row k in of A by the row i. So B_k has two identical rows, hence $\det B_k = 0$. On the other hand, expand B_k by the row k

to obtain that $0 = \det B_k = \sum_{j=1}^n a_{ij} A_{kj}$. This shows that $A(\text{adj } A) = (\det A) I_n$. Similarly one shows that $(\text{adj } A)A = (\det A) I_n$.

Theorem 1.13 part 6 shows that A is invertible iff $\det A \neq 0$. Hence for invertible A, $A^{-1} = \frac{1}{\det A} \text{adj } A$. □

Proposition 1.15 *(Cramer's rule.) Let $A \in \text{GL}(n, \mathbb{F})$ and consider the system $A\mathbf{x} = \mathbf{b}$, where $\mathbf{x} = (x_1, \ldots, x_n)^\top$. Denote by C_k the matrix obtained from A by replacing the column k in A by $\mathbf{b}$. Then $x_k = \frac{\det B_k}{\det A}$ for $k = 1, \ldots, n$.*

Proof. Clearly $\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det A}(\text{adj } A)\mathbf{b}$. Hence $x_k = (\det A)^{-1} \sum_{j=1}^n A_{jk} b_j$, where $\mathbf{b} = (b_1, \ldots, b_k)^\top$. Expand B_k by the column k to deduce that $\det B_k = \sum_{j=1}^n b_j A_{jk}$. □

Let B be a square submatrix of A. Then $\det B$ is called a *minor* of A. $\det B$ is called a principle minor of order m if B is an $m \times m$ principle submatrix of A.

1.7 Polynomials

For a field $\mathbb{F}$, (usually $\mathbb{F} = \mathbb{R}, \mathbb{C}$), denote by $\mathbb{F}[z]$, the *ring* of polynomials $p(z) = a_0 z^n + a_1 z^{n-1} + \ldots + a_n$ with coefficients in $\mathbb{F}$. The *degree* of p, denoted by $\deg p$, is the maximal degree $n - j$ of a monomial $a_j x^{n-j}$ which is not identically zero, i.e. $a_j \neq 0$. So $\deg p = n$ if and only if $a_0 \neq 0$, the degree of a nonzero constant polynomial $p(z) = a_0$ is zero, and the degree of the zero polynomial is agreed to be equal to $-\infty$. For two polynomials $p, q \in \mathbb{F}[z]$ and two scalars $a, b \in \mathbb{F}$ $ap(z) + bq(z)$ is a well defined polynomial. Hence $\mathbb{F}[z]$ is a vector space over $\mathbb{F}$, whose dimension is infinite. The set of polynomials of degree n at most, is $n + 1$ dimensional subspace of $\mathbb{F}[z]$. Given two polynomials $p = \sum_{i=0}^n a_i z^{n-i}, q = \sum_{j=0}^m b_j z^{m-j} \in \mathbb{F}[z]$ one can form the product

$$p(z)q(z) = \sum_{k=0}^{n+m} (\sum_{i=0}^{k} a_i b_{k-i}) z^{n+m-k}, \text{ where } a_i = b_j = 0 \text{ for } i > n \text{ and } j > m.$$

Note that $pq = qp$ and $\deg pq = \deg p + \deg q$. The addition and the product in $\mathbb{F}[z]$ satisfies all the nice distribution identities as the addition and multiplication in $\mathbb{F}$. Here the constant polynomial $p \equiv 1$ is the identity element, and the zero polynomial as the zero element. (That is the reason for the name *ring* of polynomials in one variable over $\mathbb{F}$.)

Recall that given two polynomials $p, q \in \mathbb{F}[z]$ one can divide p by $q \not\equiv 0$ with the residue r, i.e. $p = tq + r$ for some unique $t, r \in \mathbb{F}[z]$, where $\deg r < \deg q$. For $p, q \in \mathbb{F}[z]$ let (p, q) be the *greatest common divisor* of p, q. If p and q are identically zero then (p, q) is the zero polynomial. Otherwise (p, q) is a polynomial s of the highest degree that divides p and q. s is determined up to a multiple of a nonzero scalar. s can be chosen as a unique *monic* polynomial:

$$s(z) = z^l + s_1 z^{l-1} + \ldots + s_l \in \mathbb{F}[z]. \tag{1.22}$$

For $p, q \not\equiv 0$ s can be found using the *Euclid* algorithm:

$$p_i(z) = t_i(z) p_{i+1}(z) + p_{i+2}(z), \ \deg p_{i+2} < \deg p_{i+1} \quad i = 1, \ldots \tag{1.23}$$

Start this algorithm with $p_1 = p, p_2 = q$. Continue it until $p_k = 0$ the first time. (Note that $k \geq 3$.) Then $p_{k-1} = (p, q)$. It is easy to show, for example by induction, that each p_i is of the form $u_i p + v_i q$ for some polynomials u_i, v_i. Hence the Euclid algorithm yields

$$(p(z), q(z)) = u(z)p(z) + v(z)q(z), \text{ for some } u(z), v(z) \in \mathbb{F}[z]. \tag{1.24}$$

(This formula holds for any $p, q \in \mathbb{F}[z]$.) $p, q \in \mathbb{F}[z]$ are called *coprime* if $(p, q) = 1$.

Recall that if we divide $p(z)$ by $z - a$ we get the residue $p(a)$, i.e. $p(z) = (z - a)q(z) + p(a)$. So $z - a$ divides $p(z)$ if and only if $p(a) = 0$, i.e a is the root of p. A monic $p(z)$ *splits* to a product of linear factors if

$$p(z) = (z - z_1)(z - z_2) \dots (z - z_n) = \prod_{i=1}^{n} (z - z_i). \tag{1.25}$$

Note that $z_1, \dots, z_n$ are the *roots* of p.

Let $\mathbf{z} = (z_1, \dots, z_n)^\top \in \mathbb{F}^n$. Denote

$$\sigma_k(\mathbf{z}) := \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} z_{i_1} z_{i_2} \dots z_{i_k}, \quad k = 1, \dots, n. \tag{1.26}$$

$\sigma_k(\mathbf{z})$ is called the $k-th$ *elementary symmetric polynomial* in $z_1, \dots, z_n$. Observe

$$\begin{aligned} \sigma_1(\mathbf{z}) &= z_1 + z_2 + \dots + z_n, \ n \text{ summands} \\ \sigma_2(\mathbf{z}) = z_1 z_2 + \dots + z_1 z_n + z_2 z_3 + \dots z_{n-1}, \ &\frac{n(n-1)}{2} \text{ summands}, \\ \sigma_n(\mathbf{z}) &= z_1 z_2 \dots z_n, \ n \text{ terms in the product.} \end{aligned}$$

A straightforward calculation shows

$$\prod_{i=1}^{n} (z - z_i) = z^n + \sum_{i=1}^{n} (-1)^i \sigma_i(\mathbf{z}) z^{n-i}. \tag{1.27}$$

1.7.1 Finite extension of fields

A polynomial $p(z) \in \mathbb{F}[z]$ is called irreducible, if all polynomials q that divide p are either constant nonzero polynomials, or polynomials of the form $ap(z)$, where $a \in \mathbb{F} \setminus \{0\}$. $\mathbb{F}$ is called an *algebraically closed field* if any monic polynomial $p(z) \in \mathbb{F}[z]$ splits to linear factors in $\mathbb{F}[z]$. It is easy to see that $\mathbb{F}$ is algebraically closed if and only if the only irreducible monic polynomials are $z - a$ for all $a \in \mathbb{F}$. So $\mathbb{F}$ is not algebraically closed if and only if there exists an irreducible monic polynomial in $\mathbb{F}[z]$ of degree greater than 1.

Theorem 1.16 *Let $\mathbb{F}$ be a field. Assume that $p(z) = z^d + \sum_{i=1}^{d} a_i z^{d-i}$ be an irreducible polynomial, where $d > 1$. Denote by $\mathbb{F}[z]/(p(z)\mathbb{F}[z])$ the set of all polynomials modulo $p(z)$. That is $f(z) \equiv g(z)$ if the polynomial $f(z) - g(z)$ is divided by $p(z)$. Then this set is a field, denoted by $\mathbb{F}_{p(z)}$, under the addition and multiplication modulo $p(z)$. $\mathbb{F}_{p(z)}$ is a vector space over $\mathbb{F}$ of dimension d. The set of all constant polynomials in $\mathbb{F}_{p(z)}$ is isomorphic to $\mathbb{F}$. ($\mathbb{F}_{p(z)}$ is called a finite extension of $\mathbb{F}$, and more precisely a an extension of degree d.)*

Problems
Show

1. Prove Theorem 1.16.

2. Show that there is only one monic irreducible polynomial of degree two over $\mathbb{Z}_2$. Describe the extension of $\mathbb{Z}_2$ of degree 2.

3. Show that any finite extension of $\mathbb{Z}_p$, where $p \geq 2$ is prime, has p^d elements.

4. Recall that the characteristic polynomial of $A \in \mathbb{F}^{n\times n}$ is defined as $\det(zI_n - A)$. Show that the characteristic polynomial is a monic polynomial of degree n. Prove that the coefficient of z^{n-k} of the characteristic polynomial of A is the sum of all minors $\det A[\alpha, \alpha]$ where α runs over all subsets of $[n]$ of cardinality k.

5. Recall that $\lambda \in \mathbb{F}$ is an eigenvalue of $A \in \mathbb{F}^{n\times n}$ if there exists $\mathbf{0} \neq \mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$. This $\mathbf{x} \neq \mathbf{0}$ is called an eigenvector of A corresponding to λ. Show that λ is an eigenvalue of A if and only if λ is a zero of the characteristic polynomial of A.

6. Find a $A \in \mathbb{Z}_2^{2\times 2}$ which does not have eigenvalues in $\mathbb{Z}_2$.

7. Show that $\mathbb{C}$ is a 2-extension of $\mathbb{R}$. What is the corresponding irreducible polynomial in $\mathbb{R}[z]$?

8. Let $p \geq 3$ be a prime and consider the polynomial $f_p = \sum_{i=0}^{p-1} z^i \in \mathbb{Q}[z]$. Show that this polynomial is irreducible over $\mathbb{Q}[z]$.

1.8 Complex numbers

$\mathbb{C}$ is the field of complex numbers. A complex number z can be written in the form $z = x + \mathbf{i}y$, where $x, y \in \mathbb{R}$. Here $\mathbf{i}^2 = -1$. Sometimes in this notes we denote $\mathbf{i}$ by $\sqrt{-1}$. So $\mathbb{C}$ can be viewed as $\mathbb{R}^2$, where the vector $(x, y)^\top$ represents z. Recall that $x = \Re z$, the real part of $\mathbf{z}$, and $y = \Im z$, the imaginary part of z. $\bar{z} = x - \mathbf{i}y$ is the conjugate of z. Note that $|z| = \sqrt{x^2 + y^2}$, is the absolute value of z or the modulus of z. For $z \neq 0$, the argument of z is defined as $\arctan \frac{y}{x}$. The polar representation of z is $z = re^{\mathbf{i}\theta} = r(\cos\theta + \mathbf{i}\sin\theta)$. Here r and θ are the modulus and the argument of $z(\neq 0)$. Let $w = u + \mathbf{i}v = R(\cos\psi + \mathbf{i}\sin\psi)$, where $u, v \in \mathbb{R}$. Then

$$z + w = (x + u) + \mathbf{i}(y + v), \quad zw = (xu - yv) + \mathbf{i}(xv + yu) = rRe^{i(\theta+\psi)},$$
$$\frac{w}{z} = \frac{1}{z\bar{z}} w\bar{z} = \frac{R}{r} e^{\mathbf{i}(\psi-\theta)} \text{ if } z \neq 0.$$

For a complex number $w = Re^{\mathbf{i}\psi}$ and a positive integer $n \geq 2$, the equation $z^n - w = 0$ has n-complex distinct roots for $w \neq 0$, which are $R^{\frac{1}{n}} e^{\mathbf{i}\frac{\psi+2k\pi}{n}}$ for $k = 0, 1, \ldots, n-1$.

The fundamental theorem of algebra states that any monic polynomial $p(z) \in \mathbb{C}[z]$ of degree $n \geq 2$ splits to linear factors, i.e. $p(z) = \prod_{i=1}^{n}(z - z_i)$.

Problems
Show

1. Show that $p(z) = z^2 + bz + c \in \mathbb{R}[z]$ is irreducible over $\mathbb{R}$ if and only if $b^2 < 4c$.

2. Show that any monic polynomial $p(z) \in \mathbb{R}[z]$ of degree 2 at least splits to a product of irreducible linear and quadratic monic polynomials over $\mathbb{R}[z]$.

3. Deduce from the previous problem that any $p(z) \in \mathbb{R}[z]$ of odd degree must have a real root.

1.9 Linear transformations

Let $\mathbf{V}, \mathbf{U}$ be two finite dimensional subspaces, where $\dim \mathbf{V} = n, \dim \mathbf{U} = m$. $T : \mathbf{V} \to \mathbf{U}$ is called a linear transformation if $T(a\mathbf{u}+b\mathbf{v}) = aT(\mathbf{u})+bT(\mathbf{v})$. Assume that $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}, \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ be two bases in $\mathbf{V}, \mathbf{U}$ respectively. Then T is completely determined by $T(\mathbf{v}_j) = \sum_{i=1}^m a_{ij}\mathbf{u}_i, j = 1, \dots, n$. Let $A = [a_{ij}]_{j=i=1}^{m,n} \in \mathbb{F}^{m\times n}$. Then the above equality is equivalent to

$$T[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] = [T\mathbf{v}_1, T\mathbf{v}_2, \dots, T\mathbf{v}_n] = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]A. \tag{1.28}$$

A is called the representation matrix of T. Assume that

$$[\mathbf{y}_1, \dots, \mathbf{y}_n] = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]Y,\ [\mathbf{x}_1, \dots, \mathbf{x}_m] = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]X, \tag{1.29}$$

where $Y \in \mathrm{GL}(n, \mathbb{F}), X \in \mathrm{GL}(m, \mathbb{F})$, are another bases in $\mathbf{V}$ and $\mathbf{U}$ respectively. Then

$$T[\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n] = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]X^{-1}AY. \tag{1.30}$$

Denote by $\mathrm{L}(\mathbf{V}, \mathbf{U})$ is the set of linear transformations $T : \mathbf{V} \to \mathbf{U}$. $\mathrm{L}(\mathbf{V}, \mathbf{U})$ is a vector space by defining $(aT+bS)(\mathbf{v}) = aT(\mathbf{v})+bS(\mathbf{v})$. Then $\mathrm{L}(\mathbf{V}, \mathbf{U})$ is isomorphic to $\mathbb{F}^{m\times n}$. (Choose bases $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n], [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]$ in $\mathbf{V}, \mathbf{U}$ respectively, and identify T with its representation matrix.)

Definition 1.17 *Let $\mathbf{U}, \mathbf{V}$ be two vector spaces over $\mathbb{F}$. $T \in \mathrm{L}(\mathbf{V}, \mathbf{U})$ is called an isomorphism if it is $1-1$, i.e. $T(\mathbf{x}) = T(\mathbf{y}) \Rightarrow \mathbf{x} = \mathbf{y}$ and onto, i.e. $T(\mathbf{V}) = \mathbf{U}$. $\mathbf{U}, \mathbf{V}$ are called isomorphic if there exists an isomorphism $T \in \mathrm{L}(\mathbf{U}, \mathbf{V})$.*

Lemma 1.18 *Let $\mathbf{U}, \mathbf{V}$ be two finite dimensional vector spaces. Then $\mathbf{U}$ and $\mathbf{V}$ isomorphic if and only if $\dim \mathbf{U} = \dim \mathbf{V}$.*

Proof. Assume first that $\mathbf{U}$ and $\mathbf{V}$ are isomorphic. That is, there exists $T \in L(\mathbf{V}, \mathbf{U})$ which is an isomorphism. Assume that $[\mathbf{v}_1, \dots, \mathbf{v}_n]$ is a basis of $\mathbf{V}$. We claim that $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ spans $\mathbf{U}$. Indeed, since T is onto for each $\mathbf{u} \in \mathbf{U}$ there exists $\mathbf{v} = \sum_{i=1}^n a_i\mathbf{v}_i$ such that $\mathbf{u} = T(\mathbf{v}) = \sum_{i=1}^n a_iT(\mathbf{v}_i)$. So $\mathrm{span}(T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)) = \mathbf{U}$. Next we claim that $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ are linearly independent. Suppose $\mathbf{0} = \sum_{i=1}^n a_iT(\mathbf{v}_i)$. Let $\mathbf{v} = \sum_{i=1}^n a_i\mathbf{v}_i$. Then $\mathbf{0} = T(\mathbf{v})$. As T linear, then $T(\mathbf{0}) = \mathbf{0}$. As T is one-to-one this implies that $\mathbf{v} = \mathbf{0}$. Since $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent it follows that $a_1 = \dots = a_n = 0$. Hence $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ are linearly independent. Therefore $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ is a basis of $\mathbf{U}$. So $\dim \mathbf{U} = \dim \mathbf{V}$.

Vice versa, suppose that $\dim \mathbf{U} = \dim \mathbf{V} = n$. So $\mathbf{V}$ has a basis $[\mathbf{v}_1, \dots, \mathbf{v}_n]$ and $\mathbf{U}$ has a basis $[\mathbf{u}_1, \dots, \mathbf{u}_n]$. Define a linear transformation by $T(v_i) = \mathbf{u}_i$ for $i \in [n]$. The previous arguments show: Since $\mathrm{span}(\mathbf{u}_1, \dots, \mathbf{u}_n) = \mathbf{U}$ it follows that T is onto. Suppose that $T(\mathbf{x}) = T(\mathbf{y}) = \sum_{i=1}^n a_i\mathbf{u}_i$. Then $\mathbf{x} = \mathbf{y} = \sum_{i=1}^n \mathbf{v}_i$. So T

is one-to-one. □

Denote by $\mathbf{V}' := \mathrm{L}(\mathbf{V}, \mathbb{F})$. $\mathbf{V}'$ is the set of all linear functions on $\mathbf{V}$. $\mathbf{V}'$ is called the *dual space*. So if $\mathbf{v}, \mathbf{w} \in \mathbf{U}'$ then $a\mathbf{v} + b\mathbf{w}$ is the linear functional defined as follows. $(a\mathbf{v} + b\mathbf{w})(\mathbf{u}) = a\mathbf{v}(\mathbf{u}) + b\mathbf{w}(\mathbf{u})$. It is straightforward to show that $\mathbf{U}'$ is a vector space.

Proposition 1.19 *Let* $\mathbf{U}$ *be a finite dimensional vector space. Then* $\mathbf{U}'$ *is isomorphic to* $\mathbf{U}$.

Proof. Choose a basis $[\mathbf{u}_1, \ldots, \mathbf{u}_n]$. Let $\mathbf{v}_i \in \mathbf{U}'$ be the following linear functional $\mathbf{v}_i(\mathbf{u}_j) = \delta_{ij}$ for $i, j \in [n]$. (This basis is called a dual basis in $\mathbf{U}$.) It is straightforward to show that $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is a basis in $\mathbf{U}'$. □

If $\mathbf{U}$ is finite dimensional then all three spaces $\mathbf{U}, \mathbf{U}', \mathbf{U}'' := (\mathbf{U}')'$ are isomorphic. There is a *natural* isomorphism $\phi : \mathbf{U} \to \mathbf{U}''$. Namely

$$\phi(\mathbf{u})(\mathbf{v}) := \mathbf{v}(\mathbf{u}), \quad \mathbf{u} \in \mathbf{U},\ \mathbf{v} \in \mathbf{U}^*.$$

Let $T \in \mathrm{L}(\mathbf{V}, \mathbf{U})$. Then there exists a unique $T' \in \mathrm{L}(\mathbf{U}', \mathbf{V}')$ defined as follows.

$$(T'\mathbf{w})(\mathbf{v}) = \mathbf{w}(T\mathbf{v}), \quad \mathbf{w} \in \mathbf{U}'. \tag{1.31}$$

Assume that $\mathbf{U}, \mathbf{V}$ are finite dimensional. Choose bases $[\mathbf{u}_1, \ldots, \mathbf{u}_m]$ and $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ in $\mathbf{U}$ and $\mathbf{V}$ respectively. Assume that T represented in these bases by $A \in \mathbb{F}^{m \times n}$. Choose dual bases in $\mathbf{U}'$ and $\mathbf{V}'$ respectively. Then T' is represented in these dual bases by $A^\top$.

Denote by $\mathrm{L}(\mathbf{V}) := \mathrm{L}(\mathbf{V}, \mathbf{V})$. Any $T \in \mathrm{L}(\mathbf{V})$ is represented by a matrix $A \in \mathbb{F}^{n \times n}$ in a basis $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ of $\mathbf{V}$ as follows:

$$T[\mathbf{v}_1, \ldots, \mathbf{v}_n] = [v_1, \ldots, \mathbf{v}_n], \text{ i.e. } T(\mathbf{v}_j) = \sum_{i=1}^{n} a_{ij}\mathbf{v}_i \text{ for } j \in [n]. \tag{1.32}$$

Problems

Show

1. Any n-dimensional space $\mathbf{V}$ over $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.
2. Any finite dimensional space $\mathbf{V}$ over $\mathbb{F}$ is isomorphic to $(\mathbf{V}')'$. Define an explicit isomorphism from $\mathbf{V}$ to $(\mathbf{V}')'$.
3. Any $T \in \mathrm{L}(\mathbf{V}, \mathbf{U})$ is an isomorphism if and only if T is represented by an invertible matrix in some bases of $\mathbf{V}$ and $\mathbf{U}$.
4. Let $T \in \mathrm{L}(\mathbf{V}, \mathbf{U})$. What are the conditions on the dimensions of $\mathbf{V}$ and $\mathbf{U}$ that the following holds: T is $1-1$; T is onto; T is $1-1$ and onto.
5. $A, B \in \mathbb{F}^{n \times n}$ are called similar if $B = Q^{-1}AQ$ for some $Q \in \mathrm{GL}(n, \mathbb{F})$. Show that similarity is an equivalence relation on $\mathbb{F}^{n \times n}$.
6. $A, B \in \mathbb{F}^{n \times n}$ are similar if and only if they represent the same linear transformation $T \in \mathrm{L}(\mathbf{V})$ in different bases, for a given n-dimensional vector space $\mathbf{V}$ over $\mathbb{F}$.

7. If $A, B \in \mathbb{F}^{n\times n}$ are similar then A and B have the same characteristic polynomial. Give an example where the opposite claim does not hold.

8. Suppose that $A, B \in \mathbb{F}^n$ have the same characteristic polynomial, which has n distinct roots in $\mathbb{F}$. Show that A and B are similar.

2 Jordan canonical form

2.1 Eigenvalues and eigenvectors

Let $A \in \mathbb{F}^{n\times n}$. $p(z) := \det(zI_n - A)$ is called the characteristic polynomial of A.

Lemma 2.1 *Let $A \in \mathbb{F}^{n\times n}$. Then $\det(zI_n - A) = z^n + \sum_{i=1}^n a_i z^{n-i}$. $(-1)^i a_i$ is the sum of all $i \times i$ principle minors of A. Assume that $\det(zI_n - A) = (z - z_1)\dots(z - z_n)$, where $z_1, \dots, z_n \in \mathbb{F}^n$. Then $(-1)^i a_i = \sigma_i(\mathbf{z})$ for $i = 1, \dots, n$.*

Proof. Consider $\det(zI - A)$. To obtain the coefficient of z^{n-i} we need to take the product of some $n-i$ diagonal elements of $zI_n - A$: $(z - a_{j_1 j_1})\dots(z - a_{j_{n-i} j_{n-i}})$. We take z^{n-i} in this product. Then this product is multiplied by the $\det(-A[\boldsymbol{\alpha}, \boldsymbol{\alpha}])$, where $\boldsymbol{\alpha}$ is the complement of $\{j_1, \dots, j_{n-i}\}$ in the set $[n]$. This shows that $(-1)^i a_i$ is the sum of all principal minors of A of order i.

Suppose that $\det(zI_n - A)$ splits to linear factors in $\mathbb{F}[z]$. Then (1.27) implies that $(-1)^i a_i = \sigma_i(\mathbf{z})$. □

Corollary 2.2 *Let $A \in \mathbb{F}^{n\times n}$ and assume that $\det(zI_n - A) = \prod_{i=1}^n (z - z_i)$. Then*

$$\operatorname{tr} A := \sum_{i=1}^n a_{ii} = \sum_{i=1}^n z_i, \quad \det A = \prod_{i=1}^n z_i.$$

Definition 2.3 *Let $\mathrm{GL}(n, \mathbb{F}) \subset \mathbb{F}^{n\times n}$ denote the set (group) of all $n \times n$ invertible matrices with entries in a given field $\mathbb{F}$. $A, B \in \mathbb{F}^{n\times n}$ are called* similar*, and this is denoted by $A \sim B$, if $B = UAU^{-1}$ for some $U \in \mathrm{GL}(n, \mathbb{F})$. The set of all $B \in \mathbb{F}^{n\times n}$ similar to a fixed $A \in \mathbb{F}^{n\times n}$ is called the similarity class corresponding to A, or simply a similarity class.*

The following proposition is straightforward:

Proposition 2.4 *Let $\mathbb{F}$ be a field, ($\mathbb{F} = \mathbb{R}, \mathbb{C}$). Then the similarity relation on $\mathbb{F}^{n\times n}$ is an equivalence relation:*

$$A \sim A,\ A \sim B \iff B \sim A,\ A \sim B \text{ and } B \sim C \Rightarrow A \sim B.$$

Furthermore if $B = UAU^{-1}$ then

1. $$\det(zI_n - B) = \det(U(zI_n - A)U^{-1}) = \det U \det(zI_n - A) \det(U^{-1}) = \det(zI_n - A),$$

   i.e. A and B have the same characteristic polynomial.

2. *For any integer $m \geq 2$ $B^m = UA^mU^{-1}$.*

3. *If in addition A is invertible, then B is invertible and $B^m = UA^mU^{-1}$ for any integer m.*

Corollary 2.5 *Let $\mathbf{V}$ be n-dimensional vector space over $\mathbb{F}$. Assume that $T : \mathbf{V} \to \mathbf{V}$ is a linear transformation. Then the set of all representation matrices of T in different bases $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a similarity class. (Use (1.29) and (1.30), where $m = n$, $\mathbf{x}_i = \mathbf{y}_i, i = 1, \ldots, n, X = Y$.) Hence, the characteristic polynomial of T is defined as $\det(zI_n - A) = z^n + \sum_{i=0}^{n-1} a_i z^i$, where A is the representation matrix of T in any basis $[\mathbf{u}_1, \ldots, \mathbf{u}_n]$, and this definition is independent of the choice of a basis. In particular $\det T := \det A$, and $\operatorname{trace} T^m = \operatorname{trace} A^m$ for any nonnegative integer. (T^0 is the identity operator, i.e $T^0\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathbf{V}$, and $A^0 = I$. Here by the trace of $B \in \mathbb{F}^{n\times n}$, denoted by $\operatorname{trace} B$, we mean the sum of all diagonal elements of B.)*

Problem 2.6 *(The representation problem.) Let $\mathbf{V}$ be n-dimensional vector space over $\mathbb{F}$. Assume that $T : \mathbf{V} \to \mathbf{V}$ is a linear transformation. Find a basis $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ in which T has the simplest form. Equivalently, given $A \in \mathbb{F}^{n\times n}$ find $B \sim A$ of the simplest form.*

In the following case the answer is well known. Recall that $\mathbf{v} \in \mathbf{V}$ is called an *eigenvector* of T corresponding to the *eigenvalue* $\lambda \in \mathbb{F}$, if $\mathbf{v} \neq \mathbf{0}$ and $T\mathbf{v} = \lambda\mathbf{v}$. This is equivalent to the existence $\mathbf{0} \neq \mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Hence $(\lambda I - A)\mathbf{x} = \mathbf{0}$ which implies that $\det(\lambda I - A) = 0$. Hence λ is the zero of the characteristic polynomial of A and T. The assumption λ is a zero of the characteristic polynomial yields that the system $(\lambda I - A)\mathbf{x}$ has a nontrivial solution $\mathbf{x} \neq \mathbf{0}$.

Corollary 2.7 *Let $A \in \mathbb{F}^{n\times n}$. Then λ is an eigenvalue of A if and only if λ is a zero of the characteristic polynomial of A: $\det(zI - A)$. Let $\mathbf{V}$ be n-dimensional vector space over $\mathbb{F}$. Assume that $T : \mathbf{V} \to \mathbf{V}$ is a linear transformation. Then λ is an eigenvalue of T if and only if λ is a zero of the characteristic polynomial of T.*

Proposition 2.8 *Let $\mathbf{V}$ be n-dimensional vector space over $\mathbb{F}$. Assume that $T : \mathbf{V} \to \mathbf{V}$ is a linear transformation. Then there exists a basis in V such that T is represented in this basis dy a diagonal matrix*

$$\operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) := \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix},$$

if and only if the characteristic polynomial of T is $(z - \lambda_1)(z - \lambda_2)\ldots(z - \lambda_n)$, and $\mathbf{V}$ has a basis consisting of eigenvectors of T.

Equivalently, $A \in \mathbb{F}^{n\times n}$ is similar to a diagonal matrix $\operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ if and only if $\det(zI - A) = (z - \lambda_1)(z - \lambda_2)\ldots(z - \lambda_n)$, and A has n-linearly independent eigenvectors.

Proof. Assume that there exists a basis $[\mathbf{u}_1, \ldots, \mathbf{u}_n]$ in V such that T is represented in this basis dy a diagonal matrix $\Lambda := \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then the characteristic polynomial of T is $\det(zI - \Lambda) = \prod_{i=1}^{n}(z - \lambda_i)$. From the definition

of the representation matrix of T, it follows that $T\mathbf{u}_i = \lambda_i \mathbf{u}_i$ for $i = 1, \ldots, n$. Since each $\mathbf{u}_i \neq \mathbf{0}$, we deduce that each $\mathbf{u}_i$ is an eigenvector of T. By our assumption $\mathbf{u}_1, \ldots, \mathbf{u}_n$ for a basis in $\mathbf{V}$.

Assume now that $\mathbf{V}$ has a basis $[\mathbf{u}_1, \ldots, \mathbf{u}_n]$ consisting eigenvectors of T. So $T\mathbf{u}_i = \lambda_i \mathbf{u}_i$ for $i = 1, \ldots, n$. Hence Λ is the representation matrix of T in the basis $[\mathbf{u}_1, \ldots, \mathbf{u}_n]$.

To prove the corresponding results for $A \in \mathbb{F}^{n \times n}$, let $\mathbf{V} := \mathbb{F}^n$ and define the linear operator $T\mathbf{x} := A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$. □

Lemma 2.9 *Let $A \in \mathbb{F}^{n \times n}$ and assume that $\mathbf{x}_1, \ldots, \mathbf{x}_k$ be k eigenvectors corresponding to k eigenvalues $\lambda_1, \ldots, \lambda_k$. Suppose that $\lambda_i \neq \lambda_j$ for $i \neq j$. Then $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly independent.*

Proof. We prove by induction on k. For $k = 1$ $\mathbf{x}_1 \neq \mathbf{0}$, hence $\mathbf{x}_1$ is linearly independent. Assume that the lemma holds for $k = m - 1$. Suppose that $k = m$. Assume that $\sum_{i=1}^m a_i \mathbf{x}_i = \mathbf{0}$. So

$$\mathbf{0} = A\mathbf{0} = A \sum_{i=1}^m a_i \mathbf{x}_i = \sum_{i=1}^m a_i A\mathbf{x}_i = \sum_{i=1}^m a_i \lambda_i \mathbf{x}_i.$$

Multiply the equality $\sum_{i=1}^m a_i \mathbf{x}_i = \mathbf{0}$ by λ_m and subtract if from the above inequality to deduce that $\sum_{i=1}^{m-1} a_i(\lambda_i - \lambda_m)\mathbf{x}_i = \mathbf{0}$. Since $\mathbf{x}_1, \ldots, \mathbf{x}_{m-1}$ are linearly independent by the induction hypothesis, we deduce that $a_i(\lambda_i - \lambda_m) = 0$ for $i = 1, \ldots, m-1$. As $\lambda_i - \lambda_m \neq 0$ for $i < m$ we get that $a_i = 0$ for $i = 1, \ldots, m-1$. The assumption that $\sum_{i=1}^m a_i \mathbf{x}_i = \mathbf{0}$ yields that $a_m \mathbf{x}_m = \mathbf{0}$. Since $\mathbf{x}_m \neq \mathbf{0}$ we obtain that $a_m = 0$. Hence $a_1 = \ldots = a_m = 0$. □

Theorem 2.10 *Let $\mathbf{V}$ be n-dimensional vector space over $\mathbb{F}$. Assume that $T : \mathbf{V} \to \mathbf{V}$ is a linear transformation. Assume that the characteristic polynomial of T $p(z)$ has n distinct roots over $\mathbb{F}$, i.e. $p(z) = \prod_{i=1}^n (z - \lambda_i)$ where $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$, and $\lambda_i \neq \lambda_j$ for each $i \neq j$. Then there exists a basis in $\mathbf{V}$ in which T is represented by a diagonal matrix.*

Similarly, let $A \in \mathbb{F}^{n \times n}$ and assume that $\det(zI - A)$ has n distinct roots in $\mathbb{F}$. Then A is similar to a diagonal matrix.

Proof. It is enough to consider the case of the linear transformation T. Recall that each root of the characteristic polynomial of T is an eigenvalue of T (Corollary 2.7). Hence to each λ_i corresponds an eigenvector $\mathbf{u}_i$: $T\mathbf{u}_i = \lambda_i \mathbf{u}_i$. Then the proof of the theorem follows Lemma 2.9 and Proposition 2.8. □

Given $A \in \mathbb{F}^{n \times n}$ it may happen that $\det(zI - A)$ does not have n roots in $\mathbb{F}$. (See for example Problem 2 of this section.) Hence we can not *diagonalize* A, i.e. A is not similar to a diagonal matrix. If $\mathbb{F}$ is *algebraically closed*, i.e. any $\det(zI - A)$ has n roots in $\mathbb{F}$ we can apply Proposition 2.8 in general and Theorem 2.10 in particular to see if A is diagonable.

Since $\mathbb{R}$ is not algebraically closed and $\mathbb{C}$ is, that is the reason that we sometimes we view a real valued matrix $A \in \mathbb{R}^{n \times n}$ as a complex valued matrix $A \in \mathbb{C}^{n \times n}$. (See Problem 2 of this section.)

Corollary 2.11 *Let $A \in \mathbb{C}^{n\times n}$ be nondiagonable. Then its characteristic polynomial must have a multiple root.*

Definition 2.12 *1. Let k be a positive integer and $\lambda \in \mathbb{F}$. Then $J_k(\lambda) \in \mathbb{F}^{k\times k}$ be a $k \times k$ be an upper diagonal matrix, with λ on the main diagonal, 1 on the next sub-diagonal and other entries are equal to 0 for $k > 1$:*

$$J_k(\lambda) := \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{bmatrix},$$

($J_1(\lambda) = [\lambda]$.)

2. Let $A_i \in \mathbb{F}^{n_i\times n_i}$ for $i = 1, \dots, l$. Denote by

$$\oplus_{i=1}^k A_i = A_1 \oplus A_2 \oplus \dots \oplus A_k = \operatorname{diag}(A_1, A_2, \dots, A_k) := \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & A_k \end{bmatrix} \in \mathbb{F}^{n\times n}, \quad n = n_1 + n_2 + \dots + n_k,$$

the $n \times n$ block diagonal matrix, whose blocks are $A_1, A_2, \dots, A_k$.

Theorem 2.13 ***(The Jordan Canonical Form)*** *Let $A \in \mathbb{C}^{n\times n}$, ($A \in \mathbb{F}^{n\times n}$, where $\mathbb{F}$ is an algebraically closed field.) Then A is similar to its Jordan canonical form $\oplus_{i=1}^k J_{n_i}(\lambda_i)$ for some $\lambda_1, \dots, \lambda_k \in \mathbb{C}$, ($\lambda_1, \dots, \lambda_k \in \mathbb{F}$), and positive integers $n_1, \dots, n_k$. The Jordan canonical form is unique up to the permutations of the Jordan blocks $J_{n_1}(\lambda_1), \dots, J_{n_k}(\lambda_k)$.*

Equivalently, let $T : \mathbf{V} \to \mathbf{V}$ be a linear transformation of an n-dimensional space over $\mathbb{C}$, or any other algebraically closed field. Then there exists a basis in $\mathbf{V}$, such that $\oplus_{i=1}^k J_{n_i}(\lambda_i)$ is the representation matrix of T in this basis. The blocks $J_{n_i}(\lambda_i), i = 1, \dots, k$ are unique.

Note that $A \in \mathbb{C}^{n\times n}$ is diagonable if and only in its Jordan canonical form $k = n$, i.e. $n_1 = \dots = n_n = 1$. For $k < n$, the *Jordan canonical form* is the simplest form of the similarity class of a nondiagonable $A \in \mathbb{C}^{n\times n}$.

We will prove Theorem 2.13 in the next several sections.

Problems

1. Let $\mathbf{V}$ be a vector space over $\mathbb{F}$. (You may assume that $\mathbb{F} = \mathbb{C}$.) Let $T : \mathbf{V} \to \mathbf{V}$ be a linear transformation. Suppose that $\mathbf{u}_i$ is an eigenvector of T with the corresponding eigenvalue λ_i for $i = 1, \dots, m$. Show by induction on m that if $\lambda_1, \dots, \lambda_m$ are m distinct scalars then $\mathbf{u}_1, \dots, \mathbf{u}_m$ are linearly independent.
2. Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}$.

(a) Show that A is not diagonable over the real numbers $\mathbb{R}$.

(b) Show that A is diagonable over the complex numbers $\mathbb{C}$. Find $U \in C^{2\times 2}$ and a diagonal $\Lambda \in \mathbb{C}^{2\times 2}$ such that $A = U\Lambda U^{-1}$.

3. Let $A = \oplus_{i=1}^k J_{n_i}(\lambda_i)$. Show that $\det(zI - A) = \prod_{i=1}^k (z - \lambda_i)^{n_i}$. (You may use the fact that the determinant of an upper triangular matrix is the product of its diagonal entries.)

4. Let $A = \oplus_{i=1}^k A_i$ where $A_i \in \mathbb{C}^{n_i \times n_i}, i = 1, \ldots, k$. Show that $\det(zI_n - A) = \prod_{i=1}^k \det(zI_{n_i} - A_i)$. (First show the identity for $k = 2$ using the determinant expansion by rows. Then use induction for $k > 2$.

5. (a) Show that any eigenvector of $J_n(\lambda) \in \mathbb{C}^{n\times n}$ is in the subspace spanned by $\mathbf{e}_1$. Conclude that $J_n(\lambda)$ is not diagonable unless $n = 1$.

(b) What is the rank of $zI_n - J_n(\lambda)$ for a fixed $\lambda \in \mathbb{C}$ and for each $z \in \mathbb{C}$?

(c) What is the rank of $zI - \oplus_{i=1}^k J_{n_i}(\lambda_i)$ for fixed $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ and for each $z \in \mathbb{C}$?

6. Let $A \in \mathbb{C}^{n\times n}$ and assume that $\det(zI_n - A) = z^n + a_1 z^{n-1} + \ldots + a_{n-1}z + a_n$ has n distinct complex roots. Show that $A^n + a_1 A^{n-1} + \ldots a_{n-1}A + a_n I_n = \mathbf{0}$, where $\mathbf{0} \in \mathbb{C}^{n\times n}$ denotes the zero matrix, i.e. the matrix whose all entries are 0. (This is a special case of the Cayley-Hamilton theorem, which claims that the above identity holds for *any* $A \in \mathbb{C}^{n\times n}$.) **Hint:** Use the fact that A is diagonable.

2.2 Matrix polynomials

Let $P(z) = (p_{ij}(z))_{i=j=1}^{m,n}$ be an $m \times n$ matrix whose entries are polynomials in $\mathbb{F}[z]$. The set of all such $m \times n$ matrices is denoted by $\mathbb{F}[z]^{m\times n}$. Clearly $\mathbb{F}[z]^{m\times n}$ is a vector space over $\mathbb{F}$, of infinite dimension. Given $p(z) \in \mathbb{F}[z]$ and $P(z) \in \mathbb{F}[z]^{m\times n}$ one can define $p(z)P(z) := (p(z)p_{ij}) \in \mathbb{F}[z]$. Again, this product satisfies nice distribution properties. Thus $\mathbb{F}[z]$ is a *module* over the ring $\mathbb{F}[z]$. (Note $\mathbb{F}[z]$ is not a field!)

Let $P(z) = (p_{ij}(z)) \in \mathbb{F}[z]^{m\times n}$. Then $\deg P(z) := \max_{i,j} \deg p_{ij}(z) = l$. Write

$$p_{ij}(z) = \sum_{k=0}^{l} p_{ij,k} z^{l-k}, \; P_k := (p_{ij,k})_{i=j=1}^{m,n} \in \mathbb{F}^{m\times n} \text{ for } k = 0, \ldots, l.$$

Then

$$P(z) = P_0 z^l + P_1 z^{l-1} + \ldots + P_l, \quad P_i \in \mathbb{F}^{m\times n}, \; i = 0, \ldots, l, \tag{2.1}$$

is a matrix polynomial with coefficients in $\mathbb{F}^{m\times n}$.

Assume that $P(z), Q(z) \in \mathbb{F}[z]^{n\times n}$. Then we can define $P(z)Q(z) \in \mathbb{F}[z]$. Note that in general $P(z)Q(z) \neq Q(z)P(z)$. Hence $\mathbb{F}[z]^{n\times n}$ is a *noncommutative* ring. For $P(z) \in \mathbb{F}^{n\times n}$ of the form (2.1) and any $A \in \mathbb{F}^{n\times n}$ we define

$$P(A) = \sum_{i=0}^{l} P_i A^{l-i} = P_0 A^l + P_1 A^{l-1} + \ldots + P_l, \text{ where } A^0 = I_n.$$

Recall that given two polynomials $p, q \in \mathbb{F}[z]$ one can divide p by $q \not\equiv 0$ with the residue r, i.e. $p = tq + r$ for some unique $t, r \in \mathbb{F}[z]$, where $\deg r < \deg q$. One can trivially generalize that to polynomial matrices:

Proposition 2.14 *Let $p(z), q(z) \in \mathbb{F}[z]$ and assume that $q(z) \not\equiv 0$. Let $p(z) = t(z)q(z) + r(z)$, where $t(z), r(z) \in \mathbb{F}[z]$ are unique polynomials with $\deg r(z) < \deg q(z)$. Let $n > 1$ be an integer, and define the following scalar polynomials: $P(z) := p(z)I_n, Q(z) := q(z)I_n, T(z) := t(z)I_n, R(z) := r(z)I_n \in \mathbb{F}[z]^{n\times n}$. Then $P(A) = T(A)Q(A) + R(A)$ for any $A \in \mathbb{F}^{n\times n}$.*

Proof. Since $A^iA^j = A^{i+j}$ for any nonnegative integer, with $A^0 = I_n$, the equality $P(A) = T(A)Q(A) + R(A)$ follows trivially from the equality $p(z) = t(z)q(z) + r(z)$. □

Recall that p is divisible by q, denoted as $q|p$, if $p = tq$, i.e. r is the zero polynomial. Note that if $q(z) = (z - a)$ then $p(z) = t(z)(z - a) + p(a)$. Thus $(z-a)|p$ if and only if $p(a) = 0$. Similar results hold for square polynomial matrices, which are not scalar.

Lemma 2.15 *Let $P(z) \in \mathbb{F}[z]^{n\times n}, A \in \mathbb{F}^{n\times n}$ Then there exists a unique $T_{left}(z)$, of degree $degP - 1$ if $\deg P > 0$ or degree $-\infty$ if $\deg P \le 0$, such that*

$$P(z) = T_{left}(z)(zI - A) + P(A). \tag{2.2}$$

In particular, $P(z)$ is divisible from the right by $zI - A$ if and only if $P(A) = 0$.

Proof. We prove the lemma by induction on $\deg P$. If $\deg P \le 0$, i.e. $P(z) = P_0 \in \mathbb{F}^{n\times n}$ then $T_{left} = \mathbf{0}, P(A) = P_0$ and the lemma trivially holds. Suppose that the lemma holds for all P with $\deg P \le l - 1$, where $l \ge 1$. Let $P(z)$ be of degree $l \ge 1$ of the form (2.1). Then $P(z) = P_0z^l + \tilde{P}(z)$, where $\tilde{P}(z) = \sum_{i=1}^{l} P_iz^{l-1}$. By the induction assumption $\tilde{P}(z) = \tilde{T}_{left}(z)(zI_n - A) + \tilde{P}(A)$, where $\tilde{T}_{left}(z)$ is unique. A straightforward calculation shows that

$$P_0z^l = \hat{T}_{left}(z)(zI_n - A) + P_0A^l, \text{ where } \hat{T}_{left}(z) = \sum_{i=0}^{l-1} P_0A^iz^{l-i-1},$$

and $\hat{T}_{left}$ is unique. Hence $T_{left}(z) = \hat{T}_{left}(z) + \tilde{T}_{left}$ is unique, $P(A) = P_0A^l + \tilde{P}(A)$ and (2.2) follows.

Suppose that $P(A) = \mathbf{0}$. Then $P(z) = T_{left}(z)(zI - A)$, i.e. $P(z)$ is divisible by $zI_n - A$ from the right. Assume that $P(z)$ is divisible by $(zI_n - A)$ from the right, i.e. there exists $T(z) \in \mathbb{F}[z]^{n\times n}$ such that $P(z) = T(z)(zI_n - A)$. Subtract (2.2) from $P(z) = T(z)(zI_n - A)$ to deduce that $\mathbf{0} = (T(z) - T_{left}(z))(zI_n - A) - P(A)$. Hence $T(z) = T_{left}(z)$ and $P(A) = 0$. □

The above lemma can generalized to any $Q(z) = Q_0z^l + Q_1z^{l-1} + \ldots + Q_l \in \mathbb{F}[z]$, where $Q_0 \in \mathrm{GL}(n, \mathbb{F})$: There exists unique $T_{left}(z), R_{left}(z) \in \mathbb{F}[z]$ such that

$$P(z) = T_{left}(z)Q(z) + R_{left}(z),\ \deg R_{left} < \deg Q,\ Q(z) = \sum_{i=0}^{l} Q_iz^{l-i},\ Q_0 \in \mathrm{GL}(n, \mathbb{F}). \tag{2.3}$$

Here we agree that $(Az^i)(Bz^j) = (AB)z^{i+j}$ for any $A, B \in \mathbb{F}^{n\times n}$ and nonnegative integers i, j.

Theorem 2.16 *(Cayley-Hamilton theorem.) Let $A \in \mathbb{F}^{n\times n}$ and $p(z) = \det(zI_n - A)$ be the characteristic polynomial of A. Let $P(z) = p(z)I_n \in \mathbb{F}[z]^{n\times n}$. Then $P(A) = \mathbf{0}$.*

Proof. Let $A(z) = zI_n - A$. Fix $z \in \mathbb{F}$ and let $B(z) = (b_{ij}(z))$ be the adjoint matrix of $A(z)$, whose entries are the cofactors of $A(z)$. That is $b_{ij}(z)$ is $(-1)^{i+j}$ times the determinant of the matrix obtained from $A(z)$ by deleting row j and column i. If one views z as indeterminate then $B(z) \in \mathbb{F}[z]^{n\times n}$. Recall the identity

$$A(z)B(z) = B(z)A(z) = \det A(z)I_n = p(z)I_n = P(z).$$

Hence $(zI_n - A)$ divides from the right $P(z)$. Lemma 2.15 yields that $P(A) = \mathbf{0}$. □

For $p, q \in \mathbb{F}[z]$ let (p, q) be the *greatest common divisor* of p, q. If p and q are identically zero then (p, q) is the zero polynomial. Otherwise (p, q) is a polynomial s of the highest degree that divides p and q. s is determined up to a multiple of a nonzero scalar. s can be chosen as a unique *monic* polynomial:

$$s(z) = z^l + s_1 z^{l-1} + \ldots + s_l \in \mathbb{F}[z]. \tag{2.4}$$

Equality (1.24) yields.

Corollary 2.17 *Let $p, q \in \mathbb{F}[z]$ be coprime. Then there exists $u, v \in \mathbb{F}[z]$ such that $1 = up + vq$. Let $n > 1$ be an integer and define $P(z) := p(z)I_n, Q(z) := q(z)I_n, U(z) := u(z)I_n, V(z) := u(z)I_n \in \mathbb{F}[z]^{n\times n}$. Then for any $A \in \mathbb{F}^{n\times n}$ we have the identity $I_n = U(A)P(A) + V(A)Q(A)$, where $U(A)P(A) = P(A)U(A)$ and $V(A)Q(A) = Q(A)V(A)$.*

Let us consider that case where $p, q \in \mathbb{F}[z]$ are both nonzero polynomials that split (to linear factors) over $\mathbb{F}$. So

$$p(z) = p_0(z - \alpha_1)\ldots(z - \alpha_i), p_0 \neq 0, \quad q(z) = q_0(z - \beta_1)\ldots(z - \beta_j), q_0 \neq 0.$$

In that case $(p, q) = 1$, if p and q do not have a common root. If p and q have a common zero then (p, q) is is a nonzero polynomial that has the maximal number of common roots of p and q counting with multiplicities.

From now on for any $p \in \mathbb{F}[z]$ and $A \in \mathbb{F}^{n\times n}$ we identify $p(A)$ with $P(A)$, where $P(z) = p(z)I_n$.

2.3 Minimal polynomial and decomposition to invariant subspaces

Recall that $\mathbb{F}^{n\times n}$ is a vector space over $\mathbb{F}$ of dimension n^2. Let $A \in \mathbb{F}^{n\times n}$ and consider the powers $A^0 = I_n, A, A^2, \ldots, A^m$. Let m be the smallest positive integer such that these $m+1$ matrices are linearly dependent as vectors in $\mathbb{F}^{n\times n}$. (*Note* that $A^0 \neq \mathbf{0}$.) So $\sum_{i=0}^{m} b_i A^{m-i} = 0$, and $(b_0, \ldots, b_m)^\top \neq \mathbf{0}$. If $b_0 = 0$ then $A^0, \ldots, A^{m-1}$ are linearly dependent, which contradicts the definition of m. Hence $b_0 \neq 0$. Divide the linear dependence by b_0 to obtain.

$$\psi(A) = 0, \quad \psi(z) = z^m + \sum_{i=1}^{m} a_i z^{m-i} \in \mathbb{F}[z], \quad a_i = \frac{b_i}{b_0} \text{ for } i = 1, \ldots, m. \tag{2.5}$$

ψ is called the *minimal polynomial* of A. In principle $m \leq n^2$, but in reality $m \leq n$:

Theorem 2.18 *Let $A \in \mathbb{F}^{n\times n}$ and $\psi(z)$ be its minimal polynomial. Assume that $p(z) \in \mathbb{F}[z]$ is an annihilated polynomial of A, i.e. $p(A) = 0_{n\times n}$. Then ψ divides p. In particular, the characteristic polynomial $p(z) = \det(zI_n - A)$ is divisible by $\psi(z)$. Hence $\deg \psi \le \deg p = n$.*

Proof. Divide the annihilating polynomial p by ψ to obtain $p(z) = t(z)\psi(z) + r(z)$, where $\deg r < \deg \psi = m$. Proposition 2.14 yields that $p(A) = t(A)\psi(A)+r(A)$ which implies that $r(A) = 0$. Assume that $l = \deg r(z) \ge 0$, i.e. r is not identically the zero polynomial. So $A^0, \dots, A^l$ are linearly dependent, which contradicts the definition of m. Hence $r(z) \equiv 0$.

The Cayley-Hamilton theorem yields that the characteristic polynomial $p(z)$ of A annihilates A. Hence $\psi | p$ and $\deg \psi \le \deg p = n$. □

Theorem 2.19 *Let $A \in \mathbb{F}^{n\times n}$. Denote by $q(z)$ the g.c.d., (monic polynomial), of all the entries of* adj $(zI_n - A) \in \mathbb{F}[z]^{n\times n}$*, which is the g.c.d of all $(n-1)\times(n-1)$ minors of $(zI_n - A)$. Then $q(z)$ divides the characteristic polynomial $p(z) = \det(zI_n - A)$ of A. Furthermore, the minimal polynomial of A is equal to $\frac{p(z)}{q(z)}$.*

Proof. Expand $\det(zI_n - A)$ by the first column. Since $q(z)$ divides each $(n-1)\times(n-1)$ minor of $zI_n - A$ we deduce that $q(z)$ divides $p(z)$. Let $\phi(z) = \frac{p(z)}{q(z)}$. As $q(z)$ divides each entry of adj $(zI_n - A)$ we deduce that adj $(zI_n - A) = q(z)C(z)$ for some $C(z) \in \mathbb{F}[z]^{n\times n}$. Divide the equality $p(z)I_n =$ adj $(zI_n - A)(zI_n - A)$ by $q(z)$ to deduce that $\phi(z)I_n = C(z)(zI_n - A)$. Lemma 2.15 yields that $\phi(A) = 0$.

Let $\psi(z)$ be the minimal polynomial of A. Theorem 2.18 yields that ψ divides ϕ. We now show that ϕ divides ψ. Theorem 2.18 implies that $p(z) = s(z)\psi(z)$ for some monic polynomial $s(z)$. Since $\psi(A) = 0$ Lemma 2.15 yields that $\psi(z)I_n = D(z)(zI_n - A)$ for some $D(z) \in \mathbb{F}^{n\times n}[z]$. So $p(z)I_n = s(z)\psi(z)I_n = s(z)D(z)(zI_n - A)$. Since $p(z)I_n =$ adj $(zI_n - A)(zI_n - A)$ we deduce that $s(z)D(z) =$ adj $(zI_n - A)$. As all the entries of $D(z)$ are polynomials, it follows that $s(z)$ divides all the entries of adj $(zI_n - A)$. Since $q(z)$ is the g.c.d. of all entries of adj $(zI_n - A)$ we deduce that $s(z)$ divides $q(z)$. Consider the equality $p(z) = s(z)\psi(z) = q(z)\phi(z)$. Thus $\psi(z) = \frac{q(z)}{s(z)}\phi(z)$. Hence $\phi(z)$ divides $\psi(z)$. As $\psi(z)$ and $\phi(z)$ are monic we deduce that $\psi(z) = \phi(z)$. □

Proposition 2.20 *Let $A \in \mathbb{F}^{n\times n}$ and assume that $\lambda \in \mathbb{F}$ is an eigenvalue of A with the corresponding eigenvector $\mathbf{x} \in \mathbb{F}^n$. Then for any $h(z) \in \mathbb{F}[z]$, $h(A)\mathbf{x} = h(\lambda)\mathbf{x}$. In particular, λ is a root of the minimal polynomial $\psi(z)$ of A, i.e. $\psi(\lambda) = 0$.*

Proof. Clearly, $A^m\mathbf{x} = \lambda^m\mathbf{x}$. Hence, $h(A)\mathbf{x} = h(\lambda)\mathbf{x}$. Assume that $h(A) = 0$. As $\mathbf{x} \neq \mathbf{0}$ we deduce that $h(\lambda) = 0$. Hence $\psi(\lambda) = 0$. □

Definition 2.21 *A matrix $A \in \mathbb{F}^{n\times n}$ is called nonderogatory if the minimal polynomial of A is equal to its characteristic polynomial.*

Definition 2.22 *Let $\mathbf{V}$ be a finite dimensional vector space over $\mathbb{F}$, and assume that $\mathbf{V}_1, \dots, \mathbf{V}_i$ nonzero subspaces of $\mathbf{V}$. Then $\mathbf{V}$ is a direct sum of $\mathbf{V}_1, \dots, \mathbf{V}_i$,*

denoted as $\mathbf{V} = \oplus_{j=1}^{i} \mathbf{V}_j$ *if any vector* $\mathbf{v} \in \mathbf{V}$ *has a unique representation as* $\mathbf{v} = \mathbf{v}_1 + \ldots + \mathbf{v}_i$, *where* $\mathbf{v}_j \in \mathbf{V}_j$ *for* $j = 1, \ldots, i$. *Equivalently, let* $[\mathbf{v}_{j1}, \ldots, \mathbf{v}_{jl_j}]$ *be a basis of* $\mathbf{V}_j$ *for* $j = 1, \ldots, i$. *Then* $\dim \mathbf{V} = \sum_{j=1}^{i} \dim \mathbf{V}_j = \sum_{j=1}^{i} l_j$ *and the* $\dim \mathbf{V}$ *vectors* $\mathbf{v}_{11}, \ldots, \mathbf{v}_{1l_1}, \ldots, \mathbf{v}_{i1}, \ldots, \mathbf{v}_{il_i}$ *are linearly independent.*

Let $T : \mathbf{V} \to \mathbf{V}$ be a linear operator. A subspace $\mathbf{U}$ of $\mathbf{V}$ is called a *T-invariant subspace*, or simply an invariant subspace when there is no ambiguity about T, if $T\mathbf{u} \in \mathbf{U}$ for each $\mathbf{u} \in \mathbf{U}$. We denote this fact by $T\mathbf{U} \subseteq \mathbf{U}$. Denote by $T|_{\mathbf{U}}$ the restriction of T to the invariant subspace of T. Clearly, $T|_{\mathbf{U}}$ is a linear operator on $\mathbf{U}$.

Note $\mathbf{V}$ and the zero subspace $\{\mathbf{0}\}$, (which consist only of the zero element), are invariant subspaces. Those are called *trivial* invariant subspaces. $\mathbf{U}$ is called a *nontrivial* invariant subspace if $\mathbf{U}$ is an invariant subspace such that $0 < \dim \mathbf{U} < \dim \mathbf{V}$.

Since the representation matrices of T in different bases form a similarity class we can define the *minimal polynomial* $\psi(z) \in \mathbb{F}[z]$ of T, as the minimal polynomial of any representation matrix of T. (See Problem 1 in the end of this section.) Equivalently $\psi(z)$ is the monic polynomial of the minimal degree which annihilates T: $\psi(T) = \mathbf{0}$.

Theorem 2.23 *Let* $T : \mathbf{V} \to \mathbf{V}$ *be a linear operator on a finite dimensional space* $\dim \mathbf{V} > 0$. *Let* $\psi(z)$ *be the minimal polynomial of* T. *Assume that* $\psi(z)$ *decomposes to* $\psi(z) = \psi_1(z) \ldots \psi_k(z)$, *where each* $\psi_i(z)$ *is a monic polynomial of degree at least* 1. *Suppose furthermore that for each pair* $i \neq j$ $\psi_i(z)$ *and* $\psi_j(z)$ *are coprime. Then* $\mathbf{V}$ *is a direct sum of* $\mathbf{V}_1, \ldots, \mathbf{V}_k$, *where each* $\mathbf{V}_i$ *is a nontrivial invariant subspace of* T. *Furthermore the minimal polynomial of* $T|_{\mathbf{V}_i}$ *is equal to* $\psi_i(z)$ *for* $i = 1, \ldots, k$. *Moreover, each* $\mathbf{V}_i$ *is uniquely determined by* $\psi_i(z)$ *for* $i = 1, \ldots, k$.

Proof. We prove the theorem by induction on $k \geq 2$. Let $k = 2$. So $\psi(z) = \psi_1(z)\psi_2(z)$. Let $\mathbf{V}_1 := \psi_2(T)\mathbf{V}, \mathbf{V}_2 = \psi_1(T)\mathbf{V}$ be the ranges of the operators $\psi_2(T), \psi_1(T)$ respectively. Observe that

$$T\mathbf{V}_1 = T(\psi_2(T)\mathbf{V}) = (T\psi_2(T))\mathbf{V} = (\psi_2(T)T)\mathbf{V} = \psi_2(T)(T\mathbf{V}) \subseteq \psi_2(T)\mathbf{V} = \mathbf{V}_1.$$

Thus $\mathbf{V}_1$ is a T-invariant subspace. Assume that $\mathbf{V}_1 = \{\mathbf{0}\}$. This is equivalent to that $\psi_2(T) = 0$. By Theorem 2.18 ψ divides ψ_2 which is impossible since $\deg \psi = \deg \psi_1 + \deg \psi_2 > \deg \psi_1$. Thus $\dim \mathbf{V}_1 > 0$. Similarly $\mathbf{V}_2$ is a nonzero T-invariant subspace. Let $T_i = T|_{\mathbf{V}_i}$ for $i = 1, 2$. Clearly

$$\psi_1(T_1)\mathbf{V}_1 = \psi_1(T)\mathbf{V}_1 = \psi_1(T)(\psi_2(T)\mathbf{V}) = (\psi_1(T)\psi_2(T))\mathbf{V} = \{\mathbf{0}\},$$

since ψ is the minimal polynomial of T. Hence $\psi_1(T_1) = \mathbf{0}$, i.e. ψ_1 is an annihilating polynomial of T_1. Similarly, $\psi_2(T_2) = \mathbf{0}$.

Let $\mathbf{U} = \mathbf{V}_1 \cap \mathbf{V}_2$. Then $\mathbf{U}$ is an invariant subspace of T. We claim that $\mathbf{U} = \{\mathbf{0}\}$, i.e. $\dim \mathbf{U} = 0$. Assume to the contrary that $\dim \mathbf{U} \geq 1$. Let $Q := T|_{\mathbf{U}}$ and denote by $\phi \in \mathbb{F}[z]$ the minimal polynomial of Q. Clearly $\deg \phi \geq 1$. Since $\mathbf{U} \subseteq \mathbf{V}_i$ it follows that ψ_i is an annihilating polynomial of Q for $i = 1, 2$. Hence $\phi | \psi_1$ and $\phi | \psi_2$, i.e. ϕ is a nontrivial factor of ψ_1 and ψ_2. This contradicts the assumption that ψ_1 and ψ_2 are coprime. Hence $\mathbf{V}_1 \cap \mathbf{V}_2 = \{\mathbf{0}\}$.

Since $(\psi_1, \psi_2) = 1$ there exists polynomials $f, g \in \mathbb{F}[z]$ such that $\psi_1 f + \psi_2 g = 1$. Hence $I = \psi_1(T)f(T) + \psi_2(T)g(T)$, where I is the identity operator $I\mathbf{v} = \mathbf{v}$ on $\mathbf{V}$. In particular for any $\mathbf{v} \in \mathbf{V}$ we have $\mathbf{v} = \mathbf{v}_2 + \mathbf{v}_1$, where $\mathbf{v}_1 = \psi_2(T)(g(T)\mathbf{v}) \in \mathbf{V}_1, \mathbf{v}_2 = \psi_1(T)(f(T)\mathbf{v}) \in \mathbf{V}_2$. Since $\mathbf{V}_1 \cap \mathbf{V}_2 = \{\mathbf{0}\}$ it follows that $\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2$. Let $\tilde{\psi}_i$ be the minimal polynomial of T_i. Then $\tilde{\psi}_i | \psi_i$ for $i = 1, 2$. Hence $\tilde{\psi}_1\tilde{\psi}_2 | \psi_1\psi_2$. Let $\mathbf{v} \in \mathbf{V}$. Then $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, where $\mathbf{v}_i \in \mathbf{V}_i, i = 1, 2$. Using the facts that $\tilde{\psi}_1(T)\tilde{\psi}_2(T) = \tilde{\psi}_2(T)\tilde{\psi}_1(T)$, $\tilde{\psi}_i$ is the minimal polynomial of T_i, and the definition of T_i we deduce

$$\tilde{\psi}_1(T)\tilde{\psi}_2(T)\mathbf{v} = \tilde{\psi}_2(T)\tilde{\psi}_1(T)\mathbf{v}_1 + \tilde{\psi}_1(T)\tilde{\psi}_2(T)\mathbf{v}_2 = \mathbf{0}.$$

Hence the monic polynomial $\theta(z) := \tilde{\psi}_1(z)\tilde{\psi}_2(z)$ is an annihilating polynomial of T. Thus $\psi(z) | \theta(z)$ which implies that $\psi(z) = \theta(z)$, hence $\tilde{\psi}_i = \tilde{\psi}$ for $i = 1, 2$.

It is left to show that $\mathbf{V}_1$ and $\mathbf{V}_2$ are unique. Let $\bar{\mathbf{V}}_i := \{\mathbf{v} \in \mathbf{V} : \psi_i(T)\mathbf{v} = \mathbf{0}\}$ for $i = 1, 2$. So $\bar{\mathbf{V}}_i$ is a subspace that contains $\mathbf{V}_i$ for $i = 1, 2$. If $\psi_i(T)\mathbf{v} = \mathbf{0}$ then

$$\psi_i(T)(T\mathbf{v}) = (\psi_i(T)T)\mathbf{v} = (T\psi_i(T)\mathbf{v}) = T(\psi_i(T)\mathbf{v}) = T\mathbf{0} = \mathbf{0}.$$

Hence $\bar{\mathbf{V}}_i$ is T-invariant subspace. We claim that $\bar{\mathbf{V}}_i = \mathbf{V}_i$. Suppose to the contrary that $\dim \bar{\mathbf{V}}_i > \dim \mathbf{V}_i$ for some $i \in \{1, 2\}$. Let $j \in \{1, 2\}$ and $j \neq i$. Then $\dim(\bar{\mathbf{V}}_i \cap \mathbf{V}_j) \geq 0$. As before we conclude that $\mathbf{U} := \bar{\mathbf{V}}_i \cap \mathbf{V}_j$ is T-invariant subspace. As above, the minimal polynomial of $T|_{\mathbf{U}}$ must divide $\psi_1(z)$ and $\psi_2(z)$, which contradicts the assumption that $(\psi_1, \psi_2) = 1$. This concludes the proof of the theorem for $k = 2$.

Assume that $k \geq 3$. Let $\hat{\psi}_2 := \psi_2 \dots \psi_k$. Then $(\psi_1, \hat{\psi}_2) = 1$ and $\psi = \psi_1\hat{\psi}_2$. Then $\mathbf{V} = \mathbf{V}_1 \oplus \hat{\mathbf{V}}_2$, where $T : \mathbf{V}_1 \to \mathbf{V}_1$, has the minimal polynomial ψ_1, and $T : \hat{\mathbf{V}}_2 \to \hat{\mathbf{V}}_2$ has the minimal polynomial $\hat{\psi}_2$. Note that $\mathbf{V}_1$ and $\hat{\mathbf{V}}_2$ are unique. Apply the induction hypothesis to $T|_{\hat{V}_2}$ to deduce the theorem. □

Problems

1. Let $A, B \in \mathbb{F}^{n\times n}$ and $p(z) \in \mathbb{F}[z]$. Show
   (a) If $B = UAU^{-1}$, for some $U \in \mathrm{GL}(n, \mathbb{F})$, then $p(B) = Up(A)U^{-1}$.
   (b) If $A \sim B$ then A and B have the same minimal polynomial.
   (c) Let $A\mathbf{x} = \lambda\mathbf{x}$. Then $p(A)\mathbf{x} = p(\lambda)\mathbf{x}$. Deduce that each eigenvalue of A is a root of the minimal polynomial of A.
   (d) Assume that A has n distinct eigenvalues. Then A is nonderogatory.
2. (a) Show that the Jordan block $J_k(\lambda) \in \mathbb{F}^{k\times k}$ is nonderogatory.
   (b) Let $\lambda_1, \dots, \lambda_k \in \mathbb{F}$ be k distinct elements. Let

$$A = \oplus_{i=j=1}^{k,l_i} J_{m_{ij}}(\lambda_i), \text{ where } m_i = m_{i1} \geq \dots \geq m_{il_i} \geq 1, \text{ for } i = 1, \dots, k. \tag{2.6}$$

   Here m_{ij} and l_i are positive integers be integers. Find the minimal polynomial of A. When A is nonderogatory?

3. Find the characteristic and the minimal polynomials of

$$C := \begin{bmatrix} 2 & 2 & -2 & 4 \\ -4 & -3 & 4 & -6 \\ 1 & 1 & -1 & 2 \\ 2 & 2 & -2 & 4 \end{bmatrix},$$

4. Let $A := \begin{bmatrix} x & y \\ u & v \end{bmatrix}$. Then A is a point in four dimensional space $\mathbb{R}^4$.

   (a) What is the condition that A has a multiple eigenvalue ($\det(zI_2 - A) = (z-\lambda)^2$) ? Conclude that the set (variety) all 2×2 matrices with a multiple eigenvalue is a quadratic hypersurface in $\mathbb{R}^4$, i.e. it satisfies a polynomial equation in (x, y, u, v) of degree 2. Hence its dimension is 3.

   (b) What is the condition that A has a multiple eigenvalue and it is a diagonable matrix, i.e. similar to a diagonal matrix? Show that this is a line in $\mathbb{R}^4$. Hence its dimension is 1.

   (c) Conclude that the set (variety) of 2×2 matrices which have multiple eigenvalues and diagonable is "much smaller" then the variety of matrices with multiple eigenvalue.
   This fact holds for any $n \times n$ matrices $\mathbb{R}^{n\times n}$ or $\mathbb{C}^{n\times n}$.

5. **Programming Problem**

   Spectrum and pseudo spectrum: Let $A = (a_{ij})_{i,j=1}^n \in \mathbb{C}^{n\times n}$. Then $\det(zI_n - A) = (z-\lambda_1)\dots(z-\lambda_n)$ and the *spectrum* of A is given as $\text{spec } A := \{\lambda_1, \dots, \lambda_n\}$. In computations, the entries of A are known or given up to a certain precision. Say, in regular precision each a_{ij} is known with precision to eight digits: $a_1.a_2 \dots a_8 \times 10^m$ for some integer m., e.g. $1.2345678 \times 10^{-12}$, in floating point notation. Thus, with a given matrix A, we associate a whole class of matrices $\mathcal{C}(A) \subset \mathbb{C}^{n\times n}$ of matrices $B \in \mathbb{C}^{n\times n}$ that are represented by A. For each $B \in \mathcal{C}(A)$ we have the spectrum spec B. Then the *pseudo spectrum* of A is the union of all the spectra of $B \in \mathcal{C}(A)$: $\text{pspec } A := \cup_{B\in\mathcal{C}(A)}\text{spec } (B)$. spec A and pspec A are subsets of the complex plane $\mathbb{C}$ and can be easily plotted by computer. The shape of pspec A gives an idea of our real knowledge of the spectrum of A, and to changes of the spectrum of A under perturbations. The purpose of this programming problems to give the student a taste of this subject.

   In all the computations use double precision.

   (a) Choose at random $A = (a_{ij}) \in \mathbb{R}^{5\times 5}$ as follows: each entry a_{ij} is chosen at random from the interval $[-1, 1]$, using uniform distribution. Find the spectrum of A and plot the eigenvalues of A on the $X-Y$ axis as complex numbers, marked say as $+$, where the center of $+$ is at each eigenvalue.

      i. For each $\epsilon = 0.1, 0.01, 0.0001, 0.000001$ do the following:
      For $i = 1, \dots, 100$ choose $B_i \in \mathbb{R}^{5\times 5}$ at random as A in the item (a) and find the spectrum of $A + \epsilon B_i$. Plot these spectra, each eigenvalue of $A + \epsilon B_i$ plotted as $\cdot$ on the $X - Y$ axis, together with the plot of the spectrum of A. (Altogether you will have 4 graphs.)

(b) Let $A := \text{diag}(0.1C, [-0.5])$, i.e. $A \in \mathbb{R}^{5\times 5}$ be a block diagonal matrix where the first 4×4 block is $0.1C$, where the matrix C is given in Problem 3 above, and the second block is 1×1 matrix with the entry -0.5. Repeat part (i) of part (a) above with this specific A. (Again you will have 4 graphs.)

(c) Repeat (a) by choosing at random a symmetric matrix $A = (a_{ij}) \in \mathbb{R}^{5\times 5}$. That is choose at random a_{ij} for $1 \le i \le j$, and let $a_{ji} = a_{ij}$ for $i < j$.

i. Repeat the part (i) of (a). (B_j are not symmetric!) You will have 4 graphs.

ii. Repeat part (i) of (a), with the restriction that each B_j is a random symmetric matrix, as explained in (c). You will have 4 graphs.

(d) Can you draw some conclusions about these numerical experiments?

2.4 Quotient of vectors space and induced linear operators

Let $\mathbf{V}$ be a finite dimensional vector space over $\mathbb{F}$. Assume that $\mathbf{U}$ is a subspace of $\mathbf{V}$. Then by $\mathbf{V}/\mathbf{U}$ we denote the set of cosets $\mathbf{x}+\mathbf{U} := \{\mathbf{y} \in \mathbf{V},\ \mathbf{y} = \mathbf{x}+\mathbf{u} \text{ for all } \mathbf{u} \in \mathbf{U}\}$. An element in $\hat{\mathbf{V}} := \mathbf{V}/\mathbf{U}$ is denoted by $\hat{\mathbf{x}} := \mathbf{x} + \mathbf{U}$.

Lemma 2.24 *Let $\mathbf{V}$ be a finite dimensional vector space over $\mathbb{F}$. Assume that $\mathbf{U}$ is a subspace of $\mathbf{V}$. Then $\mathbf{V}/\mathbf{U}$ is a finite dimensional vectors space over $\mathbb{F}$ where $(\mathbf{x}+\mathbf{U}) + (\mathbf{y}+\mathbf{U}) := (\mathbf{x}+\mathbf{y}+\mathbf{U})$ and $a(\mathbf{x}+\mathbf{U}) := a\mathbf{x}+\mathbf{U}$. In particular, the neutral element is identified with the coset $\mathbf{0}+\mathbf{U} = \mathbf{U}$. In particular, $\dim \mathbf{V}/\mathbf{U} = \dim \mathbf{V} - \dim \mathbf{U}$. More precisely, if $\mathbf{U}$ has a basis $\mathbf{u}_1, \dots, \mathbf{u}_m$ and $\mathbf{V}$ has a basis $\mathbf{u}_1, \dots, \mathbf{u}_m, \mathbf{u}_{m+1}, \dots, \mathbf{u}_{m+l}$ then $\hat{\mathbf{V}}$ has a basis $\hat{\mathbf{u}}_{m+1}, \dots, \hat{\mathbf{u}}_{m+l}$. (Note that if $\mathbf{U} = \{\mathbf{0}\}$ then $\hat{\mathbf{V}} = \mathbf{V}$ and if $\mathbf{U} = \mathbf{V}$ then $\hat{\mathbf{V}}$ is the trivial subspace consisting of the zero element.)*

The proof is straightforward and is left as an exercise.

Lemma 2.25 *Assume that the assumptions of Lemma 2.24 holds. Let $T \in L(\mathbf{V}) := L(\mathbf{V}, \mathbf{V})$. Assume furthermore that $\mathbf{U}$ is an invariant subspace of T. Then T induces a linear operator $\hat{T} : \hat{\mathbf{V}} \to \hat{\mathbf{V}}$ by the equality: $\hat{T}(\mathbf{x}+\mathbf{U}) := T\mathbf{x}+\mathbf{U}$, i.e. $\hat{T}\hat{\mathbf{x}} = \widehat{T\mathbf{x}}$.*

Proof. We first need to show that $\widehat{T\mathbf{x}}$ is independent of the coset representative. Namely $T(\mathbf{x}+\mathbf{u}) + \mathbf{U} = T\mathbf{x} + (T\mathbf{u}+\mathbf{U}) = T\mathbf{x}+\mathbf{U}$, for each $\mathbf{u} \in \mathbf{U}$. Since $T\mathbf{u} \in \mathbf{U}$ we deduce that $T\mathbf{u}+\mathbf{U} = \mathbf{U}$. The linearity of $\hat{T}$ follows from the linearity of T. □

2.5 Existence and uniqueness of the Jordan canonical form

Definition 2.26 *$A \in \mathbb{F}^{n\times n}$ or a linear transformation $T : \mathbf{V} \to \mathbf{V}$ is called nilpotent respectively, if $A^m = \mathbf{0}$ or $T^m = \mathbf{0}$. The minimal $m \ge 1$ for which $A^m = \mathbf{0}$ or $T^m = \mathbf{0}$ is called the index of nilpotency of A and T respectively, and denoted by* index A *or* index T *respectively.*

Assume that A or T are nilpotent, then the s-numbers are defined as

$$s_i(A) := \text{rank } A^{i-1} - 2\text{rank } A^i + \text{rank } A^{i+1}, \tag{2.7}$$
$$s_i(T) := \text{rank } T^{i-1} - 2\text{rank } T^i + \text{rank } T^{i+1},\ i = 1, \dots$$

Note that A or T are nilpotent with the index of nilpotency m if and only z^m is the minimal polynomial of A or T respectively. Furthermore if A or T are nilpotent then the maximal l for which $s_l > 0$ is equal to the index of nilpotency of A or T respectively.

Proposition 2.27 *Let $T : \mathbf{V} \to \mathbf{V}$ be a nilpotent operator, with the index of nilpotency m, on the finite dimensional vector $\mathbf{V}$. Then*

$$\text{rank } T^i = \sum_{j=i+1}^{m} (j-i)s_j = (m-i)s_m + (m-i-1)s_{m-1} + \ldots + s_{i+1}, \quad (2.8)$$
$$i = 0, \ldots, m-1.$$

Proof. Since $T^l = \mathbf{0}$ for $l \geq m$ it follows that $s_m(T) = \text{rank } T^{m-1}$ and $s_{m-1} = \text{rank } T^{m-2} - 2\text{rank } T^{m-1}$ if $m > 1$. This proves (2.8) for $i = m-1, m-2$. For other values of i (2.8) follows straightforward from (2.7) by induction on $m - i \geq 2$. □

Theorem 2.28 *Let $T : \mathbf{V} \to \mathbf{V}$ be a linear transformation on a finite dimensional space. Assume that T is nilpotent with the index of nilpotency m. Then $\mathbf{V}$ has a basis of the form*

$$\mathbf{x}_j, T\mathbf{x}_j, \ldots T^{l_j-1}\mathbf{x}_j, \ j = 1, \ldots, i, \ \textit{where } l_1 = m \geq \ldots \geq l_i \geq 1, \quad (2.9)$$
$$\textit{and } T^{l_j}\mathbf{x}_j = 0, \ j = 1, \ldots, i.$$

More precisely, the number of l_j, which are equal to an integer $l \in [1, m]$, is equal to $s_l(T)$ given in (2.7).

Proof. We first show by induction on the dimension of $\mathbf{V}$ that there exists a basis in $\mathbf{V}$ given by (2.9). If $\dim \mathbf{V} = 1$ and T is nilpotent then any $\mathbf{x} \in \mathbf{V} \setminus \{\mathbf{0}\}$ is basis satisfying (2.9). So suppose that for $\dim \mathbf{V} \leq N$ any nilpotent T has a basis of the form (2.9). Assume now that $\dim \mathbf{V} = N + 1$.

Assume that $T^m = 0$ and $T^{m-1} \neq 0$. Hence $\text{rank } T^{m-1} = \dim \text{range } T^{m-1} > 0$. Thus there exists $\mathbf{0} \neq \mathbf{y}_1 \in \text{range } T^{m-1}$. So $T^{m-1}\mathbf{x}_1 = \mathbf{y}_1$ for some $\mathbf{x}_1 \in \mathbf{V}$. Clearly $T^m\mathbf{x}_1 = \mathbf{0}$. We claim that $\mathbf{x}_1, T\mathbf{x}_1, \ldots, T^{m-1}\mathbf{x}_1$ are linearly independent. Suppose that

$$\sum_{i=1}^{m} a_i T^{i-1}\mathbf{x}_i = \mathbf{0}. \quad (2.10)$$

Apply T^{m-1} to this equality to deduce that $\mathbf{0} = a_1 T^{m-1}\mathbf{x}_1 = a_1\mathbf{y}_1$. Since $\mathbf{y}_1 \neq \mathbf{0}$ it follows that $a_1 = 0$. Now apply to (2.10) T^{m-2} too deduce that $a_2 = 0$. Continue this process to deduce that $a_1 = \ldots = a_m = 0$.

Let $\mathbf{U} = \text{span}(\mathbf{x}_1, T\mathbf{x}_1, \ldots, T^{m-1}\mathbf{x}_1)$. Clearly $T\mathbf{U} \subset \mathbf{U}$. If $\mathbf{U} = \mathbf{V}$ we are done. So assume that $m = \dim \mathbf{U} < \dim \mathbf{V}$. We now consider $\hat{\mathbf{V}} := \mathbf{V}/\mathbf{U}$ and the corresponding $\hat{T} : \hat{\mathbf{V}} \to \hat{\mathbf{V}}$. Since $T^m = 0$ it follows that $\hat{T}^m = 0$. So $\hat{T}$ is nilpotent. Assume that $\hat{T}^{m'} = 0$ and $\hat{T}^{m'-1} \neq 0$. So $m \geq m'$. We now apply the induction hypothesis to $\hat{T}$. So $\hat{\mathbf{V}}$ has a basis of the form $\hat{\mathbf{x}}_j, \hat{T}\hat{\mathbf{x}}_j, \ldots \hat{T}^{l_j-1}\hat{\mathbf{x}}_j$ for $j = 2, \ldots, i$. Here $m' = l_2 \geq \ldots \geq l_i \geq 1$. Furthermore $\hat{T}^{l_j}\hat{\mathbf{x}}_j = \mathbf{0}$ in $\hat{\mathbf{V}}$ for $j = 2, \ldots, i$. Assume that $\hat{\mathbf{x}}_j = \mathbf{z}_j + \mathbf{U}$ for $j = 2, \ldots, i$. As $\hat{T}^{l_j}\hat{\mathbf{x}}_j = \mathbf{0}$ it follows

that $T^{l_j}\mathbf{z}_j = \sum_{i=1}^{m} a_i T^{i-1}\mathbf{x}_1$. Apply T^{m-l_j} to both sides of this identity. Since $T^m = 0$ we deduce that $\mathbf{0} = \sum_{i=1}^{l_j} a_i T^{m-l_{j-1}+i-1}\mathbf{x}_1$. As $\mathbf{x}_1, \dots, T^{m-1}\mathbf{x}_1$ are linearly independent we deduce that $a_1 = \dots = a_{l_j} = 0$. Let $\mathbf{x}_j := \mathbf{z}_j - \sum_{i=1}^{m-l_j} a_{l_j+i} T^{i-1}\mathbf{x}_1$. Observe that $T^{l_j}\mathbf{x}_j = \mathbf{0}$ for $j = 2, \dots, i$.

We claim that the vectors in (2.9) form a basis in $\mathbf{V}$ as claimed. Assume that there is a linear combinations of these vectors that gives zero vector in $\mathbf{V}$:

$$\sum_{j=1}^{i}\sum_{k=1}^{l_j} b_{jk} T^{k-1}\mathbf{x}_j = \mathbf{0}. \tag{2.11}$$

Consider this linear combination in $\hat{\mathbf{V}}$. As $\hat{\mathbf{x}}_1 = \mathbf{0}$ in $\hat{\mathbf{V}}$ we deduce that the above equality reduces to:

$$\sum_{j=2}^{i}\sum_{k=1}^{l_j} b_{jk} \hat{T}^{k-1}\hat{\mathbf{x}}_j = \mathbf{0}.$$

The induction hypothesis yield that $b_{jk} = 0$ for each $j \geq 2$ and each corresponding k. Hence (2.11) reduces to $\sum_{k=1}^{m} b_{1k} T^{k-1}\mathbf{x}_1 = \mathbf{0}$. As $\mathbf{x}_1, \dots, T^{m-1}\mathbf{x}_1$ are linearly independent we deduce that $b_{1k} = 0$ for $k \in [m]$. Hence the vectors in (2.9) are linearly independent. Note that the number of these vectors is $m + \dim \hat{V} = \dim V = N + 1$. There the vectors in (2.9) form a basis in $\mathbf{V}$.

It is left to show that the number of l_j, which are equal to an integer $l \in [1, m]$, is equal to $s_l(T)$ given in (2.7). Since $T^l = 0$ for $l \geq m$ it follows that $s_m = \operatorname{rank} T^{m-1} = \dim \operatorname{range} T^{m-1}$. Note that if $l_j < m$ then $T^{m-1}\mathbf{x}_j = \mathbf{0}$. Hence $\operatorname{range} T^{m-1} = \operatorname{span}(T^{m-1}\mathbf{x}_1, \dots, T^{m-1}\mathbf{x}_{s_m(T)})$. That is, $s_m(T)$ is the number of l_j that are equal to m. Use Lemma 2.27 and the basis of $\mathbf{V}$ given by (2.9) to deduce that the number of l_j, which are equal to an integer $l \in [1, m]$, is equal to $s_l(T)$ for $l = m-1, \dots, 1$. □

Corollary 2.29 *Let T satisfies the assumption of Theorem 2.28 hold. Denote $\mathbf{V}_j := \operatorname{span}(T^{l_j-1}\mathbf{x}_j, \dots, T\mathbf{x}_j, \mathbf{x}_j)$ for $j = 1, \dots, i$. Then each V_j is a T-invariant subspace, $T|_{\mathbf{V}_j}$ is represented by $J_{l_j}(0) \in \mathbb{C}^{l_j \times l_j}$ in the basis $[T^{l_j-1}\mathbf{x}_j, \dots, T\mathbf{x}_j, \mathbf{x}_j]$, and $\mathbf{V} = \oplus_{j=1}^{i}\mathbf{V}_j$. Each l_j is uniquely determined by the sequence $s_i(T), i = 1, \dots,$. Namely, the index m of the nilpotent T is the largest $i \geq 1$ such that $s_i(T) \geq 1$. Let $k_1 = s_m(T), l_1 = \dots = l_{k_1} = p_1 = m$ and define recursively $k_r := k_{r-1} + s_{p_r}(T)$, $l_{k_{r-1}+1} = \dots = l_{k_r} = p_r$, where $2 \leq r, p_r \in [1, m-1], s_{p_r}(T) > 0$ and $k_{r-1} = \sum_{j=1}^{m-p_r} s_{m-j+1}(T)$.*

Definition 2.30 *$T : \mathbf{V} \to \mathbf{V}$ be a nilpotent operator. Then the sequence $(l_1, \dots, l_i)$ defined in Theorem 2.28, which gives the lengths of the corresponding Jordan blocks of T in a decreasing order, is called the Segré characteristic of T. The Weyr characteristic of T is the dual to Segre's characteristic. That is consider an $m \times i$ $0-1$ matrix $B = (b_{pq}) \in \{0,1\}^{m \times i}$. The j-th column of B has 1 in the rows $1, \dots, l_j$ and 0 in the rest of the rows. Let ω_p be the p-th row sum of B for $p = 1, \dots, m$. Then $\omega_1 \geq \dots \geq \omega_m \geq 1$ is the Weyr characteristic.*

Proof of Theorem 2.13 (The Jordan Canonical Form)

Let $p(z) = \det(zI_n - A)$ be the characteristic polynomial of $A \in \mathbb{C}^{n\times n}$. Since $\mathbb{C}$ is algebraically closed $p(z) = \prod_{j=1}^{k}(z-\lambda_j)^{n_j}$. Here $\lambda_1, \dots, \lambda_k$ are k distinct roots, (eigenvalues of A), where $n_j \geq 1$ is the multiplicity of λ_j in $p(z)$. Note that $\sum_{j=1}^{k} n_j = n$. Let $\psi(z)$ be the minimal polynomial of A. By Theorem 2.18 $\psi(z)|p(z)$. Problem 1(c) of §2.3 we deduce that $\psi(\lambda_j) = 0$ for $j = 1, \dots, k$. Hence

$$\det(zI_n - A) = \prod_{j=1}^{k}(z-\lambda_j)^{n_j},\ \psi(z) = \prod_{j=1}^{k}(z-\lambda_j)^{m_j}, \tag{2.12}$$
$$1 \leq m_j \leq n_j,\ \lambda_j \neq \lambda_i \text{ for } j \neq i,\ i,j = 1, \dots, k.$$

Let $\psi_j := (z-\lambda_j)^{m_j}$ for $j = 1, \dots, k$. Then $(\psi_j, \psi_i) = 1$ for $j \neq i$. Let $\mathbf{V} := \mathbb{C}^n$ and $T : \mathbf{V} \to \mathbf{V}$ be given by $T\mathbf{x} := A\mathbf{x}$ for any $\mathbf{x} \in \mathbb{C}^n$. Then $\det(zI_n - A)$ and $\psi(z)$ are the characteristic and the minimal polynomial of T respectively. Use Theorem 2.23 to obtain the decomposition $\mathbf{V} = \oplus_{i=1}^{k}\mathbf{V}_i$, where each $\mathbf{V}_i$ is a nontrivial T-invariant subspace such that the minimal polynomial of $T_i := T|_{\mathbf{V}_i}$ is ψ_i for $i = 1, \dots, k$. That is $T_i - \lambda_i I_i$, where I_i is the identity operator, i.e. $I_i\mathbf{v} = v$ for all $\mathbf{v} \in \mathbf{V}_i$, is a nilpotent operator on $\mathbf{V}_i$ and index $(T_i - \lambda_i I_i) = m_i$. Let $Q_i := T_i - \lambda_i I_i$. Then Q_i is nilpotent and index $Q_i = m_i$. Apply Theorem 2.28 and Corollary 2.29 to deduce that $\mathbf{V}_i = \oplus_{j=1}^{q_j}\mathbf{V}_{i,j}$, where each $V_{i,j}$ is Q_i-invariant subspace, and each $V_{i,j}$ has a basis in which Q_i is represented by a Jordan block $J_{m_{ij}}(0)$ for $j = 1, \dots, q_j$. According to Corollary 2.29

$$m_i = m_{i1} \geq \dots m_{iq_i} \geq 1, \quad i = 1, \dots, k. \tag{2.13}$$

Furthermore, the above sequence is completely determined by rank $Q_i^j, j = 0, 1, \dots$ for $i = 1, \dots, k$. Noting that $T_i = Q_i + \lambda_i I_i$ it easily follows that each $V_{i,j}$ is a T_i-invariant subspace, hence T-invariant subspace. Moreover, in the same basis of $\mathbf{V}_{i,j}$ that Q_i represented by $J_{m_{ij}}(0)$ T_i is represented by $J_{m_{ij}}(\lambda_i)$ for $j = 1, \dots, q_i$ and $i = 1, \dots, k$. This shows the existence of the Jordan canonical form.

We now show that the Jordan canonical form is unique, up to a permutation of factors. Note that the minimal polynomial of A is completely determined by its Jordan canonical form. Namely $\psi(z) = \prod_{i=1}^{k}(z-z_i)^{m_{i1}}$, where m_{i1} is the biggest Jordan block with the eigenvalues λ_i. (See Problems 1,2 in §2.3.) Thus $m_{i1} = m_i$ for $i = 1, \dots, k$. Theorem 2.23 yields that the subspaces $\mathbf{V}_1, \dots, \mathbf{V}_k$ are uniquely determined by ψ. So each T_i and $Q_i = T - \lambda_i I_i$ are uniquely determined. Theorem 2.28 yields that rank $Q_i^j, j = 0, 1, \dots$ determines the sizes of the Jordan blocks of Q_i. Hence all the Jordan blocks corresponding to λ_i are uniquely determined for each $i \in [1, k]$. □

Corollary 2.31 *Let $A \in \mathbb{F}^{n\times n}$ and assume that the characteristic polynomial of $p(z) = \det(zI_n - A)$ splits to linear factors, i.e. (2.12) holds. Let B be the Jordan canonical form of A. Then*

1. *The multiplicity of the eigenvalue λ_i in the minimal polynomial $\psi(z)$ of A is the size of the biggest Jordan block corresponding to λ_i in B.*

2. *The number of Jordan blocks in B corresponding to λ_i is the nullity of $A-\lambda_i I_n$, i.e. the number of Jordan block in B corresponding to $A - \lambda_i I_n$ is the number of linearly independent eigenvectors of A corresponding to the eigenvalue λ_i.*

3. *Let λ_i be an eigenvalue of A. Then the number of the Jordan blocks of order i corresponding to λ_i in B is given in (2.14).*

Proof. 1. Since $J_n(0)^n = 0$ and $J_n(0)^{n-1} \neq 0$, it follows that the minimal polynomial of $J_n(\lambda)$ is $(z-\lambda_n)^n = \det(zI_n - J_n(\lambda))$. Use Problem 4a to deduce the first part of the corollary. 2. Since $J_n(0)$ has one independent eigenvector, use Problem 4b to deduce the second part of the corollary. 3. Use 3a to establish the last part of the corollary. □

Problems

1. Let $T : \mathbf{V} \to \mathbf{V}$ be nilpotent with $m = \operatorname{index} T$. Let $(\omega_1, \ldots, \omega_m)$ be the Weyr characteristic. Show that $\operatorname{rank} T^j = \sum_{p=j+1}^{m} \omega_p$ for $j = 1, \ldots, m$.

2. Let $A \in \mathbb{C}^{n\times n}$. Show that A is diagonable if and only if all the zeros of the minimal polynomial ψ of A are simple, i.e. ψ does not have multiple roots.

3. Let $A \in \mathbb{C}^{n\times n}$ and assume that $\det(zI_n - A) = \prod_{i=1}^{k}(z-\lambda_i)^{n_i}$, where $\lambda_1, \ldots, \lambda_k$ are k distinct eigenvalues of A. Let

$$s_i(A, \lambda_j) := \operatorname{rank}(A - \lambda_j I_n)^{i-1} - 2\operatorname{rank}(A - \lambda_j I_n)^{i} + \operatorname{rank}(A - \lambda_j I_n)^{i+1} \quad (2.14)$$
$$i = 1, \ldots, n_j, j = 1, \ldots, k.$$

   (a) Show that $s_i(A, \lambda_j)$ is the number of Jordan blocks of order i corresponding to λ_j for $i = 1, \ldots, n_j$.

   (b) Show that in order to find all Jordan blocks of A corresponding to λ_j one can stop computing $s_i(A, \lambda_j)$ at the smallest $i \in [1, n_j]$ such that $1s_1(A, \lambda_j) + 2s_2(A, \lambda_j) \ldots + is_i(A, \lambda_j) = n_j$.

4. Let $C = F \oplus G = \operatorname{diag}(F, G), F \in \mathbb{F}^{l\times l}, G \in \mathbb{F}^{m\times m}$.

   (a) Assume that ψ_F, ψ_G are the minimal polynomials of F, G respectively. Show that ψ_C, the minimal polynomial of C, is equal to $\frac{\psi_F \psi_G}{(\psi_F, \psi_G)}$.

   (b) Show that $\operatorname{nul}(C) = \operatorname{nul}(F) + \operatorname{nul}(G)$. In particular, if G is invertible, i.e. 0 is not eigenvalue of G then $\operatorname{nul}(C) = \operatorname{nul}(F)$.

2.6 Cyclic subspaces

Let $T : \mathbf{V} \to \mathbf{V}$ be a linear operator, and assume that $\mathbf{V}$ is finite dimensional. Let $\mathbf{0} \neq \mathbf{u} \in \mathbf{V}$. Consider the sequence of vectors $\mathbf{u} = T^0\mathbf{u}, T\mathbf{u}, T^2\mathbf{u}, \ldots$. Since $\dim \mathbf{V} < \infty$ there exists a positive integer $l \geq 1$ such that $\mathbf{u}, T\mathbf{u}, \ldots, T^{l-1}\mathbf{u}$ linearly independent, and $\mathbf{u}, T\mathbf{v}, \ldots, T^l\mathbf{u}$ are linearly dependent. Hence l is smallest integer such that

$$T^l\mathbf{u} = -\sum_{i=1}^{l} a_i T^{l-i}\mathbf{u}. \quad (2.15)$$

Clearly, $l \leq \dim \mathbf{V}$. The polynomial $\psi_{\mathbf{u}}(z) := z^l + \sum_{i=1}^{l} a_i z^{l-i}$ is the called the *minimal* polynomial of $\mathbf{u}$, with respect to T . It is a monic polynomial of the minimal degree such that $\phi(T)\mathbf{u} = \mathbf{0}$. Its property is similar to the property of the minimal polynomial of T. Namely if a polynomial $\phi \in \mathbb{F}[z]$ annihilates $\mathbf{u}$, i.e.

$\phi(T)\mathbf{u} = 0$ then $\psi_{\mathbf{u}}|\phi$. In particular, the minimal polynomial $\psi(z)$ of T is divisible by $\psi_{\mathbf{u}}$, since $\psi(T)\mathbf{u} = 0\mathbf{u} = \mathbf{0}$. Clearly, every vector $\mathbf{w} \in \mathbf{U}$ can be uniquely represented as $\phi(T)\mathbf{u}$, where $\phi \in \mathbb{F}[z]$ and $\deg \phi \leq l-1$. Hence $\mathbf{U}$ is T-invariant subspace. The subspace $\mathbf{U}$ spanned by $\mathbf{u}, T\mathbf{u}, \ldots,$ is called a *cyclic* subspace, generated by $\mathbf{u}$. Note that in the basis $\mathbf{u}_1 = \mathbf{u}, \mathbf{u}_2 = T\mathbf{u}, \ldots, \mathbf{u}_l = T^{l-1}\mathbf{u}$ the linear transformation $T|\mathbf{U}$ is given by the matrix

$$\begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & -a_l \\ 1 & 0 & 0 & \ldots & 0 & -a_{l-1} \\ 0 & 1 & 0 & \ldots & 0 & -a_{l-2} \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -a_1 \end{bmatrix} \in \mathbb{F}^{l \times l}. \tag{2.16}$$

The above matrix is called the *companion* matrix, corresponding to the polynomial $\psi_{\mathbf{u}}(z)$. (Sometimes the transpose of the above matrix is called the companion matrix.)

Lemma 2.32 *Let $T : \mathbf{V} \to \mathbf{V}, \dim \mathbf{V} < \infty, \mathbf{0} \neq \mathbf{u}, \mathbf{w}$. Assume that $(\psi_{\mathbf{u}}, \psi_{\mathbf{w}}) = 1$. Then $\psi_{\mathbf{u}+\mathbf{w}}(z) = \psi_{\mathbf{u}}(z)\psi_{\mathbf{w}}(z)$.*

Proof. Let $\mathbf{U}, \mathbf{W}$ be the cyclic invariant subspaces generated by $\mathbf{u}, \mathbf{w}$ respectively. We claim that the T-invariant subspace $\mathbf{X} := \mathbf{U} \cap \mathbf{W}$ is the trivial subspace $\{\mathbf{0}\}$. Suppose to the contrary, there exists $\mathbf{0} \neq \mathbf{x} \in \mathbf{X}$. Let $\mathbf{X}_1 \subset \mathbf{X}$ be a nontrivial cyclic subspace generated by $\mathbf{x}$ and $\psi_{\mathbf{x}}$ be the minimal polynomial corresponding to $\mathbf{x}$. Since $\mathbf{X}_1 \subset \mathbf{U}$ it follows that $\mathbf{x} = \phi(T)\mathbf{u}$. Hence $\psi_{\mathbf{u}}(T)\mathbf{x} = \mathbf{0}$. Thus $\psi_{\mathbf{x}}|\psi_{\mathbf{u}}$. Similarly $\psi_{\mathbf{x}}|\psi_{\mathbf{w}}$. This contradicts the assumption that $(\psi_{\mathbf{u}}, \psi_{\mathbf{w}}) = 1$. Hence $\mathbf{U} \cap \mathbf{W} = \{\mathbf{0}\}$. Let $\phi = \psi_{\mathbf{u}}\psi_{\mathbf{w}}$. Clearly

$$\phi(T)(\mathbf{u}+\mathbf{w}) = \psi_{\mathbf{w}}(T)(\psi_{\mathbf{u}}(T)\mathbf{u}) + \psi_{\mathbf{u}}(T)(\psi_{\mathbf{w}}(T)\mathbf{w}) = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

So ϕ is an annihilating polynomial of $\mathbf{u}+\mathbf{w}$. Let $\theta(z)$ be an annihilating polynomial of $\mathbf{u}+\mathbf{w}$. So $\mathbf{0} = (\theta(T)\mathbf{u}) + (\theta(T)\mathbf{w})$. Since $\mathbf{U}, \mathbf{V}$ are T-invariant subspaces and $\mathbf{U} \cap \mathbf{W} = \{\mathbf{0}\}$, it follows that $\mathbf{0} = \theta(T)\mathbf{u} = \theta(T)\mathbf{w}$. Hence $\psi_{\mathbf{u}}|\theta, \psi_{\mathbf{w}}|\theta$. As $(\psi_{\mathbf{u}}, \psi_{\mathbf{w}}) = 1$ it follows that $\phi|\theta$. Hence $\psi_{\mathbf{u}+\mathbf{w}} = \phi$. □

Theorem 2.33 *Let $T : \mathbf{V} \to \mathbf{V}$ be a linear operator and $1 \leq \dim V < \infty$. Let $\psi(z)$ be the minimal polynomial of T. Then there exists $\mathbf{0} \neq \mathbf{u} \in \mathbf{V}$ such that $\psi_{\mathbf{u}} = \psi$.*

Proof. Assume first that $\psi(z) = (\phi(z))^l$, where $\phi(z)$ is irreducible in $\mathbb{F}[z]$, i.e. $\phi(z)$ is not divisible by any polynomial θ such $1 \leq \deg \theta < \deg \phi$, and l is a positive integer. Let $0 \neq \mathbf{w} \in \mathbf{V}$. Recall that $\psi_{\mathbf{w}}|\psi$. Hence $\psi_{\mathbf{w}} = (\phi)^{l(\mathbf{w})}$, where $l(\mathbf{w}) \in [1, l]$ is a positive integer. Assume to the contrary that $1 \leq l(\mathbf{w}) \leq l-1$ for each $0 \neq \mathbf{w} \in \mathbf{V}$. Then $(\phi(T))^{l-1}(\mathbf{w}) = \mathbf{0}$. As $(\phi(T))^{l-1}(\mathbf{0}) = \mathbf{0}$ we deduce that $(\phi(z))^{l-1}$ is annihilating polynomial of T. So $\psi|\phi^{l-1}$ which is impossible. Hence there exists $\mathbf{u} \neq \mathbf{0}$ such that $l(\mathbf{u}) = l$, i.e. $\psi_{\mathbf{u}} = \psi$.

Assume now the general case

$$\psi(z) = \prod_{i=1}^{k} (\phi_i(z))^{l_i},\ l_i \in \mathbb{N},\ \phi_i \text{ irreducible },(\phi_i, \phi_j) = 1 \text{ for } i \neq j, \tag{2.17}$$

where $k > 1$. Theorem 2.23 implies that $V = \oplus_{i=1}^{k} \mathbf{V}_i$, where $T : \mathbf{V}_i \to \mathbf{V}_i$ and $\phi_i^{l_i}$ is the minimal polynomial of $T|V_i$. The first case of the proof yields the existence $\mathbf{0} \neq \mathbf{u}_i \in \mathbf{V}_i$ such that $\psi_{\mathbf{u}_i} = \phi_i^{l_i}$ for $i = 1, \ldots, k$. Let $\mathbf{w}_j = \mathbf{u}_1 + \ldots + \mathbf{u}_j$. Use Lemma 2.32 to deduce that $\psi_{\mathbf{w}_j} = \prod_{i=1}^{j} \phi_i^{l_i}$. Hence $\psi_{\mathbf{u}} = \psi$ for $\mathbf{u} := \mathbf{w}_k$. □

Let $\mathbf{U} \subset \mathbf{V}$ be a subspace of $\mathbf{V}$. Then $\mathbf{V}/\mathbf{U}$ is the following object. We introduce the following relation $\sim$ on the elements of V: $\mathbf{x} \sim \mathbf{y}$, which is also denoted as $\mathbf{x} \equiv \mathbf{y} \mod \mathbf{U}$, if $\mathbf{x} - \mathbf{y} \in \mathbf{U}$. It is straightforward to show that $\sim$ is an equivalence relation. Denote by $[\mathbf{x}]$ the equivalence class corresponding to $\mathbf{x}$. Then $\mathbf{V}/\mathbf{U} := \{[\mathbf{x}],\ \mathbf{x} \in \mathbf{V}\}$. It is straightforward to show that the following definition are valid $a[\mathbf{x}] := [a\mathbf{x}]$, $[\mathbf{x}] + [\mathbf{y}] := [\mathbf{x} + \mathbf{y}]$, i.e. these definitions are independent of the choice of the representatives of the equivalence classes. Furthermore, under these two operations $\mathbf{V}/\mathbf{U}$ is a vector space. Then $V/\mathbf{U}$ is called the quotient vector space. Moreover, $\dim \mathbf{V}/\mathbf{U} = \dim \mathbf{V} - \dim \mathbf{U}$. Observe next that if T is a linear operator $T : \mathbf{V} \to \mathbf{V}$ and $\mathbf{U}$ is an T-invariant subspace of $\mathbf{V}$, then T induces a linear transformation $\hat{T} : \mathbf{V}/\mathbf{U} \to \mathbf{V}/\mathbf{U}$. See Problem 4a.

Theorem 2.34 *Let $T : \mathbf{V} \to \mathbf{V}$ be a linear operator on a finite dimensional space $\dim V > 0$. Then there exists a decomposition of V to a direct sum of r T-cyclic subspaces $V = \oplus_{i=1}^{r} \mathbf{U}_i$ with the following properties. Assume that ψ_i is the minimal polynomial of $T|\mathbf{U}_i$ then ψ_1 is the minimal polynomial of T. Furthermore, $\psi_{i+1}|\psi_i$ for $i = 1, \ldots, r-1$.*

Proof. We prove the theorem by induction on $\dim \mathbf{V}$. For $\dim V = 1$, any $0 \neq \mathbf{u}$ is an eigenvector of T: $T\mathbf{u} = \lambda\mathbf{u}$, so $\mathbf{V}$ is cyclic, and $\psi(z) = \psi_1(z) = z - \lambda$. Assume now that the theorem holds of all $\mathbf{V}$ with $\dim V \leq n$. Let $\dim \mathbf{V} = n + 1$. Let $\psi(z)$ be the minimal polynomial of T. Assume that $m = \deg \psi$ Suppose first that T is nonderogatory, i.e. $m = n + 1$, which is the degree of the characteristic polynomial of T. Theorem 2.33 implies that $\mathbf{V}$ is a cyclic subspace. So $r = 1$, and the theorem holds in this case.

Assume now that T is derogatory. So $m < n + 1$. Theorem 2.33 implies the existence of $\mathbf{0} \neq \mathbf{u}_1$ such that $\psi_{\mathbf{u}_1} = \psi$. Let $\mathbf{U}_1$ be the cyclic subspace generated by $\mathbf{u}_1$. Let $\hat{V} := \mathbf{V}/\mathbf{U}_1$. So $1 \leq \dim \hat{V} = \dim \mathbf{V} - m \leq n$. We now apply the induction hypothesis to $\hat{T} : \hat{\mathbf{V}} \to \hat{\mathbf{V}}$. So $\hat{V} = \oplus_{i=2}^{r} \hat{U}_i$, where $\hat{\mathbf{U}}_i$ is a cyclic subspace generated by $[\hat{\mathbf{u}}_i], i = 2, \ldots, r$. The minimal polynomial of $[\hat{\mathbf{u}}_i]$ is $\psi_i, i = 2, \ldots, r$. ψ_2 is the minimal polynomial of $\hat{T}$, and $\psi_{i+1}|\psi_i$ for $i = 2, \ldots, r-1$. Observe first that for any polynomial $\phi(z)$ we have the identity $\phi(\hat{T})[\mathbf{x}] = [\phi(T)\mathbf{x}]$. Since $\psi(T)\mathbf{x} = \mathbf{0}$ it follows that $\psi_1 := \psi(z)$ is an annihilating polynomial of $\hat{T}$. Hence $\psi_2|\psi_1$.

Observe next that since $\psi_{i+1}|\psi_i$ for $i = 1, \ldots, r-1$, it follows that $\psi_i|\psi_1 = \psi$. So $\psi = \theta_i\psi_i$, where θ_i is a monic polynomial and $i = 2, ldots, r$. Since $[\mathbf{0}] = \psi_i(\hat{T})[\hat{\mathbf{u}}_i] = [\psi_i(T)\hat{\mathbf{u}}_i]$ it follows that $\psi_i(T)\hat{\mathbf{u}}_i \in \mathbf{U}_1$. Hence $\psi_i(T)\hat{\mathbf{u}}_i = \omega_i(T)\mathbf{u}_1$ for some $\omega(z) \in \mathbb{F}[z]$. Hence $\mathbf{0} = \psi(T)\hat{\mathbf{u}}_i = \theta_i(T)\psi_i(T)\hat{\mathbf{u}}_i = \theta_i(T)\omega_i(T)\mathbf{u}_1$. Therefore, $\psi_{\mathbf{u}_1} = \psi$ divides $\theta_i\omega_i$. So $\psi_i|\omega_i$. Thus $\omega_i = \psi_i\alpha_i$. Define $\mathbf{u}_i = \hat{\mathbf{u}}_i - \alpha_i(T)\mathbf{u}_1$ for $i = 2, \ldots, r$. The rest of the proof theorem follows from Problem 5. □

Theorem 2.35 *Let $A \in \mathbb{F}^{n \times n}$. Then there exists r monic polynomials $\psi_1, \ldots, \psi_r$ of degree one at least, such that the following conditions. First $\psi_{i+1}|\psi_i$ for $i = 1, \ldots, r$. Second, $\psi_1 = \psi$ is the minimal polynomial of A. Then A is similar to*

$\oplus_{i=1}^r C(\psi_i)$, *where* $C(\psi_i)$ *is the companion matrix of the form (2.16) corresponding to the polynomial* ψ_i.

Decompose each $\psi_i = \psi_{i,1} \dots \psi_{i,t_i}$ *to the product of its irreducible components, as in the decomposition of* $\psi = \psi_1$ *given in (2.17). Then* A *is similar to* $\oplus_{i=l=1}^{r,t_i} C(\psi_{i,l})$.

Suppose finally that $\psi(z) = \prod_{j=1}^k (z - \lambda_j)^{m_j}$ *where* $\lambda_1, \dots, \lambda_k$ *are the* k *distinct eigenvalues of* A. *Then* A *is similar to the Jordan canonical form given in Theorem 2.13.*

Proof. Identify A with $T : \mathbb{F}^n \to \mathbb{F}^n$, where $T(\mathbf{x}) = A\mathbf{x}$. Use Theorem 2.34 to decompose $\mathbb{F}^n = \oplus_{i=1}^r \mathbf{U}_i$, where each $\mathbf{U}_i$ is a cyclic invariant subspaces such that $T|\mathbf{U}_i$ has the minimal polynomial ψ_i. Since $\mathbf{U}_i$ is cyclic then $T|\mathbf{U}_i$ is represented by $C(\psi_i)$ in the suitable basis of $\mathbf{U}_i$ as shown in the beginning of this section. Hence A is similar to $\oplus_{i=1}^r C(\psi_i)$.

Consider next $T_i := T\mathbf{U}_i$. Use Theorem 2.23 to deduce that $\mathbf{U}_i$ decomposes to a direct sum of T_i invariant subspaces $\oplus_{l=1}^{t_i} \mathbf{U}_{i,l}$, where the minimal polynomial of $T_{i,l} := T_i|\mathbf{U}_{i,l}$ is $\psi_{i,l}$. Since $\mathbf{U}_i$ was cyclic, i.e., T_i was nonderogatory, it follows that each $T_{i,l}$ must be nonderogatory, i.e. $\mathbf{U}_{i,l}$ cyclic. (See Problem 3.) Recall that each $T_{i,l}$ is represented in a corresponding basis by $C(\psi_{i,l})$. Hence A is similar to $\oplus_{i=l=1}^{r,t_i} C(\psi_{i,l})$.

Suppose finally that $\psi(z)$ splits to linear factors. Hence $\psi_{i,l} = (z - \lambda_l)^{m_{i,l}}$. Hence $T_{i,l} - \lambda_l I$ is nilpotent of index $m_{i,l}$. So there exists a basis in $\mathbf{U}_{i,l}$ such that $T_{i,l}$ is represented by the Jordan block $J_{m_{i,l}}(\lambda_l)$. Therefore A is similar to a sum of corresponding Jordan blocks. □

The polynomials $\psi_1, \dots, \psi_k$ appearing in Theorem 2.34 are called *invariant polynomials* of T or its representation matrix A, in any basis. The polynomials $\psi_{i,1}, \dots, \psi_{i,t_i}, i = 1, \dots, k$ appearing in Theorem 2.35 are called the *elementary divisors* of A, or the corresponding linear transformation T represented by A. We now show that these polynomial are uniquely defined.

Lemma 2.36 *(Cauchy-Binet formula) For two positive integers* $1 \le p \le m$ *denote by* $\mathrm{Q}_{p,m}$ *the set of all subsets* $\boldsymbol{\alpha}$ *of* $\{1, \dots, m\}$ *of cardinality* p. *(See bottom of page 5 of these notes.) Let* $A \in \mathbb{F}^{m \times n}, B \in \mathbb{F}^{n \times l}$ *and denote* $C = AB \in \mathbb{F}^{m \times l}$. *Then for any integer* $p \in [1, \min(m, n, p)], \boldsymbol{\alpha} \in \mathrm{Q}_{p,m}, \boldsymbol{\beta} \in \mathrm{Q}_{p,l}$ *the following identity holds.*

$$\det C[\boldsymbol{\alpha}, \boldsymbol{\beta}] = \sum_{\boldsymbol{\gamma} \in \mathrm{Q}_{p,n}} \det A[\boldsymbol{\alpha}, \boldsymbol{\gamma}] \det B[\boldsymbol{\gamma}, \boldsymbol{\beta}]. \tag{2.18}$$

Proof. It is enough to assume the case where $\boldsymbol{\alpha} = \boldsymbol{\beta} = \{1, \dots, p\}$. This is equivalent to the assumption that $p = m = l \le n$. For $p = m = n = l$ the Cauchy-Binet formula reduces to $\det AB = (\det A)(\det B)$. So it is enough to consider the case $p = m = l < n$. Let $C = [c_{ij}]_{i=j=1}^p$. Then $c_{ij} = \sum_{t_j=1}^n a_{it_j} b_{t_j j}$ for $i, j = 1, \dots, p$. Use multilinearity of the determinant to deduce

$$\det C = \det[\sum_{t_j=1}^n a_{it_j} b_{t_j j}]_{i=j=1}^n = \sum_{t_1, \dots, t_p = 1}^n \det[a_{it_j} b_{t_j j}]_{i=j=1}^n =$$

$$\sum_{t_1, \dots, t_p = 1}^n \det A[\{1, \dots, p\}, \{t_1, \dots, t_p\}] b_{t_1 1} b_{t_2 2} \dots b_{t_p p}.$$

Observe next that $\det A[\{1,\dots,p\},\{t_1,\dots,t_p\}]=0$ if $t_i=t_j$ for some $1\le i<j\le p$, since the the columns t_i and t_j in $A[\{1,\dots,p\},\{t_1,\dots,t_p\}]$ are equal. Consider the sum

$$\sum_{\{t_1,t_2,\dots,t_p\}=\boldsymbol{\gamma}\in Q_{p,n}} A[\{1,\dots,p\},\{t_1,\dots,t_p\}]b_{t_11}b_{t_22}\dots b_{t_pp}.$$

The above arguments yield that this sum is $\det(A[\langle p\rangle,\boldsymbol{\gamma}]C[\boldsymbol{\gamma},\langle p\rangle])=(\det A[\langle p\rangle,\boldsymbol{\gamma}])(\det C[\boldsymbol{\gamma},\langle p\rangle])$. This establishes (2.18). □

Proposition 2.37 *Let $A(z)\in\mathbb{F}^{m\times n}[z]$. Denote by the rank of $A(z)$, denoted by $r=\operatorname{rank}A(z)$ the size of the biggest minor of $A(z)$ which is not a zero polynomial. For an integer $k\in[1,r]$ let $\delta_k(z)$ be the greatest common divisor of all $k\times k$ minors of $A(z)$, which is assumed to be a monic polynomial. Assume that $\delta_0=1$. Then $\delta_i|\delta_{i+1}$ for $i=0,\dots,r-1$.*

Proof. Expand a nonzero $(k+1)\times(k+1)$ minor of $A(z)$ to deduce that $\delta_k(z)|\delta_{k+1}(z)$. In particular, $\delta_i|\delta_j$ for $1\le i<j\le r$. □

Proposition 2.38 *Let $A(z)\in\mathbb{F}[z]^{m\times n}$. Assume that $P\in\mathrm{GL}(m,\mathbb{F}),Q\in\mathrm{GL}(n,\mathbb{F})$. Let $B(z)=PA(z)Q$. Then*

1. $\operatorname{rank}A(z)=\operatorname{rank}B(z)=r$.
2. *$\delta_k(A(z))=\delta_k(B(z))$ for $k=0,\dots,r$.*

Proof. Use Cauchy-Binet to deduce that $\operatorname{rank}B(z)\le\operatorname{rank}A(z)$. Since $A(z)=P^{-1}B(z)Q^{-1}$ it follows that $\operatorname{rank}A(z)\le\operatorname{rank}B(z)$. Hence $\operatorname{rank}A(z)=\operatorname{rank}B(z)=r$. Use the Cauchy-Binet to deduce that $\delta_k(A(z))|\delta_k(B(z))$. As $A(z)=P^{-1}B(z)Q^{-1}$ we deduce also that $\delta_k(B(z))|\delta_k(A(z))$. □

Definition 2.39 *Let $A\in\mathbb{F}^{n\times n}$ and define $A(z):=zI_n-A$. Then the polynomials the polynomials of degree 1 at least in the sequence $\phi_i=\frac{\delta_{n-i+1}}{\delta_{n-i}(z)},i=1,\dots,n$ are called the invariant polynomials of A.*

Note that the product of all invariant polynomials is $\det(zI_n-A)$. In view of Proposition 2.38 we obtain that similar matrices have the same invariant polynomials. Hence for linear transformation $T:\mathbf{V}\to\mathbf{V}$ we can define its invariant polynomials of T by any representation matrix of T in a basis of $\mathbf{V}$.

Theorem 2.40 *Let $T:\mathbf{V}\to\mathbf{V}$. Then the invariant polynomials of T are the polynomials $\psi_1,\dots,\psi_r$ appearing in Theorem 2.34.*

To prove this theorem we need the following lemma:

Lemma 2.41 *Let $A=\operatorname{diag}(B,C)(=B\oplus C)\in\mathbb{F}^{n\times n}$, where $B\in\mathbb{F}^{l\times l},C\in\mathbb{F}^{m\times m},m=n-l$. Let $p\in[n-1]$, and assume that $\boldsymbol{\alpha}=\{\alpha_1,\dots,\alpha_p\},\boldsymbol{\beta}=\{\beta_1,\dots,\beta_p\}\in Q_{p,n}$:*

$$1\le\alpha_1<\dots<\alpha_p\le n,\quad 1\le\beta_1<\dots<\beta_p\le n.$$

If $|\boldsymbol{\alpha}\cap[l]|\ne|\boldsymbol{\beta}\cap[l]|$ then $\det A[\boldsymbol{\alpha},\boldsymbol{\beta}]=0$.

Proof. Let $k := |\boldsymbol{\alpha} \cap [l]|$, i.e. $\alpha_k \le l, \alpha_{k+1} > l$. Consider a term in $\det A[\boldsymbol{\alpha}, \boldsymbol{\beta}]$. Up to a sign it is $\prod_{j=1}^{p} a_{\alpha_j \beta_{\sigma(j)}}$ for some bijection $\sigma : [p] \to [p]$. Then this product is zero unless $\sigma([k]) = [k]$ and $\beta_k \le l, \beta_{k+1} > l$. □

Proof of Theorem 2.40. From the proof of Theorem 2.34 it follows that $\mathbf{V}$ has a basis in which T is represented by $C := \oplus_{i=1}^{r} C(\psi_i)$. Assume that $C(\psi_i) \in \mathbb{F}^{l_i \times l_i}, i \in [r]$, where $n = \sum_{i=1}^{r} l_i$.

Let $\theta := z^l + a_1 z^{m-1} + \ldots + a_l$. Then the companion matrix $C(\theta)$ is given by (2.16). Consider the submatrix of $B(z)$ of $zI_l - C(\theta)$ obtained by deleting the first row an the last column. It is an upper triangular matrix with -1 on the diagonal. Hence $\det B(z) = (-1)^{l-1}$. Therefore $\delta_{l-1}(zI_m - C(\theta)) = 1$. Hence $\delta_{l-j}(zI_l - C(\theta)) = 1$ for $j \ge 1$.

We claim that $\delta_{n-j}(zI_n - C) = \prod_{i=j+1}^{r} \psi_j$ for $j \in [r]$. Consider a minor of order $n-j$ of the block diagonal matrix $zI_n - C$ of the form $(zI_n - C)[\boldsymbol{\alpha}, \boldsymbol{\beta}]$, where $\boldsymbol{\alpha}, \boldsymbol{\beta} \in Q_{n-j,n}$. Lemma 2.41 claims that $\det(zI_n - C)[\boldsymbol{\alpha}, \boldsymbol{\beta}] = 0$, unless we choose our minor such that in each submatrix $zI_{l_i} - C(\psi_i))$ we delete the same number of rows and columns. By deleting the first row and last column in $zI_{l_i} - C(\psi_i))$ we get a minor with value $(-1)^{l_i - 1}$. Recall also that $\delta_{l_i - k}(zI_{l_i} - C(\psi_i)) = 1$ for $k \ge 1$. Since $\psi_{i+1} | \psi_i$ it follows that $\delta_{n-j}(zI_n - C)$ is obtained by deleting the first row and the last column in $zI_{l_i} - C(\psi_i)$ for $i = 1, \ldots, j$ and $j = 1, \ldots, r$. ($\psi_{r+1} := 1$.) Thus $\delta_{n-j}(zI_n - C) = \prod_{i=j+1}^{r} \psi_j$ for $j \in [r]$. Hence $\psi_i = \frac{\delta_{n-i+1}(zI_n - C)}{\delta_{n-i}(zI_n - C)}$. □

Note that the above theorem implies that the invariant polynomials of A satisfy $\phi_{i+1} | \phi_i$ for $i = 1, \ldots, r-1$.

The irreducible factors $\phi_{i,1}, \ldots, \phi_{i,t_i}$ of ψ_i given in Theorem 2.35 for $i = 1, \ldots, r$ are called the *elementary divisors* of A. The matrices $\oplus_{i=1}^{k} C(\psi_i), \oplus_{i=l=1}^{k,t_i} C(\psi_{i,l})$ are called the *rational canonical forms of* A. Those are the canonical forms in the case that the characteristic polynomial of A does not split to linear factors over $\mathbb{F}$.

Problems

1. Let $\phi(z) = z^l + \sum_{i=1}^{l} a_i z^{l-i}$. Denote by $C(\phi)$ the matrix (2.16). Show
   (a) $\phi(z) = \det(zI - C(\phi))$.
   (b) Show that the minimal polynomial of $C(\phi)$ is ϕ, i.e. $C(\phi)$ is nonderogatory.
   (c) Assume that $A \in \mathbb{F}^{n \times n}$ is nonderogatory. Show that A is similar to $C(\phi)$, where ϕ is a characteristic polynomial of A.

2. Let $T : \mathbf{V} \to \mathbf{V}, \dim \mathbf{V} < \infty, \mathbf{0} \ne \mathbf{u}, \mathbf{w}$. Show that $\psi_{\mathbf{u}+\mathbf{w}} = \frac{\psi_{\mathbf{u}} \psi_{\mathbf{w}}}{(\psi_{\mathbf{u}}, \psi_{\mathbf{w}})}$.

3. Let $T : V \to V$ and assume that V is a cyclic space, i.e. T is nondegenerate. Let ψ be the minimal and the characteristic polynomial. Assume that $\psi = \psi_1 \psi_2$, where $\deg \psi_1, \deg \psi_2 \ge 1$, and $(\psi_1, \psi_2) = 1$. Show that there exists $\mathbf{0} \ne \mathbf{u}_1, \mathbf{u}_2$ such that $\psi_i = \psi_{\mathbf{u}_i}, i = 1, 2$. Furthermore $\mathbf{V} = \mathbf{U}_1 \oplus \mathbf{U}_2$, where $\mathbf{U}_i$ is the cyclic subspace generated by $\mathbf{u}_i$.

4. Let T is a linear operator $T : \mathbf{V} \to \mathbf{V}$ and $\mathbf{U}$ is a subspace of $\mathbf{V}$. Show

(a) Assume that $T\mathbf{U} \subseteq \mathbf{U}$. Show that T induces a linear transformation $\hat{T} : \mathbf{V}/\mathbf{U} \to \mathbf{V}/\mathbf{U}$, i.e. $\hat{T}[\mathbf{x}] := [T\mathbf{x}]$.

(b) Assume that $T\mathbf{U} \not\subseteq \mathbf{U}$. Show that $\hat{T}[\mathbf{x}] := [T\mathbf{x}]$ does not make sense.

5. In this problem we finish the proof of Theorem 2.34.

(a) Show that ψ_i is an annihilating polynomial of $\mathbf{u}_i$ for $i \geq 2$.

(b) By considering the vector $[\hat{\mathbf{u}}_i] = [\mathbf{u}_i]$ show that $\psi_{\mathbf{u}_i} = \psi_i$ for $i \geq 2$.

(c) Show that the vectors $\mathbf{u}_i, T\mathbf{u}_i, \ldots, T^{\deg \psi_i - 1}\mathbf{u}_i, i = 1, \ldots, r$ are linearly independent. (**Hint**: Assume linear dependence of all vectors, and then quotient this dependence by $\mathbf{U}_1$. Then use the induction hypothesis on $\hat{V}$.)

(d) Let $\mathbf{U}_i$ be the cyclic subspace generated by $\mathbf{u}_i$ for $i = 2, \ldots, r$. Conclude the proof of Theorem 2.34.

6. Let the assumptions of Theorem 2.34 hold. Show that the characteristic polynomial of T is equal to $\prod_{i=1}^{r} \psi_i$.

3 Applications of Jordan Canonical form

3.1 Functions of Matrices

Let $A \in \mathbb{C}^{n \times n}$. Consider the iterations

$$\mathbf{x}_l = A\mathbf{x}_{l-1},\ \mathbf{x}_{l-1} \in \mathbb{C}^n, \quad l = 1, \ldots \tag{3.1}$$

Clearly $\mathbf{x}_l = A^l \mathbf{x}_0$. To compute $\mathbf{x}_l$ from $\mathbf{x}_{l-1}$ one need to perform $n(2n-1)$ flops, (operations: n^2 multiplications and $n(n-1)$ additions). If we want to compute $\mathbf{x}_{10^8}$ we need to $10^8 n(2n-1)$ operations, if we simply program the iterations (3.1). If $n = 10$ it will take us some time to do these iterations, and we will probably run to the roundoff error, which will render our computations meaningless. Is there any better way to find $\mathbf{x}_{10^8}$? The answer is *yes*, and this is the purpose of this section. To do that we need to give the correct way to find directly A^{10^8}, or for that matter any $f(A)$, where $f(z)$ is either polynomial, or more complex functions as $e^z, \cos z, \sin z$, an entire function $f(z)$, or even more special functions.

Theorem 3.1 *Let $A \in \mathbb{C}^{n \times n}$ and*

$$\det(zI_n - A) = \prod_{i=1}^{k} (z - \lambda_i)^{n_i},\ \psi(z) = \prod_{i=1}^{k} (z - \lambda_i)^{m_i}, \tag{3.2}$$
$$1 \leq m := \deg \psi = \sum_{i=1}^{k} m_i \leq n = \sum_{i=1}^{k} n_i,$$
$$1 \leq m_i \leq n_i,\ \lambda_i \neq \lambda_j \text{ for } i \neq j,\ i, j = 1, \ldots, k,$$

where $\psi(z)$ is the minimal polynomial of A. Then there exists unique m linearly independent matrices $Z_{ij} \in \mathbb{C}^{n \times n}$ for $i = 1, \ldots, k$ and $j = 0, \ldots, m_i - 1$, which depend on A, such that for any polynomial $f(z)$ the following identity holds

$$f(A) = \sum_{i=1}^{k} \sum_{j=0}^{m_i - 1} \frac{f^{(j)}(\lambda_i)}{j!} Z_{ij}. \tag{3.3}$$

($Z_{ij}, i = 1, \ldots, k, j = 1, \ldots, m_i$ are called the A-components.)

Proof. We start first with $A = J_n(\lambda)$. So $J_n(\lambda) = \lambda I_n + H_n$, where $H_n := J_n(0)$. Thus H_n is a nilpotent matrix, with $H_n^n = \mathbf{0}$ and H_n^j has 1's on the j-th subdiagonal and all other elements are equal 0 for $j = 0, 1, \ldots, n-1$. Hence $I_n = H_n^0, H_n, \ldots, H_n^{n-1}$ are linearly independent.

Let $f(z) = z^l$. Then

$$A^l = (\lambda I_n + H_n)^l = \sum_{j=0}^{l} \binom{l}{j} \lambda^{l-j} H_n^j = \sum_{j=0}^{\min(l,n-1)} \binom{l}{j} \lambda^{l-j} H_n^j.$$

The last equality follows from the equality $H^j = \mathbf{0}$ for $j \geq n$. Note that $\psi(z) = \det(zI_n - J_n(\lambda)) = (z-\lambda)^n$, i.e. $k = 1$ and $m = m_1 = n$. From the above equality we conclude that $Z_{1j} = H_n^j$ for $j = 0, \ldots$ if $f(z) = z^l$ and $l = 0, 1, \ldots$. With this definition of Z_{1j} (3.3) holds for $K_l z^l$, where $K_l \in \mathbb{C}$ and $l = 0, 1, \ldots$. Hence (3.3) holds for any polynomial $f(z)$ for this choice of A.

Assume now that A is a direct sum of Jordan blocks as in (2.6): $A = \oplus_{i=j=1}^{k,l_i} J_{m_{ij}}(\lambda_i)$. Here $m_i = m_{i1} \geq \ldots \geq m_{il_i} \geq 1$ for $i = 1, \ldots, k$, and $\lambda_i \neq \lambda_j$ for $i \neq j$. Thus (3.2) holds with $n_i = \sum_{j=1}^{l_i} m_{ij}$ for $i = 1, \ldots, k$. Let $f(z)$ be a polynomial. Then $f(A) = \oplus_{i=j=1}^{k,l_i} f(J_{m_{ij}}(\lambda_i))$. Use the results for $J_n(\lambda)$ to deduce

$$f(A) = \oplus_{i=j=1}^{k,l_i} \sum_{r=0}^{m_{ij}-1} \frac{f^{(r)}(\lambda_i)}{r!} H_{m_{ij}}^r.$$

Let $Z_{ij} \in \mathbb{C}^{n \times n}$ be a block diagonal matrix of the following form. For each integer $l \in [1, k]$ with $l \neq i$ all the corresponding blocks to $J_{lr}(\lambda_l)$ are equal to zero. In the block corresponding to $J_{m_{ir}}(\lambda_i)$ Z_{ij} has the block matrix $H_{m_{ir}}^j$ for $j = 0, \ldots, m_i - 1$. Note that each Z_{ij} is a nonzero matrix with $0-1$ entries. Furthermore, two different Z_{ij} and $Z_{i'j'}$ do not have a common 1 entry. Hence $Z_{ij}, i = 1, \ldots, k, j = 0, \ldots, m_i - 1$ are linearly independent. It is straightforward to deduce (3.3) from the above identity.

Let $B \in \mathbb{C}^{n \times n}$. Then $B = UAU^{-1}$ where A is the Jordan canonical form of B. Recall that A and B have the same characteristic polynomial. Let $f(z) \in \mathbb{C}[z]$. Then (3.3) holds. Clearly

$$f(B) = Uf(A)U^{-1} = \sum_{i=1}^{k} \sum_{j=0}^{m_i-1} \frac{f^{(j)}(\lambda_i)}{j!} UZ_{ij}U^{-1}.$$

Hence (3.3) holds for B, where $UZ_{ij}U^{-1}, i = 1, \ldots, k, j = 0, \ldots, m_{ij-1}$ are the B-components.

The uniqueness of the A-components follows from the existence and uniqueness of the Lagrange-Sylvester interpolation polynomial as explained below.

□

Theorem 3.2 ***(The Lagrange-Sylvester interpolation polynomial).*** *Let $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ be k-distinct numbers. Let $m_1, \ldots, m_k$ be k positive integers and let $m = m_1 + \ldots + m_k$. Let $s_{ij}, i = 1, \ldots, k, j = 0, \ldots, m_i - 1$ be any m complex numbers.*

Then there exists a unique polynomial $\phi(z)$ of degree at most $m-1$ satisfying the conditions $\phi^{(j)}(\lambda_i) = s_{ij}$ for $i = 1, \dots, k, j = 0, \dots, m_i - 1$ satisfying the conditions. (For $m_i = 1, i = 1, \dots, k$ ϕ is the Lagrange interpolating polynomial.)

Proof. The Lagrange interpolating polynomial is given by the formula

$$\phi(z) = \sum_{i=1}^{k} \frac{(z-\lambda_1)\dots(z-\lambda_{i-1})(z-\lambda_{i+1})\dots(z-\lambda_k)}{(\lambda_i-\lambda_1)\dots(\lambda_i-\lambda_{i-1})(\lambda_i-\lambda_{i+1})\dots(\lambda_i-\lambda_k)} s_{i0}.$$

In the general case one determines $\phi(z)$ as follows. Let $\psi(z) := \prod_{i=1}^{k}(z-\lambda_i)^{m_i}$. Then

$$\phi(z) = \psi(z) \sum_{i=1}^{k} \sum_{j=0}^{m_i-1} \frac{t_{ij}}{(z-\lambda_i)^{m_i-j}} = \sum_{i=1}^{k} \sum_{j=0}^{m_i-1} t_{ij}(z-\lambda_i)^j \theta_i(z). \tag{3.4}$$

Here

$$\theta_i = \frac{\psi(z)}{(z-\lambda_i)^{m_i}} = \prod_{j \neq i} (z-\lambda_j)^{m_j} \text{ for } i = 1, \dots, k. \tag{3.5}$$

Observe

$$\frac{d^l \theta_i}{dz^l}(\lambda_r) = 0 \text{ for } l = 0, \dots, m_r - 1 \text{ and } r \neq i. \tag{3.6}$$

Now start to determine $t_{i0}, t_{i1}, \dots, t_{i(m_i-1)}$ recursively for each fixed value of i. This is done by using the values of $\phi(\lambda_i), \phi'(\lambda_i), \dots, \phi^{(m_i-1)}(\lambda_i)$ in the above formula for $\phi(z)$. Note that $\deg \phi \leq m-1$. It is straightforward to show that

$$t_{i0} = \frac{\phi(\lambda_i)}{\theta_i(\lambda_i)},\ t_{i1} = \frac{\phi'(\lambda_i) - t_{i0}\theta_i'(\lambda_i)}{\theta_i(\lambda_i)},\ t_{i2} = \frac{\phi''(\lambda_i) - t_{i0}\theta_i''(\lambda_i) - 2t_{i1}\theta_i'(\lambda_i)}{2\theta_i(\lambda_i)}. \tag{3.7}$$

The uniqueness ϕ is shown as follows. Assume that $\theta(z)$ is another Lagrange-Sylvester polynomial of degree less than m. Then $\omega(z) := \phi(z) - \theta(z)$ must be divisible by $(z-\lambda_i)^{m_i}$, since $\omega^{(j)}(\lambda_i) = 0$ for $j = 0, \dots, m_i - 1$, for each $i = 1, \dots, k$. Hence $\psi(z) | \omega(z)$. As $\deg \omega(z) \leq m-1$ it follows that $\omega(z)$ is the zero polynomial, i.e. $\phi(z) = \theta(z)$. □

How to find the components of A:

Assume that the minimal polynomial of A - $\psi(z)$ is given by 3.2. Then

$Z_{ij} = \phi_{ij}(A)$, where $\phi_{ij}(z)$ is L-S polynomial of degree m at most satisfying

$$\phi_{ij}(\lambda_p) = 0 \text{ for } p \neq i \text{ and } \phi_{ij}^{(q)}(\lambda_i) = j!\delta_{pj}\ j = 0, \dots, m_i - 1. \tag{3.8}$$

Indeed, assume that $\phi_{ij}(z)$ satisfies (3.8). Use (3.3) to deduce that $\phi_{ij}(A) = Z_{ij}$.

To find ϕ_{ij} do the following steps. Set

$$\phi_{ij}(z) = \frac{\theta_i(z)(z-\lambda_i)^j}{\theta_i(\lambda_i)}(1 + \sum_{l=j+1}^{m_i-1} a_l(z-\lambda_i)^{l-j}), j = 0, \dots, m_i - 2, \tag{3.9}$$

$$\phi_{i(m_i-1)}(z) = \frac{\theta_i(z)(z-\lambda_i)^{m_i-1}}{\theta_i(\lambda_i)}. \tag{3.10}$$

To find $a_{j+1}, \dots, a_{m_i-1}$ in (3.9) use the conditions

$$\phi_{ij}^{(q)}(\lambda_i) = 0 \text{ for } q = j+1, \dots, m_i - 1. \tag{3.11}$$

Namely, the condition for $q = j+1$ determines a_{q+1}. Next the condition for $q = j+2$ determines a_{q+2}. Continue in this manner to find all $a_{j+1}, \dots, a_{m_i-1}$.

We now explain briefly why these formulas hold. Since $\theta_i(z)|\phi_{ij}(z)$ it follows that $\phi_{ij}^{(q)}(\lambda_p) = 0$ for $p \neq i$ and $q = 0, \dots, m_q - 1$. (That is, the first condition of (3.8) holds.) Since $(z-\lambda_i)^j|\phi_{ij}(z)$ it follows that $\phi_{ij}^{(q)}(\lambda_i) = 0$ for $q = 0, \dots, j-1$. A straightforward calculation shows that $\phi_{ij}^{(j)}(\lambda_i) = j!$. The values of $a_{j+1}, \dots, a_{m_i-1}$ are determined by (3.11).

Example: Let A be the 4×4 matrix given in Problem 3 of §2.3. Recall that the minimal polynomial of A is $\psi(z) = z(z-1)^2$. See Problem 1b in §3.1. Hence

$$\lambda_1 = 0,\ \theta_1(z) = (z-1)^2,\ \theta_1(\lambda_1) = 1, \quad \lambda_2 = 1,\ \theta_2(z) = z,\ \theta_2(\lambda_2) = 1.$$

(3.10) yield

$$\phi_{10} = \frac{\theta_1(z)z^0}{\theta_1(0)} = (z-1)^2, \quad \phi_{21} = \frac{\theta_2(z)(z-1)^1}{\theta_2(1)} = z(z-1).$$

(3.9) yields that $\phi_{20} = z(1 + a_1(z-1))$. The value of a_1 is determined by the condition $phi'_{20}(1) = 0$. Hence $a_1 = -1$. So $\phi_{20} = z(2-z)$. To summarize

$$Z_{10} = (A-I)^2, \quad Z_{20} = A(2I-A), \quad Z_{21} = A(A-I).$$

Use (3.3) to deduce

$$f(A) = f(0)(A-I)^2 + f(1)A(A-2I) + f'(1)A(A-I), \tag{3.12}$$

for any polynomial $f(z)$.

Proof of the uniqueness of A-components. Let $\phi_{ij}(z)$ be the Lagrange-Sylvester polynomial given by (3.8). Then (3.3) yields that $Z_{ij} = \phi_{ij}(A)$. □

The following Proposition gives a computer algorithm to find Z_{ij}. (Not recommended for hand calculations!)

Proposition 3.3 *Let $A \in \mathbb{C}^{n\times n}$. Assume that the minimal polynomial $\psi(z)$ be given by (3.2) and denote $m = \deg \psi$. Then for each integers $u, v \in [1, n]$ denote by $a_{uv}^{(l)}$ and $(Z_{ij})_{uv}$ the (u,v) entries of A^l and of the A-component Z_{ij} respectively. Then $(Z_{ij})_{uv}, i = 1, \dots, k, j = 0, \dots, m_i - 1$ are the unique solutions of the following system with m unknowns*

$$\sum_{i=1}^{k} \sum_{j=0}^{m_i-1} \binom{l}{j} \lambda_i^{\max(l-j,0)} (Z_{ij})_{uv} = a_{uv}^{(l)}, \quad l = 0, \dots, m-1. \tag{3.13}$$

(Note that $\binom{l}{j} = 0$ for $j > l$.)

Proof. Consider the equality (3.3) for $f(z) = z^l$ where $l = 0, \dots, m-1$. Restricting these equalities to (u, v) entries we deduce that $(Z_{ij})_{uv}$ satisfy the system (3.13). Thus the systems (3.13) are solvable for each pair $(u, v), u, v = 1, \dots, n$. Let $X_{ij} \in \mathbb{C}^{n\times n}, i = 1, \dots, k, j = 1, \dots, m_i - 1$ such that $((X_{ij})_{uv}$ satisfy the system (3.13) for each $u, v \in [1, n]$. Hence $f(A) = \sum_{i=1}^{k} \sum_{j=0}^{m_i-1} \frac{f^{(j)}(\lambda_i)}{j!} T_{ij}$ for $f(z) = z^l$ and $l = 0, \dots, m-1$. Hence the above equality holds for any polynomial $f(z)$ of degree less than m. Apply the above formula to the Lagrange-Sylvester polynomial ϕ_{ij} as given in the proof of the uniqueness of the A-components. Then $\phi_{ij}(A) = X_{ij}$. So $X_{ij} = Z_{ij}$. Thus each system (3.13) has a unique solution. □

The algorithm for finding the A-components and its complexity.

1. (a) Set $i = 1$.
   (b) Compute and store A^i. Check if $I_n, A, \dots, A^i$ are linearly independent. If independent, set $i = i + 1$ and go to (b).
   (c) $m = i$ and express $A^m = \sum_{i=1}^{m} a_i A^{m-i}$. Then $\psi(z) = z^m - \sum_{i=1}^{m} a_i z^{m-i}$ is the minimal polynomial.
   (d) Find the k roots of $\psi(z)$ and their multiplicities: $\psi(z) = \prod_{i=1}^{k} (z - \lambda_i)^{m_i}$.
   (e) Find the A-components by solving n^2 systems (3.13).

2. The maximum complexity to find $\psi(z)$ happens when $m = n$. Then we need to compute and store $I_n, A, A^2, \dots, A^n$. So we need n^3 storage space. Viewing $I_n, A, \dots, A^i$ as row vectors arranged as $i \times n^2$ matrix $B_i \in \mathbb{C}^{i\times n^2}$, we bring B_i to a row echelon form: $C_i = U_i B_i, U_i \in C^{i\times i}$. Note that C_i is essentially upper triangular. Then we add $i+1$-th row: A^{i+1} to the B_i to obtain $C_{i+1} = U_{i+1} B_{i+1}$. ($C_i$ is $i \times i$ submatrix of C_{i+1}.) To get C_{i+1} from C_i we need $2in^2$ flops. In the case $m = n$ C_{n^2+1} has las row zero. So to find $\psi(z)$ we need at most Kn^4 flops. ($K \le 2$?). The total storage space is around $2n^3$.

   Now to find the roots of $\psi(z)$ with certain precision will take a polynomial time, depending on the precision.

   To solve n^2 systems with n variables, given in (3.13), use Gauss-Jordan for the augmented matrix $[S\ T]$. Here $S \in \mathbb{C}^{n\times n}$ stands for the coefficient of the system (3.13), depending on $\lambda_1, \dots, \lambda_k$. $T \in \mathbb{C}^{n\times n^2}$ given the "left-hand side" of n^2 systems of (3.13). One needs around n^3 storage space. Bring $[S\ T]$ to $[I_n\ Q]$ using Gauss-Jordan to find A-components. To do that we need about n^4 flops.

   In summary, we need storage of $2n^3$ and around $4n^4$ flops. (This would suffice to find the roots of $\psi(z)$ with good enough precision.)

Problems

1. Let $A \in \mathbb{C}^{4\times 4}$ be given as in Problem 3 of Section 2.3. Assume that the characteristic polynomial of A is $z^2(z-1)^2$.

   (a) Use Problem 3 of Section 2.6
   to find the Jordan canonical form of A.

(b) Assume that the minimal polynomial of A is $z(z-1)^2$. Find all the A-components.

(c) Give the explicit formula for any A^l.

2. Let $A \in \mathbb{C}^{n\times n}$ and assume that $\det(zI_n - A) = \prod_{i=1}^k (z-\lambda_i)^{n_i}$, and the minimal polynomial $\psi(z) = \prod_{i=1}^k (z-\lambda_i)^{m_i}$ where $\lambda_1, \dots, \lambda_k$ are k distinct eigenvalues of A. Let $Z_{ij}, j = 0, \dots, m_i - 1, i = 1, \dots, k$ are the A-components.

(a) Show that $Z_{ij}Z_{pq} = \mathbf{0}$ for $i \neq p$.

(b) What is the exact formula for $Z_{ij}Z_{ip}$?

3.2 Power stability, convergence and boundedness of matrices

Corollary 3.4 *Let $A \in \mathbb{C}^{n\times n}$ Assume that the minimal polynomial $\psi(z)$ be given by (3.2) and denote by $Z_{ij}, i = 1, \dots, k, j = 0, \dots, m_j - 1$ the A-components. Then for each positive integer l*

$$A^l = \sum_{i=1}^{k} \sum_{j=0}^{m_i-1} \binom{l}{j} \lambda_i^{\max(l-j,0)} Z_{ij}. \tag{3.14}$$

If we know the A-components then to compute A^l we need only around $2mn^2 \leq 2n^3$ flops! Thus we need at most $4n^4$ flops to compute A^l, including the computations of A-components, without dependence on l! (Note that $\lambda_i^j = e^{\log j\lambda_i}$.) So to find $\mathbf{x}_{10^8} = A^{10^8}\mathbf{x}_0$ discussed in the beginning of the previous section we need about 10^4 flops. So to compute $\mathbf{x}_{10^8}$ we need about $10^4 10^2$ flops compared with $10^8 10^2$ flops using the simple minded algorithm explained in the beginning of the previous section. There are much simpler algorithms to compute A^l which are roughly of the order $(\log_2 l)^2 n^3$ of computations and $(\log_2 l)^2 n^2$ ($4n^2$?) storage. See Problem ? However roundoff error remains a problem for large l.

Definition 3.5 *Let $A \in \mathbb{C}^{n\times n}$. A is called power stable if $\lim_{l\to\infty} A^l = \mathbf{0}$. A is called power convergent if $\lim_{l\to\infty} A^l = B$ for some $B \in \mathbb{C}^{n\times n}$. A is called power bounded if there exists $K > 0$ such that the absolute value of every entry of every $A^l, l = 1, \dots$ is bounded above by K.*

Theorem 3.6 *Let $A \in \mathbb{C}^{n\times n}$. Then*

1. *A is power stable if and only if each eigenvalue of A is in the interior of the unit disk: $|z| < 1$.*

2. *A is power convergent if and only if each eigenvalue λ of A satisfies one of the following conditions*

   (a) $|\lambda| < 1$;

   (b) $\lambda = 1$ and each Jordan block of the JCF of A with an eigenvalue 1 is of order 1, i.e. 1 is a simple zero of the minimal polynomial of A.

3. *A is power bounded if and only if each eigenvalue λ of A satisfies one of the following conditions*

(a) $|\lambda| < 1$;

(b) $|\lambda| = 1$ *and each Jordan block of the JCF of* A *with an eigenvalue* λ *is of order* 1, *i.e.* λ *is a simple zero of the minimal polynomial of* A.

Proof. Consider the formula (3.4). Since the A-components $Z_{ij}, i = 1, \dots, k, j = 0, \dots, m_i - 1$ are linearly independent we need to satisfy the conditions of the theorem for each term in (3.4), which is $\binom{l}{j}\lambda_i^{l-j} Z_{ij}$ for $l >> 1$. Note that for a fixed j $\lim_{l\to\infty} \binom{l}{j}\lambda_i^{l-j} = 0$ if and only if $|\lambda_i| < 1$. Hence we deduce the condition *1* of the theorem.

Note that the sequence $\binom{l}{j}\lambda_i^{l-j}, l = j, j+1, \dots,$ converges if and only if either $|\lambda_i| < 1$ or $\lambda_i = 1$ and $j = 0$. Hence we deduce the condition *2* of the theorem.

Note that the sequence $\binom{l}{j}\lambda_i^{l-j}, l = j, j+1, \dots,$ is bounded if and only if either $|\lambda_i| < 1$ or $|\lambda_i| = 1$ and $j = 0$. Hence we deduce the condition *3* of the theorem. □

Corollary 3.7 *Let* $A \in \mathbb{C}^{n\times n}$ *and consider the iterations* $\mathbf{x}_l = A\mathbf{x}_{l-1}$ *for* $l = 1, \dots$. *Then for any* $\mathbf{x}_0$

1. $\lim_{l\to\infty} \mathbf{x}_l = \mathbf{0}$ *if and only if* A *is power stable.*
2. $\mathbf{x}_l, l = 0, 1, \dots$ *converges if and only if* A *is power convergent.*
3. $\mathbf{x}_l, l = 0, 1, \dots$ *is bounded if and only if* A *is power bounded.*

Proof. If A satisfies the conditions of an item i Theorem 3.6 then the corresponding condition i of the corollary clearly holds. Assume that the conditions of an item i of the corollary holds. Choose $\mathbf{x}_0 = \mathbf{e}_j = (\delta_{1j}, \dots, \delta_{nj})^\top$ for $j = 1, \dots, n$ to deduce the corresponding condition i of Theorem 3.6. □

Theorem 3.8 *Let* $A \in \mathbb{C}^{n\times n}$ *and consider the nonhomogeneous iterations*

$$\mathbf{x}_l = A\mathbf{x}_{l-1} + \mathbf{b}_l, \quad l = 0, \dots \tag{3.15}$$

Then

1. $\lim_{l\to\infty} \mathbf{x}_l = \mathbf{0}$ *for any* $\mathbf{x}_0 \in \mathbb{C}^n$ *and any sequence* $\mathbf{b}_0, \mathbf{b}_1, \dots$ *satisfying the condition* $\lim_{l\to\infty} \mathbf{b}_l = 0$ *if and only if* A *is power stable.*
2. *The sequence* $\mathbf{x}_l, l = 0, 1, \dots$ *converges for any* $\mathbf{x}_0$ *and any sequence* $\mathbf{b}_0, \mathbf{b}_1, \dots$ *satisfying the condition* $\sum_{l=0}^{l} \mathbf{b}_l$ *converges if and only if* A *is power convergent.*
3. *The sequence* $\mathbf{x}_l, l = 0, 1, \dots$ *is bounded for any* $\mathbf{x}_0$ *and any sequence* $\mathbf{b}_0, \mathbf{b}_1, \dots$ *satisfying the condition* $\sum_{l=0}^{l} ||\mathbf{b}_l||_\infty$ *converges if and only if* A *is power bounded. (Here* $||(x_1, \dots, x_n)||_\infty = \max_{i\in[1,n]} |x_i|$.*)*

Proof. Assume that $\mathbf{b}_l = 0$. Since $\mathbf{x}_0$ is arbitrary we deduce the necessity of all the conditions from Theorem 3.6. The sufficiency of the above conditions follow from the Jordan Canonical Form of A as follows.

Let $J = U^{-1}AU$ where U is an invertible matrix and J is the Jordan canonical form of A. By letting $\mathbf{y}_l := U^{-1}\mathbf{x}_l$ and $\mathbf{c}_l = U^{-1}\mathbf{b}_l$ it is enough to prove the sufficiency part of the theorem for the case where A is sum of Jordan blocks. In this case system (3.15) reduces to independent systems of equations for each Jordan block. Thus it is left to prove the theorem when $A = J_n(\lambda)$.

1. We show that if $A = J_n(\lambda)$ and $|\lambda| < 1$, then $\lim_{l\to\infty} \mathbf{x}_l = \mathbf{0}$ for any $\mathbf{x}_0$ and $\mathbf{b}_l, l = 1, \dots$ if $\lim_{l\to\infty} \mathbf{b}_l = 0$. We prove this claim by the induction on n. For $n = 1$ (3.15) reduces to

$$x_l = \lambda x_{l-1} + b_l, \quad x_0, x_l, b_l \in \mathbb{C} \text{ for } l = 1, \dots \tag{3.16}$$

It is straightforward to show, e.g. use induction that

$$x_l = \sum_{i=0}^{l} \lambda^i b_{l-i} = b_l + \lambda b_{l-1} + \ldots + \lambda^l b_0 \quad l = 1, \dots, \text{ were } b_0 := x_0. \tag{3.17}$$

Let $\beta_m = \sup_{i\geq m} |b_i|$. Since $\lim_{l\to\infty} b_l = 0$, it follows that each β_m is finite, the sequence $\beta_m, m = 0, 1, \dots$ decreasing and $\lim_{m\to\infty} \beta_m = 0$. Fix m. Then for $l > m$

$$|x_l| \leq \sum_{i=0}^{l} |\lambda|^i |b_{l-i}| = \sum_{i=0}^{l-m} |\lambda|^i |b_{l-i}| + |\lambda|^{l-m} \sum_{j=1}^{m} |\lambda|^j ||b_{m-j}| \leq$$

$$\beta_m \sum_{i=0}^{l-m} |\lambda|^i + |\lambda|^{l-m} \sum_{j=1}^{m} |\lambda|^j ||b_{m-j}| \leq \beta_m \sum_{i=0}^{\infty} |\lambda|^i + |\lambda|^{l-m} \sum_{j=1}^{m} |\lambda|^j ||b_{m-j}| =$$

$$\frac{\beta_m}{1-|\lambda|} + |\lambda|^{l-m} \sum_{j=1}^{m} |\lambda|^j ||b_{m-j}| \to \frac{\beta_m}{1-|\lambda|} \text{ as } l \to \infty.$$

That is $\limsup_{l\to\infty} |x_l| \leq \frac{\beta_m}{1-|\lambda|}$. As $\lim_{m\to\infty} \beta_m = 0$ it follows that $\limsup_{l\to\infty} |x_l| = 0$, which is equivalent to the statement $\lim_{l\to\infty} x_l = 0$. This proves the case $n = 1$.

Assume that the theorem holds for $n = k$. Let $n = k + 1$. View $\mathbf{x}_l^\top := (x_{1,l}, \mathbf{y}_l^\top,)^\top, \mathbf{b}_l = (b_{1,l}, \mathbf{c}_l^\top,)^\top$, where $\mathbf{y}_l = (x_{2,l}, \dots, x_{k+1,l})^\top, \mathbf{c}_l \in \mathbb{C}^k$ are the vectors composed of the last k coordinates of $\mathbf{x}_l$ and $\mathbf{b}_l$ respectively. Then (3.15) for $A = J_{k+1}(\lambda)$ for the last k coordinates of $\mathbf{x}_l$ is given by the system $\mathbf{y}_l = J_k(\lambda)\mathbf{y}_{l-1} + \mathbf{c}_l$ for $l = 1, 2, \dots$. Since $\lim_{l\to\infty} \mathbf{c}_l = \mathbf{0}$ the induction hypothesis yields that $\lim_{l\to\infty} \mathbf{y}_l = \mathbf{0}$. The system (3.15) for $A = J_{k+1}(\lambda)$ for the first coordinate is $x_{1,l} = \lambda x_{1,l-1} + (x_{2,l-1} + b_{1,l})$ for $l = 1, ldots$. From induction hypothesis and the assumption that $\lim_{l\to\infty} \mathbf{b}_l = \mathbf{0}$ we deduce that $\lim_{l\to\infty} x_{2,l-1} + b_{1,l} = 0$. Hence from the case $k = 1$ we deduce that $\lim_{l\to\infty} x_{1,l} = 0$. Hence $\lim_{l\to\infty} \mathbf{x}_l = 0$. The proof of this case is concluded.

2. Assume that A satisfies the each eigenvalue λ of A satisfies the following conditions: either $|\lambda| < 1$, or $\lambda = 1$ and each Jordan block corresponding to 1 is of order 1. As we pointed out we assume that A is a direct sum of its Jordan form. So first we consider $A = J_k(\lambda)$ with $|\lambda| < 1$. Since we assumed

that $\sum_{l=1}^{\infty} \mathbf{b}_l$ converges we deduce that $\lim_{l\to\infty} \mathbf{b}_l = \mathbf{0}$. Thus, by part *1* we get that $\lim_{l\to\infty} \mathbf{x}_l = 0$.

Assume now that $A = (1) \in \mathbb{C}^{1\times 1}$. Thus we consider (3.16) with $\lambda = 1$. (3.17) yields that $x_l = \sum_{i=0}^{l} b_l$. By the assumption of the theorem $\sum_{i=1}^{\infty} \mathbf{b}_l$ converges, hence the sequence $x_l, l = 1, \dots$ converges.

3. As in the part *2* it is enough to consider the case $J_1(\lambda)$ with $|\lambda| = 1$. Note that (3.17) yields that $|x_l| \le \sum_{i=0}^{l} |b_i|$. The assumption that $\sum_{i=1}^{\infty} |\mathbf{b}_i|$ converges imply that $|x_l| \le \sum_{i=0}^{\infty} |b_i| < \infty$.

□

Remark 3.9 *The stability, convergence and boundedness of the nonhomogeneous systems:*

$$\mathbf{x}_l = A_l \mathbf{x}_{l-1},\ A_l \in \mathbb{C}^{n\times n}, \quad l = 1, \dots,$$
$$\mathbf{x}_l = A_l \mathbf{x}_{l-1} + \mathbf{b}_l,\ A_l \in \mathbb{C}^{n\times n},\ \mathbf{b}_l \in \mathbb{C}^n \quad l = 1, \dots,$$

are much harder to analyze. (If time permits we revisit these problems later on in the course.)

Problems

1. Consider the nonhomogeneous system $\mathbf{x}_l = A_l \mathbf{x}_{l-1},\ A_l \in \mathbb{C}^{n\times n}, \quad l = 1, \dots$. Assume that the sequence $A_l, l = 1, \dots$, is periodic, i.e. $A_{l+p} = A_l$ for all $l = 1, \dots$, and a fixed positive integer p.
   (a) Show that for each $\mathbf{x}_0 \in \mathbb{C}^n$ $\lim_{l\to\infty} \mathbf{x}_l = \mathbf{0}$ if and only if $B := A_p A_{p-1} \dots A_1$ is power stable.
   (b) Show that for each $\mathbf{x}_0 \in \mathbb{C}^n$ the sequence $\mathbf{x}_l, l = 1, \dots$, converges if and only if the following conditions satisfied. First, B is power convergent, i.e. $\lim_{l\to\infty} B^l = C$. Second, $A_i C = C$ for $i = 1, \dots, p$.
   (c) Find a necessary and sufficient conditions such that for each $\mathbf{x}_0 \in \mathbb{C}^n$ the sequence $\mathbf{x}_l, l = 1, \dots,$, is bounded.

3.3 e^{At} and stability of certain systems of ODE

Recall that the exponential function e^z has the MacLaurin expansion

$$e^z = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \dots = \sum_{l=0}^{\infty} \frac{z^l}{l!}.$$

Hence for each $A \in \mathbb{C}^{n\times n}$ one defines

$$e^A := I_n + A + \frac{A^2}{2} + \frac{A^3}{6} + \dots = \sum_{l=0}^{\infty} \frac{A^l}{l!}.$$

More generally, if $t \in \mathbb{C}$ then

$$e^{At} := I_n + At + \frac{A^2t^2}{2} + \frac{A^3t^3}{6} + \ldots = \sum_{l=0}^{\infty} \frac{A^l t^l}{l!}.$$

Hence e^{At} satisfies the matrix differential equation

$$\frac{e^{At}}{dt} = Ae^{At} = e^{At}A. \tag{3.18}$$

Also one has the standard identity $e^{At}e^{Au} = e^{A(t+u)}$ for any complex numbers t, u.

Proposition 3.10 *Let $A \in \mathbb{C}^{n\times n}$ and consider the system of linear system of n ordinary differential equations with constant coefficients $\frac{d\mathbf{x}(t)}{dt} = A\mathbf{x}(t)$, where $\mathbf{x}(t) \in \mathbb{C}^n$, satisfying the initial conditions $\mathbf{x}(t_0) = \mathbf{x}_0$. Then $\mathbf{x}(t) = e^{A(t-t_0)}\mathbf{x}_0$ is the unique solution to the above system. More generally, let $\mathbf{b}(t) \in \mathbb{C}^n$ be any continuous vector function on $\mathbb{R}$ and consider the nonhomogeneous system of n ordinary differential equations with the initial condition:*

$$\frac{d\mathbf{x}(t)}{dt} = A\mathbf{x}(t) + \mathbf{b}(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0. \tag{3.19}$$

Then this system has a unique solution of the form

$$\mathbf{x}(t) = e^{A(t-t_0)}\mathbf{x}_0 + \int_{t_0}^{t} e^{A(t-u)}\mathbf{b}(u)du. \tag{3.20}$$

Proof. The uniqueness of the solution of (3.19) follows from the uniqueness of solutions to system of ODE (Ordinary Differential Equations). The first part of the proposition follows from (3.18). To deduce the second part one does the *variations of parameters*. Namely one tries a solution $x(t) = e^{A(t-t_0)}\mathbf{y}(t)$ where $\mathbf{y}(t) \in \mathbb{C}^n$ is unknown vector function. Hence

$$\mathbf{x}' = (e^{A(t-t_0)})'\mathbf{y}(t) + e^{A(t-t_0)}\mathbf{y}'(t) = Ae^{A(t-t_0)}\mathbf{y}(t) + e^{A(t-t_0)}\mathbf{y}'(t) = A\mathbf{x}(t) + e^{A(t-t_0)}\mathbf{y}'(t).$$

Substitute this expression of $\mathbf{x}(t)$ to (3.19) to deduce the differential equation $\mathbf{y}' = e^{-A(t-t_0)}\mathbf{b}(t)$. Since $\mathbf{y}(t_0) = \mathbf{x}_0$ this simple equation have a unique solution $\mathbf{y}(t) = \mathbf{x}_0 + \int_{t_0}^{u} e^{A(u-t_0)}\mathbf{b}(u)du$. Now multiply by $e^{A(t-t_0)}$ and use the fact that $e^{At}e^{Au} = e^{A(u+v)}$ to deduce (3.20). □

Note: The second term in the formula (3.20) can be considered as a perturbation term to the solution $\frac{d\mathbf{x}(t)}{dt} = A\mathbf{x}(t), \mathbf{x}(t_0) = \mathbf{x}_0$, i.e. to the system (3.19) with $\mathbf{b}(t) \equiv \mathbf{0}$.

Use (3.3) for e^{zt} and the observation that $\frac{d^j e^{zt}}{dz^j} = t^j e^{zt}, j = 0, 1, \ldots$ to deduce:

$$e^{At} = \sum_{j=1}^{k} \sum_{j=0}^{m_i-1} \frac{t^j e^{\lambda_i t}}{j!} Z_{ij}. \tag{3.21}$$

We can substitute this expression for e^{At} in (3.20) to get a simple expression of the solution $\mathbf{x}(t)$ of (3.19).

Definition 3.11 *Let $A \in \mathbb{C}^{n\times n}$. A is called exponentially stable, or simple stable, if $\lim_{t\to\infty} e^{At} = \mathbf{0}$. A is called exponentially convergent if $\lim_{t\to\infty} e^{At} = B$ for some $B \in \mathbb{C}^{n\times n}$. A is called exponentially bounded if there exists $K > 0$ such that the absolute value of every entry of every $e^{At}, t \in [0,\infty)$ is bounded above by K.*

Theorem 3.12 *Let $A \in \mathbb{C}^{n\times n}$. Then*

1. *A is stable if and only if each eigenvalue of A is in the left half of the complex plane: $\Re z < 0$.*
2. *A is exponentially convergent if and only if each eigenvalue λ of A satisfies one of the following conditions*
   (a) *$\Re\lambda < 0$;*
   (b) *$\lambda = 2\pi l\sqrt{-1}$ for some integer l, and each Jordan block of the JCF of A with an eigenvalue λ is of order 1, i.e. λ is a simple zero of the minimal polynomial of A.*
3. *A is exponentially bounded if and only if each eigenvalue λ of A satisfies one of the following conditions*
   (a) *$\Re\lambda < 0$;*
   (b) *$\Re\lambda = 0$ and each Jordan block of the JCF of A with an eigenvalue λ is of order 1, i.e. λ is a simple zero of the minimal polynomial of A.*

Proof. Consider the formula (3.21). Since the A-components $Z_{ij}, i = 1,\dots,k, j = 0,\dots,m_i-1$ are linearly independent we need to satisfy the conditions of the theorem for each term in (3.21), which is $\frac{t^j}{j!}e^{\lambda_i t}Z_{ij}$. Note that for a fixed j $\lim_{t\to\infty}\frac{t^j}{j!}e^{\lambda_i t} = 0$ if and only if $\Re\lambda_i < 0$. Hence we deduce the condition *1* of the theorem.

Note that the function $\frac{t^j}{j!}e^{\lambda_i t}$ converges as $t \to \infty$ if and only if either $\Re\lambda_i < 0$ or $e^{\lambda_i} = 1$ and $j = 0$. Hence we deduce the condition *2* of the theorem.

Note that the function $\frac{t^j}{j!}e^{\lambda_i t}$ is bounded for $t \geq 0$ if and only if either $\Re\lambda_i < 0$ or $|e^{\lambda_i}| = 1$ and $j = 0$. Hence we deduce the condition *3* of the theorem.

□

Corollary 3.13 *Let $A \in \mathbb{C}^{n\times n}$ and consider the system of differential equations $\frac{d\mathbf{x}(t)}{dt} = A\mathbf{x}(t), \mathbf{x}(t_0) = \mathbf{x}_0$. Then for any $\mathbf{x}_0$*

1. *$\lim_{t\to\infty}\mathbf{x}(t) = \mathbf{0}$ if and only if A is stable.*
2. *$\mathbf{x}(t)$ converges as $t \to \infty$ if and only if A is exponentially convergent.*
3. *$\mathbf{x}(t), t \in [0,\infty)$ is bounded if and only if A is exponentially bounded.*

Theorem 3.14 *Let $A \in \mathbb{C}^{n\times n}$ and consider the system of differential equations (3.19). Then for any $\mathbf{x}_0 \in \mathbb{C}^n$*

1. *$\lim_{t\to\infty}\mathbf{x}(t) = \mathbf{0}$ for any continuous function $\mathbf{b}(t)$, such that $\lim_{t\to\infty}\mathbf{b}(t) = \mathbf{0}$, if and only if A is stable.*

2. $\mathbf{x}(t)$ *converges as* $t \to \infty$ *for any continuous function* $\mathbf{b}(t)$*, such that* $\int_{t_0}^{\infty} \mathbf{b}(u)du$ *converges, if and only if* A *is exponentially convergent.*

3. $\mathbf{x}(t), t \in [0, \infty)$ *is bounded for any continuous function* $\mathbf{b}(t)$*, such that* $\int_{t_0}^{\infty} |\mathbf{b}(u)|du$ *converges if and only if* A *is exponentially bounded.*

Proof. The necessity of the conditions of the theorem follow from Corollary 3.13 by choosing $\mathbf{b}(t) \equiv \mathbf{0}$.

1. Suppose that A is stable. Then Corollary 3.13 yields that $\lim_{t\to\infty} e^{At}\mathbf{x}_0 = \mathbf{0}$. Thus show that $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{0}$, it is enough to show that the second term in (3.20) tends to $\mathbf{0}$. Use (3.21) to show that it is enough to demonstrate that

$$\lim_{t\to\infty} \int_{t_0}^{t} (t-u)^j e^{\lambda(t-u)} g(u)du = 0, \text{ where } \Re\lambda < 0,$$

for any continuous $g(t) \in [t_0, \infty)$, such that $\lim g(t) = 0$. For $\epsilon > 0$ there exists $T = T(\epsilon)$ such that $|g(t)| \le \epsilon$ for $t \ge T(\epsilon)$. Let $t > T(\epsilon)$. Then

$$\begin{aligned} |\int_{t_0}^{t} (t-u)^j e^{\lambda(t-u)} g(u)du| &= |\int_{t_0}^{T(\epsilon)} (t-u)^j e^{\lambda(t-u)} g(u)du + \int_{T(\epsilon)}^{t} (t-u)^j e^{\lambda(t-u)} g(u)du| \\ &\le |\int_{t_0}^{T(\epsilon)} (t-u)^j e^{\lambda(t-u)} g(u)du| + |\int_{T(\epsilon)}^{t} (t-u)^j e^{\lambda(t-u)} g(u)du| \\ &\le |\int_{t_0}^{T(\epsilon)} (t-u)^j e^{\lambda(t-u)} g(u)du| + \epsilon \int_{T(\epsilon)}^{t} (t-u)^j e^{\Re\lambda(t-u)} du. \end{aligned}$$

Consider the first term in the last inequality. Since $\lim_{t\to\infty} t^j e^{\lambda t} = 0$ it follows that the first term converges to zero. The second term bounded by ϵK for $K := \int_0^{\infty} t^j e^{\Re\lambda t} dt$. Hence as $\epsilon \to 0$ we deduce that $\lim_{t\to\infty} \int_{t_0}^{t} (t-u)^j e^{\lambda(t-u)} g(u)du = 0$.

2. Using part *1* we deduce the result for any eigenvalue λ with $\Re\lambda < 0$. It is left to discuss the case $\lambda = 0$. We assume that the Jordan blocks of A correspond to $\lambda = 0$ are of length one. So the A-component corresponding to $\lambda = 0$ is Z_{10}. The corresponding term is

$$\dots$$

4 Inner product spaces

4.1 Inner product

Definition 4.1 *Let* $\mathbb{F} = \mathbb{R}, \mathbb{C}$ *and let* $\mathbf{V}$ *be a vector space over* $\mathbb{F}$*. Then* $\langle \cdot, \cdot \rangle : \mathbf{V} \times \mathbf{V} \to \mathbb{F}$ *is called an inner product if the following conditions hold:*

(a) $\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a\langle \mathbf{x}, \mathbf{z} \rangle + b\langle \mathbf{y}, \mathbf{z} \rangle$, for all $a, b \in \mathbb{F}$, $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{V}$,

(br) for $\mathbb{F} = \mathbb{R}$ $\langle \mathbf{y}, \mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$, for all $\mathbf{x}, \mathbf{y} \in \mathbf{V}$;

(bc) for $\mathbb{F} = \mathbb{C}$ $\langle \mathbf{y}, \mathbf{x} \rangle = \overline{\langle \mathbf{x}, \mathbf{y} \rangle}$, for all $\mathbf{x}, \mathbf{y} \in \mathbf{V}$;

(c) $\langle \mathbf{x}, \mathbf{x} \rangle > 0$ for all $\mathbf{x} \in \mathbf{V} \backslash \{\mathbf{0}\}$.

$||\mathbf{x}|| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ *is called the norm (length) of* $\mathbf{x} \in \mathbf{V}$.

Other standard properties of inner products are mentioned in Problems 1-2. We will use the abbreviation IPS for inner product space. In this chapter we assume that $\mathbb{F} = \mathbb{R}, \mathbb{C}$ unless stated otherwise.

Proposition 4.2 *Let $\mathbf{V}$ be a vector space over $\mathbb{R}$. Identify $\mathbf{V}_c$ with the set of pairs $(\mathbf{x}, \mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \mathbf{V}$. Then $\mathbf{V}_c$ is a vector space over $\mathbb{C}$ with*

$$(a + \sqrt{-1}b)(\mathbf{x}, \mathbf{y}) := a(\mathbf{x}, \mathbf{y}) + b(-\mathbf{y}, \mathbf{x}), \quad \text{for all } a, b \in \mathbb{R},\ \mathbf{x}, \mathbf{y} \in \mathbf{V}.$$

If $\mathbf{V}$ has a basis $\mathbf{e}_1, ..., \mathbf{e}_n$ over $\mathbb{F}$ then $(\mathbf{e}_1, \mathbf{0}), ..., (\mathbf{e}_n, \mathbf{0})$ is a basis of $\mathbf{V}_c$ over $\mathbb{C}$. Any inner product $\langle \cdot, \cdot \rangle$ on $\mathbf{V}$ over $\mathbb{R}$ induces the following inner product on $\mathbf{V}_c$:

$$\langle (\mathbf{x}, \mathbf{y}), (\mathbf{u}, \mathbf{v}) \rangle = \langle \mathbf{x}, \mathbf{u} \rangle + \langle \mathbf{y}, \mathbf{v} \rangle + \sqrt{-1}(\langle \mathbf{y}, \mathbf{u} \rangle - \langle \mathbf{x}, \mathbf{v} \rangle),\ \mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v} \in \mathbf{V}.$$

We leave the proof of this proposition to the reader (Problem 3).

Definition 4.3 *Let $\mathbf{V}$ be an IPS. Then*
(a) $\mathbf{x}, \mathbf{y} \in \mathbf{V}$ are called orthogonal if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.
(b) $S, T \subset \mathbf{V}$ are called orthogonal if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ for any $\mathbf{x} \in S,\ \mathbf{y} \in T$.
(d) For any $S \subset \mathbf{V}$, $S^{\perp} \subset \mathbf{V}$ is the maximal orthogonal set to S.
(e) $\mathbf{x}_1, ..., \mathbf{x}_m$ is called an orthonormal set if

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \delta_{ij}, \quad i, j = 1, ..., m.$$

(f) $\mathbf{x}_1, ..., \mathbf{x}_n$ is called an orthonormal basis if it is an orthonormal set which is a basis in $\mathbf{V}$.

Definition 4.4 *(**Gram-Schmidt algorithm**.) Let $\mathbf{V}$ be an IPS and $S = \{\mathbf{x}_1, ..., \mathbf{x}_m\} \subset \mathbf{V}$ a finite (possibly empty) set ($m \geq 0$). Then $\tilde{S} = \{\mathbf{e}_1, ..., \mathbf{e}_p\}$ is the orthonormal set ($p \geq 1$) or the empty set ($p = 0$) obtained from S using the following recursive steps:*
(a) If $\mathbf{x}_1 = 0$ remove it from S. Otherwise replace $\mathbf{x}_1$ by $||\mathbf{x}_1||^{-1}\mathbf{x}_1$.
(b) Assume that $\mathbf{x}_1, ..., \mathbf{x}_k$ is an orthonormal set and $1 \leq k < m$. Let $\mathbf{y}_{k+1} = \mathbf{x}_{k+1} - \sum_{i=1}^{k} \langle \mathbf{x}_{k+1}, \mathbf{x}_i \rangle \mathbf{x}_i$. If $\mathbf{y}_{k+1} = 0$ remove $\mathbf{x}_{k+1}$ from S. Otherwise replace $\mathbf{x}_{k+1}$ by $||\mathbf{y}_{k+1}||^{-1}\mathbf{y}_{k+1}$.

Corollary 4.5 *Let $\mathbf{V}$ be an IPS and $S = \{\mathbf{x}_1, ..., \mathbf{x}_n\} \subset \mathbf{V}$ be n linearly independent vectors. Then the Gram-Schmidt algorithm on S is given as follows:*

$$\begin{aligned}
&\mathbf{y}_1 := \mathbf{x}_1,\ r_{11} := ||\mathbf{y}_1||,\ \mathbf{e}_1 := \frac{\mathbf{y}_1}{r_{11}}, \\
&r_{ji} := \langle \mathbf{x}_i, \mathbf{e}_j \rangle,\ j = 1, ..., i-1, \\
&\mathbf{p}_{i-1} := \sum_{j=1}^{i-1} r_{ji}\mathbf{e}_j, \quad \mathbf{y}_i := \mathbf{x}_i - \mathbf{p}_{i-1}, \\
&r_{ii} := ||\mathbf{y}_i||,\ \mathbf{e}_i := \frac{\mathbf{y}_i}{r_{ii}},\ i = 2, ..., n.
\end{aligned} \tag{4.1}$$

In particular, $\mathbf{e}_i \in S_i$ and $||\mathbf{y}_i|| = \operatorname{dist}(\mathbf{x}_i, S_{i-1})$, where $S_i = \operatorname{span}(\mathbf{x}_1, ..., \mathbf{x}_i)$ for $i = 1, ..., n$ and $S_0 = \{\mathbf{0}\}$. (See Problem 4 for the definition of $\operatorname{dist}(\mathbf{x}_i, S_{i-1})$.)

Corollary 4.6 *Any (ordered) basis in a finite dimensional IPS $\mathbf{V}$ induces an orthonormal basis by the Gram-Schmidt algorithm.*

See Problem 4 for some known properties related to the above notions.

4.2 Explanation of G-S process in standard Euclidean space

First observe that $r_{i(k+1)} := \langle \mathbf{x}_{k+1}, \mathbf{e}_i \rangle$ is the scalar projection of $\mathbf{x}_{k+1}$ on $\mathbf{e}_i$. Next observe that $\mathbf{p}_k$ is the projection of $\mathbf{x}_{k+1}$ on $\text{span}(\mathbf{e}_1, \ldots, \mathbf{e}_k) = \text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$. Hence $\mathbf{y}_{k+1} = \mathbf{x}_{k+1} - \mathbf{p}_k \perp \text{span}(\mathbf{e}_1, \ldots, \mathbf{e}_k)$. Thus $r_{(k+1)(k+1)} = ||\mathbf{x}_{k+1} - \mathbf{p}_k||$ is the distance of $\mathbf{x}_{k+1}$ to $\text{span}(\mathbf{e}_1, \ldots, \mathbf{e}_k) = \text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$. The assumption that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent yields that $r_{(k+1)(k+1)} > 0$. Hence $\mathbf{e}_{k+1} = r^{-1}_{(k+1)(k+1)}(\mathbf{x}_{k+1} - \mathbf{p}_k)$ is a vector of unit length orthogonal to $\mathbf{e}_1, \ldots, \mathbf{e}_k$.

4.3 An example of G-S process

Let $\mathbf{x}_1 = (1,1,1,1)^\top, \mathbf{x}_2 = (-1,4,4,-1)^\top$, $\mathbf{x}_3 = (4,-2,2,0)^\top$. Then

$$
\begin{aligned}
&r_{11} = ||\mathbf{x}_1|| = 2, \mathbf{e}_1 = \frac{1}{r_{11}}\mathbf{x}_1 = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})^\top, \\
&r_{12} = \mathbf{e}_1^\top \mathbf{x}_2 = 3, \mathbf{p}_1 = r_{12}\mathbf{e}_1 = 3\mathbf{e}_1 = (\frac{3}{2}, \frac{3}{2}, \frac{3}{2}, \frac{3}{2})^\top, \\
&\mathbf{y}_2 = \mathbf{x}_2 - \mathbf{p}_1 = (-\frac{5}{2}, \frac{5}{2}, \frac{5}{2}, -\frac{5}{2})^\top, r_{22} = ||\mathbf{y}_2|| = 5, \\
&\mathbf{e}_2 = \frac{1}{r_{22}}(\mathbf{y}_2) = (-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2})^\top, \\
&r_{13} = \mathbf{e}_1^\top \mathbf{x}_3 = 2, r_{23} = \mathbf{e}_2^\top \mathbf{x}_3 = -2, \\
&\mathbf{p}_2 = r_{13}\mathbf{e}_1 + r_{23}\mathbf{e}_2 = (2,0,0,2)^\top, \\
&\mathbf{y}_3 = \mathbf{x}_3 - \mathbf{p}_2 = (2,-2,2,-2)^\top, r_{33} = ||\mathbf{y}_3|| = 4, \\
&\mathbf{e}_3 = \frac{1}{r_{33}}(\mathbf{x}_3 - \mathbf{p}_2) = (\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2})^\top.
\end{aligned}
$$

4.4 QR Factorization

Let $A = [\mathbf{a}_1\ \mathbf{a}_2 \ldots \mathbf{a}_n] \in \mathbb{R}^{m \times n}$ and assume that rank $A = n$. (I.e. the columns of A are linearly independent.) Perform G-S process with the book keeping as above:

- $r_{11} := ||\mathbf{a}_1||$, $\mathbf{e}_1 := \frac{1}{r_{11}}\mathbf{a}_1$.
- Assume that $\mathbf{e}_1, \ldots, \mathbf{e}_{k-1}$ were computed. Then $r_{ik} := \mathbf{e}_i^\top \mathbf{a}_k$ for $i = 1, \ldots, k-1$. $\mathbf{p}_{k-1} := r_{1k}\mathbf{e}_1 + r_{2k}\mathbf{e}_2 + \ldots r_{(k-1)k}\mathbf{e}_{k-1}$ and $r_{kk} := ||\mathbf{a}_k - \mathbf{p}_{k-1}||$, $\mathbf{e}_k := \frac{1}{r_{kk}}(\mathbf{a}_k - \mathbf{p}_{k-1})$ for $k = 2, ..., n$.

Let $Q = [\mathbf{e}_1\ \mathbf{e}_2 \ldots \mathbf{e}_n] \in \mathbb{R}^{m \times n}$ and $R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \ldots & r_{1n} \\ 0 & r_{22} & r_{23} & \ldots & r_{2n} \\ 0 & 0 & r_{33} & \ldots & r_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & r_{nn} \end{bmatrix}$

Then $A = QR$, $Q^\top Q = I_n$ and $A^\top A = R^\top R$. The Least Squares Solution of $A\mathbf{x} = \mathbf{b}$ is given by the upper triangular system $R\hat{\mathbf{x}} = Q^\top \mathbf{b}$ which can be solved by back substitution. Formally $\hat{\mathbf{x}} = R^{-1}Q^\top \mathbf{b}$.

Proof $A^\top A\mathbf{x} = R^\top Q^\top QR\mathbf{x} = R^\top R\mathbf{x} = A^\top \mathbf{b} = R^\top Q^\top \mathbf{b}$. Multiply from left by $(R^\top)^{-1}$ to get $R\hat{\mathbf{x}} = Q^\top \mathbf{b}$

Note: $QQ^\top \mathbf{b}$ is the projection of $\mathbf{b}$ on the columns space of A. The matrix $P := QQ^\top$ is called an orthogonal projection. It is symmetric and $P^2 = P$, as $(QQ^\top)(QQ^\top) = Q(Q^\top Q)Q^\top = Q(I)Q^\top = QQ^\top$. Note $QQ^\top : \mathbb{R}^m \to \mathbb{R}^m$ is the orthogonal projection.

Equivalently: The assumption that rank $A = n$ is equivalent to the assumption that $A^\top A$ is invertible. So the LSS $A^\top A\hat{x} = A^\top \mathbf{b}$ has unique solution $\hat{\mathbf{x}} = (A^\top A)^{-1}\mathbf{b}$. Hence the projection of $\mathbf{b}$ on the column space of A is $P\mathbf{b} = A\hat{\mathbf{x}} = A(A^\top A)^{-1}A^\top \mathbf{b}$. Hence $P = A(A^\top A)^{-1}A^\top$.

4.5 An example of QR algorithm

Let $A = [\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3] = \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix}$ be the matrix corresponding to the Example of G-S algorithm §4.3. Then

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix} = \begin{bmatrix} 2 & 3 & 2 \\ 0 & 5 & -2 \\ 0 & 0 & 4 \end{bmatrix},$$

$$Q = [\mathbf{q}_1\ \mathbf{q}_2\ \mathbf{q}_3] = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

(*Explain why in this example* $A = QR$!) **Note** $QQ^\top : \mathbb{R}^4 \to \mathbb{R}^4$ is the projection on $\text{span}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$.

Remark 4.7 *It is known, e.g. [1] that the Gram-Schmidt process as described in (4.1) is numerically unstable. That is, there is a severe loss of orthogonality of* $\mathbf{y}_1, \ldots$ *as we proceed to compute* $\mathbf{y}_i$. *In computations one uses either a modified GSP or Householder orthogonalization [1].*

Definition 4.8 *(***Modified Gram-Schmidt algorithm***.) Let* $\mathbf{V}$ *be an IPS and* $S = \{\mathbf{x}_1, ..., \mathbf{x}_m\} \subset \mathbf{V}$ *a finite (possibly empty) set (*$m \geq 0$*). Then* $\tilde{S} = \{\mathbf{e}_1, ..., \mathbf{e}_p\}$ *is the orthonormal set (*$p \geq 1$*) or the empty set (*$p = 0$*) obtained from* S *using the following recursive steps:*

1. *Initialize* $j = 1$ *and* $p = m$.
2. *If* $\mathbf{x}_j \neq \mathbf{0}$ *goto* 3. *Otherwise replace* p *by* $p - 1$. *If* $j > p$ *exit. Replace* $\mathbf{x}_i$ *by* $\mathbf{x}_{i+1}$ *for* $i = j, \ldots, p$. *Repeat.*
3. $\mathbf{e}_j := \frac{1}{\|\mathbf{x}_j\|}\mathbf{x}_j$.
4. $j = j + 1$. *If* $j > p$ *exit.*
5. *For* $i = 1, \ldots, j - 1$ *let* $\mathbf{x}_j := \mathbf{x}_j - \langle \mathbf{x}_j, \mathbf{e}_i \rangle \mathbf{e}_i$.
6. *Goto* 2.

MGS algorithm is stable, needs mn^2 flops, which is more time consuming then GS algorithm.

Lemma 4.9 *Let $\mathbf{V}$ be a finite dimensional IPS over $\mathbb{F} = \mathbb{R}$. Let $\mathbf{U}$ be a subspace of $\mathbf{V}$. Then*

$$\mathbf{V} = \mathbf{U} \oplus \mathbf{U}^{\perp} \tag{4.2}$$

Proof. If $\mathbf{U} = \mathbf{V}$ or $\mathbf{U} = \{\mathbf{0}\}$ the lemma is trivial. So we assume that $n = \dim \mathbf{V} > m = \dim \mathbf{U} \geq 1$. Choose a basis $\mathbf{u}_1, \ldots, \mathbf{u}_m$. Complete this basis to a basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Perform the Gram-Schmidt process on $\mathbf{u}_1, \ldots, \mathbf{u}_n$ to obtain an orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Recall that $\operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_i) = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_i)$ for $i = 1, \ldots, n$. Hence $\mathbf{U} = \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_m) = \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$. I.e. $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is an orthonormal basis in $\mathbf{U}$. Clearly, $\mathbf{y} = \sum_{i=1}^{n} \langle \mathbf{y}, \mathbf{v}_i \rangle \mathbf{v}_i \in \mathbf{U}^{\perp}$ if and only if $\mathbf{y} = \sum_{i=m+1}^{n} a_i \mathbf{v}_i$. I.e. $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n$ is a basis in $\mathbf{U}^{\perp}$. So when we write a vector $\mathbf{z} = \sum_{i=1}^{n} z \mathbf{v}_i$ it is of unique form $\mathbf{u} + \mathbf{w}, \mathbf{u} = \sum_{i=1}^{m} z_i \mathbf{v}_i \in \mathbf{U}, \mathbf{w} = \sum_{i=m+1}^{n} z_i \mathbf{v}_i \in \mathbf{U}^{\perp}$. □

Corollary 4.10 *Let the assumptions of Lemma 4.9 hold. Then $(\mathbf{U}^{\perp})^{\perp} = \mathbf{U}$.*

Lemma 4.11 *(Fredholm alternative) Let $A \in \mathbb{C}^{m \times n}$. Then the system*

$$A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \in \mathbb{C}^n, \ \mathbf{b} \in \mathbb{C}^m \tag{4.3}$$

is solvable if and only if for each $\mathbf{y} \in \mathbb{C}^m$ satisfying $\mathbf{y}^{\top} A = \mathbf{0}$ the equality $\mathbf{y}^{\top} \mathbf{b} = 0$ holds.

Proof. Suppose that (4.3) solvable. Then $\mathbf{y}^{\top} \mathbf{b} = \mathbf{y}^{\top} A \mathbf{x} = \mathbf{0}^{\top} \mathbf{x} = 0$. Observe next that $\mathbf{y}^{\top} A = \mathbf{0}^{\top}$ if and only if $\bar{\mathbf{y}} \in R(A)^{\perp}$. As $(R(A)^{\perp})^{\perp} = R(A)$ it follows that if $\mathbf{y}^{\top} \mathbf{b} = 0$ for each $\mathbf{y}^{\top} A = \mathbf{0}$ then $\mathbf{b} \in R(A)$, i.e. (4.3) is solvable. □

Theorem 4.12 *(The least squares theorem) Consider the system* (4.3)*. Multiply the both sides of this system by A^* to obtain the least squares system corresponding to* (4.3)*:*

$$A^* A \mathbf{x} = A^* \mathbf{b}, \quad A \in \mathbb{C}^{m \times n}, \mathbf{b} \in \mathbb{C}^m. \tag{4.4}$$

Then (4.4) *is always solvable. For each $\mathbf{x}_0$ the vector $A\mathbf{x}_0$ is the orthogonal projection of $\mathbf{b}$ on $R(A)$. That is $A\mathbf{x}_0 \in R(A)$ and $\mathbf{b} - A\mathbf{x}_0 \in R(A)^{\perp}$. Furthermore,*

$$\|\mathbf{b} - \mathbf{z}\| \geq \|\mathbf{b} - A\mathbf{x}_0\| \text{ for any } \mathbf{z} \in R(A). \tag{4.5}$$

Equality holds if and only if $\mathbf{z} = A\mathbf{x}_0$.

Proof. Suppose that $\mathbf{z}^{\top} A^* A = \mathbf{0}$. Then $0 = \mathbf{z}^{\top} A^* A \bar{\mathbf{z}} = \|\mathbf{z}^{\top} A^*\|^2$. Hence $\mathbf{z}^{\top} A^* = \mathbf{0}$. In particular $\mathbf{z}^{\top} A^* \mathbf{b} = 0$. Lemma 4.11 yields that (4.4) is solvable. Let $\mathbf{x}_0$ be a solution of (4.4). So $A^*(\mathbf{b} - A\mathbf{x}_0) = \mathbf{0}$. Since the columns of A span $R(A)$ it follows that $\mathbf{b} - A\mathbf{x}_0 \in R(A)^{\perp}$. Clearly, $A\mathbf{x}_0 \in R(A)$. Hence $A\mathbf{x}_0$ is the orthogonal projection on the range of A. Let $\mathbf{z} \in R(A)$. Use the condition that $\mathbf{b} - A\mathbf{x}_0 \in R(A)^{\perp}$ to deduce

$$\|\mathbf{b} - \mathbf{z}\|^2 = \|(\mathbf{b} - A\mathbf{x}_0) + (A\mathbf{x}_0 - \mathbf{z})\|^2 = \|\mathbf{b} - A\mathbf{x}_0\|^2 + \|A\mathbf{x}_0 - \mathbf{z}\|^2 \geq \|\mathbf{b} - A\mathbf{x}_0\|^2. \tag{4.6}$$

Equality holds if and only if $\mathbf{z} = A\mathbf{x}_0$. $\square$

Problems

1. Let $\mathbf{V}$ be an IPS over $\mathbb{F}$. Show

$$\begin{aligned}
&\langle \mathbf{0}, \mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{0} \rangle = 0, \\
&\text{for } \mathbb{F} = \mathbb{R} \ \langle \mathbf{z}, a\mathbf{x} + b\mathbf{y} \rangle = a\langle \mathbf{z}, \mathbf{x} \rangle + b\langle \mathbf{z}, \mathbf{y} \rangle, \text{ for all } a, b \in \mathbb{R}, \ \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{V}, \\
&\text{for } \mathbb{F} = \mathbb{C} \ \langle \mathbf{z}, a\mathbf{x} + b\mathbf{y} \rangle = \bar{a}\langle \mathbf{z}, \mathbf{x} \rangle + \bar{b}\langle \mathbf{z}, \mathbf{y} \rangle, \text{ for all } a, b \in \mathbb{C}, \ \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{V}.
\end{aligned}$$

2. Let $\mathbf{V}$ be an IPS. Show

   (a) $||a\mathbf{x}|| = |a| \ ||\mathbf{x}||$ for $a \in \mathbb{F}$ and $\mathbf{x} \in \mathbf{V}$.

   (b) The Cauchy-Schwarz inequality:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq ||\mathbf{x}|| \ ||\mathbf{y}||,$$

   and equality holds if and only if $\mathbf{x}, \mathbf{y}$ are linearly dependent (collinear).

   (c) The triangle inequality

$$||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||,$$

   and equality holds if either $\mathbf{x} = 0$ or $\mathbf{y} = a\mathbf{x}$ for $a \in \mathbb{R}_+$.

3. Prove Proposition 4.2.

4. Let $\mathbf{V}$ be a finite dimensional IPS of dimension n. Assume that $S \subset \mathbf{V}$. Show

   (a) If $\mathbf{x}_1, ..., \mathbf{x}_m$ is an orthonormal set then $\mathbf{x}_1, ..., \mathbf{x}_m$ are linearly independent.

   (b) Assume that $\mathbf{e}_1, ..., \mathbf{e}_n$ is an orthonormal basis in $\mathbf{V}$. Show that for any $\mathbf{x} \in \mathbf{V}$ the orthonormal expansion holds

$$\mathbf{x} = \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{e}_i \rangle \mathbf{e}_i. \tag{4.7}$$

   Furthermore for any $\mathbf{x}, \mathbf{y} \in \mathbf{V}$

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{e}_i \rangle \overline{\langle \mathbf{y}, \mathbf{e}_i \rangle}. \tag{4.8}$$

   (c) Assume that S is a finite set. Let $\tilde{S}$ be the set obtained by the Gram-Schmidt process. Show that $\tilde{S} = \emptyset \iff \text{span} S = \{\mathbf{0}\}$. Show that if $\tilde{S} \neq \emptyset$ then $\mathbf{e}_1, ..., \mathbf{e}_p$ is an orthonormal basis in span S.

   (d) There exists an orthonormal basis $\mathbf{e}_1, ..., \mathbf{e}_n$ in $\mathbf{V}$ and $0 \leq m \leq n$ such that

$$\begin{aligned}
&\mathbf{e}_1, ..., \mathbf{e}_m \in S, \quad \text{span } S = \text{span}(\mathbf{e}_1, ..., \mathbf{e}_m), \\
&S^{\perp} = \text{span}(\mathbf{e}_{m+1}, ..., \mathbf{e}_n), \\
&(S^{\perp})^{\perp} = \text{span} S.
\end{aligned}$$

(e) Assume from here to the end of the problem that S is a subspace. Show $\mathbf{V} = S \oplus S^{\perp}$.

(f) Let $\mathbf{x} \in \mathbf{V}$ and let $\mathbf{x} = \mathbf{u} + \mathbf{v}$ for unique $\mathbf{u} \in S,\ \mathbf{v} \in S^{\perp}$. Let $P(\mathbf{x}) := \mathbf{u}$ be the projection of $\mathbf{x}$ on S. Show that $P : \mathbf{V} \to \mathbf{V}$ is a linear transformation satisfying

$$P^2 = P, \quad \text{Range } P = S, \quad \text{Ker } P = S^{\perp}.$$

(g) Show

$$\begin{aligned} &\text{dist}(\mathbf{x}, S) := ||\mathbf{x} - P\mathbf{x}|| \leq ||\mathbf{x} - \mathbf{w}|| \text{ for any } \mathbf{w} \in S \\ &\text{and equality } \iff \mathbf{w} = P\mathbf{x}. \end{aligned} \tag{4.9}$$

(h) Show that $\text{dist}(\mathbf{x}, S) = ||\mathbf{x} - \mathbf{w}||$ for some $\mathbf{w} \in S$ if and only if $\mathbf{x} - \mathbf{w}$ is orthogonal to S.

(i) Let $\mathbf{e}_1, \ldots, \mathbf{e}_m$ be an orthonormal basis of S. Show that for each $\mathbf{x} \in \mathbf{V}$ $P\mathbf{x} = \sum_{i=1}^{p} \langle \mathbf{y}, \mathbf{e}_i \rangle \mathbf{e}_i$.

(**Note**: $P\mathbf{x}$ is called *the least square approximation* to $\mathbf{x}$ in the subspace S.)

5. Let $X \in \mathbb{C}^{m \times n}$ and assume that $m \geq n$ and rank $X = n$. Let $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{C}^m$ be the columns of X, i.e. $X = (\mathbf{x}_1, ..., \mathbf{x}_n)$. Assume that $\mathbb{C}^m$ is an IPS with the standard inner product $< \mathbf{x}, \mathbf{y} >= \mathbf{y}^* \mathbf{x}$. Perform the Gram-Schmidt algorithm (4.5) to obtain the matrix $Q = (\mathbf{e}_1, ..., \mathbf{e}_n) \in \mathbb{C}^{m \times n}$. Let $R = (r_{ji})_1^n \in \mathbb{C}^{n \times n}$ be the upper triangular matrix with $r_{ji},\ j \leq i$ given by (4.1). Show that $\bar{Q}^T Q = I_n$ and $X = QR$. (This is the QR algorithm.) Show that if in addition $X \in \mathbb{R}^{m \times n}$ then Q and R are real valued matrices.

6. Let $C \in \mathbb{C}^{n \times n}$ and assume that $\{\lambda_1, ..., \lambda_n\}$ are n eigenvalues of C counted with their multiplicities. View C as an operator $C : \mathbb{C}^n \to \mathbb{C}^n$. View $\mathbb{C}^n$ as $2n$-dimensional vector space over $\mathbb{R}^{2n}$. Let $C = A + \sqrt{-1}B,\ A, B \in \mathrm{M}_n(\mathbb{R})$.

   a. Then $\hat{C} := \begin{bmatrix} A & -B \\ B & A \end{bmatrix} \in \mathrm{M}_{2n}(\mathbb{R})$ represents the operator $C : \mathbb{C}^n \to \mathbb{C}^n$ as an operator over $\mathbb{R}$ in suitably chosen basis.

   b. Show that $\{\lambda_1, \bar{\lambda}_1, ..., \lambda_n, \bar{\lambda}_n\}$ are the $2n$ eigenvalues of $\hat{C}$ counting with multiplicities.

   c. Show that the Jordan canonical form of $\tilde{C}$, is obtained by replacing each Jordan block $\lambda I + H$ in C by two Jordan blocks $\lambda I + H$ and $\bar{\lambda} I + H$.

4.6 Geometric interpretation of the determinant

Definition 4.13 *Let $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^n$ be m given vectors. Then the parallelepiped $\mathrm{P}(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ is defined as follows. The 2^m vertices of $\mathrm{P}(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ are of the form $\mathbf{v} := \sum_{i=1}^{m} a_i \mathbf{x}_i$, where $a_i = 0, 1$ for $i = 1, \ldots, m$. Two vertices $\mathbf{v} = \sum_{i=1}^{m} a_i \mathbf{x}_i$ and $\mathbf{w} = \sum_{i=1}^{m} b_i \mathbf{x}_i$ of $\mathrm{P}(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ are adjacent, i.e. connected by an edge in $\mathrm{P}(\mathbf{x}_1, \ldots, \mathbf{x}_m)$, if $||(a_1, \ldots, a_m)^\top - (b_1, \ldots, b_m)^\top|| = 1$, i.e. the $0-1$ coordinates of $(a_1, \ldots, a_m)^\top$ and $(b_1, \ldots, b_m)^\top$ differ only at one coordinate k, for some $k \in [1, m]$.*

Note that if $\mathbf{e}_1, \dots, \mathbf{e}_n$ is the standard basis in $\mathbb{R}^n$, i.e. $\mathbf{e}_i = (\delta_{1i}, \dots, \delta_{ni})^\top, i = 1, \dots, n$, then $\mathrm{P}(\mathbf{e}_1, \dots, \mathbf{e}_m)$ is the m-dimensional unit cube, whose edges are parallel to $\mathbf{e}_1, \dots, \mathbf{e}_m$ and its center (of gravity) is $\frac{1}{2}(\underbrace{1, \dots, 1}, 0, \dots, 0)^\top$, where 1 appears m times for $1 \leq m \leq n$.

For $m > n$ $\mathrm{P}(\mathbf{x}_1, \dots, \mathbf{x}_m)$ is "flattened" parallelepiped, since $\mathbf{x}_1, \dots, \mathbf{x}_m$ are always linearly dependent in $\mathbb{R}^n$ for $m > n$.

Proposition 4.14 *Let $A \in \mathbb{R}^{n \times n}$ and view $A = [\mathbf{c}_1 \ \mathbf{c}_2 \dots \mathbf{c}_n]$ as an ordered set of n vectors, (columns), $\mathbf{c}_1, \dots, \mathbf{c}_n$. Then $|\det A|$ is the n-dimensional volume of the parallelepiped $\mathrm{P}(\mathbf{c}_1, \dots, \mathbf{c}_n)$. If $\mathbf{c}_1, \dots, \mathbf{c}_n$ are linearly independent then the orientation in $\mathbb{R}^n$ induced by $\mathbf{c}_1, \dots, \mathbf{c}_n$ is the same as the orientation induced by $\mathbf{e}_1, \dots, \mathbf{e}_n$ if $\det A > 0$, and is the opposite orientation if $\det A < 0$.*

Proof. $\det A = 0$ if and only if the columns of A are linearly dependent. If $\mathbf{c}_1, \dots, \mathbf{c}_n$ are linearly dependent, then $\mathrm{P}(\mathbf{c}_1, \dots, \mathbf{c}_n)$ lies in a subspace of $\mathbb{R}^n$, i.e. some $n-1$ dimensional subspace, and hence the n-dimensional volume of $\mathrm{P}(\mathbf{c}_1, \dots, \mathbf{c}_n)$ is zero.

Assume now that $\det A \neq 0$, i.e. $\mathbf{c}_1, \dots, \mathbf{c}_n$ are linearly independent. Perform that Gram-Schmidt process 4.4. Then $A = QR$, where $Q = [\mathbf{e}_1 \ \mathbf{e}_2 \dots \mathbf{e}_n]$ is an orthogonal matrix and $R = (r_{ji}) \in \mathbb{R}^{n \times n}$ is an upper diagonal matrix. (See Problem 5.) So $\det A = \det Q \det R$. Since $Q^\top Q = I_n$ we deduce that $1 = \det I_n = \det Q^\top \det Q = \det Q \det Q = (\det Q)^2$. So $\det Q = \pm 1$ and the sign of $\det Q$ is the sign of $\det A$.

Hence $|\det A| = \det R = r_{11} r_{22} \dots r_{nn}$. Recall that r_{11} is the length of the vector $\mathbf{c}_1$, and r_{ii} is the distance of the vector $\mathbf{e}_i$ to the subspace spanned by $\mathbf{e}_1, \dots, \mathbf{e}_{i-1}$ for $i = 2, \dots, n$. (See Problem 4 parts (f-i).) Thus the length of $\mathrm{P}(\mathbf{c}_1)$ is r_{11}. The distance of $\mathbf{c}_2$ to $\mathrm{P}(\mathbf{c}_1)$ is r_{22}. Hence the area, i.e 2-dimensional volume of $\mathrm{P}(\mathbf{c}_1, \mathbf{c}_2)$ is $r_{11} r_{22}$. Continuing in this manner we deduce that the $i-1$ dimensional volume of $\mathrm{P}(\mathbf{c}_1, \dots, \mathbf{c}_{i-1})$ is $r_{11} \dots r_{(i-1)(i-1)}$. As the distance of $\mathbf{c}_i$ to $\mathrm{P}(\mathbf{c}_1, \dots, \mathbf{c}_{i-1})$ is r_{ii} it follows that the i-dimensional volume of $\mathrm{P}(\mathbf{c}_1, \dots, \mathbf{c}_i)$ is $r_{11} \dots r_{ii}$. For $i = n$ we get that $|\det A| = r_{11} \dots r_{nn}$ which is equal to the n-dimensional volume of $\mathrm{P}(\mathbf{c}_1, \dots, c_n)$.

As we already pointed out the sign of $\det A$ is equal to the sign of $\det Q = \pm 1$. If $\det Q = 1$ it is possible to "rotate" the standard basis in $\mathbb{R}^n$ to the basis given by the columns of an orthogonal matrix Q with $\det Q = 1$. If $\det Q = -1$, we need one reflection, i.e. replace the standard basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the new basis $-\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ and the rotate the new basis $-\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ to the basis consisting of the columns of an orthogonal matrix Q, where $\det Q = -1$. □

Theorem 4.15 ***(The Hadamard determinantal inequality)*** *Let $A = [\mathbf{c}_1, \dots, \mathbf{c}_n] \in \mathbb{C}^{n \times n}$. Then $|\det A| \leq ||\mathbf{c}_1|| \, ||\mathbf{c}_2|| \dots ||\mathbf{c}_n||$. Equality holds if and only if either $\mathbf{c}_i = \mathbf{0}$ for some i or $\langle \mathbf{c}_i, \mathbf{c}_j \rangle = 0$ for all $i \neq j$, i.e. $\mathbf{c}_1, \dots, \mathbf{c}_n$ is an orthogonal system.*

Proof. Assume first that $\det A = 0$. Clearly the Hadamard inequality holds. Equality in Hadamard inequality if and only if $\mathbf{c}_i = \mathbf{0}$ for some i.

Assume now that $\det A \neq 0$ and perform the Gram-Schmidt process. From (4.1) it follows that $A = QR$ where Q is a unitary matrix, i.e. $Q^* Q = I_n$ and $R = (r_{ji}) \in$

$\mathbb{C}^{n\times n}$ upper triangular with r_{ii} real and positive numbers. So $\det A = \det Q \det R$. Thus

$$1 = \det I_n = \det Q^*Q = \det Q^* \det Q = \overline{\det Q} \det Q = |\det Q|^2 \Rightarrow |\det Q| = 1.$$

Hence $|\det A| = \det R = r_{11}r_{22}\dots r_{nn}$. According to Problem 4 and the proof of Proposition 4.14 we know that $||\mathbf{c}_i|| \geq \operatorname{dist}(\mathbf{c}_i, \operatorname{span}(\mathbf{c}_1, \dots, \mathbf{c}_{i-1})) = r_{ii}$ for $i = 2, \dots, n$. Hence $|\det A| = \det R \leq ||\mathbf{c}_1||\ ||\mathbf{c}_2|| \dots ||\mathbf{c}_n||$. Equality holds if $||\mathbf{c}_i|| = \operatorname{dist}(\mathbf{c}_i, \operatorname{span}(\mathbf{c}_1, \dots, \mathbf{c}_{i-1}))$ for $i = 2, \dots, n$. Use Problem 4 to deduce that $||\mathbf{c}_i|| = \operatorname{dist}(\mathbf{c}_i, \operatorname{span}(\mathbf{c}_1, \dots, \mathbf{c}_{i-1}))$ if an only if $\langle \mathbf{c}_i, \mathbf{c}_j \rangle = 0$ for $j = 1, \dots, i-1$. Use these conditions for $i = 2, \dots$ to deduce that equality in Hadamard inequality holds if and only if $\mathbf{c}_1, \dots, \mathbf{c}_n$ is an orthogonal system. □

Problems

1. Let $A = (a_{ij})_{i,j} \in \mathbb{C}^{n\times n}$. Assume that $|a_{ij}| \leq K$ for all $i, j = 1, \dots, n$. Show that $|\det A| \leq K^n n^{\frac{n}{2}}$.

2. Let $A = (a_{ij})_{i,j=1}^n \in \mathbb{C}^{n\times n}$ such that $|a_{ij}| \leq 1$ for $i, j = 1, \dots, n$. Show that $|\det A| = n^{\frac{n}{2}}$ if and only if $A^*A = AA^* = nI_n$. In particular, if $|\det A| = n^{\frac{n}{2}}$ then $|a_{ij}| = 1$ for $i, j = 1, \dots, n$.

3. Show that for each n there exists a matrix $A = (a_{ij})_{i,j=1}^n \in \mathbb{C}^{n\times n}$ such that $|a_{ij}| = 1$ for $i, j = 1, \dots, n$ and $|\det A| = n^{\frac{n}{2}}$.

4. Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ and assume that $a_{ij} = \pm 1, i, j = 1, \dots, n$. Show that if $n > 2$ then the assumption that $|\det A| = n^{\frac{n}{2}}$ yields that n is divisible by 4.

5. Show that for any $n = 2^m$, $m = 0, 1, \dots$ there exists $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ such that $a_{ij} = \pm 1, i, j = 1, \dots, n$ and $|\det A| = n^{\frac{n}{2}}$. (*Hint*: Try to prove by induction on m that $A \in \mathbb{R}^{2^m \times 2^m}$ can be chosen symmetric, and then construct $B \in \mathbb{R}^{2^{m+1}\times 2^{m+1}}$ using A.)

Note: A matrix $A = (a_{ij})_{i,j=1}^n \in \mathbb{R}^{n\times n}$ such that $a_{ij} = \pm 1$ for $i, j = 1, \dots, n$ and $|\det A| = n^{\frac{n}{2}}$ is called a *Hadamard* matrix. *It is conjectured* that for each n divisible by 4 there exists a Hadamard matrix.

4.7 Special transformations in IPS

Proposition 4.16 *Let* $\mathbf{V}$ *be an IPS and* $T : \mathbf{V} \to \mathbf{V}$ *a linear transformation Then there exists a unique linear transformation* $T^* : \mathbf{V} \to \mathbf{V}$ *such that* $\langle T\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, T^*\mathbf{y} \rangle$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbf{V}$.

See Problems 1-2.

Definition 4.17 *Let* $\mathbf{V}$ *be an IPS and let* $T : \mathbf{V} \to \mathbf{V}$ *be a linear transformation. Then*
(a) T *is called self-adjoint if* $T^* = T$*;*
(b) T *is called anti self-adjoint if* $T^* = -T$*;*
(c) T *is called unitary if* $T^*T = TT^* = I$*;*

(d) T is called normal if $T^*T = TT^*$.

Denote by $\mathbf{S}(\mathbf{V})$, $\mathbf{AS}(\mathbf{V})$, $\mathbf{U}(\mathbf{V})$, $\mathbf{N}(\mathbf{V})$ the sets of self-adjoint, anti self-adjoint, unitary and normal operators on $\mathbf{V}$ respectively.

Proposition 4.18 *Let $\mathbf{V}$ be an IPS over $\mathbb{F} = \mathbb{R}, \mathbb{C}$ with an orthonormal basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. Let $T : \mathbf{V} \to \mathbf{V}$ be a linear transformation. Let $A = (a_{ij}) \in \mathbb{F}^{n\times n}$ be the representation matrix of T in the basis E:*

$$a_{ij} = \langle T\mathbf{e}_j, \mathbf{e}_i \rangle, \quad i, j = 1, ..., n. \tag{4.1}$$

Then for $\mathbb{F} = \mathbb{R}$:

(a) T^* is represented by $A^\top$,
(b) T is selfadjoint $\iff$ $A = A^\top$,
(c) T is anti selfadjoint $\iff$ $A = -A^\top$,
(d) T is unitary $\iff$ A is orthogonal $\iff$ $AA^\top = A^\top A = I$,
(e) T is normal $\iff$ A is normal $\iff$ $AA^\top = A^\top A$,

and for $\mathbb{F} = \mathbb{C}$:

(a) T^* is represented by A^* $(:= \bar{A}^\top)$,
(b) T is selfadjoint $\iff$ A is hermitian $\iff$ $A = A^*$,
(c) T is anti selfadjoint $\iff$ A is anti hermitian $\iff$ $A = -A^*$,
(d) T is unitary $\iff$ A is unitary $\iff$ $AA^* = A^*A = I$,
(e) T is normal $\iff$ A is normal $\iff$ $AA^* = A^*A$.

See Problem 3.

Proposition 4.19 *Let $\mathbf{V}$ be an IPS over $\mathbb{R}$, and let $T \in \mathrm{Hom}(\mathbf{V})$. Let $\mathbf{V}_c$ be the complexification of $\mathbf{V}$. Show that there exists a unique $T_c \in \mathrm{Hom}(\mathbf{V}_c)$ such that $T_c|\mathbf{V} = T$. Furthermore T is self-adjoint, unitary or normal if and only if T_c is self-adjoint, unitary or normal respectively.*

See Problem 4

Definition 4.20 *For a field $\mathbb{F}$ let*

$$\begin{aligned}
&\mathbf{S}(n, \mathbb{F}) := \{A \in \mathbb{F}^{n\times n} : \quad A = A^\top\},\\
&\mathbf{AS}(n, \mathbb{F}) := \{A \in \mathbb{F}^{n\times n} : \quad A = -A^\top\},\\
&\mathbf{O}(n, \mathbb{F}) := \{A \in \mathbb{F}^{n\times n} : \quad AA^\top = A^\top A = I\},\\
&\mathbf{SO}(n, \mathbb{F}) := \{A \in \mathbf{O}(n, \mathbb{F}) : \quad \det A = 1\},\\
&\mathbf{DO}(n, \mathbb{F}) := \mathbf{D}(n, \mathbb{F}) \cap \mathbf{O}(n, \mathbb{F}),\\
&\mathbf{N}(n, \mathbb{R}) := \{A \in \mathbb{R}^{n\times n} : \quad AA^\top = A^\top A\},\\
&\mathbf{N}(n, \mathbb{C}) := \{A \in \mathbb{C}^{n\times n} : \quad AA^* = A^*A\},\\
&\mathbf{H}_n := \{A \in \mathbb{C}^{n\times n} : \quad A = A^*\},\\
&\mathbf{AH}_n := \{A \in \mathbb{C}^{n\times n} : \quad A = -A^*\},\\
&\mathbf{U}_n := \{A \in \mathbb{C}^{n\times n} : \quad AA^* = A^*A = I\},\\
&\mathbf{SU}_n := \{A \in \mathbf{U}_n : \quad \det A = 1\},\\
&\mathbf{DU}_n := \mathbf{D}(n, \mathbb{C}) \cap \mathbf{U}_n.
\end{aligned}$$

See Problem 5 for relations between these classes.

Theorem 4.21 *Let* $\mathbf{V}$ *be an IPS over* $\mathbb{C}$ *of dimension* n*. Then a linear transformation* $T : \mathbf{V} \to \mathbf{V}$ *is normal if and only if* $\mathbf{V}$ *has an orthonormal basis consisting of eigenvectors of* T*.*

Proof. Suppose first that $\mathbf{V}$ has an orthonormal basis $\mathbf{e}_1, ..., \mathbf{e}_n$ such that $T\mathbf{e}_i = \lambda_i \mathbf{e}_i,\ i = 1, ..., n$. From the definition of T^* it follows that $T^*\mathbf{e}_i = \bar{\lambda}_i \mathbf{e}_i,\ i = 1, ..., n$. Hence $TT^* = T^*T$.

Assume now T is normal. Since $\mathbb{C}$ is algebraically closed T has an eigenvalue λ_1. Let $\mathbf{V}_1$ be the subspace of $\mathbf{V}$ spanned by all eigenvectors of T corresponding to the eigenvalue λ_1. Clearly $T\mathbf{V}_1 \subset \mathbf{V}_1$. Let $\mathbf{x} \in \mathbf{V}_1$. Then $T\mathbf{x} = \lambda_1 \mathbf{x}$. Thus

$$T(T^*\mathbf{x}) = (TT^*)\mathbf{x} = (T^*T)\mathbf{x} = T^*(T\mathbf{x}) = \lambda_1 T^*\mathbf{x} \Rightarrow T^*\mathbf{V}_1 \subset \mathbf{V}_1.$$

Hence $T\mathbf{V}_1^\perp, T^*\mathbf{V}_1^\perp \subset \mathbf{V}_1^\perp$. Since $\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_1^\perp$ it is enough to prove the theorem for $T|\mathbf{V}_1$ and $T|\mathbf{V}_1^\perp$.

As $T|\mathbf{V}_1 = \lambda_1 I_{\mathbf{V}_1}$ it is straightforward to show $T^*|\mathbf{V}_1 = \bar{\lambda}_1 I_{\mathbf{V}_1}$ (see Problem 2). Hence for $T|\mathbf{V}_1$ the theorem trivially holds. For $T|\mathbf{V}_1^\perp$ the theorem follows by induction. □

Corollary 4.22 *Assume that* $A \in \mathbb{C}^{n\times n}$ *is a normal matrix. Then* A *is unitarily similar to a diagonal matrix. That is, there exists a unitaty matrix* $U \in \mathbb{C}^{n\times n}$ *and a diagonal matrix* $\Lambda \in \mathbb{C}^{n\times n}$ *such that* $A = U\Lambda U^* = U\Lambda U^{-1}$*. (The columns of* U *is an orthonormal basis of* $\mathbb{C}^n$ *consisting of eigenvectors of* A*, and the diagonal entries of* Λ *are the corresponding eigenvalues of* A*.)*

The proof of Theorem 4.21 yields:

Corollary 4.23 *Let* $\mathbf{V}$ *be an IPS over* $\mathbb{R}$ *of dimension* n*. Then the linear transformation* $T : \mathbf{V} \to \mathbf{V}$ *with a real spectrum is normal if and only if* $\mathbf{V}$ *has an orthonormal basis consiting of eigenvectors of* T*.*

Proposition 4.24 *Let* $\mathbf{V}$ *be an IPS over* $\mathbb{C}$*. Let* $T \in \mathbf{N}(\mathbf{V})$*. Then*

$$T \text{ is self} - \text{adjoint} \iff \text{spec}\,(T) \subset \mathbb{R},$$
$$T \text{ is unitary} \iff \text{spec}\,(T) \subset S^1 = \{z \in \mathbb{C} : \quad |z| = 1\}.$$

Proof. Since T is normal there exists an orthonormal basis $\mathbf{e}_1, ..., \mathbf{e}_n$ such that $T\mathbf{e}_i = \lambda_i \mathbf{e}_i,\ i = 1, ..., n$. Hence $T^*\mathbf{e}_i = \bar{\lambda}_i \mathbf{e}_i$. Then

$$T = T^* \iff \lambda_i = \bar{\lambda}_i,\ i = 1, ..., n,$$
$$TT^* = T^*T = I \iff |\lambda_i| = 1,\ i = 1, ..., n.$$

□

Combine Proposition 4.19 and Corollary 4.23 with the above proposition to deduce:

Corollary 4.25 *Let* $\mathbf{V}$ *be an IPS over* $\mathbb{R}$ *and let* $T \in \mathbf{S}(\mathbf{V})$. *Then* spec $(T) \subset \mathbb{R}$ *and* $\mathbf{V}$ *has an orthonormal basis consisting of the eigenvectors of* T.

Proposition 4.26 *Let* $\mathbf{V}$ *be an IPS over* $\mathbb{R}$ *and let* $T \in \mathbf{U}(\mathbf{V})$. *Then* $\mathbf{V} = \oplus_{i\in\{-1,1,2,...,k\}}\mathbf{V}_i$, *where* $k \geq 1$, $\mathbf{V}_i$ *and* $\mathbf{V}_j$ *are orthogonal for* $i \neq j$, *such that*
(a) $T|\mathbf{V}_{-1} = -I_{\mathbf{V}_{-1}}$ $\dim \mathbf{V}_{-1} \geq 0$,
(b) $T|\mathbf{V}_1 = I_{\mathbf{V}_1}$ $\dim \mathbf{V}_1 \geq 0$,
(c) $T\mathbf{V}_i = \mathbf{V}_i$, $\dim \mathbf{V}_i = 2$, spec $(T|\mathbf{V}_i) \subset S^1\backslash\{-1,1\}$ *for* $i = 2,...,k$.

See Problem 7.

Proposition 4.27 *Let* $\mathbf{V}$ *be an IPS over* $\mathbb{R}$ *and let* $T \in \mathbf{AS}(\mathbf{V})$. *Then* $\mathbf{V} = \oplus_{i\in\{1,2,...,k\}}\mathbf{V}_i$, *where* $k \geq 1$, $\mathbf{V}_i$ *and* $\mathbf{V}_j$ *are orthogonal for* $i \neq j$, *such that*
(a) $T|\mathbf{V}_1 = 0_{\mathbf{V}_1}$ $\dim \mathbf{V}_0 \geq 0$,
(b) $T\mathbf{V}_i = \mathbf{V}_i$, $\dim \mathbf{V}_i = 2$, spec $(T|\mathbf{V}_i) \subset \sqrt{-1}\mathbb{R}\backslash\{0\}$ *for* $i = 2,...,k$.

See Problem 8.

Theorem 4.28 *Let* $\mathbf{V}$ *be an IPS over* $\mathbb{C}$ *of dimension* n. *Let* $T \in \mathrm{Hom}(\mathbf{V})$. *Let* $\lambda_1,...,\lambda_n \in \mathbb{C}$ *be* n *eigenvalues of* T *counted with their multiplicities. Then there exists an orthonormal basis* $\mathbf{g}_1,...,\mathbf{g}_n$ *of* $\mathbf{V}$ *with the following properties:*

$$T\mathrm{span}(\mathbf{g}_1,...,\mathbf{g}_i) \subset \mathrm{span}(\mathbf{g}_1,...,\mathbf{g}_i),\ \langle T\mathbf{g}_i,\mathbf{g}_i\rangle = \lambda_i,\ i = 1,...,n. \tag{4.2}$$

Let $\mathbf{V}$ *be an IPS over* $\mathbb{R}$ *of dimension* n. *Let* $T \in \mathrm{Hom}(\mathbf{V})$ *and assume that* spec $(T) \subset \mathbb{R}$. *Let* $\lambda_1,...,\lambda_n \in \mathbb{R}$ *be* n *eigenvalues of* T *counted with their multiplicities. Then there exists an orthonormal basis* $\mathbf{g}_1,...,\mathbf{g}_n$ *of* $\mathbf{V}$ *such that (4.2) holds.*

Proof. Assume first that $\mathbf{V}$ is IPS over $\mathbb{C}$ of dimension n. The proof is by induction on n. For $n = 1$ the theorem is trivial. Assume that $n > 1$. Since $\lambda_1 \in$ spec (T) it follows that there exists $\mathbf{g}_1 \in \mathbf{V}$, $\langle \mathbf{g}_1,\mathbf{g}_1\rangle = 1$ such that $T\mathbf{g}_1 = \lambda_1\mathbf{g}_1$. Let $\mathbf{U} := \mathrm{span}(\mathbf{g}_1)^\perp$. Let P be the orthogonal projection on $\mathbf{U}$. Observe that $P\mathbf{v} = \mathbf{v} - \langle \mathbf{v},\mathbf{g}_1\rangle\mathbf{g}_1$ for any vector $\mathbf{v} \in \mathbf{V}$. Let $T_1 := PT|_{\mathbf{U}}$. Clearly, $T_1 \in \mathrm{Hom}(\mathbf{U})$. Let $\tilde{\lambda}_2,...,\tilde{\lambda}_n$ be the eigenvalues of T_1 counted with their multiplicities. The induction hypothesis yields the existence of an orthonormal basis $\mathbf{g}_2,...,\mathbf{g}_n$ of $\mathbf{U}$ such that

$$T_1\mathrm{span}(\mathbf{g}_2,...,\mathbf{g}_i) \subset \mathrm{span}(\mathbf{g}_2,...,\mathbf{g}_i),\ \langle T_1\mathbf{g}_i,\mathbf{g}_i\rangle = \tilde{\lambda}_i,\ i = 1,...,n.$$

As $T_1\mathbf{u} = T\mathbf{u} - \langle T\mathbf{u},\mathbf{g}_1\rangle\mathbf{g}_1$ it follows that $T\mathrm{span}(\mathbf{g}_1,...,\mathbf{g}_i) \subset \mathrm{span}(\mathbf{g}_1,...,\mathbf{g}_i)$ for $i = 1,...,n$. Hence in the orthonormal basis $\mathbf{g}_1,...,\mathbf{g}_n$ T is presented by an upper diagonal matrix $B = [b_{ij}]_1^n$, with $b_{11} = \lambda_1$ and $b_{ii} = \tilde{\lambda}_i$, $i = 2,...,n$. Therefore $\lambda_1,\tilde{\lambda}_2,...,\tilde{\lambda}_n$ are the eigenvalues of T counted with their multiplicities. This establishes the theorem in this case. The real case is treated similarly. □

Combine the above results with Problems 6 and 12 to deduce:

Corollary 4.29 *Let $A \in \mathbb{C}^{n\times n}$. Let $\lambda_1, ..., \lambda_n \in \mathbb{C}$ be n eigenvalues of A counted with their multiplicities. Then there exist an upper triangular matrix $B = (b_{ij})_1^n \in \mathbb{C}^{n\times n}$, such that $b_{ii} = \lambda_i,\ i = 1, ..., n$, and a unitary matrix $U \in \mathbf{U}_n$ such that $A = UBU^{-1}$. If $A \in \mathbf{N}(n, \mathbb{C})$ then B is a diagonal matrix.*

Let $A \in \mathbb{R}^{n\times n}$ and assume that spec $(T) \subset \mathbb{R}$. *Then $A = UBU^{-1}$ where U can be chosen a real orthogonal matrix and B a real upper triangular matrix. If $A \in \mathbf{N}(n, \mathbb{R})$ and* spec $(A) \subset \mathbb{R}$ *then B is a diagonal matrix.*

It is easy to show that U in the above Corollary can be chosen in $\mathbf{SU}_n$ or $\mathbf{SO}(n, \mathbb{R})$ respectively (Problem 11).

Definition 4.30 *Let $\mathbf{V}$ be a vector space and assume that $T : \mathbf{V} \to \mathbf{V}$ is a linear operator. Let $0 \neq \mathbf{v} \in \mathbf{V}$. Then $\mathbf{W} = \text{span}(\mathbf{v}, T\mathbf{v}, T^2\mathbf{v}, \ldots)$ is called a cyclic invariant subspace of T generated by $\mathbf{v}$. (It is also referred as a Krylov subspace of T generated by $\mathbf{v}$.) Sometimes we will call $\mathbf{W}$ just a cyclic subspace, or Krylov subspace.*

Theorem 4.31 *Let $\mathbf{V}$ be a finite dimensional IPS. Let $T : \mathbf{V} \to \mathbf{V}$ be a linear operator. For $0 \neq \mathbf{v} \in \mathbf{V}$ let $\mathbf{W} = \text{span}(\mathbf{v}, T\mathbf{v}, ..., T^{r-1}\mathbf{v})$ be a cyclic T-invariant subspace of dimension r generated by $\mathbf{v}$. Let $\mathbf{u}_1, ..., \mathbf{u}_r$ be an orthonormal basis of $\mathbf{W}$ obtained by the Gram-Schmidt process from the basis $[\mathbf{v}, T\mathbf{V}, ..., T^{r-1}\mathbf{v}]$ of $\mathbf{W}$. Then $\langle T\mathbf{u}_i, \mathbf{u}_j\rangle = 0$ for $1 \leq i \leq j-2$, i.e. the representation matrix of $T|\mathbf{W}$ in the basis $[\mathbf{u}_1, \ldots, \mathbf{u}_r]$ is upper Hessenberg. If T is self-adjoint then the representation matrix of $T|\mathbf{W}$ in the basis $[\mathbf{u}_1, \ldots, \mathbf{u}_r]$ is a tridiagonal hermitian matrix.*

Proof. Let $\mathbf{W}_j = \text{span}(\mathbf{v}, \ldots, T^{j-1}\mathbf{v})$ for $j = 1, ..., r+1$. Clearly $T\mathbf{W}_j \subset \mathbf{W}_{j+1}$ for $j = 1, ..., r$. The assumption that $\mathbf{W}$ is T-invariant subspace yields $\mathbf{W} = \mathbf{W}_r = \mathbf{W}_{r+1}$. Since $\dim \mathbf{W} = r$ it follows that $\mathbf{v}, ..., T^{r-1}\mathbf{v}$ are linearly independent. Hence $[\mathbf{v}, \ldots, T^{r-1}\mathbf{v}]$ is a basis for $\mathbf{W}$. Recall that $\text{span}(\mathbf{u}_1, ..., \mathbf{u}_j) = \mathbf{W}_j$ for $j = 1, \ldots, r$. Let $r \geq j \geq i+2$. Then $T\mathbf{u}_i \in T\mathbf{W}_i \subset \mathbf{W}_{i+1}$. As $\mathbf{u}_j \perp \mathbf{W}_{i+1}$ it follows that $\langle T\mathbf{u}_i, \mathbf{u}_j\rangle = 0$. Assume that $T^* = T$. Let $r \geq i \geq j+2$. Then $\langle T\mathbf{u}_i, \mathbf{u}_j\rangle = \langle \mathbf{u}_i, T\mathbf{u}_j\rangle = 0$. Hence the representation matrix of $T|\mathbf{W}$ in the basis $[\mathbf{u}_1, \ldots, \mathbf{u}_r]$ is a tridiagonal hermitian matrix. □

Problems

1. Prove Proposition 4.16.
2. Let $P, Q \in \text{Hom}(\mathbf{V}), \mathbf{a}, b \in \mathbb{F}$. Show that $(aP + bQ)^* = \bar{a}P^* + \bar{b}Q^*$.
3. Prove Proposition 4.18.
4. Prove Proposition 4.19 for finite dimensional $\mathbf{V}$. (*Hint*: Choose an orthonormal basis in $\mathbf{V}$.)

5. Show the following

$$\mathbf{SO}(n,\mathbb{F}) \subset \mathbf{O}(n,\mathbb{F}) \subset \mathrm{GL}(n,\mathbb{F}),$$
$$\mathbf{S}(n,\mathbb{R}) \subset \mathbf{H}_n \subset \mathbf{N}(n,\mathbb{C}),$$
$$\mathbf{AS}(n,\mathbb{R}) \subset \mathbf{AH}_n \subset \mathbf{N}(n,\mathbb{C}),$$
$$\mathbf{S}(n,\mathbb{R}), \mathbf{AS}(n,\mathbb{R}) \subset \mathbf{N}(n,\mathbb{R}) \subset \mathbf{N}(n,\mathbb{C}),$$
$$\mathbf{O}(n,\mathbb{R}) \subset \mathbf{U}_n \subset \mathbf{N}(n,\mathbb{C}),$$
$$\mathbf{SO}(n,\mathbb{F}),\ \mathbf{O}(n,\mathbb{F}),\ \mathbf{SU}_n,\ \mathbf{U}_n \text{ are groups}$$
$$\mathbf{S}(n,\mathbb{F}) \text{ is a } - \text{vector space of dimension } \binom{n+1}{2},$$
$$\mathbf{AS}(n,\mathbb{F}) \text{ is a } - \text{vector space of dimension } \binom{n}{2},$$
$$\mathbf{H}_n \text{ is an } \mathbb{R} - \text{ vector space of dimension } n^2.$$
$$\mathbf{AH}_n = \sqrt{-1}\,\mathbf{H}_n$$

6. Let $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be an orthonormal basis in IPS $\mathbf{V}$ over $\mathbb{F}$. Let $G = \{\mathbf{g}_1, ..., \mathbf{g}_n\}$ be another basis in $\mathbf{V}$. Show that F is an orthonormal basis if and only if the tranfer matrix either from E to G or from G to E is a unitary matrix.

7. Prove Proposition 4.26

8. Prove Proposition 4.27

9. a. Show that $A \in \mathbf{SO}(2,\mathbb{R})$ is of the form $A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}, \theta \in \mathbb{R}$.

   b. Show that $\mathbf{SO}(2,\mathbb{R}) = e^{\mathbf{AS}(2,\mathbb{R})}$. That is for any $B \in \mathbf{AS}(2,\mathbb{R})$ $e^B \in \mathbf{SO}(2,\mathbb{R})$ and any $A \in \mathbf{SO}(n,\mathbb{R})$ is e^B for some $B \in \mathbf{AS}(2,\mathbb{R})$. (*Hint*: Consider the power series for e^B, $B = \begin{bmatrix} 0 & \theta \\ -\theta & 0 \end{bmatrix}$.)

   c. Show that $\mathbf{SO}(n,\mathbb{R}) = e^{\mathbf{AS}(n,\mathbb{R})}$. (*Hint*: Use Propositions 4.26 and 4.27 and part b.)

   d. Show that $\mathbf{SO}(n,\mathbb{R})$ is a path connected space. (See part e.)

   e. Let $\mathbf{V}$ be an $n(>1)$-dimensional IPS over $\mathbb{F} = \mathbb{R}$. Let $p \in \langle n-1 \rangle$. Assume that $\mathbf{x}_1, ..., \mathbf{x}_p$ and $\mathbf{y}_1, ..., \mathbf{y}_p$ be two orthonormal systems in $\mathbf{V}$. Show that these two o.n.s. are path connected. That is there are p continuous mappings $\mathbf{z}_i(t) : [0,1] \to \mathbf{V},\ i = 1, ..., p$ such that for each $t \in [0,1]$ $\mathbf{z}_1(t), ..., \mathbf{z}_p(t)$ is an o.n.s. and $\mathbf{z}_i(0) = \mathbf{x}_i, \mathbf{z}_i(1) = \mathbf{y}_i, i = 1, ..., p$.

10. a. Show that $\mathbf{U}_n = e^{\mathbf{AH}_n}$. (*Hint*: Use Proposition 4.24 and its proof.)

    b. Show that $\mathbf{U}_n$ is path connected.

    c. Prove Problem 9e for $\mathbb{F} = \mathbb{C}$.

11. Show

    (a) $D_1 D D_1^* = D$ for any $D \in \mathbf{D}(n,\mathbb{C}),\ D_1 \in \mathbf{DU}_n$.

(b) $A \in \mathbf{N}(n,\mathbb{C}) \iff A = UDU^*,\ U \in \mathbf{SU}_n,\ D \in \mathbf{D}(n,\mathbb{C})$.

(c) $A \in \mathbf{N}(n,\mathbb{R}),\ \sigma(A) \subset \mathbb{R} \iff A = UDU^\top,\ U \in \mathbf{SO}_n,\ D \in \mathbf{D}(n,\mathbb{R})$.

12. Show that an upper triangular or a lower triangular matrix $B \in \mathbb{C}^{n\times n}$ is normal if and only if B is diagonal. (**Hint**: consider the equality $(BB^*)_{11} = (B^*B)_{11}$.)

13. Let the assumptions of Theorem 4.31 hold. Show that instead of performing the Gram-Schmidt process on $\mathbf{v}, T\mathbf{v}, ..., T^{r-1}\mathbf{v}$ one can perform the following process. Let $\mathbf{w}_1 := \frac{1}{\|\mathbf{v}\|}\mathbf{v}$. Assume that one already obtained i orthonormal vectors $\mathbf{w}_1, ..., \mathbf{w}_i$. Let $\tilde{\mathbf{w}}_{i+1} := T\mathbf{w}_i - \sum_{j=1}^{i}\langle T\mathbf{w}_i, \mathbf{w}_j\rangle \mathbf{w}_j$. If $\tilde{\mathbf{w}}_{i+1} = 0$ then stop the process, i.e. one is left with i orthonormal vectors. If $\mathbf{w}_{i+1} \neq 0$ then $\mathbf{w}_{i+1} := \frac{1}{\|\tilde{\mathbf{w}}_{i+1}\|}\tilde{\mathbf{w}}_{i+1}$ and continue the process. Show that the process ends after obtaining r orthonormal vectors $\mathbf{w}_1, \ldots, \mathbf{w}_r$ and $\mathbf{u}_i = \mathbf{w}_i$ for $i = 1, ..., r$. (This is a version of *Lanczos tridiagonalization* process.)

4.8 Symmetric bilinear and hermitian forms

Definition 4.32 *Let $\mathbf{V}$ be a vector space over $\mathbb{F}$ and $Q : \mathbf{V} \times \mathbf{V} \to \mathbb{F}$. Q is called a symmetric bilinear form (on $\mathbf{V}$) if the following conditions are satisfied:*
(a) $Q(\mathbf{x},\mathbf{y}) = Q(\mathbf{y},\mathbf{x})$ *for all $\mathbf{x},\mathbf{y} \in \mathbf{V}$ (symmetricity);*
(b) $Q(a\mathbf{x}+b\mathbf{z}, y) = aQ(\mathbf{x},\mathbf{y}) + bQ(\mathbf{z},\mathbf{y})$ *for all $a, b \in \mathbb{F}$ and $\mathbf{x},\mathbf{y},\mathbf{z} \in \mathbf{V}$ (bilinearity).*

For $\mathbb{F} = \mathbb{C}$ Q is called hermitian form or sesquilinear (on $\mathbf{V}$) if Q satisfies the conditions (a') and (b) where
(a') $Q(\mathbf{x},\mathbf{y}) = \overline{Q(\mathbf{y},\mathbf{x})}$ *for all $\mathbf{x},\mathbf{y} \in \mathbf{V}$ (barsymmetricity).*

The following results are elementary (see Problems 1-2):

Proposition 4.33 *Let $\mathbf{V}$ be a vector space over $\mathbb{F}$ with a basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. Then there is $1-1$ correspondence between a symmetric bilinear form Q on $\mathbf{V}$ and $A \in \mathbf{S}(n,\mathbb{F})$:*

$$Q(\mathbf{x},\mathbf{y}) = \eta^\top A\xi,$$
$$\mathbf{x} = \sum_{i=1}^{n}\xi_i\mathbf{e}_i,\ \mathbf{y} = \sum_{i=1}^{n}\eta_i\mathbf{e}_i,\ \xi = (\xi_1, ..., \xi_n)^\top, \eta = (\eta_1, ..., \eta_n)^\top \in \mathbb{F}^n.$$

Let $\mathbf{V}$ be a vector space over $\mathbb{C}$ with a basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$. Then there is $1-1$ correspondence between a hermitian form Q on $\mathbf{V}$ and $A \in \mathbf{H}_n$:

$$Q(\mathbf{x},\mathbf{y}) = \eta^* A\xi,$$
$$\mathbf{x} = \sum_{i=1}^{n}\xi_i\mathbf{e}_i,\ \mathbf{y} = \sum_{i=1}^{n}\eta_i\mathbf{e}_i,\ \xi = (\xi_1, ..., \xi_n)^\top, \eta = (\eta_1, ..., \eta_n)^\top \in \mathbb{C}^n.$$

Definition 4.34 *Let the assumptions of Proposition 4.33 hold. Then A is called the representation matrix of Q in the basis E.*

Proposition 4.35 *Let the assumptions of Proposition 4.33 Let $F = \{\mathbf{f_1}, ..., \mathbf{f_n}\}$ be another basis of $\mathbf{V}$. Then the symmetric bilinear form Q is represented by $B \in \mathbf{S}(n,\mathbb{F})$ in the basis F, where B is congruent A:*

$$B = U^\top AU, \quad U \in \mathrm{GL}(n,\mathbb{F})$$

and U is the matrix corresponding to the basis change from F to E. For $\mathbb{F}=\mathbb{C}$ the hermitian form Q is presented by $B\in\mathbf{H}_n$ in the basis F, where B hermicongruent to A:

$$B=U^*AU,\quad U\in \mathrm{GL}(n,\mathbb{C})$$

and U is the matrix corresponding to the basis change from F to E.

In what follows we assume that $\mathbb{F}=\mathbb{R},\mathbb{C}$.

Proposition 4.36 *Let $\mathbf{V}$ be an n dimensional vector space over $\mathbb{R}$. Let $Q:\mathbf{V}\times\mathbf{V}\to\mathbb{R}$ be a symmetric bilinear form. Let $A\in\mathbf{S}(n,\mathbb{R})$ the representation matrix of Q with respect to a basis E in $\mathbf{V}$. Let $\mathbf{V}_c$ be the extension of $\mathbf{V}$ over $\mathbb{C}$. Then there exists a unique hermitian form $Q_c:\mathbf{V}_c\times\mathbf{V}_c\to\mathbb{C}$ such that $Q_c|_{\mathbf{V}\times\mathbf{V}}=Q$ and Q_c is presented by A with respect to the basis E in $\mathbf{V}_c$.*

See Problem 3

Normalization 4.37 *Let $\mathbf{V}$ is a finite dimensional IPS over $\mathbb{F}$. Let $Q:\mathbf{V}\times\mathbf{V}\to\mathbb{F}$ be either a symmetric bilinear form for $\mathbb{F}=\mathbb{R}$ or a hermitian form for $\mathbb{F}=\mathbb{C}$. Then a representation matrix A of Q is chosen with respect to an orthonormal basis E.*

The following proposition is straightforward (see Problem 4).

Proposition 4.38 *Let $\mathbf{V}$ is an n-dimensional IPS over $\mathbb{F}$. Let $Q:\mathbf{V}\times\mathbf{V}\to\mathbb{F}$ be either a symmetric bilinear form for $\mathbb{F}=\mathbb{R}$ or a hermitian form for $\mathbb{F}=\mathbb{C}$. Then there exists a unique $T\in\mathbf{S}(\mathbf{V})$ such that $Q(\mathbf{x},\mathbf{y})=\langle T\mathbf{x},\mathbf{y}\rangle$ for any $\mathbf{x},\mathbf{y}\in\mathbf{V}$. In any orthonormal basis of $\mathbf{V}$ Q and T represented by the same matrix A. In particular the characteristic polynomial $p(\lambda)$ of T is called the characteristic polynomial of Q. Q has only real roots:*

$$\lambda_1(Q)\geq ...\geq\lambda_n(Q),$$

which are called the eigenvalues of Q. Furthermore there exists an orthonormal basis $F=\{\mathbf{f}_1,...,\mathbf{f}_n\}$ in $\mathbf{V}$ such that $D=\mathrm{diag}(\lambda_1(Q),...,\lambda_n(Q))$ is the representation matrix of Q in F.

Vice versa, for any $T\in\mathbf{S}(\mathbf{V})$ and any subspace $\mathbf{U}\subset\mathbf{V}$ the form $Q(T,\mathbf{U})$ defined by

$$Q(T,\mathbf{U})(\mathbf{x},\mathbf{y}):=\langle T\mathbf{x},\mathbf{y}\rangle\quad\text{for }\mathbf{x},\mathbf{y}\in\mathbf{U}$$

is either a symmetric bilinear form for $\mathbb{F}=\mathbb{R}$ or a hermitian form for $\mathbb{F}=\mathbb{C}$.

In the rest of the book we use the following normalization unless stated otherwise.

Normalization 4.39 *Let $\mathbf{V}$ is an n-dimensional IPS over $\mathbb{F}$. Assume that $T\in\mathbf{S}(\mathbf{V})$. Then arrange the eigenvalues of T counted with their multiplicities in the decreasing order*

$$\lambda_1(T)\geq ...\geq\lambda_n(T).$$

Same normalization applies to real symmetric matrices and complex hermitian matrices.

Problems

1. Prove Proposition 4.33.

2. Prove Proposition 4.35.

3. Prove Proposition 4.36.

4. Prove Proposition 4.38.

4.9 Max-min characterizations of eigenvalues

Definition 4.40 *Let* $\mathbf{V}$ *be a finite dimensional space over the field* $\mathbb{F}$*. Denote by* $\mathrm{Gr}(m, \mathbf{V})$ *be the space of all* m*-dimensional subspaces in* $\mathbf{U}$ *of dimension* $m \in [0, n] \cap \mathbb{Z}_+$*.*

Theorem 4.41 *(The convoy principle) Let* $\mathbf{V}$ *be an* n*-dimensional IPS. Let* $T \in \mathbf{S}(\mathbf{V})$*. Then*

$$\lambda_k(T) = \max_{\mathbf{U} \in \mathrm{Gr}(k, \mathbf{V})} \min_{\mathbf{0} \neq \mathbf{x} \in \mathbf{U}} \frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \max_{\mathbf{U} \in \mathrm{Gr}(k, \mathbf{V})} \lambda_k(Q(T, \mathbf{U})), \quad k = 1, ..., n, \tag{4.1}$$

where the quadratic form $Q(T, \mathbf{U})$ *is defined in Proposition 4.38. For* $k \in [1, n] \cap \mathbb{N}$ *let* $\mathbf{U}$ *be an invariant subspace of* T *spanned by eigenvectors* $\mathbf{e}_1, ..., \mathbf{e}_k$ *corresponding to the eigenvalues* $\lambda_1(T), ..., \lambda_k(T)$*. Then* $\lambda_k(T) = \lambda_k(Q(T, \mathbf{U}))$*. Let* $\mathbf{U} \in \mathrm{Gr}(k, \mathbf{V})$ *and assume that* $\lambda_k(T) = \lambda_k(Q(T, \mathbf{U}))$*. Then* $\mathbf{U}$ *contains and eigenvector of* T *corresponding to* $\lambda_k(T)$*.*

In particular

$$\lambda_1(T) = \max_{\mathbf{0} \neq \mathbf{x} \in \mathbf{V}} \frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle}, \quad \lambda_n(T) = \min_{\mathbf{0} \neq \mathbf{x} \in \mathbf{V}} \frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \tag{4.2}$$

Moreover for any $\mathbf{x} \neq \mathbf{0}$

$$\lambda_1(T) = \frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \iff T\mathbf{x} = \lambda_1(T)\mathbf{x},$$
$$\lambda_n(T) = \frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \iff T\mathbf{x} = \lambda_n(T)\mathbf{x},$$

The quotient $\frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle}$, $\mathbf{0} \neq \mathbf{x} \in \mathbf{V}$ is called *Rayleigh quotient.* The characterization (4.2) is called *convoy principle.*

Proof. Choose an orthonormal basis $E = \{\mathbf{e}_1, ..., \mathbf{e}_n\}$ such that

$$T\mathbf{e}_i = \lambda_i(T)\mathbf{e}_i, \ < \mathbf{e}_i, \mathbf{e}_j > = \delta_{ij} \quad i, j = 1, ..., n. \tag{4.3}$$

Then

$$\frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \frac{\sum_{i=1}^n \lambda_i(T)|x_i|^2}{\sum_{i=1}^n |x_i|^2}, \quad \mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i \neq \mathbf{0}. \tag{4.4}$$

The above equality yields straightforward (4.2) and the equality cases in these characterizations. Let $\mathbf{U} \in \mathrm{Gr}(k, \mathbf{V})$. Then the minimal characterization of $\lambda_k(Q(T, \mathbf{U}))$ yields the equality

$$\lambda_k(Q(T, \mathbf{U})) = \min_{\mathbf{0} \neq \mathbf{x} \in \mathbf{U}} \frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \quad \text{for any } \mathbf{U} \in \mathrm{Gr}(k, \mathbf{U}). \tag{4.5}$$

Next there exists $\mathbf{0} \neq \mathbf{x} \in \mathbf{U}$ such that $\langle \mathbf{x}, \mathbf{e}_i \rangle = 0$ for $i = 1, ..., k-1$. (For $k = 1$ this condition is void.) Hence

$$\frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \frac{\sum_{i=k}^n \lambda_i(T)|x_i|^2}{\sum_{i=k}^n |x_i|^2} \leq \lambda_k(T) \Rightarrow \lambda_k(T) \geq \lambda_k(Q(T, \mathbf{U})).$$

Let

$$\begin{aligned} &\lambda_1(T) = ... = \lambda_{n_1}(T) > \lambda(T)_{n_1+1}(T) = ... = \lambda_{n_2}(T) > ... > \\ &\lambda_{n_{r-1}+1}(T) = ... = \lambda_{n_r}(T) = \lambda_n(T), \quad n_0 = 0 < n_1 < ... < n_r = n. \end{aligned} \tag{4.6}$$

Assume that $n_{j-1} < k \leq n_j$. Suppose that $\lambda_k(Q(T, \mathbf{U})) = \lambda_k(T)$. Then for $\mathbf{x} \in \mathbf{U}$ such that $\langle \mathbf{x}, \mathbf{e}_i \rangle = 0$ we have equality $\lambda_k(Q(T, \mathbf{U})) = \lambda_k(T)$ if and only if $\mathbf{x} = \sum_{i=k}^{n_j} x_i \mathbf{e}_i$. Thus $T\mathbf{x} = \lambda_k(T)\mathbf{x}$.

Let $\mathbf{U}_k = \mathrm{span}(\mathbf{e}_1, ..., \mathbf{e}_k)$. Let $\mathbf{0} \neq \mathbf{x} = \sum_{i=1}^k \in \mathbf{U}_k$. Then

$$\frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \frac{\sum_{i=1}^k \lambda_i(T)|x_i|^2}{\sum_{i=1}^k |x_i|^2} \geq \lambda_k(T) \Rightarrow \lambda_k(Q(T, \mathbf{U}_k)) \geq \lambda_k(T).$$

Hence $\lambda_k(Q(T, \mathbf{U}_k)) = \lambda_k(T)$. □

It can be shown that for $k > 1$ and $\lambda_1(T) > \lambda_k(T)$ there exist $\mathbf{U} \in \mathrm{Gr}(k, \mathbf{V})$ such that $\lambda_k(T) = \lambda_k(T, \mathbf{U})$ and $\mathbf{U}$ is not an invariant subspace of T, in particular $\mathbf{U}$ does not contain all $\mathbf{e}_1, ..., \mathbf{e}_k$ satisfying (4.3). (See Problem 1.)

Corollary 4.42 *Let the assumptions of Theorem 4.41 hold. Let* $1 \leq \ell \leq n$. *Then*

$$\lambda_k(T) = \max_{\mathbf{W} \in \mathrm{Gr}(\ell, \mathbf{V})} \lambda_k(Q(T, \mathbf{W})), \quad k = 1, ..., \ell. \tag{4.7}$$

Proof. For $k \leq \ell$ apply Theorem 4.41 to $\lambda_k(Q(T, \mathbf{W}))$ to deduce that $\lambda_k(Q(T, \mathbf{W})) \leq \lambda_k(T)$. Let $\mathbf{U}_\ell = \mathrm{span}(\mathbf{e}_1, ..., \mathbf{e}_\ell)$. Then

$$\lambda_k(Q(T, \mathbf{U}_\ell)) = \lambda_k(T), \quad k = 1, ..., \ell.$$

□

Theorem 4.43 *(Courant-Fisher principle) Let* $\mathbf{V}$ *be an* n*-dimensional IPS and* $T \in \mathbf{S}(\mathbf{V})$. *Then*

$$\lambda_k(T) = \min_{\mathbf{W} \in \mathrm{Gr}(k-1, \mathbf{V})} \max_{\mathbf{0} \neq \mathbf{x} \in \mathbf{W}^\perp} \frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle}, \quad k = 1, ..., n.$$

See Problem 2 for the proof of the theorem and the following corollary.

Corollary 4.44 *Let* $\mathbf{V}$ *be an* n*-dimensional IPS and* $T \in \mathbf{S}(\mathbf{V})$*. Let* $k, \ell \in [1, n]$ *be integers satisfying* $k \leq l$*. Then*

$$\lambda_{n-\ell+k}(T) \leq \lambda_k(Q(T, \mathbf{W})) \leq \lambda_k(T), \quad \text{for any } \mathbf{W} \in \mathrm{Gr}(\ell, \mathbf{V}).$$

Theorem 4.45 *Let* $\mathbf{V}$ *be an* n*-dimensional IPS and* $S, T \in \mathbf{S}(\mathbf{V})$*. Then for any* $i, j \in \mathbb{N}, i + j - 1 \leq n$ *the inequality* $\lambda_{i+j-1}(S + T) \leq \lambda_i(S) + \lambda_j(T)$ *holds.*

Proof. Let $\mathbf{U}_{i-1}, \mathbf{V}_{j-1} \subset \mathbf{V}$ be eigenspaces of S, T spanned by the first $i - 1, j - 1$ eigenvectors of S, T respectively. So

$$\langle S\mathbf{x}, \mathbf{x}\rangle \leq \lambda_i(S)\langle \mathbf{x}, \mathbf{x}\rangle,\ \langle T\mathbf{y}, \mathbf{y}\rangle \leq \lambda_j(T)\langle \mathbf{y}, \mathbf{y}\rangle \text{ for all } \mathbf{x} \in \mathbf{U}_{i-1}^{\perp}, \mathbf{y} \in \mathbf{V}_{j-1}^{\perp}.$$

Note that $\dim \mathbf{U}_{i-1} = i - 1, \dim \mathbf{V}_{j-1} = j - 1..$ Let $\mathbf{W} = \mathbf{U}_{i-1} + \mathbf{V}_{j-1}$. Then $\dim \mathbf{W} = l - 1 \leq i + j - 2$. Assume that $\mathbf{z} \in \mathbf{W}^{\perp}$. Then $\langle (S + T)\mathbf{z}, \mathbf{z}\rangle = \langle S\mathbf{z}, \mathbf{z}\rangle + \langle T\mathbf{z}, \mathbf{z}\rangle \leq (\lambda_i(S) + \lambda_j(T))\langle \mathbf{z}, \mathbf{z}\rangle$. Hence $\max_{\mathbf{0} \neq \mathbf{z} \in \mathbf{W}^{\perp}} \frac{\langle (S+T)\mathbf{z}, \mathbf{z}\rangle}{\langle \mathbf{z}, \mathbf{z}\rangle} \leq \lambda_i(S) + \lambda_j(T)$. Use Theorem 4.43 to deduce that $\lambda_{i+j-1}(S + T) \leq \lambda_l(S + T) \leq \lambda_i(S) + \lambda_j(T)$. □

Definition 4.46 *Let* $\mathbf{V}$ *be an* n*-dimensional IPS. Fix an integer* $k \in [1, n]$*. Then* $F_k = \{\mathbf{f}_1, ..., \mathbf{f}_k\}$ *is called an orthonormal* k*-frame if* $< \mathbf{f}_i, \mathbf{f}_j >= \delta_{ij}$ *for* $i, j = 1, ..., k$*. Denote by* $\mathrm{Fr}(k, \mathbf{V})$ *the set of all orthonormal* k*-frames in* $\mathbf{V}$*.*

Note that each $F_k \in \mathrm{Fr}(k, \mathbf{V})$ induces $\mathbf{U} = \mathrm{span} F_k \in \mathrm{Gr}(k, \mathbf{V})$. Vice versa, any $\mathbf{U} \in \mathrm{Gr}(k, \mathbf{V})$ induces the set $\mathrm{Fr}(k, \mathbf{U})$ of all orthonormal k-frames which span $\mathbf{U}$.

Theorem 4.47 *Let* $\mathbf{V}$ *be an* n*-dimensional IPS and* $T \in \mathbf{S}(\mathbf{V})$*. Then for any integer* $k \in [1, n]$

$$\sum_{i=1}^{k} \lambda_i(T) = \max_{\{\mathbf{f}_1, ..., \mathbf{f}_k\} \in \mathrm{Fr}(k, \mathbf{V})} \sum_{i=1}^{k} \langle T\mathbf{f}_i, \mathbf{f}_i\rangle.$$

Furthermore

$$\sum_{i=1}^{k} \lambda_i(T) = \sum_{i=1}^{k} \langle T\mathbf{f}_i, \mathbf{f}_i\rangle$$

for some k*-orthonormal frame* $F_k = \{\mathbf{f}_1, ..., \mathbf{f}_k\}$ *if and only if* $\mathrm{span} F_k$ *is spanned by* $\mathbf{e}_1, ..., \mathbf{e}_k$ *satisfying (4.3).*

Proof. Define

$$\begin{aligned} \mathrm{tr}\, Q(T, \mathbf{U}) &:= \sum_{i=1}^{k} \lambda_i(Q(T, \mathbf{U})) \quad \text{for } \mathbf{U} \in \mathrm{Gr}(k, \mathbf{V}), \\ \mathrm{tr}_k T &:= \sum_{i=1}^{k} \lambda_i(T). \end{aligned} \tag{4.8}$$

Let $F_k = \{\mathbf{f}_1, ..., \mathbf{f}_k\} \in \mathrm{Fr}(k, \mathbf{V})$. Set $\mathbf{U} = \mathrm{span} F_k$. Then in view of Corollary 4.42

$$\sum_{i=1}^{k} \langle T\mathbf{f}_i, \mathbf{f}_i\rangle = \mathrm{tr}\, Q(T, \mathbf{U}) \leq \sum_{i=1}^{k} \lambda_i(T).$$

Let $E_k := \{\mathbf{e}_1, ..., \mathbf{e}_k\}$ where $\mathbf{e}_1, ..., \mathbf{e}_n$ are given by (4.3). Clearly $\operatorname{tr}_k T = \operatorname{tr} Q(T, \operatorname{span} E_k)$. This shows the maximal characterization of $\operatorname{tr}_k T$.

Let $\mathbf{U} \in \operatorname{Gr}(k, \mathbf{V})$ and assume that $\operatorname{tr}_k T = \operatorname{tr} Q(T, \mathbf{U})$. Hence $\lambda_i(T) = \lambda_i(Q(T, \mathbf{U}))$ for $i = 1, ..., k$. Then there exists $G_k = \{\mathbf{g}_1, ..., \mathbf{g}_k\} \in \operatorname{Fr}(k, \mathbf{U}))$ such that

$$\min_{\mathbf{0} \neq \mathbf{x} \in \operatorname{span}(\mathbf{g}_1, ..., \mathbf{g}_i\}} \frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \lambda_i(Q(T, \mathbf{U})) = \lambda_i(T), \ i = 1, ..., k.$$

Use Theorem 4.41 to deduce that $T\mathbf{g}_i = \lambda_i(T)\mathbf{g}_i$ for $i = 1, ..., k$. □

Theorem 4.48 *Let $\mathbf{V}$ be an n-dimensional IPS and $T \in \mathbf{S}(\mathbf{V})$. Then for any integer $k, l \in [1, n]$, such that $k + l \leq n$*

$$\sum_{i=l+1}^{l+k} \lambda_i(T) = \min_{\mathbf{W} \in \operatorname{Gr}(l, \mathbf{V})} \max_{\{\mathbf{f}_1, ..., \mathbf{f}_k\} \in \operatorname{Fr}(k, \mathbf{V} \cap \mathbf{W}^\perp)} \sum_{i=1}^{k} \langle T\mathbf{f}_i, \mathbf{f}_i \rangle.$$

Proof. Let $\mathbf{W}_j := \operatorname{span}(\mathbf{e}_1, ..., \mathbf{e}_j), j = 1, \ldots, n$, where $\mathbf{e}_1, ..., \mathbf{e}_n$ are given by (4.3). Then $\mathbf{V}_1 := \mathbf{V} \cap \mathbf{W}_l$ is an invariant subspace of T. Let $T_1 := T|\mathbf{V}_1$. Then $\lambda_i(T_1) = \lambda_{l+i}(T)$ for $i = 1, \ldots, n - l$. Theorem 4.47 for T_1 yields

$$\max_{\{\mathbf{f}_1, ..., \mathbf{f}_k\} \in \operatorname{Fr}(k, \mathbf{V} \cap \mathbf{W}_l^\perp)} \sum_{i=1}^{k} \langle T\mathbf{f}_i, \mathbf{f}_i \rangle = \sum_{i=l+1}^{l+k} \lambda_i(T).$$

Let $T_2 := T|\mathbf{W}_{l+k}$ and $\mathbf{W} \in \operatorname{Gr}(l, \mathbf{V})$. Set $\mathbf{U} := \mathbf{W}_{l+k} \cap \mathbf{W}^\perp$. Then $\dim \mathbf{U} \geq k$. Apply Theorem 4.47 to $-T_2$ to deduce

$$\sum_{i=1}^{k} \lambda_i(-T_2) \geq \sum_{i=1}^{k} \langle -T\mathbf{f}_i, \mathbf{f}_i \rangle \text{ for } \{\mathbf{f}_1, ..., \mathbf{f}_k\} \in \operatorname{Fr}(k, \mathbf{U}).$$

The above inequality is equal to the inequality

$$\sum_{i=l+1}^{l+k} \lambda_i(T) \leq \sum_{i=1}^{k} \langle T\mathbf{f}_i, \mathbf{f}_i \rangle \text{ for } \{\mathbf{f}_1, ..., \mathbf{f}_k\} \in \operatorname{Fr}(k, \mathbf{U}) \leq$$

$$\max_{\{\mathbf{f}_1, ..., \mathbf{f}_k\} \in \operatorname{Fr}(k, \mathbf{V} \cap \mathbf{W}^\perp)} \sum_{i=1}^{k} \langle T\mathbf{f}_i, \mathbf{f}_i \rangle.$$

The above inequalities yield the theorem. □

Problems

1. Let $\mathbf{V}$ be 3 dimensional IPS and $T \in \operatorname{Hom}(\mathbf{V})$ be self-adjoint. Assume that

$$\lambda_1(T) > \lambda_2(T) > \lambda_3(T), \quad T\mathbf{e}_i = \lambda_i(T)\mathbf{e}_i, \ i = 1, 2, 3.$$

Let $\mathbf{W} = \operatorname{span}(\mathbf{e}_1, \mathbf{e}_3)$.

(a) Show that for each $t \in (\lambda_3(T), \lambda_1(T))$ there exist two $\mathbf{W}(t) \in \operatorname{Gr}(1, \mathbf{W})$ such that $\lambda_1(Q(T, \mathbf{W}(t))) = t$.

(b) Let $t \in [\lambda_2(T), \lambda_1(T)]$. Let $\mathbf{U}(t) = \operatorname{span}(\mathbf{W}(t), \mathbf{e}_2) \in \operatorname{Gr}(2, \mathbf{V})$. Show that $\lambda_2(T) = \lambda_2(Q(T, \mathbf{U}(t))$.

2. (a) Let the assumptions of Theorem 4.43 hold. Let $\mathbf{W} \in \mathrm{Gr}(k-1, \mathbf{V})$. Show that there exists $\mathbf{0} \neq \mathbf{x} \in \mathbf{W}^{\perp}$ such that $\langle \mathbf{x}, \mathbf{e}_i \rangle = 0$ for $k+1, ..., n$, where $\mathbf{e}_1, ..., \mathbf{e}_n$ satisfy (4.3). Conclude that $\lambda_1(Q(T, \mathbf{W}^{\perp})) \geq \frac{\langle T\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \geq \lambda_k(T)$.

   (b) Let $\mathbf{U}_\ell = \mathrm{span}(\mathbf{e}_1, ..., \mathbf{e}_\ell)$. Show that $\lambda_1(Q(T, \mathbf{U}_\ell^{\perp})) = \lambda_{\ell+1}(T)$ for $\ell = 1, ..., n-1$.

   (c) Prove Theorem 4.43.

   (d) Prove Corollary 4.44. (**Hint**: Choose $\mathbf{U} \in \mathrm{Gr}(k, \mathbf{W})$ such that $\mathbf{U} \subset \mathbf{W} \cap \mathrm{span}(\mathbf{e}_{n-\ell+k+1}, ..., \mathbf{e}_n)^{\perp}$. Then $\lambda_{n-\ell+k}(T) \leq \lambda_k(Q(T, \mathbf{U})) \leq \lambda_k(Q(T, \mathbf{W}))$.)

3. Let $B = [b_{ij}]_{i,j=1}^n \in \mathbf{H}_n$ and denote by $A \in \mathbf{H}_{n-1}$ the matrix obtained from B by deleting the $j-th$ row and column.

   (a) Show the Cauchy interlacing inequalities

   $$\lambda_i(B) \geq \lambda_i(A) \geq \lambda_{i+1}(B), \text{ for } i = 1, ..., n-1.$$

   (b) Show that inequality $\lambda_1(B) + \lambda_n(B) \leq \lambda_1(A) + b_{ii}$.
   Hint. Express the traces of B and A respectively in terms of eigenvalues to obtain

   $$\lambda_1(B) + \lambda_n(B) = b_{ii} + \lambda_1(A) + \sum_{i=2}^{n-1} (\lambda_i(A) - \lambda_i(B)).$$

   Then use the Cauchy interlacing inequalities.

4. Show the following generalization of Problem 3.b ([**?**, p.56]). Let $B \in \mathbf{H}_n$ be the following 2×2 block matrix $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{bmatrix}$. Show that

   $$\lambda_1(B) + \lambda_n(B) \leq \lambda_1(B_{11}) + \lambda_1(B_{22}).$$

   Hint. Assume that $B\mathbf{x} = \lambda_1(B)\mathbf{x}, \mathbf{x}^{\top} = (\mathbf{x}_1^{\top}, \mathbf{x}_2^{\top})$, partitioned as B. Consider $\mathbf{U} = \mathrm{span}((\mathbf{x}_1^{\top}, \mathbf{0})^{\top}, (\mathbf{0}, \mathbf{x}_2^{\top})^{\top})$. Analyze $\lambda_1(Q(T, \mathbf{U})) + \lambda_2(Q(T, \mathbf{U}))$.

5. Let $B = (b_{ij})_1^n \in \mathbf{H}_n$. Show that $B > 0$ if and only if $\det(b_{ij})_1^k > 0$ for $k = 1, ..., n$.

6. Let $T \in \mathbf{S}(\mathbf{V})$. Denote by $\iota_+(T), \iota_0(T), \iota_-(T)$ the number of positive, negative and zero eigenvalues among $\lambda_1(T) \geq ... \geq \lambda_n(T)$. The triple $\iota(T) := (\iota_+(T), \iota_0(T), \iota_-(T))$ is called the inertia of T. For $B \in \mathbf{H}_n$ let $\iota(B) := (\iota_+(B), \iota_0(B), \iota_-(B))$ be the inertia of B, where $\iota_+(B), \iota_0(B), \iota_-(B)$ is the number of positive, negative and zero eigenvalues of B respectively. Let $\mathbf{U} \in \mathrm{Gr}(k, \mathbf{V})$. Show

   (a) Assume that $\lambda_k(Q(T, \mathbf{U})) > 0$, i.e. $Q(T, \mathbf{U}) > 0$. Then $k \leq \iota_+(T)$. If $k = \iota_+(T)$ then one can choose $\mathbf{U}$ to be an invariant subspace of $\mathbf{V}$ spanned by the eigenvectors of T corresponding to positive eigenvalues of T. (Usually such a subspace is not unique.)

   (b) Assume that $\lambda_k(Q(T, \mathbf{U})) \geq 0$, i.e. $Q(T, \mathbf{U}) \geq 0$. Then $k \leq \iota_+(T) + \iota_0(T)$. If $k = \iota_+(T) + \iota_0(T)$ then $\mathbf{U}$ is the unique invariant subspace of $\mathbf{V}$ spanned by the eigenvectors of T corresponding to nonnegative eigenvalues of T.

(c) Assume that $\lambda_1(Q(T,\mathbf{U})) < 0$, i.e. $Q(T,\mathbf{U}) < 0$. Then $k \leq \iota_-(T)$. If $k = \iota_-(T)$ then $\mathbf{U}$ can be chosen to be an invariant subspace of $\mathbf{V}$ spanned by the eigenvectors of T, corresponding to negative eigenvalues of T. (Usually such a subspace may not be unique.)

(d) Assume that $\lambda_1(Q(T,\mathbf{U})) \leq 0$, i.e. $Q(T,\mathbf{U}) \leq 0$. Then $k \leq \iota_-(T) + \iota_0(T)$. If $k = \iota_-(T) + \iota_0(T)$ then $\mathbf{U}$ is a unique invariant subspace of $\mathbf{V}$ spanned by the eigenvectors of T corresponding to nonpositive eigenvalues of T.

7. Let $B \in \mathbf{H}_n$ and assume that $A = PBP^*$ for some $P \in \mathrm{GL}(n,\mathbb{C})$. Then $\iota(A) = \iota(B)$.

4.10 Positive definite operators and matrices

Definition 4.49 *Let $\mathbf{V}$ be a finite dimensional IPS over $\mathbb{F} = \mathbb{C}, \mathbb{R}$. Let $S, T \in \mathbf{S}(\mathbf{V})$. Then $T > S$, $(T \geq S)$ if $\langle T\mathbf{x}, \mathbf{x}\rangle > \langle S\mathbf{x}, \mathbf{x}\rangle$, $(\langle T\mathbf{x}, \mathbf{x}\rangle \geq \langle S\mathbf{x}, \mathbf{x}\rangle)$ for all $\mathbf{0} \neq \mathbf{x} \in \mathbf{V}$. T is called positive (nonnegative) definite if $T > 0$ $(T \geq 0)$, where 0 is the zero operator in $\mathrm{Hom}(\mathbf{V})$.*

Denote by $\mathbf{S}_+(\mathbf{V})^o \subset \mathbf{S}_+(\mathbf{V}) \subset \mathbf{S}(\mathbf{V})$ the open set of positive definite self adjoint operators and the closed set of nonnegative selfadjoint operators respectively.

Let P, Q be either quadratic forms if $\mathbb{F} = \mathbb{R}$ or hermitian forms if $\mathbb{F} = \mathbb{C}$. Then $Q > P$, $(Q \geq P)$ if $Q(\mathbf{x}, \mathbf{x}) > P(\mathbf{x}, \mathbf{x})$, $(Q(\mathbf{x}, \mathbf{x}) \geq P(\mathbf{x}, \mathbf{x}))$ for all $\mathbf{0} \neq \mathbf{x} \in \mathbf{V}$. Q is called positive (nonnegative) definite if $Q > 0$ $(Q \geq 0)$, where 0 is the zero operator in $\mathrm{Hom}(\mathbf{V})$.

For $A, B \in \mathbf{H}_n$ $B > A$ $(B \geq A)$ if $\mathbf{x}^* B\mathbf{x} > \mathbf{x}^* A\mathbf{x}$ $(\mathbf{x}^* B\mathbf{x} \geq \mathbf{x}^* A\mathbf{x})$ for all $\mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n$. $B \in \mathbf{H}_n$ is called is called positive (nonnegative) definite if $B > 0$ $(B \geq 0)$. Denote by $\mathbf{H}^o_{n,+} \subset \mathbf{H}_{n,+} \subset \mathbf{H}_n$ the open set of positive definite $n \times n$ hermitian matrices and the closed set of $n \times n$ nonnegative hermitian matrices respectively. Let $\mathbf{S}_+(n,\mathbb{R}) := \mathbf{S}(n,\mathbb{R}) \cap \mathbf{H}_{n,+}$, $\mathbf{S}_+(n,\mathbb{R})^o := \mathbf{S}(n,\mathbb{R}) \cap \mathbf{H}^o_{n,+}$.

Use (4.1) to deduce.

Corollary 4.50 *Let $\mathbf{V}$ be n-dimensional IPS. Let $T \in \mathbf{S}(\mathbf{V})$. Then $T > 0$ $(T \geq 0)$ if and only if $\lambda_n(T) > 0$ $(\lambda_n(T) \geq 0)$. Let $S \in \mathbf{S}(\mathbf{V})$ and assume that $T > S$ $(T \geq S)$. Then $\lambda_i(T) > \lambda_i(S)$ $(\lambda_i(T) \geq \lambda_i(S))$ for $i = 1, ..., n$.*

Proposition 4.51 *Let $\mathbf{V}$ be a finite dimensional IPS. Assume that $T \in \mathbf{S}(\mathbf{V})$. Then $T \geq 0$ if and only if there exists $S \in \mathbf{S}(\mathbf{V})$ such that $T = S^2$. Furthermore $T > 0$ if and only if S is invertible. For $0 \leq T \in \mathbf{S}(\mathbf{V})$ there exists a unique $0 \leq S \in \mathbf{S}(\mathbf{V})$ such that $T = S^2$. This S is called the square root of T and is denoted by $T^{\frac{1}{2}}$.*

Proof. Assume first that $T \geq 0$. Let $\mathbf{e}_1, ..., \mathbf{e}_n$ be an orthonormal basis consisting of eigenvectors of T as in (4.3). Since $\lambda_i(T) \geq 0$, $i = 1, ..., n$ we can define $P \in \mathrm{Hom}(\mathbf{V})$ as follows

$$P\mathbf{e}_i = \sqrt{\lambda_i(T)}\mathbf{e}_i, \quad i = 1, ..., n.$$

Clearly P is self-adjoint nonnegative and $T = P^2$.

Suppose now that $T = S^2$ for some $S \in \mathbf{S}(\mathbf{V})$. Then $T \in \mathbf{S}(\mathbf{V})$ and $\langle T\mathbf{x}, \mathbf{x}\rangle = \langle S\mathbf{x}, S\mathbf{x}\rangle \geq 0$. Hence $T \geq 0$. Clearly $\langle T\mathbf{x}, \mathbf{x}\rangle = 0 \iff S\mathbf{x} = 0$. Hence

$T > 0 \iff S \in \mathrm{GL}(\mathbf{V})$. Suppose that $S \geq 0$. Then $\lambda_i(S) = \sqrt{\lambda_i(T)},\ i = 1, ..., n$. Furthermore each eigenvector of S is an eigenvector of T. It is straightforward to show that $S = P$, where P is defined above. Clearly $T > 0$ if and only if $\sqrt{\lambda_n(T)} > 0$, i.e. if and only if S is invertible. □

Corollary 4.52 *Let $B \in \mathbf{H}_n$ $(\mathbf{S}(n, \mathbb{R}))$. Then $B \geq 0$ if and only there exists $A \in \mathbf{H}_n$ $(\mathbf{S}(n, \mathbb{R}))$ such that $B = A^2$. Furthermore $B > 0$ if and only if A is invertible. For $B \geq 0$ there exists a unique $A \geq 0$ such that $B = A^2$. This A is denoted by $B^{\frac{1}{2}}$.*

Theorem 4.53 *Let $\mathbf{V}$ be an IPS over $\mathbb{F} = \mathbb{C}, \mathbb{R}$. Let $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbf{V}$. Then the grammian matrix $G(\mathbf{x}_1, ..., \mathbf{x}_n) := (\langle \mathbf{x}_i, \mathbf{x}_j \rangle)_1^n$ is a hermitian nonnegative definite matrix. (If $\mathbb{F} = \mathbb{R}$ then $G(\mathbf{x}_1, ..., \mathbf{x}_n)$ is real symmetric nonnegative definite.) $G(\mathbf{x}_1, ..., \mathbf{x}_n) > 0$ if and only $\mathbf{x}_1, ..., \mathbf{x}_n$ are linearly independent. Furthemore for any integer $k \in [1, n-1]$*

$$\det G(\mathbf{x}_1, ..., \mathbf{x}_n) \leq \det G(\mathbf{x}_1, ..., \mathbf{x}_k)\ \det G(\mathbf{x}_{k+1}, ..., \mathbf{x}_n). \tag{4.1}$$

Equality holds if and only if either $\det G(\mathbf{x}_1, ..., \mathbf{x}_k)\ \det G(\mathbf{x}_{k+1}, ..., \mathbf{x}_n) = 0$ or $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0$ for $i = 1, ..., k$ and $j = k+1, ..., n$.

Proof. Clearly $G(\mathbf{x}_1, ..., \mathbf{x}_n) \in \mathbf{H}_n$. If $\mathbf{V}$ is an IPS over $\mathbb{R}$ then $G(\mathbf{x}_1, ..., \mathbf{x}_n) \in \mathbf{S}(n, \mathbb{R})$. Let $\mathbf{a} = (a_1, ..., a_n)^\top \in \mathbb{F}^n$. Then

$$\mathbf{a}^* G(\mathbf{x}_1, ..., \mathbf{x}_n)\mathbf{a} = \langle \sum_{i=1}^n a_i \mathbf{x}_i, \sum_{j=1}^n a_j \mathbf{x}_j \rangle \geq 0.$$

Equality holds if and only if $\sum_{i=1}^n a_i \mathbf{x}_i = 0$. Hence $G(\mathbf{x}_1, ..., \mathbf{x}_n) \geq 0$ and $G(\mathbf{x}_1, ..., \mathbf{x}_n) > 0$ if and only if $\mathbf{x}_1, ..., \mathbf{x}_n$ are linearly independent. In particular $\det G(\mathbf{x}_1, ..., \mathbf{x}_n) \geq 0$ and $\det G(\mathbf{x}_1, ..., \mathbf{x}_n) > 0$ if and only if $\mathbf{x}_1, ..., \mathbf{x}_n$ are linearly independent.

We now prove the inequality (4.1). Assume first that the right-hand side of (4.1) is zero. Then either $\mathbf{x}_1, ..., \mathbf{x}_k$ or $\mathbf{x}_{k+1}, ..., \mathbf{x}_n$ are linearly dependent. Hence $\mathbf{x}_1, ..., \mathbf{x}_n$ are linearly dependent and $\det G = 0$.

Assume now that the right-hand side of (4.1) is positive. Hence $\mathbf{x}_1, ..., \mathbf{x}_k$ and $\mathbf{x}_{k+1}, ..., \mathbf{x}_n$ are linearly independent. If $\mathbf{x}_1, ..., \mathbf{x}_n$ are linearly dependent then $\det G = 0$ and strict inequality holds in (4.1). It is left to show the inequality (4.1) and the equality case when $\mathbf{x}_1, ..., \mathbf{x}_n$ are linearly independent. Perform the Gram-Schmidt algorithm on $\mathbf{x}_1, ..., \mathbf{x}_n$ as given in (4.1). Let $S_j = \mathrm{span}(\mathbf{x}_1, ..., \mathbf{x}_j)$ for $j = 1, ..., n$. Corollary 4.1 yields that $\mathrm{span}(\mathbf{e}_1, ..., \mathbf{e}_{n-1}) = S_{n-1}$. Hence $\mathbf{y}_n = \mathbf{x}_n - \sum_{j=1}^{n-1} b_j \mathbf{x}_j$ for some $b_1, ..., b_{n-1} \in \mathbb{F}$. Let G' be the matrix obtained from $G(\mathbf{x}_1, ..., \mathbf{x}_n)$ by subtracting from the n-th row b_j times j-th row. Thus the last row of G' is $(\langle \mathbf{y}_n, \mathbf{x}_1 \rangle, ..., \langle \mathbf{y}_n, \mathbf{x}_n \rangle) = (0, ..., 0, \|\mathbf{y}_n\|^2)$. Clearly $\det G(\mathbf{x}_1, ..., \mathbf{x}_n) = \det G'$. Expand $\det G'$ by the last row to deduce

$$\begin{aligned} &\det G(\mathbf{x}_1, ..., \mathbf{x}_n) = \det G(\mathbf{x}_i, ..., \mathbf{x}_{n-1})\ \|\mathbf{y}_n\|^2 = ... = \\ &\det G(\mathbf{x}_1, ..., \mathbf{x}_k) \prod_{i=k+1}^n \|\mathbf{y}_i\|^2 = \\ &\det G(\mathbf{x}_1, ..., \mathbf{x}_k) \prod_{i=k+1}^n \mathrm{dist}(\mathbf{x}_i, S_{i-1})^2, \quad k = n-1, ..., 1. \end{aligned} \tag{4.2}$$

Perform the Gram-Schmidt process on $\mathbf{x}_{k+1}, ..., \mathbf{x}_n$ to obtain the orthogonal set of vectors $\hat{\mathbf{y}}_{k+1}, ..., \hat{\mathbf{y}}_n$ such that

$$\hat{S}_j := \text{span}(\mathbf{x}_{k+1}, ..., \mathbf{x}_j) = \text{span}(\hat{\mathbf{y}}_{k+1}, ..., \hat{\mathbf{y}}_j),\ \text{dist}(\mathbf{x}_j, \hat{S}_{j-1}) = \|\hat{\mathbf{y}}_j\|,$$

for $j = k+1, ..., n$, where $\hat{S}_k = \{\mathbf{0}\}$. Use (4.2) to deduce that $\det G(\mathbf{x}_{k+1}, ..., \mathbf{x}_n) = \prod_{j=k+1}^n \|\hat{\mathbf{y}}_j\|^2$. As $\hat{S}_{j-1} \subset S_{j-1}$ for $j > k$ it follows that

$$\|\mathbf{y}_j\| = \text{dist}(\mathbf{x}_j, S_{j-1}) \le \text{dist}(\mathbf{x}_j, \hat{S}_{j-1}) = \|\hat{\mathbf{y}}_j\|,\ j = k+1, ..., n.$$

This shows (4.1). Assume now equality holds in (4.1). Then $\|\mathbf{y}_j\| = \|\hat{\mathbf{y}}_j\|$ for $j = k+1, ..., n$. Since $\hat{S}_{j-1} \subset S_{j-1}$ and $\hat{\mathbf{y}}_j - \mathbf{x}_j \in \hat{S}_{j-1} \subset S_{j-1}$ it follows that $\text{dist}(\mathbf{x}_j, S_{j-1}) = \text{dist}(\hat{\mathbf{y}}_j, S_{j-1}) = \|\mathbf{y}_j\|$. Hence $\|\hat{\mathbf{y}}_j\| = \text{dist}(\hat{\mathbf{y}}_j, S_{j-1})$. Part (h) of Problem 4.1.4 yields that $\hat{\mathbf{y}}_j$ is orthogonal on S_{j-1}. In particular each $\hat{\mathbf{y}}_j$ is orthogonal to S_k for $j = k+1, ..., n$. Hence $\mathbf{x}_j \perp S_k$ for $j = k+1, ..., n$, i.e. $\langle \mathbf{x}_j, \mathbf{x}_i \rangle = 0$ for $j > k$ and $i \le k$. Clearly, if the last condition holds then $\det G(\mathbf{x}_1, ..., \mathbf{x}_n) = \det G(\mathbf{x}_1, ..., \mathbf{x}_k)\ \det G(\mathbf{x}_{k+1}, ..., \mathbf{x}_n)$. □

$\det G(\mathbf{x}_1, ..., \mathbf{x}_n)$ has the following geometric meaning. Consider a parallelepiped Π in $\mathbf{V}$ spanned by $\mathbf{x}_1, ..., \mathbf{x}_n$ starting from the origin $\mathbf{0}$. That is Π is a convex hull spanned by the vectors $\mathbf{0}$ and $\sum_{i \in S} \mathbf{x}_i$ for all nonempty subsets $S \subset \{1, ..., n\}$. Then $\sqrt{\det G(\mathbf{x}_1, ..., \mathbf{x}_n)}$ is the n-volume of Π. The inequality (4.1) and equalities (4.2) are "obvious" from this geometrical point of view.

Corollary 4.54 *Let $0 \le B = (b_{ij})_1^n \in \mathbf{H}_{n,+}$. Then*

$$\det B \le \det(b_{ij})_1^k\ \det(b_{ij})_{k+1}^n,\ \text{for } k = 1, ..., n-1.$$

For a fixed k equality holds if and only if either the right-hand side of the above inequality is zero or $b_{ij} = 0$ for $i = 1, ..., k$ and $j = k+1, ..., n$.

Proof. From Corollary 4.52 it follows that $B = X^2$ for some $X \in \mathbf{H}_n$. Let $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{C}^n$ be the n-columns of $X^T = (\mathbf{x}_1, ..., \mathbf{x}_n)$. Let $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x}$. Since $X \in \mathbf{H}_n$ we deduce that $B = G(\mathbf{x}_1, ..., \mathbf{x}_n)$. □

Theorem 4.55 *Let $\mathbf{V}$ be an n-dimensional IPS. Let $T \in \mathbf{S}$. TFAE:*
(a) *$T > 0$.*
(b) *Let $\mathbf{g}_1, ..., \mathbf{g}_n$ be a basis of $\mathbf{V}$. Then $\det(\langle T\mathbf{g}_i, \mathbf{g}_j \rangle)_{i,j=1}^k > 0,\ k = 1, ..., n$.*

Proof. (a) ⇒ (b). According to Proposition 4.51 $T = S^2$ for some $S \in \mathbf{S}(\mathbf{V}) \cap \text{GL}(\mathbf{V})$. Then $\langle T\mathbf{g}_i, \mathbf{g}_j \rangle = \langle S\mathbf{g}_i, S\mathbf{g}_j \rangle$. Hence $\det(\langle T\mathbf{g}_i, \mathbf{g}_j \rangle)_{i,j=1}^k = \det G(S\mathbf{g}_1, ..., S\mathbf{g}_k)$. Since S is invertible and $\mathbf{g}_1, ..., \mathbf{g}_k$ linearly independent it follows that $S\mathbf{g}_1, ..., S\mathbf{g}_k$ are linearly independent. Theorem 4.1 implies that $\det G(S\mathbf{g}_1, ..., S\mathbf{g}_k) > 0$ for $k = 1, ..., n$.
(b) ⇒ (a). The proof is by induction on n. For $n = 1$ (a) is obvious. Assume that (a) holds for $n = m - 1$. Let $\mathbf{U} := \text{span}(\mathbf{g}_1, ..., \mathbf{g}_{n-1})$ and $Q := Q(T, \mathbf{U})$. Then there exists $P \in \mathbf{S}(\mathbf{U})$ such that $< P\mathbf{x}, \mathbf{y} >= Q(\mathbf{x}, \mathbf{y}) =< T\mathbf{x}, \mathbf{y} >$ for any $\mathbf{x}, \mathbf{y} \in \mathbf{U}$. By induction $P > 0$. Corollary 4.42 yields that $\lambda_{n-1}(T) \ge \lambda_{n-1}(P) > 0$. Hence T has at least $n-1$ positive eigenvalues. Let $\mathbf{e}_1, ..., \mathbf{e}_n$ be given by (4.3).

Then $\det(\langle T\mathbf{e}_i, \mathbf{e}_j\rangle)_{i,j=1}^n = \prod_{i=1}^n \lambda_i(T) > 0$. Let $A = (a_{pq})_1^n \in \mathrm{GL}(n, \mathbb{C})$ be the transformation matrix from the basis $\mathbf{g}_1, ..., \mathbf{g}_n$ to $\mathbf{e}_1, ..., \mathbf{e}_n$, i.e.

$$\mathbf{g}_i = \sum_{p=1}^n a_{pi}\mathbf{e}_p,\ i = 1, ..., n.$$

It is straightforward to show that

$$\begin{gathered}(\langle T\mathbf{g}_i, \mathbf{g}_j\rangle)_1^n = A^T(\langle T\mathbf{e}_p, \mathbf{e}_q\rangle)\bar{A} \Rightarrow \\ \det(\langle T\mathbf{g}_i, \mathbf{g}_j\rangle)_1^n = \det(\langle T\mathbf{e}_i, \mathbf{e}_j\rangle)_1^n |\det A|^2 = |\det A|^2 \prod_{i=1}^n \lambda_i(T).\end{gathered} \tag{4.3}$$

Since $\det(\langle T\mathbf{g}_i, \mathbf{g}_j\rangle)_1^n > 0$ and $\lambda_1(T) \geq ... \geq \lambda_{n-1}(T) > 0$ it follows that $\lambda_n(T) > 0$. □

Corollary 4.56 *Let $B = (b_{ij})_1^n \in \mathbf{H}_n$. Then $B > 0$ if and only if $\det(b_{ij})_1^k > 0$ for $k = 1, ..., n$.*

The following result is straightforward (see Problem 1):

Proposition 4.57 *Let $\mathbf{V}$ be a finite dimensional IPS over $\mathbb{F} = \mathbb{R}, \mathbb{C}$ with the inner product $\langle \cdot, \cdot \rangle$. Assume that $T \in \mathbf{S}(\mathbf{V})$. Then $T > 0$ if and only if $(\mathbf{x}, \mathbf{y}) := \langle T\mathbf{x}, \mathbf{y}\rangle$ is an inner product on $\mathbf{V}$. Vice versa any inner product $(\cdot, \cdot) : \mathbf{V} \times \mathbf{V} \to \mathbb{R}$ is of the form $(\mathbf{x}, \mathbf{y}) =< T\mathbf{x}, \mathbf{y} >$ for a unique self-adjoint positive definite operator $T \in \mathrm{Hom}(\mathbf{V})$.*

Example 4.58 *Each $0 < B \in \mathbf{H}_n$ induces and inner product on $\mathbb{C}^n$: $(\mathbf{x}, \mathbf{y}) = \mathbf{y}^* B\mathbf{x}$. Each $0 < B \in \mathbf{S}(n, \mathbb{R})$ induces and inner product on $\mathbb{R}^n$: $(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T B\mathbf{x}$. Furthermore any inner product on $\mathbb{C}^n$ or $\mathbb{R}^n$ is of the above form. In particular, the standard inner products on $\mathbb{C}^n$ and $\mathbb{R}^n$ are induces by the identity matrix I.*

Definition 4.59 *Let $\mathbf{V}$ be a finite dimensional IPS with the inner product $\langle \cdot, \cdot \rangle$. Let $S \in \mathrm{Hom}(V)$. Then S is called symmetrizable if there exists an inner product $(\cdot, \cdot)$ on $\mathbf{V}$ such that S is self-adjoint with respect to $(\cdot, \cdot)$.*

Problems

1. Show Proposition 4.57.

2. Recall the Hölder inequality

$$\sum_{l=1}^n x_l y_l a_l \leq (\sum_{l=1}^n x_l^p a_l)^{\frac{1}{p}} (\sum_{l=1}^n y_l^q a_l)^{\frac{1}{q}} \tag{4.4}$$

for any $\mathbf{x} = (x_1, \ldots, x_n)^\top, \mathbf{y} = (y_1, \ldots, y_n)^\top, \mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}_+^n$ and $p, q \in (1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Show

(a) Let $A \in \mathbf{H}_{n,+}, \mathbf{x} \in \mathbb{C}^n$ and $0 \le i < j < k$ be three integers. Then

$$\mathbf{x}^* A^j \mathbf{x} \le (\mathbf{x}^* A^i \mathbf{x})^{\frac{k-j}{k-i}} (\mathbf{x}^* A^k \mathbf{x})^{\frac{j-i}{k-i}}. \tag{4.5}$$

Hint: Diagonalize A.

(b) Assume that $A = e^B$ for some $B \in \mathbf{H}_n$. Show that (4.5) holds for any three real numbers $i < j < k$.

4.11 Inequalities for traces

Let $\mathbf{V}$ be a finite dimensional IPS over $\mathbb{F} = \mathbb{R}, \mathbb{C}$. Let $T : \mathbf{V} \to \mathbf{V}$ be a linear operator. Then $\operatorname{tr} T$ is the trace of the representation matrix A of with respect to any orthonormal basis of $\mathbf{V}$. See Problem 1.

Theorem 4.60 *Let $\mathbf{V}$ be an n-dimensional IPS over $\mathbb{F} = \mathbb{R}, \mathbb{C}$. Assume that $S, T \in \mathbf{S}(\mathbf{V})$. Then $\operatorname{tr} ST$ is bounded below and above by*

$$\sum_{i=1}^{n} \lambda_i(S)\lambda_{n-i+1}(T) \le \operatorname{tr} ST \le \sum_{i=1}^{n} \lambda_i(S)\lambda_i(T). \tag{4.1}$$

Equality for the upper bound holds if and only if $ST = TS$ and there exists an orthonormal basis $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbf{V}$ such that

$$S\mathbf{x}_i = \lambda_i(S)\mathbf{x}_i, \quad T\mathbf{x}_i = \lambda_i(T)\mathbf{x}_i, \quad i = 1, ..., n. \tag{4.2}$$

Equality for the lower bound holds if and only if $ST = TS$ and there exists an orthonormal basis $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbf{V}$ such that

$$S\mathbf{x}_i = \lambda_i(S)\mathbf{x}_i, \quad T\mathbf{x}_i = \lambda_{n-i+1}(T)\mathbf{x}_i, \quad i = 1, ..., n. \tag{4.3}$$

Proof. Let $\mathbf{y}_1, ..., \mathbf{y}_n$ be an orthonormal basis of $\mathbf{V}$ such that

$$\begin{aligned}
&T\mathbf{y}_i = \lambda_i(T)\mathbf{y}_i, \quad i = 1, ..., n, \\
&\lambda_1(T) = ... = \lambda_{i_1}(T) > \lambda_{i_1+1}(T) = ... = \lambda_{i_2}(T) > ... > \\
&\lambda_{i_{k-1}+1}(T) = ... = \lambda_{i_k}(T) = \lambda_n(T), \ 1 \le i_1 < ... < i_k = n.
\end{aligned}$$

If $k = 1 \iff i_1 = n$ it follows that $T = \lambda_1 I$ and the theorem is trivial in this case. Assume that $k > 1$. Then

$$\begin{aligned}
&\operatorname{tr} ST = \sum_{i=1}^{n} \lambda_i(T)\langle S\mathbf{y}_i, \mathbf{y}_i\rangle = \\
&\sum_{i=1}^{n-1} (\lambda_i(T) - \lambda_{i+1}(T))(\sum_{l=1}^{i} \langle S\mathbf{y}_l, \mathbf{y}_l\rangle) + \lambda_n(T)(\sum_{l=1}^{n} \langle S\mathbf{y}_l, \mathbf{y}_l\rangle) = \\
&\sum_{j=1}^{k-1} (\lambda_{i_j}(T) - \lambda_{i_{j+1}}(T)) \sum_{l=1}^{i_j} \langle S\mathbf{y}_l, \mathbf{y}_l\rangle + \lambda_n(T) \operatorname{tr} S.
\end{aligned}$$

Theorem 4.47 yields that $\sum_{l=1}^{i_j} \langle S\mathbf{y}_l, \mathbf{y}_l\rangle \le \sum_{l=1}^{i_j} \lambda_l(S)$. Substitute these inequalities for $j = 1, ..., k-1$ in the above identity to deduce the upper bound in (4.1).

Clearly the condition (4.2) implies that $\operatorname{tr} ST$ is equal to the upper bound in (4.1). Assume now that $\operatorname{tr} ST$ is equal to the upper bound in (4.1). Then $\sum_{l=1}^{i_j}\langle S\mathbf{y}_l, \mathbf{y}_l\rangle = \sum_{l=1}^{i_j}\lambda_l(S)$ for $j = 1, ..., k-1$. Theorem 4.47 yields that $\operatorname{span}(\mathbf{y}_1, ..., \mathbf{y}_{i_j})$ is spanned by some i_j eigenvectors of S corresponding to the first i_j eigenvalues of S for $j = 1, ..., k-1$. Let $\mathbf{x}_1, ..., \mathbf{x}_{i_1}$ be an orthonormal basis of $\operatorname{span}(\mathbf{y}_1, ..., \mathbf{y}_{i_1})$ consisting of the eigenvectors of S corresponding to the eigenvalues of $\lambda_1(S), ..., \lambda_{i_1}(S)$. Since any $0 \neq \mathbf{x} \in \operatorname{span}(\mathbf{y}_1, ..., \mathbf{y}_{i_1})$ is an eigenvector of T corresponding to the eigenvalue $\lambda_{i_1}(T)$ it follows that (4.2) holds for $i = 1, ..., i_1$. Consider $\operatorname{span}(\mathbf{y}_1, ..., \mathbf{y}_{i_2})$. The above arguments imply that this subspace contains i_2 eigenvectors of S and T corresponding to the first i_2 eigenvalues of S and T. Hence $\mathbf{U}_2$, the orthogonal complement of $\operatorname{span}(\mathbf{x}_1, ..., \mathbf{x}_{i_1})$ in $\operatorname{span}(\mathbf{y}_1, ..., \mathbf{y}_{i_2})$, spanned by $\mathbf{x}_{i_1+1}, ..., \mathbf{x}_{i_2}$, which are $i_2 - i_1$ orthonormal eigenvectors of S corresponding to the eigenvalues $\lambda_{i_1+}(S), ..., \lambda_{i_2}(S)$. Since any nonzero vector in $\mathbf{U}_2$ is an eigenvector of T corresponding to the eigenvalue $\lambda_{i_2}(T)$ we deduce that (4.2) holds for $i = 1, ..., i_2$. Continuing in the same manner we obtain (4.2).

To prove the equality case in the lower bound consider the equality in the upper bound for $\operatorname{tr} S(-T)$. □

Corollary 4.61 *Let $\mathbf{V}$ be an n-dimensional IPS over $\mathbb{F} = \mathbb{R}, \mathbb{C}$. Assume that $S, T \in \mathbf{S}(\mathbf{V})$. Then*

$$\sum_{i=1}^{n}(\lambda_i(S) - \lambda_i(T))^2 \leq \operatorname{tr}(S-T)^2. \tag{4.4}$$

Equality holds if and only if $ST = TS$ and $\mathbf{V}$ has an orthonormal basis $\mathbf{x}_1, ..., \mathbf{x}_n$ satisfying (4.2).

Proof. Note

$$\sum_{i=1}^{n}(\lambda_i(S) - \lambda_i(T))^2 = \operatorname{tr} S^2 + \operatorname{tr} T^2 - 2\sum_{i=1}^{n}\lambda_i(S)\lambda_i(T).$$

□

Corollary 4.62 *Let $S, T \in \mathbf{H}_n$. Then the inequalities (4.1) and (4.4) hold. Equalities in the upper bounds hold if and only if there exists $U \in \mathbf{U}_n$ such that $S = U \operatorname{diag}\lambda(S)U^*, T = U \operatorname{diag}\lambda(T)U^*$. Equality in the lower bound of (4.1) if and only if there exists $V \in \mathbf{U}_n$ such that $S = V \operatorname{diag}\lambda(S)V^*, -T = V \operatorname{diag}\lambda(-T)V^*$.*

Problems

1. Let $\mathbf{V}$ be a n-dimensional IPS over $\mathbb{F} = \mathbb{R}, \mathbb{C}$.

   (a) Assume that $T : \mathbf{V} \to \mathbf{V}$ be a linear transformation. Show that for any o.n. basis $\mathbf{x}_1, ..., \mathbf{x}_n$

   $$\operatorname{tr} T = \sum_{i=1}^{n}\langle T\mathbf{x}_i, \mathbf{x}_i\rangle.$$

   Furthermore, if $\mathbb{F} = \mathbb{C}$ then $\operatorname{tr} T$ is the sum of the n eigenvalues of T.

   (b) Let $S, T \in \mathbf{S}(\mathbf{V})$. Show that $\operatorname{tr} ST = \operatorname{tr} TS \in \mathbb{R}$.

4.12 Singular Value Decomposition

Let $\mathbf{U}, \mathbf{V}$, be finite dimensional IPS over $\mathbb{F} = \mathbb{R}, \mathbb{C}$, with the inner products $\langle \cdot, \cdot \rangle_{\mathbf{U}}, \langle \cdot, \cdot \rangle_{\mathbf{V}}$ respectively. Let $\mathbf{u}_1, ..., \mathbf{u}_m$ and $\mathbf{v}_1, ..., \mathbf{v}_n$ be bases in $\mathbf{U}$ and $\mathbf{V}$ respectively. Let $T : \mathbf{V} \to \mathbf{U}$ be a linear operator. In these bases T is represented by a matrix $A = [a_{ij}] \in \mathbb{F}^{m \times n}$ as given by

$$T\mathbf{v}_j = \sum_{i=1}^{m} a_{ij}\mathbf{u}_i, \quad j = 1, \dots, n.$$

Let $T^* : \mathbf{U}^* = \mathbf{U} \to \mathbf{V}^* = \mathbf{V}$. Then $T^*T : \mathbf{V} \to \mathbf{V}$ and $TT^* : \mathbf{U} \to \mathbf{U}$ are selfadjoint operators. As

$$\langle T^*T\mathbf{v}, \mathbf{v} \rangle_{\mathbf{V}} = \langle T\mathbf{v}, T\mathbf{v} \rangle_{\mathbf{V}} \geq 0, \quad \langle TT^*\mathbf{u}, \mathbf{u} \rangle_{\mathbf{U}} = \langle T^*\mathbf{u}, T^*\mathbf{u} \rangle_{\mathbf{U}} \geq 0$$

it follows that $T^*T \geq 0, TT^* \geq 0$. Let

$$\begin{aligned} &T^*T\mathbf{c}_i = \lambda_i(T^*T)\mathbf{c}_i, \ \langle \mathbf{c}_i, \mathbf{c}_k \rangle_{\mathbf{V}} = \delta_{ik}, \ i, k = 1, ..., n, && (4.1)\\ &\lambda_1(T^*T) \geq ... \geq \lambda_n(T^*T) \geq 0,\\ &TT^*\mathbf{d}_j = \lambda_j(TT^*)\mathbf{d}_j, \ \langle \mathbf{d}_j, \mathbf{d}_l \rangle_{\mathbf{U}} = \delta_{jl}, \ j, l = 1, ..., m, && (4.2)\\ &\lambda_1(TT^*) \geq ... \geq \lambda_m(TT^*) \geq 0, \end{aligned}$$

Proposition 4.63 *Let $\mathbf{U}, \mathbf{V}$, be finite dimensional IPS over $\mathbb{F} = \mathbb{R}, \mathbb{C}$. Let $T : \mathbf{V} \to \mathbf{U}$. Then* $\operatorname{rank} T = \operatorname{rank} T^* = \operatorname{rank} T^*T = \operatorname{rank} TT^* = r$. *Furthermore the selfadjoint nonnegative definite operators T^*T and TT^* have exactly r positive eigenvalues, and*

$$\lambda_i(T^*T) = \lambda_i(TT^*) > 0, \quad i = 1, ..., \operatorname{rank} T. \tag{4.3}$$

Moreover for $i \in [1, r]$ $T\mathbf{c}_i$ and $T^*\mathbf{d}_i$ are eigenvectors of TT^* and T^*T corresponding to the eigenvalue $\lambda_i(TT^*) = \lambda_i(T^*T)$ respectively. Furthermore if $\mathbf{c}_1, ..., \mathbf{c}_r$ satisfy (4.1) then $\tilde{\mathbf{d}}_i := \frac{T\mathbf{c}_i}{||T\mathbf{c}_i||}, i = 1, ..., r$ satisfy (4.2) for $i = 1, ..., r$. Similar result holds for $\mathbf{d}_1, ..., \mathbf{d}_r$.

Proof. Clearly $T\mathbf{x} = 0 \iff \langle T\mathbf{x}, T\mathbf{x} \rangle = 0 \iff T^*T\mathbf{x} = 0$. Hence

$$\operatorname{rank} T^*T = \operatorname{rank} T = \operatorname{rank} T^* = \operatorname{rank} TT^* = r.$$

Thus T^*T and TT^* have exactly r positive eigenvalues. Let $i \in [1, r]$. Then $T^*T\mathbf{c}_i \neq 0$. Hence $T\mathbf{c}_i \neq 0$. (4.1) yields that $TT^*(T\mathbf{c}_i) = \lambda_i(T^*T)(T\mathbf{c}_i)$. Similarly $T^*T(T^*\mathbf{d}_i) = \lambda_i(TT^*)(T^*\mathbf{d}_i) \neq 0$. Hence (4.3) holds. Assume that $\mathbf{c}_1, ..., \mathbf{c}_r$ satisfy (4.1). Let $\tilde{\mathbf{d}}_1, ..., \tilde{\mathbf{d}}_r$ be defined as above. By the definition $||\tilde{\mathbf{d}}_i|| = 1, i = 1, ..., r$. Let $1 \leq i < j \leq r$. Then

$$0 = \langle \mathbf{c}_i, \mathbf{c}_j \rangle = \lambda_i(T^*T)\langle \mathbf{c}_i, \mathbf{c}_j \rangle = \langle T^*T\mathbf{c}_i, \mathbf{c}_j \rangle = \langle T\mathbf{c}_i, T\mathbf{c}_j \rangle \Rightarrow \langle \tilde{\mathbf{d}}_i, \tilde{\mathbf{d}}_j \rangle = 0.$$

Hence $\tilde{\mathbf{d}}_1, ..., \tilde{\mathbf{d}}_r$ is an orthonormal system. □

Let

$$\begin{aligned} &\sigma_i(T) = \sqrt{\lambda_i(T^*T)} \text{ for } i = 1, ...r, \quad \sigma_i(T) = 0 \text{ for } i > r,\\ &\sigma_{(p)}(T) := (\sigma_1(T), ..., \sigma_p(T))^\top \in \mathbb{R}^p_{\searrow}, \quad p \in \mathbb{N}. \end{aligned} \tag{4.4}$$

Then $\sigma_i(T) = \sigma_i(T^*), i = 1, ..., \min(m, n)$ are called the singular values of T and T^* respectively. Note that the singular values are arranged in a decreasing order. The positive singular values are called principal singular values of T and T^* respectively. Note that

$$
\begin{aligned}
&||T\mathbf{c}_i||^2 = \langle T\mathbf{c}_i, T\mathbf{c}_i\rangle = \langle T^*T\mathbf{c}_i, \mathbf{c}_i\rangle = \lambda_i(T^*T) = \sigma_i^2 \Rightarrow \\
&||T\mathbf{c}_i|| = \sigma_i,\ i = 1, ..., n, \\
&||T^*\mathbf{d}_j||^2 = \langle T^*\mathbf{d}_j, T^*\mathbf{d}_j\rangle = \langle TT^*\mathbf{d}_j, \mathbf{d}_i\rangle = \lambda_i(TT^*) = \sigma_j^2 \Rightarrow \\
&||T\mathbf{d}_j|| = \sigma_j,\ j = 1, ..., m.
\end{aligned}
$$

Let $\mathbf{c}_1, ...\mathbf{c}_n$ be an orthonormal basis of $\mathbf{V}$ satisfying (4.1). Choose an orthonormal basis $\mathbf{d}_1, ..., \mathbf{d}_m$ as follows. Set $\mathbf{d}_i := \frac{T\mathbf{c}_i}{\sigma_i}, i = 1, ..., r$. Then complete the orthonormal set $\{\mathbf{d}_1, ..., \mathbf{d}_r\}$ to an orthonormal basis of $\mathbf{U}$. Since $\text{span}(\mathbf{d}_1, ..., \mathbf{d}_r)$ is spanned by all eigenvectors of TT^* corresponding to nonzero eigenvalues of TT^* it follows that $\ker T^* = \text{span}(\mathbf{d}_{r+1}, ..., \mathbf{d}_m)$. Hence (4.2) holds. In these orthonormal bases of $\mathbf{U}$ and $\mathbf{V}$ the operators T and T^* represented quite simply:

$$
\begin{aligned}
&T\mathbf{c}_i = \sigma_i(T)\mathbf{d}_i,\ i = 1, ..., n, \quad \text{where } \mathbf{d}_i = 0 \text{ for } i > m, \\
&T^*\mathbf{d}_j = \sigma_j(T)\mathbf{c}_j,\ j = 1, ..., m, \quad \text{where } \mathbf{c}_j = 0 \text{ for } j > n..
\end{aligned}
\tag{4.5}
$$

Let

$$
\Sigma = (s_{ij})_{i,j=1}^{m,n},\ s_{ij} = 0 \text{ for } i \neq j,\ s_{ii} = \sigma_i \text{ for } i = 1, ..., \min(m.n). \tag{4.6}
$$

In the case $m \neq n$ we call Σ a diagonal matrix with the diagonal $\sigma_1, ..., \sigma_{\min(m,n)}$. Then in the bases $[\mathbf{d}_1, ..., \mathbf{d}_m]$ and $[\mathbf{c}_1, ..., \mathbf{c}_n]$ T and T^* represented by the matrices Σ and $\Sigma^\top$ respectively.

Lemma 4.64 *Let $[\mathbf{u}_1, ..., \mathbf{u}_m], [\mathbf{v}_1, ..., \mathbf{v}_n]$ be orthonormal bases in the vector spaces $\mathbf{U}, \mathbf{V}$ over $\mathbb{F} = \mathbb{R}, \mathbb{C}$ respectively. Then T and T^* are presented by the matrices $A \in \mathbb{F}^{m\times n}$ and $A^* \in \mathbb{F}^{n\times m}$ respectively. Let $U \in \mathbf{U}(m)$ and $V \in \mathbf{U}(n)$ be the unitary matrices representing the change of base $[\mathbf{d}_1, ..., \mathbf{d}_m]$ to $[\mathbf{u}_1, ..., \mathbf{u}_m]$ and $[\mathbf{c}_1, ..., \mathbf{c}_n]$ to $[\mathbf{v}_1, ..., \mathbf{v}_n]$ respectively. (If $\mathbb{F} = \mathbb{R}$ then U and V are orthogonal matrices.) Then*

$$
A = U\Sigma V^* \in \mathbb{F}^{m\times n}, \quad U \in \mathbf{U}(m),\ V \in \mathbf{U}(n). \tag{4.7}
$$

Proof. By the definition $T\mathbf{v}_j = \sum_{i=1}^m a_{ij}\mathbf{u}_i$. Let $U = (u_{ip})_{i,p=1}^m, V = (v_{jq})_{j,q=1}^n$. Then

$$
T\mathbf{c}_q = \sum_{j=1}^n v_{jq}T\mathbf{v}_j = \sum_{j=1}^n v_{jq}\sum_{i=1}^m a_{ij}\mathbf{u}_i = \sum_{j=1}^n v_{jq}\sum_{i=1}^m a_{ij}\sum_{p=1}^m \bar{u}_{ip}\mathbf{d}_p.
$$

Use the first equality of (4.5) to deduce that $U^*AV = \Sigma$. □

Definition 4.65 *(4.7) is called the* singular value decomposition *(SVD) of A.*

Proposition 4.66 *Let $\mathbb{F} = \mathbb{R}, \mathbb{C}$ and denote by $\mathcal{R}_{m,n,k}(\mathbb{F}) \subset \mathbb{F}^{m\times n}$ the set of all matrices of rank $k \in [1, \min(m, n)]$ at most. Then $A \in \mathcal{R}_{m,n,k}(\mathbb{F})$ if and only if A can be expressed as a sum of at most k matrices of rank 1. Furthermore $\mathcal{R}_{m,n,k}(\mathbb{F})$ is a variety in $\mathbb{F}^{m\times n}$ given by the polynomial conditions: Each $(k + 1) \times (k + 1)$ minor of A is equal to zero.*

For the proof see Problem 2

Definition 4.67 *Let $A \in \mathbb{C}^{m\times n}$ and assume that A has the SVD given by (4.7), where $U = [\mathbf{u}_1, \dots, \mathbf{u}_m], V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$. Denote by $A_k := \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^* \in \mathbb{C}^{m\times n}$ for $k = 1, \dots, \operatorname{rank} A$. For $k > \operatorname{rank} A$ we define $A_k := A\ (= A_{\operatorname{rank} A})$.*

Note that for $1 \le k < \operatorname{rank} A$, the matrix A_k is uniquely defined if and only if $\sigma_k > \sigma_{k+1}$. (See Problem 1.)

Theorem 4.68 *For $\mathbb{F} = \mathbb{R}, \mathbb{C}$ and $A = (a_{ij}) \in \mathbb{F}^{m\times n}$ the following conditions hold:*

$$||A||_F := \sqrt{\operatorname{tr} A^*A} = \sqrt{\operatorname{tr} AA^*} = \sqrt{\sum_{i=1}^{\operatorname{rank} A} \sigma_i(A)^2}. \tag{4.8}$$

$$||A||_2 := \max_{\mathbf{x}\in\mathbb{F}^n, ||\mathbf{x}||_2=1} ||A\mathbf{x}||_2 = \sigma_1(A). \tag{4.9}$$

$$\min_{B\in\mathcal{R}_{m,n,k}(\mathbb{F})} ||A - B||_2 = ||A - A_k|| = \sigma_{k+1}(A), k = 1, ..., \operatorname{rank} A - 1. \tag{4.10}$$

$$\begin{gathered}\sigma_i(A) \ge \sigma_i((a_{i_p j_q})_{p=1,q=1}^{m',n'}) \ge \sigma_{i+(m-m')+(n-n')}(A), \\ m' \in [1, m],\ n' \in [1, n],\ 1 \le i_1 < ... < i_{m'} \le m,\ 1 \le j_1 < ... < j_{n'} \le n.\end{gathered} \tag{4.11}$$

Proof. The proof of (4.8) is left a Problem 7. We now show the equality in (4.9). View A as an operator $A : \mathbb{C}^n \to \mathbb{C}^m$. From the definition of $||A||_2$ it follows

$$||A||_2^2 = \max_{0\ne\mathbf{x}\in\mathbb{R}^n} \frac{\mathbf{x}^*A^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \lambda_1(A^*A) = \sigma_1(A)^2,$$

which proves (4.9).

We now prove (4.10). In the SVD decomposition of A (4.7) assume that $U = (\mathbf{u}_1, ..., \mathbf{u}_m)$ and $V = (\mathbf{v}_1, ..., \mathbf{v}_n)$. Then (4.7) is equivalent to the following representation of A:

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*,\ \mathbf{u}_1, ..., \mathbf{u}_r \in \mathbb{R}^m,\ \mathbf{v}_1, ..., \mathbf{v}_r \in \mathbb{R}^n,\ \mathbf{u}_i^*\mathbf{u}_j = \mathbf{v}_i^*\mathbf{v}_j = \delta_{ij},\ i, j = 1, ..., r, \tag{4.12}$$

where $r = \operatorname{rank} A$. Let $B = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^* \in \mathcal{R}_{m,n,k}$. Then in view of (4.9)

$$||A - B||_2 = ||\sum_{k+1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*||_2 = \sigma_{k+1}.$$

Let $B \in \mathcal{R}_{m,n,k}$. To show (4.10) it is enough to show that $||A - B||_2 \ge \sigma_{k+1}$. Let

$$\mathbf{W} := \{\mathbf{x} \in \mathbb{R}^n : \quad B\mathbf{x} = 0\}.$$

Then $\operatorname{codim} \mathbf{W} \ge k$. Furthermore

$$||A - B||_2^2 \ge \max_{||\mathbf{x}||_2=1, \mathbf{x}\in\mathbf{W}} ||(A - B)\mathbf{x}||^2 = \max_{||\mathbf{x}||_2=1, \mathbf{x}\in\mathbf{W}} \mathbf{x}^*A^*A\mathbf{x} \ge \lambda_{k+1}(A^*A) = \sigma_{k+1}^2,$$

where the last inequality follows from the min-max characterization of $\lambda_{k+1}(A^*A)$.

Let $C = (a_{ij_q})_{i,q=1}^{m,n'}$. Then C^*C is an a principal submatrix of A^*A of dimension n'. The interlacing inequalities between the eigenvalues of A^*A and C^*C yields (4.11) for $m' = m$. Let $D = (a_{i_p j_q})_{p,q=1}^{m',n'}$. Then DD^* is a principle submatrix of CC^*. Use the interlacing properties of the eigenvalues of CC^* and DD^* to deduce (4.11). □

We now restate the above results for linear operators.

Definition 4.69 *Let $\mathbf{U}, \mathbf{V}$ be finite dimensional vector spaces over $\mathbb{F} = \mathbb{R}, \mathbb{C}$. For $k \in \mathbb{Z}_+$ denote $\mathrm{L}_k(\mathbf{V}, \mathbf{U}) := \{T \in \mathrm{L}(\mathbf{V}, \mathbf{U}) : \operatorname{rank} T \le k\}$. Assume furthermore that $\mathbf{U}, \mathbf{V}$ are IPS. Let $T \in \mathrm{L}(\mathbf{V}, \mathbf{U})$ and assume that the orthonormal bases of $[\mathbf{d}_1, \dots, \mathbf{d}_m], [\mathbf{c}_1, \dots, \mathbf{c}_n]$ of $\mathbf{U}, \mathbf{V}$ respectively satisfy (4.5). Define $T_0 := \mathbf{0}$ and $T_k := T$ for an integer $k \ge \operatorname{rank} T$. Let $k \in [1, \operatorname{rank} T - 1] \cap \mathbb{N}$. Define $T_k \in \mathrm{L}(\mathbf{V}, \mathbf{U})$ by the equality $T_k(\mathbf{v}) = \sum_{i=1}^k \sigma_i(T)\langle \mathbf{v}, \mathbf{c}_i\rangle \mathbf{d}_i$ for any $\mathbf{v} \in \mathbf{V}$.*

It is straightforward to show that $T_k \in \mathrm{L}_k(\mathbf{V}, \mathbf{U})$ and T_k is unique if and only if $\sigma_k(T) > \sigma_{k+1}(T)$. See Problem 8. Theorem 4.68 yields:

Corollary 4.70 *Let $\mathbf{U}$ and $\mathbf{V}$ be finite dimensional IPS over $\mathbb{F} = \mathbb{R}, \mathbb{C}$. Let $T : \mathbf{V} \to \mathbf{U}$ be a linear operator. Then*

$$||T||_F := \sqrt{\operatorname{tr} T^*T} = \sqrt{\operatorname{tr} TT^*} = \sqrt{\sum_{i=1}^{\operatorname{rank} T} \sigma_i(T)^2}. \tag{4.13}$$

$$||T||_2 := \max_{\mathbf{x} \in \mathbf{V}, ||\mathbf{x}||_2 = 1} ||T\mathbf{x}||_2 = \sigma_1(T). \tag{4.14}$$

$$\min_{Q \in \mathrm{L}_k(\mathbf{V}, \mathbf{U})} ||T - Q||_2 = \sigma_{k+1}(T), \quad k = 1, ..., \operatorname{rank} T - 1. \tag{4.15}$$

Problems

1. Let $\mathbf{U}, \mathbf{V}$ be finite dimensional inner product spaces. Assume that $T \in \mathrm{L}(\mathbf{U}, \mathbf{V})$. Show that for any complex number $t \in \mathbb{C}$ $\sigma_i(tT) = |t|\sigma_i(T)$ for all i.

2. Prove Proposition 4.66. (Use SVD to prove the nontrivial part of the Proposition.)

3. Let $A \in \mathbb{C}^{m \times n}$ and assume that $U \in \mathbf{U}_m, V \in \mathbf{V}_n$. Show that $\sigma_i(UAV) = \sigma_i(A)$ for all i.

4. Let $A \in \mathrm{GL}(n, \mathbb{C})$. Show that $\sigma_1(A^{-1}) = \sigma_n(A)^{-1}$.

5. Let $\mathbf{U}, \mathbf{V}$ be IPS inner product space of dimensions m and n respectively. Assume that

$$\mathbf{U} = \mathbf{U}_1 \oplus \mathbf{U}_2, \dim \mathbf{U}_1 = m_1, \dim \mathbf{U}_2 = m_2, \mathbf{U}_1 \perp \mathbf{U}_2,$$
$$\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2, \dim \mathbf{V}_1 = n_1, \dim \mathbf{V}_2 = n_2, \mathbf{V}_1 \perp \mathbf{V}_2.$$

Assume that $T \in \mathrm{L}(\mathbf{V}, \mathbf{U})$. Suppose furthermore that $T\mathbf{V}_1 \subseteq \mathbf{U}_1, T\mathbf{V}_2 \subseteq \mathbf{U}_2$. Let $T_i \in \mathrm{L}(\mathbf{V}_i, \mathbf{U}_i)$ be the restriction of T to $\mathbf{V}_i$ for $i = 1, 2$. Then $\operatorname{rank} T = \operatorname{rank} T_1 + \operatorname{rank} T_2$ and $\{\sigma_1(T), \dots, \sigma_{\operatorname{rank} T}(T)\} = \{\sigma_1(T_1), \dots, \sigma_{\operatorname{rank} T_1}(T_1)\} \cup \{\sigma_1(T_2), \dots, \sigma_{\operatorname{rank} T_2}(T_2)\}$.

6. Let the assumptions of the Definition 4.67 hold. Show that for $1 \leq k < \operatorname{rank} A$ A_k is uniquely defined if and only if $\sigma_k > \sigma_{k+1}$.

7. Prove the equalities in (4.8).

8. Let the assumptions of Definition 4.69 hold. Show that for $k \in [1, \operatorname{rank} T - 1] \cap \mathbb{N}$ $\operatorname{rank} T_k = k$ and T_k is unique if and only if $\sigma_k(T) > \sigma_{k+1}(T)$.

9. Let $\mathbf{V}$ be an n-dimensional IPS. Assume that $T \in \mathrm{L}(\mathbf{V})$ is a normal operator. Let $\lambda_1(T), \dots, \lambda_n(T)$ be the eigenvalues of T arranged in the order $|\lambda_1(T)| \geq \dots \geq |\lambda_n(T)|$. Show that $\sigma_i(T) = |\lambda_i(T)|$ for $i = 1, \dots, n$.

4.13 Characterizations of singular values

Theorem 4.71 *Let $\mathbb{F} = \mathbb{R}, \mathbb{C}$ and assume that $A \in \mathbb{F}^{m \times n}$. Define*

$$H(A) = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \in \mathrm{H}_{m+n}(\mathbb{F}). \tag{4.1}$$

Then

$$\begin{aligned} &\lambda_i(H(A)) = \sigma_i(A),\ \lambda_{m+n+1-i}(H(A)) = -\sigma_i(A),\ i = 1, ..., \operatorname{rank} A, \\ &\lambda_j(H(A)) = 0,\ j = \operatorname{rank} A + 1, ..., n + m - \operatorname{rank} A. \end{aligned} \tag{4.2}$$

View A as an operator $A : \mathbb{F}^n \to \mathbb{F}^m$. Choose orthonormal bases $[\mathbf{d}_1, ..., \mathbf{d}_m], [\mathbf{c}_1, ..., \mathbf{c}_n]$ in $\mathbb{F}^m, \mathbb{F}^n$ respectively satisfying (4.5). Then

$$\begin{aligned} &\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \begin{bmatrix} \mathbf{d}_i \\ \mathbf{c}_i \end{bmatrix} = \sigma_i(A) \begin{bmatrix} \mathbf{d}_i \\ \mathbf{c}_i \end{bmatrix}, \ \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \begin{bmatrix} \mathbf{d}_i \\ -\mathbf{c}_i \end{bmatrix} = -\sigma_i(A) \begin{bmatrix} \mathbf{d}_i \\ -\mathbf{c}_i \end{bmatrix}, \\ &\qquad i = 1, ..., \operatorname{rank} A, \\ &\ker H(A) = \operatorname{span}((\mathbf{d}_{r+1}^*, 0)^*, ..., (\mathbf{d}_m^*, 0)^*, (0, \mathbf{c}_{r+1}^*)*, ..., (0, \mathbf{c}_n^*)^*),\ r = \operatorname{rank} A. \end{aligned} \tag{4.3}$$

Proof. It is straightforward to show the equalities (4.3). Since all the eigenvectors appearing in (4.3) are linearly independent we deduce (4.2). □

Corollary 4.72 *Let $\mathbb{F} = \mathbb{R}, \mathbb{C}$ and assume that $A \in \mathbb{F}^{m \times n}$. Let $\hat{A} := A[\alpha, \beta] \in \mathbb{F}^{p \times q}$ be a submatrix of A, formed by the set of rows and columns $\alpha \in Q_{p,m}, \beta \in Q_{q,n}$ respectively. Then*

$$\sigma_i(\hat{A}) \leq \sigma_i(A) \text{ for } i = 1, \dots. \tag{4.4}$$

For $l \in [1, \operatorname{rank} A] \cap \mathbb{N}$ the equalities $\sigma_i(\hat{A}) = \sigma_i(A), i = 1, \dots, l$ hold if and only if there exists two orthonormal systems of l right and left singular vectors $\mathbf{c}_1, \dots, \mathbf{c}_l \in \mathbb{F}^n$, $\mathbf{d}_1, \dots, \mathbf{d}_l \in \mathbb{F}^n$ satisfying (4.3) for $i = 1, \dots, l$ such that the nonzero coordinates vectors $\mathbf{c}_1, \dots, \mathbf{c}_l$ and $\mathbf{d}_1, \dots, \mathbf{d}_l$ are located at the indices β, α respectively.

See Problem 1.

Corollary 4.73 *Let $\mathbf{V}, \mathbf{U}$ be IPS over $\mathbb{F} = \mathbb{R}, \mathbb{C}$. Assume that $\mathbf{W}$ is a subspace of $\mathbf{V}$. Let $T \in \mathrm{L}(\mathbf{V}, \mathbf{U})$ and denote by $\hat{T} \in \mathrm{L}(\mathbf{W}, \mathbf{U})$ the restriction of T to $\mathbf{W}$. Then $\sigma_i(\hat{T}) \le \sigma_i(T)$ for any $i \in \mathbb{N}$. Furthermore $\sigma_i(\hat{T}) = \sigma_i(T)$ for $i = 1, \ldots, l \le \operatorname{rank} T$ if and only if $\mathbf{U}$ contains a subspace spanned by the first l right singular vectors of T.*

See Problem 2.

Define by $\mathbb{R}^n_{+,\searrow} := \mathbb{R}^n_{\searrow} \cap \mathbb{R}^n_+$. Then $D \subset \mathbb{R}^n_{+,\searrow}$ is called a strong Schur set if for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n_{+,\searrow}, \mathbf{x} \preceq \mathbf{y}$ we have the implication $\mathbf{y} \in D \Rightarrow \mathbf{x} \in D$.

Theorem 4.74 *Let $p \in \mathbb{N}$ and $D \subset \mathbb{R}^p_{\searrow} \cap \mathbb{R}^p_+$ be a regular convex strong Schur domain. Fix $m, n \in \mathbb{N}$ and let $\sigma_{(p)}(D) := \{A \in \mathbb{F}^{m\times n} : \sigma_{(p)}(A) \in D\}$. Let $h : D \to \mathbb{R}$ be a convex and strongly Schur's order preserving on D. Let $f : \sigma_{(p)} :\to \mathbb{R}$ be given as $h \circ \sigma_{(p)}$. Then f is a convex function.*

See Problem 3.

Corollary 4.75 *Let $\mathbb{F} = \mathbb{R}, \mathbb{C}$, $m, n, p \in \mathbb{N}$, $q \in [1, \infty)$ and $w_1 \ge w_2 \ge \ldots \ge w_p > 0$. Then the following function*

$$f : \mathbb{F}^{m\times n} \to \mathbb{R}, \quad f(A) := (\sum_{i=1}^{p} w_i \sigma_i(A)^q)^{\frac{1}{q}}, \ A \in \mathbb{F}^{m\times n}$$

is a convex function.

See Problem 4

We now translate Theorem 4.71 to the operator setting.

Lemma 4.76 *Let $\mathbf{U}, \mathbf{V}$ be finite dimensional IPS spaces with the inner products $\langle \cdot, \cdot \rangle_{\mathbf{U}}, \langle \cdot, \cdot \rangle_{\mathbf{V}}$ respectively. Define $\mathbf{W} := \mathbf{V} \oplus \mathbf{U}$ be the induced IPS with*

$$\langle (\mathbf{y}, \mathbf{x}), (\mathbf{v}, \mathbf{u}) \rangle_{\mathbf{W}} := \langle \mathbf{y}, \mathbf{v} \rangle_{\mathbf{V}} + \langle \mathbf{x}, \mathbf{u} \rangle_{\mathbf{U}}.$$

Let $T : \mathbf{V} \to \mathbf{U}$ be a linear operator, and $T^* : \mathbf{U} \to \mathbf{V}$ be the adjoint of T. Define the operator

$$\hat{T} : \mathbf{W} \to \mathbf{W}, \quad \hat{T}(\mathbf{y}, \mathbf{x}) := (T^*\mathbf{x}, T\mathbf{y}). \tag{4.5}$$

Then $\hat{T}$ is self-adjoint operator and $\hat{T}^2 = T^*T \oplus TT^*$. Hence the spectrum of $\hat{T}$ is symmetric with respect to the origin and $\hat{T}$ has exactly $2\operatorname{rank} T$ nonzero eigenvalues. More precisely, if $\dim \mathbf{U} = m, \dim \mathbf{V} = n$ then:

$$\begin{aligned} \lambda_i(\hat{T}) = -\lambda_{m+n-i+1}(\hat{T}) = \sigma_i(T), \ \textit{for } i = 1, \ldots, \operatorname{rank} T, \\ \lambda_j(\hat{T}) = 0, \ \textit{for } j = \operatorname{rank} T + 1, \ldots, n + m - \operatorname{rank} T. \end{aligned} \tag{4.6}$$

Let $\{\mathbf{d}_1, \ldots, \mathbf{d}_{\min(m,n)}\} \in \operatorname{Fr}(\min(m, n), \mathbf{U}), \{\mathbf{c}_1, \ldots, \mathbf{c}_{\min(m,n)}\} \in \operatorname{Fr}(\min(m, n), \mathbf{V})$ be the set of vectors satisfying (4.5). Define

$$\mathbf{z}_i := \frac{1}{\sqrt{2}}(\mathbf{c}_i, \mathbf{d}_i), \mathbf{z}_{m+n-i+1} := \frac{1}{\sqrt{2}}(\mathbf{c}_i, -\mathbf{d}_i), i = 1, \ldots, \min(m, n). \tag{4.7}$$

Then $\{\mathbf{z}_1, \mathbf{z}_{m+n}, \ldots, \mathbf{z}_{\min(m,n)}, \mathbf{z}_{m+n-\min(m,n)+1}\} \in \operatorname{Fr}(2\min(m, n), \mathbf{W})$. Furthermore $\hat{T}\mathbf{z}_i = \sigma_i(T)\mathbf{z}_i, \hat{T}\mathbf{z}_{m+n-i+1} = -\sigma_i(T)\mathbf{z}_{m+n-i+1}$ for $i = 1, \ldots, \min(m, n)$.

See Problem 5.

Theorem 4.77 *Let $\mathbf{U}, \mathbf{V}$ be m and n-dimensional IPS over $\mathbb{C}$ respectively. Let $T : \mathbf{V} \to \mathbf{U}$ be a linear operator. Then for each $k \in [1, \min(m,n)] \cap \mathbb{Z}$*

$$\sum_{i=1}^{k} \sigma_i(T) = \max_{\{\mathbf{f}_1,...,\mathbf{f}_k\} \in \mathrm{Fr}(k,\mathbf{U}), \{\mathbf{g}_1,...,\mathbf{g}_k\} \in \mathrm{Fr}(k,\mathbf{V})} \sum_{i=1}^{k} \Re \langle T\mathbf{g}_i, \mathbf{f}_i \rangle_{\mathbf{U}} = \tag{4.8}$$

$$\max_{\{\mathbf{f}_1,...,\mathbf{f}_k\} \in \mathrm{Fr}(k,\mathbf{U}), \{\mathbf{g}_1,...,\mathbf{g}_k\} \in \mathrm{Fr}(k,\mathbf{V})} \sum_{i=1}^{k} |\langle T\mathbf{g}_i, \mathbf{f}_i \rangle_{\mathbf{U}}|.$$

Furthermore $\sum_{i=1}^{k} \sigma_i(T) = \sum_{i=1}^{k} \Re \langle T\mathbf{g}_i, \mathbf{f}_i \rangle_{\mathbf{U}}$ for some two k-orthonormal frames $F_k = \{\mathbf{f}_1, ..., \mathbf{f}_k\}, G_k = \{\mathbf{g}_1, ..., \mathbf{g}_k\}$ if and only $\mathrm{span}((\mathbf{g}_1, \mathbf{f}_1), \ldots, (\mathbf{g}_k, \mathbf{f}_k))$ *is spanned by k eigenvectors of $\hat{T}$ corresponding to the first k eigenvalues of $\hat{T}$.*

Proof. Assume that $\{\mathbf{f}_1, ..., \mathbf{f}_k\} \in \mathrm{Fr}(k, \mathbf{U}), \{\mathbf{g}_1, ..., \mathbf{g}_k\} \in \mathrm{Fr}(k, \mathbf{V})$. Let $\mathbf{w}_i := \frac{1}{\sqrt{2}}(\mathbf{g}_i, \mathbf{f}_i), i = 1, \ldots, k$. Then $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\} \in \mathrm{Fr}(k, \mathbf{W})$. A straightforward calculation shows $\sum_{i=1}^{k} \langle \hat{T}\mathbf{w}_i, \mathbf{w}_i \rangle_{\mathbf{W}} = \sum_{i=1}^{k} \Re \langle T\mathbf{g}_i, \mathbf{f}_i \rangle_{\mathbf{U}}$. The maximal characterization of $\sum_{i=1}^{k} \lambda_i(\hat{T})$, (Theorem 4.47), and (4.6) yield the inequality $\sum_{i=1}^{k} \sigma_i(\hat{T}) \geq \sum_{i=1}^{k} \Re \langle T\mathbf{g}_i, \mathbf{f}_i \rangle_{\mathbf{U}}$ for $k \in [\min(m,n) \cap \mathbb{Z}$. Let $\mathbf{c}_1, \ldots, \mathbf{c}_{\min(m,n)}$, $\mathbf{d}_1, \ldots, \mathbf{d}_{\min(m,n)}$ satisfy (4.5). Then Lemma 4.76 yields that $\sum_{i=1}^{k} \sigma_i(\hat{T}) = \sum_{i=1}^{k} \Re \langle T\mathbf{c}_i, \mathbf{d}_i \rangle_{\mathbf{U}}$ for $k \in [\min(m,n) \cap \mathbb{Z}$. This proves the first equality of (4.8). The second equality of (4.8) is straightforward. (See Problem 6).)

Assume now that $\sum_{i=1}^{k} \sigma_i(T) = \sum_{i=1}^{k} \Re \langle T\mathbf{g}_i, \mathbf{f}_i \rangle_{\mathbf{U}}$ for some two k-orthonormal frames $F_k = \{\mathbf{f}_1, ..., \mathbf{f}_k\}, G_k = \{\mathbf{g}_1, ..., \mathbf{g}_k\}$. Define $\mathbf{w}_1, \ldots, \mathbf{w}_k$ as above. The above arguments yield that $\sum_{i=1}^{k} \langle \hat{T}\mathbf{w}_i, \mathbf{w}_i \rangle_{\mathbf{W}} = \sum_{i=1}^{k} \lambda_i(\hat{T})$. Theorem 4.47 yields that $\mathrm{span}((\mathbf{g}_1, \mathbf{f}_1), \ldots, (\mathbf{g}_k, \mathbf{f}_k))$ is spanned by k eigenvectors of $\hat{T}$ corresponding to the first k eigenvalues of $\hat{T}$. Vice versa, assume that $\{\mathbf{f}_1, ..., \mathbf{f}_k\} \in \mathrm{Fr}(k, \mathbf{U}), \{\mathbf{g}_1, ..., \mathbf{g}_k\} \in \mathrm{Fr}(k, \mathbf{V})$ and $\mathrm{span}((\mathbf{g}_1, \mathbf{f}_1), \ldots, (\mathbf{g}_k, \mathbf{f}_k))$ is spanned by k eigenvectors of $\hat{T}$ corresponding to the first k eigenvalues of $\hat{T}$. Define $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\} \in \mathrm{Fr}(\mathbf{W})$ as above. Then $\mathrm{span}(\mathbf{w}_1, \ldots, \mathbf{w}_k)$ contains k linearly independent eigenvectors corresponding to the the first k eigenvalues of $\hat{T}$. Theorem 4.47 and Lemma 4.76 yield that $\sigma_i(T) = \sum_{i=1}^{k} \langle \hat{T}\mathbf{w}_i, \mathbf{w}_i \rangle_{\mathbf{W}} = \sum_{i=1}^{k} \Re \langle T\mathbf{g}_i, \mathbf{f}_i \rangle_{\mathbf{U}}$. □

Theorem 4.78 $\mathbf{U}, \mathbf{V}$ *be m and n dimensional IPS spaces. Assume that Let $S, T : \mathbf{V} \to \mathbf{U}$ be linear operators. Then*

$$\Re \operatorname{tr}(S^*T) \leq \sum_{i=1}^{\min(m,n)} \sigma_i(S)\sigma_i(T). \tag{4.9}$$

Equality holds if and only if there exists two orthonormal set $\{\mathbf{d}_1, \ldots, \mathbf{d}_{\min(m,n)}\} \in \mathrm{Fr}(\min(m,n), \mathbf{U}), \{\mathbf{c}_1, \ldots, \mathbf{c}_{\min(m,n)}\} \in \mathrm{Fr}(\min(m,n), \mathbf{V})$, such that

$$S\mathbf{c}_i = \sigma_i(S)\mathbf{d}_i, T\mathbf{c}_i = \sigma_i(T)\mathbf{d}_i, S^*\mathbf{d}_i = \sigma_i(S)\mathbf{c}_i, T^*\mathbf{d}_i = \sigma_i(T)\mathbf{c}_i, i = 1, \ldots, \min(m,n). \tag{4.10}$$

Proof. Let $A, B \in \mathbb{C}^{n \times m}$. Then

$\operatorname{tr} B^*A = \overline{\operatorname{tr} AB^*}$. Hence $2\Re \operatorname{tr} AB^* = \operatorname{tr} H(A)H(B)$. Therefore $2\Re \operatorname{tr} S^*T = \operatorname{tr} \hat{S}\hat{T}$. Use Theorem 4.60 for $\hat{S}, \hat{T}$ and Lemma 4.76 to deduce (4.9). Equality in (4.9) if and only if $\operatorname{tr} \hat{S}\hat{T} = \sum_{i=1}^{m+n} \lambda_i(\hat{S})\lambda_i(\hat{T})$.

Clearly, the assumptions that $\{\mathbf{d}_1, \dots, \mathbf{d}_{\min(m,n)}\} \in \operatorname{Fr}(\min(m,n), \mathbf{U})$, $\{\mathbf{c}_1, \dots, \mathbf{c}_{\min(m,n)}\} \in \operatorname{Fr}(\min(m,n), \mathbf{V})$, and the equalities (4.10) imply equality in (4.9).

Assume equality in (4.9). Theorem 4.60 and the definitions of $\hat{S}, \hat{T}$ yields the existence $\{\mathbf{d}_1, \dots, \mathbf{d}_{\min(m,n)}\} \in \operatorname{Fr}(\min(m,n), \mathbf{U})$, $\{\mathbf{c}_1, \dots, \mathbf{c}_{\min(m,n)}\} \in \operatorname{Fr}(\min(m,n), \mathbf{V})$, such that (4.10) hold. □

Theorem 4.79 *Let* $\mathbf{U}$ *and* $\mathbf{V}$ *be finite dimensional IPS over* $\mathbb{F} = \mathbb{R}, \mathbb{C}$. *Let* $T : \mathbf{V} \to \mathbf{U}$ *be a linear operator. Then*

$$\min_{Q \in \mathrm{L}_k(\mathbf{V},\mathbf{U})} ||T - Q||_F = \sqrt{\sum_{i=k+1}^{\operatorname{rank} T} \sigma_i^2(T)}, \quad k = 1, ..., \operatorname{rank} T - 1. \tag{4.11}$$

Furthermore $||T - Q||_F = \sqrt{\sum_{i=k+1}^{\operatorname{rank} T} \sigma_i^2(T)}$ *for some* $Q \in \mathrm{L}_k(\mathbf{V}, \mathbf{U}), k < \operatorname{rank} T$, *if and only there* $Q = T_k$, *where* T_k *is defined in Definition 4.69.*

Proof. Use Theorem 4.78 to deduce that for any $Q \in \mathrm{L}(\mathbf{V}, \mathbf{U})$ one has

$$||T - Q||_F^2 = \operatorname{tr} T^*T - 2\Re \operatorname{tr} Q^*T + \operatorname{tr} Q^*Q \geq$$
$$\sum_{i=1}^{\operatorname{rank} T} \sigma_i^2(T) - 2\sum_{i=1}^{k} \sigma_i(T)\sigma_i(Q) + \sum_{i=1}^{k} \sigma_i^2(Q) =$$
$$\sum_{i=1}^{k} (\sigma_i(T) - \sigma_i(Q))^2 + \sum_{i=k+1}^{\operatorname{rank} T} \sigma_i^2(T) \geq \sum_{i=k+1}^{\operatorname{rank} T} \sigma_i^2(T).$$

Clearly $||T - T_k||_F^2 = \sum_{i=k+1}^{\operatorname{rank} T} \sigma_i^2(T)$. Hence (4.11) holds. Vice versa if $Q \in \mathrm{L}_k(\mathbf{V}, \mathbf{U})$ and $||T - Q||_F^2 = \sum_{i=k+1}^{\operatorname{rank} T} \sigma_i^2(T)$ then the equality case in Theorem 4.78 yields that $Q = T_k$. □

Corollary 4.80 *Let* $F = \mathbb{R}, \mathbb{C}$ *and* $A \in \mathbb{F}^{m \times n}$. *Then*

$$\min_{B \in \mathcal{R}_{m,n,k}(\mathbb{F})} ||A - B||_F = \sqrt{\sum_{i=k+1}^{\operatorname{rank} A} \sigma_i^2(A)}, \quad k = 1, ..., \operatorname{rank} A - 1. \tag{4.12}$$

Furthermore $||A - B||_F = \sqrt{\sum_{i=k+1}^{\operatorname{rank} A} \sigma_i^2(A)}$ *for some* $B \in \mathcal{R}_{m,n,k}(\mathbb{F}), k < \operatorname{rank} A$, *if and only there* $B = A_k$, *where* A_k *is defined in Definition 4.67.*

Theorem 4.81 *Let* $\mathbb{F} = \mathbb{R}, \mathbb{C}$ *and* $A \in \mathbb{F}^{m \times n}$. *Then*

$$\min_{B \in \mathcal{R}_{m,n,k}(\mathbb{F})} \sum_{i=1}^{j} \sigma_i(A - B) = \sum_{i=k+1}^{k+j} \sigma_i(A), \quad j = 1, ..., \min(m,n) - k, \ k = 1, \dots, \min(m,n) - 1. \tag{4.13}$$

Proof. Clearly, for $B = A_k$ we have the equality $\sum_{i=1}^{j} \sigma_i(A-B) = \sum_{i=k+1}^{k+j} \sigma_i(A)$. Let $B \in \mathcal{R}_{m,n,k}(\mathbb{F})$. Let $\mathbf{X} \in \mathrm{Gr}(k, \mathbb{C}^n)$ be an subspace which contains the columns of B. Let $\mathbf{W} = \{(\mathbf{0}^\top, \mathbf{x}^\top)^\top \in \mathbb{F}^{m+n}, \mathbf{x} \in \mathbf{X}\}$. Observe that for any $\mathbf{z} \in \mathbf{W}^\perp$ one has the equality $\mathbf{z}^* H((A-B))\mathbf{z} = \mathbf{z}^* H(A)\mathbf{z}$. Combine Theorems 4.48 and 4.71 to deduce $\sum_{i=1}^{j} \sigma_i(B-A) \geq \sum_{i=k+1}^{k+j} \sigma_i(A)$. □

Theorem 4.82 *Let $\mathbf{V}$ be an n-dimensional IPS over $\mathbb{C}$. Let $T : \mathbf{V} \to \mathbf{V}$ be a linear operator. Assume the n eigenvalues of T $\lambda_1(T), \ldots, \lambda_n(T)$ are arranged the order $|\lambda_1(T)| \geq \ldots \geq |\lambda_n(T)|$. Let $\boldsymbol{\lambda}_a(T) := (|\lambda_1(T)|, \ldots, |\lambda_n(T)|), \boldsymbol{\sigma}(T) := (\sigma_1(T), \ldots, \sigma_n(T))$. Then $\boldsymbol{\lambda}_a(T) \preceq \boldsymbol{\sigma}(T)$. That is*

$$\sum_{i=1}^{k} |\lambda_i(T)| \leq \sum_{i=1}^{k} \sigma_i(T), \quad i = 1, \ldots, n. \tag{4.14}$$

Furthermore, $\sum_{i=1}^{k} |\lambda_i(T)| = \sum_{i=1}^{k} \sigma_i(T)$ for some $k \in [1, n] \cap \mathbb{Z}$ if and only if the following conditions are satisfied. There exists an orthonormal basis $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of $\mathbf{V}$ such that:

1. *$T\mathbf{x}_i = \lambda_i(T)\mathbf{x}_i, T^*\mathbf{x}_i = \overline{\lambda_i(T)}\mathbf{x}_i$ for $i = 1, \ldots, k$.*
2. *Denote by $S : \mathbf{U} \to \mathbf{U}$ the restriction of T to the invariant subspace $\mathbf{U} = \mathrm{span}(\mathbf{x}_{k+1}, \ldots, \mathbf{x}_n)$. Then $||S||_2 \leq |\lambda_k(T)|$.*

Proof. Use Theorem 4.28 to choose an orthonormal basis $\mathbf{g}_1, \ldots, \mathbf{g}_n$ of $\mathbf{V}$, such that T is represented by an upper diagonal matrix $A = [a_{ij}] \in \mathbb{C}^{n \times n}$ such that $a_{ii} = \lambda_i(T), i = 1, \ldots, n$. Let $\epsilon_i \in \mathbb{C}, |\epsilon_i| = 1$ such that $\bar{\epsilon}_i \lambda_i(T) = |\lambda_i(T)|$ for $i = 1, \ldots, n$. Let $S \in \mathrm{L}(\mathbf{V})$ be presented in the basis $\mathbf{g}_1, \ldots, \mathbf{g}_n$ by a diagonal matrix $\mathrm{diag}(\epsilon_1, \ldots, \epsilon_k, 0, \ldots, 0)$. Clearly, $\sigma_i(S) = 1$ for $i = 1, \ldots, k$ and $\sigma_i(S) = 0$ for $i = k+1, \ldots, n$. Furthermore, $\Re \operatorname{tr} S^*C = \sum_{i=1}^{k} |\lambda_i(T)|$. Hence Theorem 4.78 yields (4.14).

Assume now that $\sum_{i=1}^{k} |\lambda_i(T)| = \sum_{i=1}^{k} \sigma_i(T)$. Hence equality sign holds in (4.9). Hence there exists two orthonormal bases $\{\mathbf{c}_1, \ldots, \mathbf{c}_n\}, \{\mathbf{d}_1, \ldots, \mathbf{d}_n\}$ in $\mathbf{V}$ such that (4.10) holds. It easily follows that $\{\mathbf{c}_1, \ldots, \mathbf{c}_k\}, \{\mathbf{d}_1, \ldots, \mathbf{d}_k\}$ are orthonormal bases of $\mathbf{W} := \mathrm{span}(\mathbf{g}_1, \ldots, \mathbf{g}_k)$. Hence $\mathbf{W}$ is an invariant subspace of T and T^*. Hence $A = A_1 \oplus A_2$, i.e. A is a block diagonal matrix. Thus $A_1 = (a_{ij})_{i,j=1}^{k} \in \mathbb{C}^{k \times k}, A_2 = (a_{ij})_{i,j=k+1}^{n} \in \mathbb{C}^{(n-k) \times (n-k)}$ represent the restriction of T to $\mathbf{W}, \mathbf{U} := \mathbf{W}^\perp$, denoted by T_1 and T_2 respectively. Hence $\sigma_i(T_1) = \sigma_i(T)$ for $i = 1, \ldots, k$. Note that the restriction of S to $\mathbf{W}$, denoted by S_1 is given by the diagonal matrix $D_1 := \mathrm{diag}(\epsilon_1, \ldots, \epsilon_k) \in \mathbf{U}(k)$. (4.10) yield that $S_1^{-1} T_1 \mathbf{c}_i = \sigma_i(T)\mathbf{c}_i$ for $i = 1, \ldots, k$,i.e. $\sigma_1(T), \ldots, \sigma_k(T)$ are the eigenvalues of $S_1^{-1}T_1$. Clearly $S_1^{-1}T_1$ is presented in the basis $[\mathbf{g}_1, \ldots, \mathbf{g}_k]$ by the matrix $D_1^{-1}A_1$, which is a diagonal matrix with $|\lambda_1(T)|, \ldots, |\lambda_k(T)|$ on the main diagonal. That is $S_1^{-1}T_1$ has eigenvalues $|\lambda_1(T)|, \ldots, |\lambda_k(T)|$. Therefore $\sigma_i(T) = |\lambda_i(T)|$ for $i = 1, \ldots, k$. Theorem 4.68 yields that

$$\operatorname{tr} A_1^* A_1 = \sum_{i,j=1}^{k} |a_{ij}|^2 = \sum_{i=1}^{k} \sigma_i^2(A_1) = \sum_{i=1}^{k} \sigma_i^2(T_1) = \sum_{i=1}^{k} |\lambda_i(T)|^2.$$

As $\lambda_1(T), \ldots, \lambda_k(T)$ are the diagonal elements of A_1 is follows from the above equality that A_1 is a diagonal matrix. Hence we can choose $\mathbf{x}_i = \mathbf{g}_i$ for $i = 1, \ldots, n$ to obtain the part *1* of the equality case.

Let $T\mathbf{x} = \lambda\mathbf{x}$ where $||\mathbf{x}|| = 1$ and $\rho(T) = |\lambda|$. Recall $||T||_2 = \sigma_1(T)$, where $\sigma_1(T)^2 = \lambda_1(T^*T)$ is the maximal eigenvalue of the self-adjoint operator T^*T. The maximum characterization of $\lambda_1(T^*T)$ yields that $|\lambda|^2 = \langle T\mathbf{x}, T\mathbf{x}\rangle = \langle T^*T\mathbf{x}, \mathbf{x}\rangle \leq \lambda_1(T^*T) = ||T||_2^2$. Hence $\rho(T) \leq ||T||_2$.

Assume now that $\rho(T) = ||T||_2$. $\rho(T) = 0$ then $||T||_2 = 0 \Rightarrow T = 0$, and theorem holds trivially n this case. Assume that $\rho(T) > 0$. Hence the eigenvector $\mathbf{x}_1 := \mathbf{x}$ is also the eigenvector of T^*T corresponding to $\lambda_1(T^*T) = |\lambda|^2$. Hence $|\lambda|^2\mathbf{x} = T^*T\mathbf{x} = T^*(\lambda\mathbf{x})$, which implies that $T^*\mathbf{x} = \bar{\lambda}\mathbf{x}$. Let $\mathbf{U} = \mathrm{span}(\mathbf{x})^\perp$ be the orthogonal complement of $\mathrm{span}(\mathbf{x})$. Since $T\mathrm{span}(\mathbf{x}) = \mathrm{span}(\mathbf{x})$ it follows that $T^*\mathbf{U} \subseteq \mathbf{U}$. Similarly, since $T^*\mathrm{span}(\mathbf{x}) = \mathrm{span}(\mathbf{x})$ $T\mathbf{U} \subseteq \mathbf{U}$. Thus $\mathbf{V} = \mathrm{span}(\mathbf{x}) \oplus \mathbf{U}$ and $\mathrm{span}(\mathbf{x}), \mathbf{U}$ are invariant subspaces of T and T^*. Hence $\mathrm{span}(\mathbf{x}), \mathbf{U}$ are invariant subspaces of T^*T and TT^*. Let T_1 be the restriction of T to $\mathbf{U}$. Then $T_1^*T_1$ is the restriction of T^*T. Therefore $||T_1||_2^2 = \lambda_1(T_1 * T_1) \geq \lambda_1(T^*T) = ||T||_2^2$. This establishes the second part of theorem, labeled (a) and (b).

The above result imply that the conditions (a) and (b) of the theorem yield the equality $\rho(T) = ||T||_2$. □

Corollary 4.83 *Let* $\mathbf{U}$ *be an n-dimensional IPS over* $\mathbb{C}$. *Let* $T : \mathbf{U} \to \mathbf{U}$ *be a linear operator. Denote by* $|\lambda(T)| = (|\lambda_1(T)|, ..., |\lambda_n(T)|)^T$ *the absolute eigenvalues of* T, *(counting with their multiplicities), arranged in a decreasing order. Then* $|\lambda(T)| = (\sigma_1(T), ..., \sigma_n(T))^T$ *if and only if* T *is a normal operator.*

Problems

1. Let the assumptions of Corollary 4.72 hold.
   (a) Since rank $\hat{A} \leq$ rank A show that the inequalities (4.4) reduce to $\sigma_i(\hat{A}) = \sigma_i(A) = 0$ for $i >$ rank A.
   (b) Since $H(\hat{A})$ is a submatrix of $H(A)$ use the Cauchy interlacing principle to deduce the inequalities (4.4) for $i = 1, \ldots,$ rank A. Furthermore, if $p' := m - \#\alpha, q' = n - \#\beta$ then the Cauchy interlacing principle gives the complementary inequalities $\sigma_i(\hat{A}) \geq \sigma_{i+p'+q'}(A)$ for any $i \in \mathbb{N}$.
   (c) Assume that $\sigma_i(\hat{A}) = \sigma_i(A)$ for $i = 1, \ldots, l \leq$ rank A. Compare the maximal characterization of the sum of the first k eigenvalues of $H(\hat{A})$ and $H(A)$ given by Theorem 4.47 for $k = 1, \ldots, l$ to deduce the last part of Corollary (4.72).
2. Prove Corollary 4.73 by choosing any orthonormal basis in $\mathbf{U}$, an orthonormal basis in $\mathbf{V}$ whose first dim $\mathbf{W}$ elements span $\mathbf{W}$, and using Problem 1.
3. Combine Theorems **??** and 4.71 to deduce Theorem 4.74.
4. (a) Prove Corollary 4.75
   (b) Recall the definition of a norm on a vector space over $\mathbb{F} = \mathbb{R}, \mathbb{C}$ **??**. Show that the function f defined in Corollary 4.75 is a norm. For $p = \min(m, n)$ and $w_1 = ... = w_p = 1$ this norm is called the $q - Schatten$ norm.

5. Prove Lemma 4.76.

6. Under the assumptions of Theorem 4.77 show the equalities.

$$\max_{\{\mathbf{f}_1,\ldots,\mathbf{f}_k\}\in \mathrm{Fr}(k,\mathbf{U}),\{\mathbf{g}_1,\ldots,\mathbf{g}_k\}\in \mathrm{Fr}(k,\mathbf{V})} \sum_{i=1}^{k} \Re\langle T\mathbf{g}_i,\mathbf{f}_i\rangle_{\mathbf{U}} =$$
$$\max_{\{\mathbf{f}_1,\ldots,\mathbf{f}_k\}\in \mathrm{Fr}(k,\mathbf{U}),\{\mathbf{g}_1,\ldots,\mathbf{g}_k\}\in \mathrm{Fr}(k,\mathbf{V})} \sum_{i=1}^{k} |\langle T\mathbf{g}_i,\mathbf{f}_i\rangle_{\mathbf{U}}|.$$

7. Under the assumptions of Theorem 4.77 is it *true* that that for $k > 1$

$$\sum_{i=1}^{k} \sigma_i(T) = \max_{\{\mathbf{f}_1,\ldots,\mathbf{f}_k\}\in \mathrm{Fr}(k,\mathbf{U})} \sum_{i=1}^{k} ||T\mathbf{f}_i||_{\mathbf{V}}.$$

   I doubt it.

8. Let $\mathbf{U},\mathbf{V}$ be finite dimensional IPS. Assume that $P,T \in \mathrm{L}(\mathbf{U},\mathbf{V})$. Show that $\Re \operatorname{tr}(P^*T) \geq -\sum_{i=1}^{\min(m,n)} \sigma_i(S)\sigma_i(T)$. Equality holds if and only if $S = -P$ and T satisfy the conditions of Theorem 4.78.

4.14 Moore-Penrose generalized inverse

Let $A \in \mathbb{C}^{m\times n}$. Then (4.12) is called the *reduced* SVD of A. It can be written as

$$\begin{gathered} A = U_r\Sigma_r V_r^*, \quad r = \operatorname{rank} A, \quad \Sigma_r := \operatorname{diag}(\sigma_1(A),\ldots,\sigma_r(A)) \in S_r(\mathbb{R}), \\ U_r = [\mathbf{u}_1,\ldots,\mathbf{u}_r] \in \mathbb{C}^{m\times r}, V_r = [\mathbf{v}_1,\ldots,\mathbf{v}_r] \in \mathbb{C}^{n\times r}, U_r^*U_r = V_r^*V_r = I_r,. \end{gathered} \tag{4.1}$$

Recall that

$$AA^*\mathbf{u}_i = \sigma_i(A)^2\mathbf{u}_i, A^*A\mathbf{v}_i = \sigma_i(A)^2\mathbf{v}_i,$$
$$\mathbf{v}_i = \frac{1}{\sigma_i(A)}A^*\mathbf{u}_i, \mathbf{u}_i = \frac{1}{\sigma_i(A)}A\mathbf{v}_i, i = 1,\ldots,r.$$

Then

$$A^\dagger := V_r\Sigma_r^{-1}U_r^* \in \mathbb{C}^{n\times m} \tag{4.2}$$

is the *Moore-Penrose* generalized inverse of A. If $A \in \mathbb{R}^{m\times n}$ then we assume that $U \in \mathbb{R}^{m\times r}$ and $V \in R^{n\times r}$, i.e. U, V are real values matrices over the real numbers $\mathbb{R}$.

Theorem 4.84 *Let $A \in \mathbb{C}^{m\times n}$ matrix. Then the Moore-Penrose generalized inverse $A^\dagger \in \mathbb{C}^{n\times m}$ satisfies the following properties.*

1. $\operatorname{rank} A = \operatorname{rank} A^\dagger$.

2. $A^\dagger AA^\dagger = A^\dagger$, $AA^\dagger A = A$, $A^*AA^\dagger = A^\dagger AA^* = A^*$.

3. $A^\dagger A$ *and* $AA^\dagger$ *are Hermitian nonnegative definite idempotent matrices, i.e.* $(A^\dagger A)^2 = A^\dagger A$ *and* $(AA^\dagger)^2 = AA^\dagger$, *having the same rank as* A.

4. *The least square solution of* $A\mathbf{x} = \mathbf{b}$*, i.e. the solution of the system* $A^*A\mathbf{x} = A^*\mathbf{b}$*, has a solution* $\mathbf{y} = A^\dagger\mathbf{b}$*. This solution has the minimal norm* $||\mathbf{y}||$*, for all possible solutions of* $A^*A\mathbf{x} = A^*\mathbf{b}$*.*

5. *If* rank $A = n$ *then* $A^\dagger = (A^*A)^{-1}A^*$*. In particular, if* $A \in \mathbb{C}^{n\times n}$ *is invertible then* $A^\dagger = A^{-1}$*.*

To prove the above theorem we need the following proposition.

Proposition 4.85 *Let* $E \in \mathbb{C}^{l\times m}, G \in \mathbb{C}^{m\times n}$*. Then* rank $EG \leq \min(\text{rank } E, \text{rank } G)$*. If* $l = m$ *and* E *is invertible then* rank $EG =$ rank G*. If* $m = n$ *and* G *is invertible then* rank $EG =$ rank E*.*

Proof. Let $\mathbf{e}_1, \ldots, \mathbf{e}_m \in \mathbb{C}^l, \mathbf{g}_1, \ldots, \mathbf{g}_n \in \mathbb{C}^m$ be the columns of E and G respectively. Then rank $E = \dim \text{span}(\mathbf{e}_1, \ldots, \mathbf{e}_l)$. Observe that $EG = [E\mathbf{g}_1, \ldots, E\mathbf{g}_n] \in \mathbb{C}^{l\times n}$. Clearly $E\mathbf{g}_i$ is a linear combination of the columns of E. Hence $E\mathbf{g}_i \in \text{span}(\mathbf{e}_1, \ldots, \mathbf{e}_l)$. Therefore $\text{span}(E\mathbf{g}_1, \ldots, E\mathbf{g}_n) \subseteq \text{span}(\mathbf{e}_1, \ldots, \mathbf{e}_l)$, which implies that rank $EG \leq$ rank E. Note that $(EG)^T = G^TE^T$. Hence
rank $EG =$ rank $(EG)^T \leq$ rank $G^T =$ rank G. Thus
rank $EG \leq \min(\text{rank } E, \text{rank } G)$. Suppose E is invertible. Then rank $EG \leq$ rank $G =$ rank $E^{-1}(EG) \leq$ rank EG. Hence rank $EG =$ rank G. Similarly rank $EG =$ rank E if G is invertible. □

Proof of Theorem 4.84.

1. Proposition 4.85 yields that rank $A^\dagger =$ rank $V_r\Sigma_r^{-1}U_r^* \leq$ rank $\Sigma_r^{-1}U_r^* \leq$ rank $\Sigma_r^{-1} = r =$ rank A. Since $\Sigma_r = V_r^*A^\dagger U_r$ Proposition 4.85 yields that rank $A^\dagger \geq$ rank $\Sigma_r^{-1} = r$. Hence rank $A =$ rank $A^\dagger$.

2. $AA^\dagger = (U_r\Sigma_rV_r^*)(V_r\Sigma_r^{-1}U_r^*) = U_r\Sigma_r\Sigma_r^{-1}U_r^* = U_rU_r^*$. Hence

$$AA^\dagger A = (U_rU_r^*)(U_r\Sigma_rV_r^*) = U_r\Sigma V_r^* = A.$$

   Hence $A^*AA^\dagger = (V_r\Sigma_rU_r^*)(U_rU_r^*) = A^*$. Similarly $A^\dagger A = V_rV_r^*$ and $A^\dagger AA^\dagger = A^\dagger, A^\dagger AA^* = A^*$.

3. Since $AA^\dagger = U_rU_r^*$ we deduce that $(AA^\dagger)^* = (U_rU_r^*)^* = (U_r^*)^*U_r^* = AA^\dagger$, i.e. $AA^\dagger$ is Hermitian. Next $(AA^\dagger)^2 = (U_rU_r^*)^2 = (U_rU_r^*)(U_rU_r^*) = (U_rU_r^*) = AA^\dagger$, i.e. $AA^\dagger$ is idempotent. Hence $AA^\dagger$ is nonnegative definite. As $AA^\dagger = U_rI_rU_r^*$, the arguments of part *1* yield that rank $AA^\dagger = r$. Similar arguments apply to $A^\dagger A = V_rV_r^*$.

4. Since $A^*AA^\dagger = A^*$ it follows that $A^*A(A^\dagger\mathbf{b}) = A^*\mathbf{b}$, i.e. $\mathbf{y} = A^\dagger\mathbf{b}$ is a least square solution. It is left to show that if $A^*A\mathbf{x} = A^*\mathbf{b}$ then $||\mathbf{x}|| \geq ||A^\dagger\mathbf{b}||$ and equality holds if and only if $\mathbf{x} = A^\dagger\mathbf{b}$.

   We now consider the system $A^*A\mathbf{x} = A^*\mathbf{b}$. To analyze this system we use the full form of SVD given in (4.7). It is equivalent to

   $(V\Sigma^TU^*)(U\Sigma V^*)\mathbf{x} = V\Sigma^TU^*\mathbf{b}$. Multiplying by V^* we obtain the system $\Sigma^T\Sigma(V^*\mathbf{x}) = \Sigma^T(U^*\mathbf{b})$. Let $\mathbf{z} = (z_1, \ldots, z_n)^T := V^*\mathbf{x}$,

$\mathbf{c} = (c_1, \ldots, c_m)^T := U^*\mathbf{b}$. Note that $\mathbf{z}^*\mathbf{z} = \mathbf{x}^*VV\mathbf{x} = \mathbf{x}^*\mathbf{x}$, i.e. $||\mathbf{z}|| = ||\mathbf{x}||$. After these substitutions the least square system in $z_1, \ldots, z_n$ variables is given in the form $\sigma_i(A)^2 z_i = \sigma_i(A)c_i$ for $i = 1, \ldots, n$. Since $\sigma_i(A) = 0$ for $i > r$ we obtain that $z_i = \frac{1}{\sigma_i(A)}c_i$ for $i = 1, \ldots, r$ while $z_{r+1}, \ldots, z_n$ are free variables. Thus $||\mathbf{z}||^2 = \sum_{i=1}^{r} \frac{1}{\sigma_i(A)^2} + \sum_{i=r+1}^{n} |z_i|^2$. Hence the least square solution with the minimal length $||\mathbf{z}||$ is the solution with $z_i = 0$ for $i = r+1, \ldots, n$. This solution corresponds the $\mathbf{x} = A^\dagger\mathbf{b}$.

5. Since rank A^*A = rank $A = n$ it follows that A^*A is an invertible matrix. Hence the least square solution is unique and is given by $\mathbf{x} = (A^*A)^{-1}A^*\mathbf{b}$. Thus for each $\mathbf{b}$ one has $(A^*A)^{-1}A^*\mathbf{b} = A^\dagger\mathbf{b}$, hence $A^\dagger = (A^*A)^{-1}A^*$.

   If A is an $n \times n$ matrix and is invertible it follows that $(A^*A)^{-1}A^* = A^{-1}(A^*)^{-1}A^* = A^{-1}$. □

Problems

1. $P \in \mathbb{C}^{n\times n}$ is called a *projection* if $P^2 = P$. Show that P is a projection if and only if the following two conditions are satisfied:
   - Each eigenvalue of P is either 0 or 1.
   - P is a diagonable matrix.
2. $P \in \mathbb{R}^{n\times n}$ is called an *orthogonal* projection if P is a projection and a symmetric matrix. Let $\mathbf{V} \subseteq \mathbb{R}^n$ be the subspace spanned by the columns of P. Show that for any $\mathbf{a} \in \mathbb{R}^n, \mathbf{b} \in P\mathbf{V}$ the following inequality holds: $||\mathbf{a} - \mathbf{b}|| \geq ||\mathbf{a} - P\mathbf{a}||$. Furthermore equality holds if and only if $\mathbf{b} = P\mathbf{a}$. That is, $P\mathbf{a}$ is the orthogonal projection of $\mathbf{a}$ on the column space of P.
3. Let $A \in \mathbb{R}^{m\times n}$ and assume that the SVD of A is given by (4.7), where $U \in \mathbf{O}(m, \mathbb{R}), V \in \mathbf{O}(n, \mathbb{R})$.
   (a) What is the SVD of A^T?
   (b) Show that $(A^T)^\dagger = (A^\dagger)^T$.
   (c) Suppose that $B \in \mathbb{R}^{l\times m}$. Is it true that $(BA)^\dagger = A^\dagger B^\dagger$? Justify!

5 Perron-Frobenius theorem

Denote by $\mathbb{R}_+^{m\times n}$, the set of nonnegative matrices $A = [a_{ij}]_{i,j=1}^{m,n}$, where each entry $a_{ij} \geq 0$. We denote a nonnegative matrix by $A \geq 0$. We say that A is a nonzero nonnegative if $A \geq 0$ and $A \neq 0$. We denote that by $A \gneq 0$. We say that A is a positive matrix if all the entries of A are positive, i.e. $a_{ij} > 0$ for all i, j. We denote that by $A > 0$. Similar notation is for vectors, i.e. $n = 1$. From now on we assume that all matrices are square matrices, unless stated otherwise. A nonnegative square matrix $A \in \mathbb{R}_+^{n\times n}$ is called *irreducible*, if $(I + A)^{n-1} > 0$. A nonnegative matrix is called *primitive* if $A^p > 0$ for some positive integer p.

Theorem 5.1 *Let $A \in \mathbb{R}_+^{n\times n}$. Denote by $\rho(A)$, the spectral radius of A, i.e. the maximum of the absolute values of the eigenvalues of A. The the following conditions hold.*

1. *$\rho(A)$ is an eigenvalues of A.*

2. *There exists $\mathbf{x} \gneq \mathbf{0}$ such that $A\mathbf{x} = \rho(A)\mathbf{x}$.*

3. *Assume that λ is an eigenvalue of A, $\lambda \neq \rho(A)$ and $|\lambda| = \rho(A)$. Then* index $(\lambda) \leq$ index $(\rho(A))$. *Furthermore $\zeta = \frac{\lambda}{\rho(A)}$ is a root of unity of order n at most, i.e. $\zeta^m = 1$ and $m \leq n$.*

4. *Suppose that A is primitive. Then $\rho(A) > 0$, $\rho(A)$ is a simple root of the characteristic polynomial, and there exists $\mathbf{x} > 0$ such that $A\mathbf{x} = \rho(A)\mathbf{x}$. Furthermore if λ is an eigenvalue of A different from $\rho(A)$ then $|\lambda| < \rho(A)$.*

5. *Suppose that A is irreducible. Then $\rho(A) > 0$, $\rho(A)$ is a simple root of the characteristic polynomial, and there exists $\mathbf{x} > 0$ such that $A\mathbf{x} = \rho(A)\mathbf{x}$. Let $\lambda_0 := \rho(A), \lambda_1, \ldots, \lambda_{m-1}$ be all distinct eigenvalues of A satisfying $|\lambda| = \rho(A)$. Then all these eigenvalues are simple roots of the characteristic polynomial of A. Furthermore $\lambda_j = \rho(A)\zeta^j$, for $j = 0, \ldots, m-1$, where $\zeta = e^{\frac{2\pi\sqrt{-1}}{m}}$ is a primitive $m-th$ root of 1. Finally, the characteristic polynomials of ζA and A are equal.*

We first prove this result for symmetric nonnegative matrix.

Proposition 5.2 *Let $A = A^\top \geq 0$ be a symmetric matrix with nonnegative entries. Then the conditions of Theorem 5.1 hold.*

Proof. Recall that A is diagonable by an orthogonal matrix. Hence the index of each eigenvalue of A is 1. Next recall that each eigenvalue is given by the Rayleigh quotient $\frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top\mathbf{x}}$. For $\mathbf{x} = (x_1, \ldots, x_n)^\top$ let $|\mathbf{x}| := (|x_1|, \ldots, |x_n|)^\top$. Note that $\mathbf{x}^\top\mathbf{x} = |\mathbf{x}|^\top|\mathbf{x}|$. Also $|\mathbf{x}^\top A\mathbf{x}| \leq |\mathbf{x}|^\top A|\mathbf{x}|$. Hence $|\frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top\mathbf{x}}| \leq \frac{|\mathbf{x}|^\top A|\mathbf{x}|}{|\mathbf{x}|^\top|\mathbf{x}|}$. Thus, the maximum of the Rayleigh quotient is achieved for some $\mathbf{x} \gneq \mathbf{0}$. This shows that $\rho(A)$ is an eigenvalues of A, and there exists an eigenvector $\mathbf{x} \gneq \mathbf{0}$ corresponding to A.

Suppose that $A > 0$. Then $A\mathbf{x} = \rho(A)\mathbf{x}, \mathbf{x} \gneq 0$. So $A\mathbf{x} > \mathbf{0}$. Hence $\rho(A) > 0$ and $\mathbf{x} > 0$. Assume that there exists a linear independent eigenvector $\mathbf{y}$ corresponding to $\rho(A)$. It can be chosen to be orthogonal to the positive eigenvector $\mathbf{x}$ corresponding to $\rho(A)$, i.e $\mathbf{x}^\top\mathbf{y} = 0$. Hence $\mathbf{y}$ has positive and negative coordinates. Therefore

$$|\frac{\mathbf{y}^\top A\mathbf{y}}{\mathbf{y}^\top\mathbf{y}}| < \frac{|\mathbf{y}|^\top A|\mathbf{y}|}{|\mathbf{y}|^\top|\mathbf{y}|} \leq \rho(A).$$

This gives the contradiction $\rho(A) < \rho(A)$. Therefore, $\rho(A)$ is a simple root of the characteristic polynomial of A. In a similar way we deduce that each eigenvalue of λ of A, different from $\rho(A)$ satisfy $|\lambda| < \rho(A)$.

Since the eigenvectors of A^p are the same of A we deduce the same results if A is a nonnegative symmetric matrix which is primitive. Suppose that A is irreducible. Then $(tI + A)^{n-1} > 0$ for any $t > 0$. Clearly $\rho(tI + A) = \rho(A) + t$. Hence the eigenvector corresponding to $\rho(A)$ is positive. Therefore $\rho(A) > 0$. Each eigenvalue of $tI + A$ is of the form $\lambda + t$ for some eigenvalue λ of A. Since $(tI + A)$ is primitive for any $t > 0$ it follows that $|\lambda + t| < \rho(A) + t$. Let $t \to 0^+$ to deduce that $|\lambda| \leq \rho(A)$. Since all the eigenvalues of A are real, we can have an equality for some eigenvalue $\lambda \neq \rho(A)$ if and only if $\lambda = -\rho(A)$. This will be the case if A has the following form

$C = \begin{bmatrix} 0 & B \\ B^\top & 0 \end{bmatrix}$ for some $B \in \mathbb{R}_+^{m\times l}$. It can be shown that if A is a symmetric irreducible matrix such that $-\rho(A)$ then $A = PCP^\top$ for some permutation matrix P. We already showed before that C and $-C$ has the same eigenvalues, counted with their multiplicities. Hence $-A$ and A have the same characteristic polynomial. □

Recall the ℓ_∞ norm on $\mathbb{C}^n$ and the corresponding operator norm on $\mathbb{C}^{n\times n}$:

$$\|(x_1,\ldots,x_n)^\top\|_\infty := \max_{i\in[n]} |x_i|, \quad \|A\|_\infty := \max_{\|\mathbf{x}\|_\infty\le 1} \|A\mathbf{x}\|_\infty,\ A \in \mathbb{C}^{n\times n}. \tag{5.1}$$

Lemma 5.3 *Let $A = [a_{ij}]_{i,j\in[n]} \in \mathbb{C}^{n\times n}$. Then*

$$\|A\|_\infty = \max_{i\in[n]} \sum_{j\in[n]} |a_{ij}|. \tag{5.2}$$

Proof. Let $a := \max_{i\in[n]} \sum_{j\in[n]} |a_{ij}|$. Assume that $\|(x_1,\ldots,x_n)^\top\| \le 1$. Then

$$|(A\mathbf{x})_i| = |\sum_{j\in[n]} a_{ij}x_j| \le \sum_{j\in[n]} |a_{ij}|\,|x_j| \le \sum_{j\in[n]} |a_{ij}| \le a,$$

for each $i \in [n]$. Hence $\|A\|_\infty \le a$.

From the definition of a it follows that $a = \sum_{j\in[n]} |a_{kj}|$ for some $k \in [n]$. Let x_j be a complex number of modulus 1, i.e. $x_j = 1$, such that $|a_{kj}| = a_{kj}x_j$ for each $j \in [n]$. Let $\mathbf{x} = (x_1,\ldots,x_n)^\top$. Clearly $\|\mathbf{x}\|_\infty = 1$. Furthermore $a = \sum_{j\in[n]} a_{kj}x_j$. Hence $\max_{i\in[n]} \|A\mathbf{x}\|_\infty \ge a$. Therefore $a = \|A\|_\infty$. □

Definition 5.4 *Let $A = [a_{ij}] \in \mathbb{C}^{n\times n}$. Denote $|A| := [|a_{ij}|] \in \mathbb{R}_+^{n\times n}$. For $\mathbf{x} = (x_1,\ldots,x_n) \ge \mathbf{0}$ define $r(A,\mathbf{x}) := \min\{t \ge 0, |A|\mathbf{x} \le t\mathbf{x}\}$.*

It is straightforward to show that

$$r(A,\mathbf{x}) := \max_{i\in[n]} \frac{\sum_{j\in[n]} |a_{ij}|x_j}{x_i} \quad \text{for } \mathbf{x} > \mathbf{0}. \tag{5.3}$$

Lemma 5.5 *Let $A = [a_{ij}] \in \mathbb{C}^{n\times n}$ and $\mathbf{1} := (1,\ldots,1)^\top \in \mathbb{C}^n$. Then*

1. *$\|A\|_\infty = r(A,\mathbf{1})$.*
2. *Assume that $\mathbf{x} = (x_1,\ldots,x_n) > \mathbf{0}$. Then each eigenvalue λ of A satisfies $|\lambda| \le r(A,\mathbf{x})$. That is, $\rho(A) \le r(A,\mathbf{x})$. In particular*

$$\rho(A) \le \|A\|_\infty. \tag{5.4}$$

Proof. *1.* follows straightforward from the definition of $\|A\|_\infty$ and $r(A,\mathbf{1})$. We now show (5.4). Assume that $A\mathbf{z} = \lambda\mathbf{z}$, where $\mathbf{z} = (z_1,\ldots,z_n)^\top \in \mathbb{C}^n\setminus\{\mathbf{0}\}$. Without loss of generality we may assume that $|z_k| = \|\mathbf{z}\|_\infty = 1$ for some $k \in [n]$. Then

$$|\lambda| = |\lambda z_k| = |(A\mathbf{z})_k| \le \sum_{j=1}^n |a_{kj}|\,|z_j| \le \sum_{j=1}^n |a_{kj}| \le \|A\|_\infty.$$

Assume that $\mathbf{x} = (x_1, \ldots, x_n) > 0$. Let $D = \operatorname{diag}(x_1, \ldots, x_n)$. Define $A_1 := D^{-1}AD$. A straightforward calculation shows that $r(A, \mathbf{x}) = r(A_1, \mathbf{1}) = \|A_1\|_\infty$. Clearly, A and A_1 are similar. Use (5.4) for A_1 to deduce that $|\lambda| \leq r(A, \mathbf{x})$. □

Lemma 5.6 *Let $A \in \mathbb{R}_+^{n\times n}$. Then $r(A, \mathbf{x}) \geq r(A, (I + A)\mathbf{x})$.*

Proof. As $A\mathbf{x} \leq r(A, \mathbf{x})\mathbf{x}$ we deduce

$$A((I + A)\mathbf{x}) = (I + A)(A\mathbf{x}) \leq (I + A)(r(A, \mathbf{x})\mathbf{x}) = r(A, \mathbf{x})A((I + A)\mathbf{x})).$$

The definition of $r(A, (I + A)\mathbf{x})$ yields the lemma. □

The next result is a basic step in the proof of Theorem 5.1, and is due to Wielandt.

Theorem 5.7 *Let $A \in \mathbb{R}_+^{n\times n}$ be an irreducible matrix. Then*

$$\rho(A) = \min_{\mathbf{x}>\mathbf{0}} r(A, \mathbf{x}) > 0. \tag{5.5}$$

Equality holds if and only if $A\mathbf{x} = \rho(A)\mathbf{x}$. There is a unique positive eigenvector $\mathbf{x} > \mathbf{0}$ corresponding to $\rho(A)$, up to a multiple by a constant. That is, rank $(\rho(A)I - A) = n - 1$. *Any other real eigenvector of A corresponding to $\lambda \neq \rho(A)$ has two coordinates with opposite signs.*

Proof. Since for any $a > 0, \mathbf{x} > \mathbf{0}$ we have that $r(A, a\mathbf{x}) = r(A, \mathbf{x})$ it is enough to assume that $\mathbf{x}$ is a probability vector, i.e. $\|\mathbf{x}\|_1 = \sum_{i=1}^n x_i = 1$. Denote by $\Pi_n \subset \mathbb{R}^n$ the set of probability vectors. Clearly Π_n is a compact set. We now consider the extremal problem

$$r(A) := \inf_{\mathbf{x}\in\Pi_n} r(A, \mathbf{x}). \tag{5.6}$$

We claim that this infimum is achieved for some $\mathbf{y} \in \Pi_n$. As $r(A, \mathbf{x}) \geq 0$ for $\mathbf{x} \geq \mathbf{0}$ it follows that there exists a sequence $\mathbf{x}_m \in \Pi_n$ such $\lim_{m\to\infty} r(A, \mathbf{x}_m) = r(A)$. Clearly $r(A, \mathbf{x}_m) \geq r(A)$. Hence by considering a subsequence if necessary we can assume that $r(A, \mathbf{x}_m)$ is a nonincreasing sequence, i.e $r(A, \mathbf{x}_l) \geq r(A, \mathbf{x}_m)$ for $m > l$. So $r(A, \mathbf{x}_l)\mathbf{x}_m \geq r(A, \mathbf{x}_m)\mathbf{x}_m \geq A\mathbf{x}_m$. As $\lim_{m\to\infty} \mathbf{x}_m = \mathbf{y}$ it follows that $r(A, \mathbf{x}_l)\mathbf{y} \geq A\mathbf{y}$. Hence $r(A, \mathbf{x}_l) \geq r(A, \mathbf{y}) \Rightarrow r(A) \geq r(A, \mathbf{y})$. From the definition of $r(A)$ we deduce that $r(A) \leq r(A, \mathbf{y})$. Hence $r(A) = r(A, \mathbf{y})$.

Assume first that $\mathbf{y}$ is not an eigenvector of A. Let $\mathbf{z} := r(A)\mathbf{y} - A\mathbf{y}$. So $\mathbf{z} \geq \mathbf{0}$ and $\mathbf{z} \neq 0$. Since A is irreducible then $(I + A)^{n-1} > 0$. Hence $(I + A)^{n-1}\mathbf{z} > \mathbf{0}$. This is equivalent to

$$r(A)((I + A)^{n-1}\mathbf{y}) > (I + A)^{n-1}(A\mathbf{y}) = A((I + A)^{n-1}\mathbf{y}).$$

Let $\mathbf{u} := (I + A)^{n-1}\mathbf{y}$. Then the above inequality yields that $r(A, \mathbf{u}) < r(A)$. Let $\mathbf{v} = b\mathbf{u} \in \Pi_n$ for a corresponding $b > 0$. So $r(A, \mathbf{v}) = r(A, \mathbf{u}) < r(A)$. This contradicts the definition of $r(A)$. Hence $A\mathbf{y} = r(A)\mathbf{y}$. Clearly, $(I + A)^{n-1}\mathbf{y} = (1 + r(A))^{n-1}\mathbf{y}$. As $\mathbf{y} \in \Pi_n$ and $(I + A)^{n-1} > 0$ it follows that $(I + A)^{n-1}\mathbf{y} > \mathbf{0}$. Hence $\mathbf{y} > 0$. Part *2.* of Lemma 5.5 yields that $\rho(A) \leq r(A, \mathbf{y}) = r(A)$. As $r(A)$ is an eigenvalue of A it follows that $r(A) = \rho(A)$. Since A is irreducible $A \neq 0$. Hence

$r(A) = r(A, \mathbf{y}) > 0$. Combine the definition of $r(A)$ with the fact that $A\mathbf{y} = r(A)\mathbf{y}$ to deduce (5.5).

We next claim that rank $(\rho(A)I - A) = n-1$, i.e. the dimension of the eigenspace of A corresponding to A is one. Assume to the contrary that $\mathbf{w} \in \mathbb{R}^n$ is a an eigenvector of A corresponding to $\rho(A)$ which is not colinear with $\mathbf{y}$. Consider $\mathbf{w} + t\mathbf{y}$. Since $\mathbf{y} > 0$ for $t \gg 0$ and $-t \gg 0$ we have that $w + t\mathbf{y} > 0$ and $w + t\mathbf{y} < 0$ respectively. Hence there exist $t_0 \in \mathbb{R}$ such that $\mathbf{u} := w + t_0\mathbf{y} \geq \mathbf{0}$ and at least one of the coordinates of $\mathbf{u} = \mathbf{0}$. As $\mathbf{w}$ and $\mathbf{y}$ are linearly independent it follows that $\mathbf{u} \neq \mathbf{0}$. Also $A\mathbf{u} = \rho(A)\mathbf{u}$. Since A is irreducible we obtain that $(1 + \rho(A))^{n-1}\mathbf{u} = (I + A)^{n-1}\mathbf{u} > \mathbf{0}$ This contradicts the assumption that $\mathbf{u}$ has at least one zero coordinates.

Suppose finally that $A\mathbf{v} = \lambda\mathbf{v}$ for some $\lambda \neq \rho(A)$ and $\mathbf{v} \neq \mathbf{0}$. Observe that $A^\top \geq 0$ and $(I + A^\top)^{n-1} = ((I + A)^{n-1})^\top > 0$. That is, $A^\top$ is irreducible. Hence there exists $\mathbf{u} > 0$ such that $A^\top\mathbf{u} = \rho(A^\top)\mathbf{u}$. Recall that $\rho(A^\top) = \rho(A)$. Consider now $\mathbf{u}^\top A\mathbf{v} = \lambda\mathbf{u}^\top\mathbf{v} = \rho(A)\mathbf{u}^\top\mathbf{v}$. As $\rho(A) \neq \lambda$ we deduce that $\mathbf{u}^\top\mathbf{v} = 0$. As $\mathbf{u} > \mathbf{0}$ it follows that $\mathbf{v}$ has two coordinates with opposite signs. □

We now give the full proof of Theorem 5.1. We recall the well known result for the roots of normalized polynomials from complex variable course:

Lemma 5.8 *Let $p(z) = z^n + a_1z^{n-1} + \ldots + a_n$ be a normalized polynomial of degree n with complex coefficients. Assume that $p(z) = \prod_{j=1}^n (z - z_j(p))$. Given $\epsilon > 0$ then there exists $\delta(\epsilon) > 0$ such that for each coefficient vector $\mathbf{b} = (b_1, \ldots, b_n)$ satisfying $\sum_{j=1}^n |b_j - a_j|^2 < \delta(\epsilon)^2$ one can rearrange the roots of $q(z) = z^n + b_1z^{n-1} + \ldots + b_n = \prod_{j=1}^n (z - z_j(q))$ such that $|z_j(q) - z_j(p)| < \epsilon$ for each $j \in [n]$.*

Corollary 5.9 *Assume that $A_m \in \mathbb{C}^{n\times n}, m \in \mathbb{N}$ converges to $A \in \mathbb{C}^{n\times n}$. Then $\lim_{m\to\infty} \rho(A_m) = \rho(A)$. That is, the function $\rho(\cdot) : \mathbb{C}^{n\times n} \to [0, \infty)$, that assigns to each A its spectral radius, is a continuous function.*

Proof of *1.* and *2.* of Theorem 5.1. Assume that $A \in \mathbb{R}_+^{n\times n}$, i.e. A has nonnegative entries. Let $J_n \in \mathbb{R}_+^{n\times n}$ be a matrix whose all entries are 1. Let $A_m = A + \frac{1}{m}J_n$ for $m \in \mathbb{N}$ so each A_m is a positive matrix and $\lim_{m\to\infty} A_m = A$. Corollary 5.9 yields that $\lim_{m\to\infty} \rho(A_m) = \rho(A)$. Clearly, each A_m is irreducible. Hence by Theorem 5.7 there exists a probability vector $\mathbf{x}_m \in \Pi_n$ such that $A_m\mathbf{x}_m = \rho(A_m)\mathbf{x}_m$. As Π_n is compact there exists a subsequence $m_k, k \in \mathbb{N}$ such that $\lim_{k\to\infty} \mathbf{x}_{m_k} = \mathbf{x} \in \Pi_n$. Consider the equalities $A_{m_k}\mathbf{x}_{m_k} = \rho(A_{m_k})\mathbf{x}_{m_k}$. Let $k \to \infty$ to deduce that $A\mathbf{x} = \rho(A)\mathbf{x}$. So $\rho(A)$ is an eigenvector of A with a corresponding probability vector $\mathbf{x}$. □

In what follows we need the following lemma:

Lemma 5.10 *Let $B \in \mathbb{C}^{n\times n}$ and assume that* rank $B = n - 1$. *Then* adj $B = \mathbf{u}\mathbf{v}^\top \neq 0$ *where $B\mathbf{u} = B^\top\mathbf{v} = \mathbf{0}$. Suppose furthermore that $A \in \mathbb{R}_+^{n\times n}$ and $\rho(A)$ is a geometrically simple eigenvalue. I.e.,* rank $B = n-1$, *where $B = \rho(A)I_n - A$. Then* adj $B = \mathbf{u}\mathbf{v}^\top$ *and $\mathbf{u}, \mathbf{v} \gneq \mathbf{0}$ are the eigenvectors of A and $A^\top$ corresponding to the eigenvalue $\rho(A)$.*

Proof. Recall that the entries of adj B are all $n-1$ minors of B. Since rank $B = n-1$ we deduce that adj $B \neq 0$. Recall next that $B(\text{adj } B) = (\text{adj } B)B = \det B\, I_n =$

0, as rank $B = n-1$. Let adj $B = [\mathbf{u}_1\ \mathbf{u}_2 \dots \mathbf{u}_n]$. As Badj $B = 0$ we deduce that $B\mathbf{u}_i = \mathbf{0}$ for $i \in [n]$. Since rank $B = n-1$ the dimension of the null space is one. So the null space of B is spanned by $\mathbf{u} \neq \mathbf{0}$. Hence $\mathbf{u}_i = v_i\mathbf{u}$ for some $v_i, i \in [n]$. Let $\mathbf{v} = (v_1, \dots, v_n)^\top$. Then adj $B = \mathbf{u}\mathbf{v}^\top$. Since adj $B \neq 0$ it follows that $\mathbf{v} \neq \mathbf{0}$. As $0 = (\text{adj } B)B = \mathbf{u}\mathbf{v}^\top B = \mathbf{u}(B^\top\mathbf{v})$ and $\mathbf{u} \neq \mathbf{0}$ it follows that $B^\top\mathbf{v} = \mathbf{0}$.

Assume now that $A \in \mathbb{R}_+^{n\times n}$. Then *1.* of Theorem 5.1 yields that $\rho(A)$ is an eigenvalue of A. Hence $B = \rho(A)I_n - A$ is singular. Suppose that rank $B = n-1$. We claim that adj $B \geq 0$. We first observe that for $t > \rho(A)$ we have that $\det(tI_n - A) = \prod_{j=1}^n (t - \lambda_j(A)) > 0$. Indeed, if $\lambda_j(A)$ is real then $t - \lambda_j(A) > 0$. If $\lambda_j(A)$ is not real, i.e. complex number then $\bar{\lambda}_j(A)$ is also an eigenvalue of A. Recall that $t > \rho(A) \geq |\lambda_j(A)|$. Hence $(t - \lambda_j(A))(t - \bar{\lambda}_j(A)) = |t - \lambda_j(A)|^2 > 0$. So $\det(tI_n - A) > 0$. Furthermore,

$$\text{adj } (tI_n - A) = \det(tI_n - A)(tI_n - A)^{-1} =$$
$$\det(tI_n - A)t^{-1}(I_n - t^{-1}A)^{-1} = \det(tI_n - A)t^{-1}\sum_{j=0}^{\infty}(t^{-1}A)^j.$$

As $A^j \geq 0$ for each nonnegative integer j it follows that adj $(tI_n - A) \geq 0$ for $t > \rho(A)$. Let $t \searrow \rho(A)$ to deduce that adj $B \gneq 0$. As $B = \mathbf{u}\mathbf{v}^\top$ we can assume that $\mathbf{u}, \mathbf{v} \gneq \mathbf{0}$. □

Lemma 5.11 *Let $A \in \mathbb{R}_+^{n\times n}$ be an irreducible matrix. Then $\rho(A)$ is an algebraically simple eigenvalue.*

Proof. Clearly, we may assume that $n > 1$. Theorem 5.7 implies that $\rho(A)$ is geometrically simple, i.e. nul $(\rho(A)I - A) = 1$. Hence rank $(\rho(A)I - A) = n-1$. Lemma 5.10 yields that adj $(\rho(A)I - A) = \mathbf{u}\mathbf{v}^\top$, where $A\mathbf{u} = \rho(A)\mathbf{u}, A^\top\mathbf{v} = \rho(A)\mathbf{v}, \mathbf{u}, \mathbf{v} \gneq \mathbf{0}$. Theorem 5.7 yileds that $\mathbf{u}, \mathbf{v} > \mathbf{0}$. Hence $\mathbf{u}\mathbf{v}^\top$ is a positive matrix. In particular, $\operatorname{tr}\mathbf{u}\mathbf{v}^\top = \mathbf{v}^\top\mathbf{u} > 0$. Since

$$(\det(\lambda I - A))'(\lambda = \rho(A)) = \operatorname{tr}\text{adj } (\rho(A)I - A) = \mathbf{v}^\top\mathbf{u} > 0,$$

we deduce that $\rho(A)$ is a simple root of the characteristic polynomial of A. □
For $A \in \mathbb{R}_+^{n\times n}$ and $\mathbf{x} \gneq \mathbf{0}$ denote $s(A, \mathbf{x}) = \max\{t \geq 0, A\mathbf{x} \geq t\mathbf{x}\}$. From the definition of $r(A, \mathbf{x})$ we immediately conclude that $r(A, \mathbf{x}) \geq s(A, \mathbf{x})$. We now give a complementary part of Theorem 5.7.

Lemma 5.12 *Let $A \in \mathbb{R}_+^{n\times n}$ be an irreducible matrix. Then*

$$\rho(A) = \max_{\mathbf{x}>\mathbf{0}} s(A, \mathbf{x}) > 0. \tag{5.7}$$

Equality holds if and only if $A\mathbf{x} = \rho(A)\mathbf{x}$.

The proof of this lemma is very similar to the proof of Theorem 5.7, and we omit the proof.

As usual, denote by $\mathrm{S}^1 := \{z \in \mathbb{C}, |z| = 1\}$ the unit circle in the complex plane

Lemma 5.13 *Let $A \in \mathbb{R}_+^{n\times n}$ be irreducible, $C \in \mathbb{C}^{n\times n}$. Assume that $|C| \le A$. Then $\rho(C) \le \rho(A)$. Equality holds, i.e. there exists $\lambda \in \operatorname{spec} C$, such that $\lambda = \zeta\rho(A)$ for some $\zeta \in \mathrm{S}^1$, if and only if there exists a complex diagonal matrix $D \in \mathbb{C}^{n\times n}$, whose diagonal entries are equal to 1, such that $C = \zeta DAD^{-1}$. The matrix D is unique up to a multiplication by $t \in \mathrm{S}^1$.*

Proof. We can assume that $n > 1$. Assume that $A = [a_{ij}], C = [c_{ij}]$. Theorem 5.7 yields that there exists $\mathbf{u} > \mathbf{0}$ such that $A\mathbf{u} = \rho(A)\mathbf{u}$. Hence $r(A, \mathbf{u}) = \rho(A)$. Since $|C| \le A$ it follows that $r(C, \mathbf{u}) \le r(A, \mathbf{u}) = \rho(A)$. Lemma 5.5 yields that $\rho(C) \le r(C, \mathbf{u})$. Hence $\rho(C) \le \rho(A)$.

Suppose that $\rho(C) = \rho(A)$. Hence we have equalities $\rho(C) = r(C, \mathbf{u}) = \rho(A) = r(A, \mathbf{u})$. So $\lambda = \zeta\rho(A)$ is an eigenvalue of C for some $\zeta \in \mathrm{S}^1$. Furthermore, for the corresponding eigenvector $\mathbf{z}$ of C we have

$$|\lambda|\ |\mathbf{z}| = |C\mathbf{z}| \le |C|\ |\mathbf{z}| \le A|\mathbf{z}|.$$

Hence $\rho(A)|z| \le A|z|$. Therefore $s(A, |z|) \ge \rho(A)$. Lemma 5.12 yields that $s(A, |z|) = \rho(A)$ and $|z|$ is the corresponding nonnegative eigenvector. So $|C\mathbf{z}| = |C||\mathbf{z}| = A|\mathbf{z}|$. Theorem 5.7 yields that $|\mathbf{z}| = \mathbf{u}$.

Let $z_i = d_i u_i, |d_i| = 1$ for $i = 1, \dots, n$. The equality $|C\mathbf{z}| = |C|\ |\mathbf{z}| = A|\mathbf{z}|$ combined with the triangle inequality and $|C| \le A$, yields first that $|C| = A$. Furthermore for each fixed i the nonzero complex numbers $c_{i1}z_1, \dots, c_{in}z_n$ have the same argument, i.e. $c_{ij} = \zeta_i a_{ij}\bar{d}_j$ for $j = 1, \dots, n$ and some complex number ζ_j, where $|\zeta_i| = 1$. Recall that $\lambda z_i = (C\mathbf{z})_i$. Hence $\zeta_i = \zeta d_i$ for $i = 1, \dots, n$. Thus $C = \zeta DAD^{-1}$, where $D = \operatorname{diag}(d_1, \dots, d_n)$. It is straightforward to see that D is unique up tD for any $t \in \mathrm{S}^1$.

Suppose now that for $D = \operatorname{diag}(d_1, \dots, d_n)$, where $|d_1| = \dots = |d_n| = 1$ and $|\zeta| = 1$ we have that $C = \zeta DAD^{-1}$. Then $\lambda_i(C) = \zeta\lambda_i(A)$ for $i = 1, \dots, n$. So $\rho(C) = \rho(A)$. Furthermore $c_{ij} = \zeta d_i c_{ij}\bar{d}_j, i, j = 1, \dots, n$. So $|C| = A$. □

Lemma 5.14 *Let $\zeta_1, \dots, \zeta_h \in \mathrm{S}^1$ be h distinct complex numbers which form a multiplicative semi-group, i.e. for any integers $i, j \in [1, h]$ $\zeta_i\zeta_j \in \{\zeta_1, \dots, \zeta_h\}$. Then the set $\{\zeta_1, \dots, \zeta_h\}$ is the set, (the group), of all h roots of 1: $e^{\frac{2\pi i\sqrt{-1}}{h}}, i = 1, \dots, h$.*

Proof. Let $\zeta \in \mathcal{T} := \{\zeta_1, \dots \zeta_h\}$. Consider the sequence $\zeta^i, i = 1, \dots$. Since $\zeta^{i+1} = \zeta\zeta^i$ for $i = 1, \dots$, and $\mathcal{T}$ is a semigroup, it follows that each ζ^i is in $\mathcal{T}$. Since $\mathcal{T}$ is a finite set, we must have two positive integers such that $\zeta^k = \zeta^l$ for $k < l$. Assume that k and l are the smallest possible positive integers. So $\zeta^p = 1$, where $p = l - k \ge 1$, and $\mathcal{T}_p := \{\zeta, \zeta^2, \dots, \zeta^{p-1}, \zeta^p = 1\}$ are all p roots of 1. ζ is called a p-primitive root of 1. I.e. $\zeta = e^{\frac{2\pi p_1\sqrt{-1}}{p}}$ where p_1 is an positive integer less than p. Furthermore p_1 and p are *coprime*, which is denoted by $(p_1, p) = 1$. Note that $\zeta^i \in \mathcal{T}$ for any integer i.

Next we choose $\zeta \in \mathcal{T}$, such that ζ is a primitive p-root of 1 of the maximal possible order. We claim that $p = h$, which is equivalent to the equality $\mathcal{T} = \mathcal{T}_p$. Assume to the contrary that $\mathcal{T}_p \subsetneq \mathcal{T}$. Let $\eta \in \mathcal{T}\backslash\mathcal{T}_p$. The previous arguments show that $\eta =$ is a q-primitive root of 1. So $\mathcal{T}_q \subset \mathcal{T}$, and $\mathcal{T}_q \subsetneq \mathcal{T}_p$. So q can not divide p. Also the maximality of p yields that $q \le p$. Let $(p, q) = r$ be the g.c.d., the greatest common divisor of p and q. So $1 \le r < q$. Recall that Euclid algorithm, which is

applied to the division of p by q with a residue, yields that there exists two integers i, j such that $ip + jq = r$. Let $l := \frac{pq}{r} > p$ be the least common multiplier of p and q. Observe that $\zeta' = e^{\frac{2\pi\sqrt{-1}}{p}} \in \mathcal{T}_p, \eta' = e^{\frac{2\pi\sqrt{-1}}{q}} \in \mathcal{T}_q$. So

$$\xi := (\eta')^i(\zeta')^j = e^{\frac{2\pi(ip+jq)\sqrt{-1}}{pq}} = e^{\frac{2\pi\sqrt{-1}}{l}} \in \mathcal{T}.$$

As ξ is an l-primitive root of 1, we obtain a contradiction to the maximality of p. So $p = h$ and $\mathcal{T}$ is the set of all h-roots of unity. □

Theorem 5.15 *Let $A \in \mathbb{R}^{n\times n}_+$ be irreducible, and assume that for a positive integer $h \geq 2$, A has $h-1$ distinct eigenvalues $\lambda_1, \ldots, \lambda_{h-1}$, which are distinct from $\rho(A)$, such that $|\lambda_1| = \ldots = |\lambda_{h-1}| = \rho(A)$. Then the following conditions hold*

1. *Assume that A is imprimitive, i.e. not primitive. Then there exists exactly $h - 1 \geq 1$ distinct eigenvalues $\lambda_1, \ldots, \lambda_{h-1}$ different from $\rho(A)$ and satisfying $|\lambda_i| = \rho(A)$. Furthermore, the following conditions hold.*
   (a) *λ_i is an algebraically simple eigenvalue of A for $i = 1, \ldots, h-1$.*
   (b) *The complex numbers $\frac{\lambda_i}{\rho(A)}, i = 1, \ldots, h-1$ and 1 are all h roots of unity, i.e. $\lambda_i = \rho(A)e^{\frac{2\pi\sqrt{-1}i}{h}}$ for $i = 1, \ldots, h-1$. Furthermore, if $A\mathbf{z}_i = \lambda_i\mathbf{z}_i, \mathbf{z}_i \neq \mathbf{0}$ then $|\mathbf{z}_i| = \mathbf{u} > 0$, the Perron-Frobenius eigenvector $\mathbf{u}$ given in Theorem 5.7.*
   (c) *Let ζ be any h-root of 1, i.e. $\zeta^h = 1$. Then the matrix ζA is similar to A. Hence, if λ is an eigenvalue of A then $\zeta\lambda$ is an eigenvalue of A having the same algebraic and geometric multiplicity as λ.*
   (d) *There exists a permutation matrix $P \in \mathcal{P}_n$ such that $P^\top AP = B$ has a block h-circulant form*

$$B = \begin{bmatrix} 0 & B_{12} & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & B_{23} & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \vdots & 0 & B_{(h-1)h} \\ B_{h1} & 0 & 0 & 0 & \vdots & 0 & 0 \end{bmatrix},$$

$$B_{i(i+1)} \in \mathbb{R}^{n_i\times n_{i+1}}, i = 1, \ldots, h, B_{h(h+1)} = B_{h1},$$
$$n_{h+1} = n_1, n_1 + \ldots + n_h = n.$$

   Furthermore, the diagonal blocks of B^h are all irreducible primitive matrices, i.e.

$$C_i := B_{i(i+1)} \ldots B_{(h-1)h}B_{h1} \ldots B_{(i-1)i} \in \mathbb{R}^{n_i\times n_i}_+, \ i = 1, \ldots, h, \qquad (5.8)$$

   are irreducible and primitive.

Proof. Assume that $\zeta_i := \frac{\lambda_i}{\rho(A)} \in \mathrm{S}^1$ for $i = 1, \ldots, h-1$ and $\zeta_h = 1$. Apply Lemma 5.13 to $C = A$ and $\lambda = \zeta_i\rho(A)$ to deduce that $A = \zeta_i D_i A D_i^{-1}$ where D_i is a diagonal matrix such that $|D| = I$ for $i = 1, \ldots, h$. Hence, if λ is an eigenvalue of

A then $\zeta_i\lambda$ is an eigenvalue of A, with an algebraic and geometrical multiplicity as λ. In particular, since $\rho(A)$ is an algebraically simple eigenvalue of A, $\lambda_i = \zeta_i\rho(A)$ is an algebraically simple of A for $i = 1, \ldots, h-1$. This establish (1a).

Let $\mathcal{T} = \{\zeta_1, \ldots, \zeta_h\}$. Note that

$$A = \zeta_i D_i A D_i^{-1} = \zeta_i D_i(\zeta_j D_j A D_j^{-1}) D_i^{-1} = (\zeta_i\zeta_j)(D_i D_j) A (D_i D_j)^{-1}. \tag{5.9}$$

So $\zeta_i\zeta_j\rho(A)$ is an eigenvalue of A. Hence $\zeta_i\zeta_j \in \mathcal{T}$, i.e. $\mathcal{T}$ is a semigroup. Lemma 5.14 yields that $\{\zeta_1, \ldots, \zeta_n\}$ are all h roots of 1. Note that if $A\mathbf{z}_i = \lambda_i\mathbf{z}_i, \mathbf{z}_i \neq \mathbf{0}$, then $\mathbf{z}_i = tD_i\mathbf{u}$ for some $0 \neq t \in \mathbb{C}$, where $\mathbf{u} > \mathbf{0}$ is the Perron-Frobenius vector given in Theorem 5.7. This establish (1b).

Let $\zeta = e^{\frac{2\pi\sqrt{-1}}{h}} \in \mathcal{T}$. Then $A = \zeta DAD^{-1}$, where D is a diagonal matrix $D = (d_1, \ldots, d_n), |D| = I$. Since D can be replaced by $\bar{d}_1 D$, we can assume that $d_1 = 1$. (5.9) yields that $A = \zeta^h D^h A D^{-h} = IAI^{-1}$. Lemma 5.13 yields that $D^h = \operatorname{diag}(d_1^h, \ldots, d_n^h) = tI$. Since $d_1 = 1$ it follows that $D^h = I$. So all the diagonal entries of D are h-roots of unity. Let $P \in \mathcal{P}_n$ be a permutation matrix such that the diagonal matrix $E = P^\top DP$ is of the following block diagonal form

$$E = I_{n_1} \oplus \mu_1 I_{n_2} \oplus \ldots \oplus \mu_{l-1} I_{n_l},\ \mu_i = e^{\frac{2\pi k_i\sqrt{-1}}{h}},$$
$$i = 1, \ldots, l-1,\ 1 \leq k_1 < k_2 < \ldots < k_{l-1} \leq h-1.$$

Note that $l \leq h$ and equality holds if and only if $k_i = i$. Let $\mu_0 = 1$.

Let $B = P^\top AP$. Partition B to a block matrix $[B_{ij}]_{i=j=1}^{l}$ where $B_{ij} \in \mathbb{R}_+^{n_i \times n_j}$ for $i, j = 1, \ldots, l$. Then the equality $A = \zeta DAD^{-1}$ yields $B = \zeta EBE^{-1}$. The structure of B and E implies the equalities

$$B_{ij} = \zeta\frac{\mu_{i-1}}{\mu_{j-1}} B_{ij}, \quad i, j = 1, \ldots, l.$$

Since all the entries of B_{ij} are nonnegative we obtain that $B_{ij} = 0$ if $\zeta\frac{\mu_{i-1}}{\mu_{j-1}} \neq 1$. Hence $B_{ii} = 0$ for $i = 1, \ldots, l$ Since B is irreducible it follows that not all $B_{i1}, \ldots, B_{il}$ are zero matrices for each $i = 1, \ldots, l$. First start with $i = 1$. Since $\mu_0 = 1$ and $j_1 \geq 1$ it follows that $\mu_j \neq \zeta$ for $j > 1$. So $B_{1j} = 0$ for $j = 3, \ldots, l$. Hence $B_{12} \neq 0$, which implies that $\mu_1 = \zeta$, i.e. $k_1 = 1$. Now let $i = 2$ and consider $j = 1, \ldots, l$. As $k_i \in [k_1 + 1, h-1]$ for $i > 1$, it follows that $B_{2j} = 0$ for $j \neq 3$. Hence $B_{23} \neq 0$ which yields that $k_2 = 2$. Applying these arguments for $i = 3, \ldots, l-1$ we deduce that $B_{ij} = 0$ for $j \neq i+1$, $B_{i(i+1)} \neq 0, k_i = i$ for $i = 1, \ldots, l-1$. It is left to consider $i = l$. Note that

$$\frac{\zeta\mu_{l-1}}{\mu_{j-1}} = \frac{\zeta^l}{\zeta^{j-1}} = \zeta^{l-(j-1)},\ \text{which is different from 1 for } j \in [2, l].$$

Hence $B_{lj} = 0$ for $j > 1$. Since B is irreducible, $B_{11} \neq 0$. So $\zeta^l = 1$. Since $l \leq h$ we deduce that $l = h$. Hence B has the block form given in (1d). □

Lemma 5.16 *Let $A \in \mathbb{R}_+^{n\times}$ be irreducible. Then A is primitive if and only if one of the following conditions hold:*

1. *$n = 1$ and $A > 0$.*

2. *$n > 1$ and each eigenvalue λ of A different from $\rho(A)$ satisfies the inequality $|\lambda| < \rho(A)$.*

Proof. If $n = 1$ then A is primitive if and only if $A > 0$. Assume now that $n > 1$. So $\rho(A) > 0$. By considering $B = \frac{1}{\rho(A)}A$, it is enough to consider the case $\rho(A) = 1$. Assume first that if $\lambda \neq 1$ is an eigenvalue of A then $|\lambda| < 1$. Theorem 5.7 implies $A\mathbf{u} = \mathbf{u}, A^\top \mathbf{w} = \mathbf{w}$ for some $\mathbf{u}, \mathbf{w} > \mathbf{0}$. So $\mathbf{w}^\top \mathbf{u} > 0$. Let $\mathbf{v} := (\mathbf{w}^\top \mathbf{u})^{-1}\mathbf{w}$. Then $A^\top \mathbf{v} = \mathbf{v}$ and $\mathbf{v}^\top \mathbf{u} = 1$. Part *2.* of Theorem 3.6 yields $\lim_{k\to\infty} A^k = Z_{10} \geq 0$ where Z_{10} is the component of A corresponding to an algebraically simple eigenvalue $1 = \rho(A)$. Problem 2(a) in §3.1 yields that $Z_{10} = \mathbf{u}\mathbf{v}^\top > 0$. So there exists integer $k_0 \geq 1$, such that $A^k > 0$ for $k \geq k_0$, i.e. A is primitive.

Assume now A is has exactly $h > 1$ distinct eigenvalues λ satisfying $|\lambda| = 1$. Lemma 5.15 implies that there exists a permutation matrix P such that $B = P^\top AP$ is of the form (1d) of Theorem 5.15 . Note that B^h is a block diagonal matrix. Hence $B^{hj} = (B^h)^j$ is a block diagonal matrix for $j = 1, \ldots, ...$ Hence, B^{hj} is never a positive matrix, so A^{hj} is never a positive matrix. Hence A is not primitive. □

Lemma 5.17 *Let $B \in \mathbb{R}_+^{n\times n}$ be an irreducible, imprimitive matrix, having $h > 1$ distinct eigenvalues λ satisfying $|\lambda| = \rho(B)$. Suppose furthermore that B has the form (1d) of Theorem 5.15 . Then B^h is a block diagonal matrix, where each diagonal block is an irreducible primitive matrix whose spectral radius is $\rho(B)^h$. In particular, the last claim of (1d) of Theorem 5.15 holds.*

Proof. Let $D(B) = ([n], E)$ be the digraph associated with B. Let $p_0 = 0, p_1 = p_0 + n_1, \ldots, p_h = p_{h-1} + n_h = n$. Denote $V_i = \{p_{i-1} + 1, \ldots, p_i\}$ for $i = 1, \ldots, h$, and let $V_{h+1} := V_1$. So $[n] = \cup_{i=1}^h V_i$. The form of B implies that $E \subset \cup_{i=1}^h V_i \times V_{i+1}$. Thus, any walk that connects vertices $j, k \in V_i$ must be divisible by h. Observe next that $B^h = \operatorname{diag}(C_1, \ldots, C_h)$, where $C_1 = [c_{jk}^{(1)}]_{j=k=1}^{n_1}, \ldots, C_h = [c_{jk}^{(h)}]_{j=k=1}^{n_h}$ are defined in (5.8). Let $D(C_i) = (V_i, E_i)$ be the digraph associated with C_i for $i = 1, \ldots, h$. Then there exists a path of length h from j to k in V_i if and only if $c_{jk}^{(i)} > 0$. Since B is irreducible, $D(B)$ is strongly connected. Hence, each $D(C_i)$ is strongly connected. Thus, each C_i is irreducible.

Recall that $B\mathbf{u} = \rho(B)\mathbf{u}$ for the Perron-Frobenius vector $\mathbf{u}^\top = (\mathbf{u}_1^\top, \ldots, \mathbf{u}_h^\top) > \mathbf{0}^\top, \mathbf{u}_i \in \mathbb{R}_+^{n_i}, i = 1, \ldots, h$. Thus, $B^h\mathbf{u} = \rho(B)^h\mathbf{u}$, which implies that $C_i\mathbf{u}_i = \rho(B)^h\mathbf{u}_i, i = 1, \ldots, h$. Since $\mathbf{u}_i > \mathbf{0}$ Proposition **??** yields that $\rho(C_i) = \rho(B)^h, i = 1, \ldots, h$. Recall that the eigenvalues of B^h are the h power of the eigenvalues of B, i.e. $\lambda_i(B^h) = \lambda_i(B)^h$ for $i = 1, \ldots, n$. Furthermore, B has h simple eigenvalues $\rho(B)e^{\frac{2\pi i\sqrt{-1}}{h}}, i = 1, \ldots, h$ with $|\lambda| = \rho(B)$, and all other eigenvalues satisfy $|\lambda| < \rho(B)$. Hence B^h has one eigenvalues $\rho(B)^h$ of an algebraic multiplicity h and all other eigenvalues μ satisfy $|\mu| < \rho(B)^h$.

Since $B^h = \operatorname{diag}(C_1, \ldots, C_h)$, we deduce that the eigenvalues of B^h are the eigenvalues of $C_1, \ldots, C_h$. As C_i is irreducible and $\rho(C_i) = \rho(B)^h$, we deduce that all other eigenvalues μ of C_i satisfy $|\mu| < \rho(C_i)$. Lemma 5.16 yields that C_i is primitive. □

References

[1] G.H. Golub and C.F. Van Loan. Matrix Computation, *John Hopkins Univ. Press, 3rd Ed.*, Baltimore, 1996.

[2] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 2013.