Lecture Notes for Advanced Linear Algebra

Vladislav Kargin

June 18, 2024

Contents

Bibliography

[Axl15] Sheldon Axler. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer, 3rd edition, 2015.

[TB97] Lloyd N. Trefethen and David Bau. *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, 1st edition, 1997.

[Tre21] Sergei Treil. *Linear Algebra Done Wrong*. 2021. Available at `https://sites.google.com/a/brown.edu/sergei-treil-homepage/linear-algebra-done-wrong?authuser=0`.

Chapter 1

Introduction

These lecture notes often refer to results in textbooks by Sergei Treil [Tre21] and Sheldon Axler [Axl15], and occasionally to a textbook by Lloyd Trefethen and David Bau [TB97].

1.1 Vector spaces

Here we give a very dry algebraist-style definition of vector spaces over a field.

1.1.1 Definition

Definition 1.1.1. A **semigroup** is a set G with a binary operation $\cdot : G \times G \to G$, which is associative.

Definition 1.1.2. A **monoid** is a semigroup with a unit element e: $ge = g = eg$.

Definition 1.1.3. A **group** is a monoid such that every element g has an inverse h, $gh = e = hg$.

Definition 1.1.4. A group is called **abelian** if $ab = ba$ for all elements a, b of G.

Definition 1.1.5. A **ring** is a set R with two binary operations $+$ and $\cdot$, such that

1. $(R, +, 0)$ is an **abelian** group,
2. $(R, \cdot)$ is a semigroup,
3. Distributive laws hold.

If $a \cdot b = b \cdot a$ for all $a, b \in R$ then the ring is called **commutative**.

Definition 1.1.6. A **field** F is a commutative ring with identity element 1, such that $(F - \{0\}, \cdot, 1)$ is an abelian group.

Definition 1.1.7. An R**-module** is an abelian group $(M, +, 0)$, such that ring R **acts** on M, that is there is a map $\cdot : R \times M \to M$, that sastisifies:

1. $r \cdot (m_1 + m_2) = (r \cdot m_1) + (r \cdot m_2)$.
2. $(r_1 + r_2) \cdot m = r_1 \cdot m + r_2 \cdot m$.
3. $(r_1 \cdot r_2) \cdot m = r_1 \cdot (r_2 \cdot m)$.

Definition 1.1.8. A **vector space** $(V, +, 0)$ over field F is an F-module. The action of F on V is called scalar multiplication.

Note: I will say a **linear space** and a vector space interchangeably in these lecture notes, meaning the same thing.

We have to distinguish $0 \in F$ and $0 \in V$, so I will (sometimes) denote 0 in V as $\vec{0}$.

Example 1.1.9 ($m \times n$ matrices with entries from F.). Let $1 \leq m, n \in \mathbb{Z}$. Let

$$F_n^m = \left\{ A = [a_{ij}] \middle| 1 \leq i \leq m, 1 \leq j \leq n, a_{ij} \in F \right\}$$

be the set of all $m \times n$ matrices. So a_{ij} is the entry in row i and column j. Define the entries of the matrix $C = A + B$ by $c_{ij} = a_{ij} + b_{ij}$ and for $\alpha \in F$, define αA as the matrix with entries αa_{ij}. Finally let $\vec{0}$ be the $m \times n$ matrix such that all entries are 0.

Then F_n^m is a vector space over F. (We will also use the notation $F_{m \times n}$ for this space.)

Example 1.1.10. Consider the set of polynomials with coefficients in F and the maximum degree n. The elements of this set are the sums

$$f(x) = \sum_{i=0}^{n} a_i x^{n-i},$$

where $a_i \in F$. This set is a vector space over F. We will denote it $P_n[x]$.

By convention we also denote F_1^m as F^m, the space of "column vectors", and F_n^1 as F_n, the space of "row vectors".

Most of the examples that we consider will be for $F = \mathbb{R}$ (real vector spaces) and $F = \mathbb{C}$, (complex vector spaces).

1.1.2 Bases and dimension

(Section 1.2 in Treil) One of the most important properties of vector spaces is the existence of a basis and the fact that one can define the dimension of a vector space.

Definition 1.1.11. Let $S \subset V$ be a set of vectors in a vector space V. A **linear combination** from S is an element $\sum_{i=1}^{m} x_i s_i \in V$ for some $x_i \in F$ and $s_i \in S$. The set of all such elements is the span of S, denoted by $\operatorname{span} S$ or $\langle S \rangle$. If $S = \emptyset$, then by convention $\langle S \rangle = \{\vec{0}\}$.

Definition 1.1.12. A set $S \subset V$ of vectors in V is **generating** if every $v \in V$ is a linear combinations of vectors in S, that is, $V = \operatorname{span} S$.

In this course, the standing assumption is that vector spaces that we consider are finitely-generated, that is, they have a finite generating set S. Such spaces are also called finitely-dimensional. For spaces which are not finitely-generated, we would need to introduce topological concepts in order to define a basis. While we consider some examples that involve infinite-dimensional spaces for illustration purposes, I will avoid proving explicit theorems for these spaces.

Definition 1.1.13. A set $S \subset V$ is **linearly independent** if the zero vector can be represented in a unique way by vectors in S, that is, if $0 = \sum_{i=1}^{m} c_i v_i$ for distinct $v_i \in S$, then all $c_i = 0$.

Definition 1.1.14. A set of vectors $S \in V$ is called a **basis** (for the vector space V) if S is generating and linearly independent.

In fact this is equivalent to the following definition. [Check it.]

A set of vectors $\mathcal{B} = \{v_1, \ldots, v_n\} \in V$ is a **basis** (for the vector space V) if any vector $v \in V$ admits a **unique** representation as a linear combination

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n = \sum_{k=1}^{n} \alpha_k v_k.$$

The coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ in this representation are called **coordinates** of the vector v (with respect to the basis $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$).

The first important result is the existence of a basis.

Theorem 1.1.15. *Any finite generating set of a vector space contains a basis.*

Proof: Exercise. [First show that if any element of a generating set is a linear combination of the other elements, then this element can be removed and the remainder set is still a generating set. Then consider a minimal finite generating set, that is a set, from which we cannot remove any element without making it non-generating. Show that this set is linearly independent and therefore a basis].

The second result states that every basis has the same number of elements.

Theorem 1.1.16. *Any two bases in a (finite-dimensional) vector space V have the same number of vectors in them.*

For the proof Exercises 1.4.11 - 1.4.14. Alternatively, see Prop 3.3. in Treil, however for this proof one needs more information about systems of linear equations, which we develop in the next Chapter.

Definition 1.1.17. The dimension $\dim V$ of a vector space V is the number of vectors in a basis.

Examples. Some vector spaces occurs frequently in practice and they often come with a specific basis, which is called the standard basis.

Example 1.1.18. Basis and dimension for F^m.

Example 1.1.19. Basis and dimension for F_n^m.

Example 1.1.20. Basis for $P_n[t]$, the vector space of polynomials with degree $\leq n$.

Example 1.1.21. The vector spaces P_5, F_2^3, and F^6 are look different but they all have the same dimension 6.

Note that in general, a vector space V does are not have a specific basis, which we could a call a standard basis. However, once a basis is chosen we can map the vector space to the vector space F^n for some $n \geq 0$ in the following way. If a basis in V is $v_1, \ldots, v_n$ and $v = a_1 v_1 + \ldots a_n v_n$, so that the coordinates of v in this basis are $\{a_1, \ldots, a_n\}$, then we map

$$v \to \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

As we will see later, this map is linear and bijective. If one vector space can be bijectively mapped to another vector space and the map is linear then we say that these vector spaces are **isomorphic**. Therefore, all vector spaces that have the same dimension are isomorphic.

Example 1.1.22. The vector spaces P_5, F_2^3, and F^6 are isomorphic.

The concept of the basis is specific for vector spaces. In general, modules over a ring R might have no basis. For example, consider the module $\mathbb{Z}/m\mathbb{Z}$ over the ring $\mathbb{Z}$ with scalar multiplication given by:

$$\alpha \cdot \overline{x} = \overline{\alpha x} \mod m,$$

where $x \in \mathbb{Z}$ is a representative for the element $\overline{x}$ in $\mathbb{Z}/m\mathbb{Z}$. Then, this module has a generating set $\{\overline{1}\}$ but this generating set is not a basis because $m \cdot \overline{1} = \overline{0}$, so this set is not linearly independent. In fact, this module does not have any basis.

1.1.3 Subspaces and spans

Definition 1.1.23. A subset W of a linear space V is called a **subspace**, if it is a vector space with respect to the same operations.

Theorem 1.1.24. *$W \subset V$ is a subspace of V if and only if*

1. *W is closed under $+$, that is, for any $w_1, w_2 \in W$, $w_1 + w_2 \in W$.*
2. *W is closed under $\cdot$, that is, for all $\alpha \in F$ and $w \in W$, $\alpha \cdot w \in W$.*
3. *$\vec{0} \in W$.*

Proof. If W is a subspace, then the conditions are necessary by the definition of a subspace. Conversely, if the conditions holds then one needs to check the axioms of the vector space for W. For example, we need to check that

$$\alpha \cdot (w_1 + w_2) = \alpha \cdot w_1 + \alpha \cdot w_2.$$

for any $w_1, w_2 \in W$. This equality holds in V because V is a vector space, and by our assumptions both the left hand side and the right hand side are in W. Therefore, the equality holds in W as well.

For another example, one needs to check that W is an abelian group, in particular, if $w \in W$ then there exists an additive inverse of w. This can be established by proving first some additional properties of vector spaces.

Lemma 1.1.25. *Suppose V is a vector space over F. Then,*

1. *$0 \cdot v = \vec{0}$ for all $v \in V$.*
2. *$\alpha \cdot \vec{0} = \vec{0}$, for all $\alpha \in F$.*
3. *$-v = (-1) \cdot v$, for all $v \in V$.*

Given that the lemma is true, one defines $-w = -1 \cdot w$, which is an inverse of w in V by the lemma. In addition, it is in W by assumed property 2. Therefore it is the inverse of w in W. □

By using Theorem 1.1.24, it is easy to check that the span of a set of vectors is a linear subspace.

Proposition 1.1.26. *For any $S \in V$, $\langle S \rangle$ is a subspace of V.*

1.1.4 Direct sums of subspaces

Reading: Axler Section 1C, Treil Section 4.2.4

Suppose $W_1, \ldots, W_t \subset V$ are subspaces of V. Then their sum is

$$W_1 + \ldots + W_t = \{w_1 + \ldots + w_t | w_i \in W_i \text{ for } 1 \leq i \leq t\}.$$

It is easy to check that this is a subspace of V.

We say that the sum is **direct** when each vector w in the sum has a **unique** expression $w = w_1 + \ldots + w_t$, $w_i \in W_i$ for $1 \leq i \leq t$.

This equivalent to the requirement that if $w_1 + \ldots + w_t = 0$ and $w_i \in W_i$ for $1 \leq i \leq t$, then $w_i = 0$ for all i. (check this!)

Theorem 1.1.27. *The sum $\sum_{i=1}^{t} W_i$ is direct if and only if*

$$W_i \cap \Big(\sum_{j \neq i} W_j \Big) = \{\vec{0}\} \text{ for each } 1 \leq i \leq t.$$

Proof. ($\Rightarrow$) Assume the sum is direct. By seeking a contradiction, suppose we can find a non-zero $w \in W_i \cap \big(\sum_{j \neq i} W_j \big)$. This means that w has two different expressions as a sum of vectors in $W_1, \ldots, W_t$. The first one is $w = w$, with $w \in W_i$ and the second is $w = \sum_{j=1}^{t} w_j$, with $w_j \in W_j$ and $w_i = 0$. These two expressions are different because the summand corresponding to W_i is non-zero in the first sum and zero in the second one. This contradict the definition of the direct sum.

($\Leftarrow$) Suppose that all those intersections are zero and suppose we have two distinct expressions for a vector w:

$$\sum_{i=1}^{t} w_i = \sum_{i=1}^{t} w_i', \quad w_i, w_i' \in W_i.$$

where for some i, $w_i \neq w_i'$. Without loss of generality, let $i = 1$, then $u := w_1 - w_1' \neq 0$ and

$$u = w_1 - w_1' = \sum_{i=2}^{t} (w_i' - w_i),$$

so $u \in W_1 \cap \sum_{i=2}^{t} W_i$, which contradict the assumption that this intersection is trivial. $\square$

Example 1.1.28. For $t = 2$, $W_1 + W_2$ is a direct sum if and only if $W_1 \cap W_2 = \{\vec{0}\}$. For $t = 3$, $W_1 + W_2 + W_3$ is direct if and only if

$$W_1 \cap (W_2 + W_3) = W_2 \cap (W_1 + W_3) = W_3 \cap (W_1 + W_2) = \{\vec{0}\}.$$

Exercise 1.1.29. Find an example (in $V = \mathbb{R}^3$) of subspaces W_1, W_2, W_3, such that $W_1 \cap W_2 = \{\vec{0}\}$, $W_1 \cap W_3 = \{\vec{0}\}$, $W_2 \cap W_3 = \{\vec{0}\}$ but the sum $W_1 + W_2 + W_3$ is not direct.

In case when a sum $W_1 + \ldots + W_t$ is direct we will write it

$$W_1 \oplus \ldots \oplus W_t = \bigoplus_{i=1}^{t} W_i.$$

Theorem 1.1.30. *Suppose $\mathcal{B}_i = \{w_{i1}, \ldots, w_{id_i}\}$ is a basis of W_i. Then $W = \sum_{i=1}^{t} W_i$ is direct if and only if $\mathcal{B}_1 \cup \ldots \cup \mathcal{B}_t$ is a basis for W.*

In particular, $\dim W = \sum_{i=1}^{t} \dim W_i$. We omit the proof. You can attempt it as an exercise or see a textbook, for example, Thm 4.2.6 in Treil's book.

If W is a subspace of V, $W \subset V$, then a subspace $U \subset V$ is called **complementary** if $W \oplus U = V$. A complementary subspace always exists but not unique. (Unless $W = V$ or $W = \{\vec{0}\}$.)

Dimension formula

Here is a formula which is especially useful when $W = W_1 + W_2$ and the sum is not necessarily direct.

Theorem 1.1.31. *Let $W_1, W_2 \in V$ be two subspaces of V. Then,*

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2). \tag{1.1}$$

Proof. Choose a basis $\mathcal{R} = \{u_1, \ldots, u_r\}$ in $W_1 \cap W_2$. Extend $\mathcal{R}$ to a basis of W_1: $\mathcal{S} = \{u_1, \ldots, u_r, v_1, \ldots, v_s\}$. Extend $\mathcal{R}$ to a basis of W_2: $\mathcal{T} = \{u_1, \ldots, u_r, w_1, \ldots w_t\}$. Then the claim is that

$$\mathcal{U} = \{u_1, \ldots, u_r, v_1, \ldots, v_s, w_1, \ldots, w_t\}$$

is a basis in $W_1 + W_2$. If we prove this claim then the claim of the theorem follows because

$$\begin{aligned} \dim(W_1 + W_2) &= r + s + t = (r + s) + (r + t) - r \\ &= \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2). \end{aligned}$$

It is easy to see that $\mathcal{U}$ is a generating set. In order to prove that $\mathcal{U}$ is linearly independent, suppose

$$\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{s} b_j v_j + \sum_{k=1}^{t} c_k w_k = 0 \tag{1.2}$$

We have

$$w := -\sum_k c_k w_k = \sum_i a_i u_i + \sum_j b_j v_j,$$

so this vector belongs both to W_1 and W_2 and so it can be written as $\sum_i d_i u_i$, since $\{u_i\}$ is a basis in $W_1 \cap W_2$. Plugging this into (1.2), we find

$$\sum_{i=1}^{r} (a_i - d_i) u_i + \sum_{j=1}^{s} b_j v_j = 0,$$

Since $\{u_1, \ldots, u_r, v_1, \ldots v_s\}$ is a basis we find that all $b_j = 0$. Then, (1.2) implies that

$$\sum_{i=1}^{r} a_i u_i + \sum_{k=1}^{t} c_k w_k = 0,$$

and since $\{u_1, \ldots, u_r, w_1, \ldots w_t\}$ is a basis, we find that all $a_i = 0$ and all $c_k = 0$. This gives linear independence of $\mathcal{T}$.

□

Example 1.1.32. If $\dim(V) = 10$, $\dim(W_1) = 7$ and $\dim(W_2) = 6$, what are all possibilities for $\dim(W_1 \cap W_2)$?

Solution:

$$\begin{gathered} \dim(W_1 + W_2) \leq \dim(V) = 10, \\ \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2) \leq 10, \text{ so} \\ \dim(W_1 \cap W_2) \geq 7 + 6 - 10 = 3. \end{gathered}$$

On the other hand $W_1 \cap W_2$ is a subspace of both W_1 and W_2, so

$$\dim(W_1 \cap W_2) \leq \min\{\dim W_1, \dim W_2\} = 6.$$

So the possibilities are $3 \leq \dim(W_1 \cap W_2) \leq 6$.

1.2 Linear systems and linear transformations

1.2.1 Linear systems

Definition 1.2.1. A **linear system** of m equations in n variables is a list of the form:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots a_{2n}x_n &= b_2, \\ &\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots a_{mn}x_n &= b_m, \end{aligned}$$

Given all coefficients a_{ij} and b_i, the values x_j for which all equations are true are called the solutions. If there are no solutions, then the system is called **inconsistent**. Otherwize, it is **consistent**.

Let $A = [a_{ij}] \in F_n^m$ be a coefficient matrix and $B = [b_i] \in F^m$ be the matrix of constants. Let us define

$$AX = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \dots \\ \sum_{j=1}^n a_{mj}x_j \end{bmatrix}$$

Then, we can write the linear system as $AX = B$. We call the operation $X \to AX$ the **matrix-vector multiplication**.

Note that we can think about AX as a linear combination of columns of matrix A with coefficients given by vector X:

$$AX = \sum_{j=1}^{n} x_j Col_j(A), \tag{1.3}$$

where $Col_j(A) = [a_{ij}] \in F^m$, $i = 1, \dots, m$.

Definition 1.2.2. A linear system $AX = B$ is **homogeneous** when $B = \vec{0}$, that is, all $b_i = 0$. Otherwise, the system is called **inhomogeneous**.

A homogeneous system $AX = \vec{0}$ is always consistent since it has the "trivial" solution $X = \vec{0}$. So the main question for $AX = \vec{0}$ is whether it has any "non-trivial" solutions $X \neq \vec{0}$.

Also note that formula (1.3) implies that the system $AX = B$ is consistent if and only if $B \in \langle \{Col_1 A, \dots, Col_n(A)\} \rangle$, that is, the right-hand side is in the span of column vectors of A.

Here are some of the properties of the matrix-vector multiplication operation:

Theorem 1.2.3. *For any $A \in F_n^m$, $X, Y \in F^n$, $\alpha \in F$, we have*

1. $A(X+Y) = AX + AY$,
2. $A(\alpha \cdot X) = \alpha \cdot (AX)$,
3. $A\vec{0} = \vec{0}$.

Then we have the following result.

Theorem 1.2.4. *For any $A \in F_n^m$, the solution set $W = \{X \in F^n | AX = \vec{0}\}$ of the homogeneous system $AX = \vec{0}$ is a subspace of F^n.*

Proof. By using the previous theorem, it is easy to check that W is closed under addition and scalar multiplication. Then the result follows from Theorem 1.1.24 that characterizes subspaces. □

The efficient algorithm for solution of linear systems is given by the Row reduction to Reduced Row Echelon Form.

1.2.2 Linear maps

Definition 1.2.5. Let V and W be any vector spaces over F. We say that a map (that is, a function) $L : V \to W$ is **linear** when

1. $L(v_1 + v_2) = L(v_1) + L(v_2)$ for all $v_1, v_2 \in V$,
2. $L(\alpha \cdot v) = \alpha \cdot L(v)$ for all $v \in V$, $\alpha \in F$.

In this case we will write $L \in \mathcal{L}(V, W)$. We will sometimes use **operator** as a synonym for **linear map**. If $L \in \mathcal{L}(V, W)$, then L is called an **endomorhism** of V. The set of all endomorphisms of vector space V is denoted $End(V)$. (More generally, an **endomorphism** of an R-module M is an R-linear map of M to itself.)

If we have a basis $\mathcal{S} = \{v_1, \ldots, v_n\}$ in V and a basis $\mathcal{T} = \{w_1, \ldots, w_2\}$ in W, then we can define a **matrix of operator** L. Namely, let $Lv_j = \sum a_{ij} w_i$. Then the matrix of L is the matrix A that has entry a_{ij} in row i and column j. Note that the matrix A depends on the choice of the bases $\mathcal{S}$ and $\mathcal{T}$.

In other words,

$$Col_j(A) = [L(\vec{v}_j)], \text{ for } 1 \leq j \leq n,$$

where $[L(\vec{v}_j)]$ is the column vector of coordinates of $[L(\vec{v}_j)]$ in the basis $\{w_1, \ldots, w_m\}$.

Example 1.2.6. For a matrix $A = [a_{ij}] \in F_n^m$ define the map $L_A : F^n \to F^m$ by $L_A(X) = AX$ for every $X \in F^n$.

Theorem 1.2.7. *The map L_A is linear. That is, it satisfies:*

1. $L_A(X + Y) = L_A(X) + L_A(Y)$,
2. $L_A(\alpha X) = \alpha \cdot L_A(X)$.

This theorem is simply a restatement of Theorem 1.2.3.

Exercise: The matrix of transformation L_A with respect to standard bases in F^n and F^m is A.

The point of this exercise is that if we identify V with F^n and W with F^m in such a way that the bases $\mathcal{S}$ and $\mathcal{T}$ are mapped to the standard bases in F^n and F^m then the linear transformation L corresponds to matrix multiplication: $x \to Ax$.

Example 1.2.8. Let V_n be the space of polynomials with coefficients in F and maximum degree n.

1. Is the differentiation operation ($D : P(x) \to P'(x)$) a linear map?
2. Is the shift operation $T : P(x) \to P(x + 1)$ a linear map?
3. Consider the integration operation $S : V_n \to V_{n+1}$, given by $S : P(x) \to \int_0^x P(t)\, dt$.

Let $\vec{e}_j = x^j$, $j = 0, \ldots, n$, be a basis of V_n. What is the matrix of the linear maps D, T, S in this basis?

Definition 1.2.9. For a linear map $L : V \to W$ define

$$\begin{aligned} \mathrm{Ker}(L) = \mathrm{Null}(L) &= \{v \in V | L(v) = \vec{0}\}, \text{ and} \\ \mathrm{range}(L) &= \{L(v) \in W | v \in V\}. \end{aligned}$$

Theorem 1.2.10. *For any linear $L : V \to W$, $\mathrm{Ker}(L)$ and $\mathrm{range}(L)$ are subspaces of V and W, respectively.*

The proof is by checking the condition of Theorem 1.1.24. For example, if we have $w_1, w_2 \in \mathrm{Ker}(L)$, then we need to check that $w_1 + w_2 \in Ker(L)$. We can write:

$$L(w_1 + w_2) = L(w_1) + L(w_2) = 0 + 0 = 0,$$

where the first equality holds by linearity of L and the second holds by assumption.

One other useful operation defined on matrices is **transposition**:

$$A = [a_{ij}] \in F_n^m \to A^t = [a_{ji}] \in F_m^n$$

It is easy to check that it is a linear map.

1.2.3 Injective and surjective maps

If we have a system of equations, then in general we want to know whether it is soluble and how many solutions it has.

Let $f : S \to T$ be a function from set S to set T. (That is, for every $s \in S$, there exists $t \in T$, such that $f(s) = t$.) We don't assume at this moment that it is linear.

Definition 1.2.11. The function f is **surjective** (onto) if range$(f) = T$, that is, for every $t \in T$, there exists $s \in S$, such that $f(s) = t$.

Definition 1.2.12. The function f is **injective** (one-to-one) when if $s_1 \neq s_2$ implies that $f(s_1) \neq f(s_2)$.

Definition 1.2.13. The function f is **bijective** if it is both injective and surgective.

Note that if f is bijective, then it is invertible in the following sense.

Definition 1.2.14. Function $f : S \to T$ is **invertible** when there exists a function $g : T \to S$ such that

1. For any $s \in S$, $g(f(s)) = s$, and
2. for any $t \in T$, $f(g(t)) = t$.

If we have two functions: $f : S \to T$ and $g : R \to S$, then we can define their **composition** $f \circ g : R \to T$: $(f \circ g)(r) = f(g(r))$ for every $r \in R$. Note the order in which we write the composition: g is written second, although we apply it first.

Theorem 1.2.15. *Composition of functions is associative.*

Now we specialize these concepts to linear maps.

Theorem 1.2.16. *Let $L : V \to W$ be linear. Then L is injective if and only if* $\mathrm{Ker}(L) = \vec{0}$. *It is surjective if and only if* $\mathrm{range}(L) = W$.

Proof. We know that for linear maps, $L(\vec{0}) = \vec{0}$. If L is injective then $L(v) = \vec{0}$ implies that $v = \vec{0}$, therefore $\mathrm{Ker}(L) = \vec{0}$.

Conversely, suppose that $\mathrm{Ker}(L) = \vec{0}$ and $v_1 \neq v_2$, then $L(v_1) - L(v_2) = L(v_1 - v_2) \neq \vec{0}$ and therefore $L(v_1) \neq L(v_2)$. It follows that L is injective.

The statement about the surjective maps follows from the definition of the range(L) and surjectivity.

□

Corollary 1.2.17. *Let $A \in F_n^m$ and $L_A : F^n \to F^m$ be the linear map map $L_A(X) = AX$. Then L_A is injective if and only if the homogeneous linear system $AX = \vec{0}$, has only the trivial solution.*

This corollary is simply reformulation of the previous theorem for the specific example of the linear map L_A.

Theorem 1.2.18. *The map $L_A : F^n \to F^m$ is surjective if and only if $AX = B$ is consistent for every $B \in F^m$.*

This theorem is essentially a reformulation of the definition of surjectivity, so it does not help much. We will see a more useful approach through the reduction of matrix A to the row echelon form.

1.3 Matrix-matrix product

Now let $L_A : F^n \to F^m$ is a linear map for matrix $A = [a_{ij}] \in F_n^m$, $L_B : F^p \to F^n$ is a linear map for matrix $B = [b_{ij}] \in F_p^n$. Then we can define the composition $L_A \circ L_B : F^p \to F^m$. A direct calculation gives the following result.

Theorem 1.3.1. *$L_A \circ L_B = L_C : X \to CX$ where $C = [c_{ij}] \in F_p^m$.*

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}. \tag{1.4}$$

Definition 1.3.2. The matrix C defined in (1.4) is called the matrix product of A and B, and denoted $C = AB$.

Since the composition of functions is associative, we can immediately conclude that the matrix multiplication is associative:

$$(AB)C = A(BC).$$

One other useful way to look at the matrix product is to think about it as matrix-vector product applied to columns of B:

$$Col_k(AB) = A\,Col_k(B) \text{ for all } k.$$

If we recall formula (1.3), we can get another interpretation: every column of B calculates a linear combination of columns of A.

Example 1.3.3.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 5 & 9 \\ 4 & 6 \end{bmatrix} = \begin{bmatrix} 21 & 37 \\ 45 & 85 \end{bmatrix}.$$

It is easy to check that matrix addition and matrix multiplication for the set F_n^n satisfies the axioms of the ring with zero matrix as the identity element for addition and the matrix

$$I_n = [\delta_{ij}] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

as the identity element for multiplication.

Theorem 1.3.4. *The set F_n^n is a ring with respect to matrix addition and matrix multiplication.*

Since every element of F_n^n corresponds to a linear map from F^n to F^n, this ring is $End(F^n)$, the ring of **endomorphisms** of F^n.

The ring $End(F^n)$ is non-commutative. The matrices in this ring (that is, all square $n \times n$ matrices) have an important characteristic, called **trace**.

$$\mathrm{Tr} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdots a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = a_{11} + a_{22} + \ldots + a_{nn} = \sum_{i=1}^{n} a_{ii}.$$

Theorem 1.3.5. *Let $A \in F_{m\times n}$ and $B \in F_{n\times m}$, so that AB and BA are both well-defined. Then,*

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA).$$

Proof. Exercise. □

Recall that we also have an additional operation, transposition, and it is easy to check that with respect to matrix-matrix multiplication it has the following property:

$$(AB)^t = B^t A^t$$

Note that the order of multiplication changes after the transposition operation is applied.

In general, linear maps $L_A : F^n \to F^m$ are not invertible and we will see later that they can be invertible only if $n = m$.

Definition 1.3.6. $A \in F_n^n$ is called invertible when there exists $B \in F_n^n$ such that $AB = I = BA$.

Theorem 1.3.7. *If $A \in F_n^n$ is invertible, there exists only one matrix B such that $AB = I_n = BA$.*

We will denote this inverse by A^{-1}.

Proof. Suppose there are two inverses, B_1 and B_2. Then,

$$B_1 I_n = B_1(AB_2) = (B_1 A)B_2 = B_2.$$

□

Example 1.3.8. For a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the inverse is

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

It exists only if $ad - bc \neq 0$.

Note also the following useful identity.

Theorem 1.3.9. *If A and B are two invertible $n \times n$ matrices, then AB is also invertible and*

$$(AB)^{-1} = B^{-1} A^{-1}.$$

The proof is left as an exercise.

Sometimes the concepts of the **left and right inverses** of a matrix A are also useful. An $m \times n$ matrix A is **left-invertible** if there exists an $n \times m$ matrix B such that $BA = I_n$. It is **right-invertible** there exists an $n \times m$ matrix C such that such that $AC = I_m$. These matrices are usually not unique. In fact, it can be proved that they are unique if and only if A is invertible.

1.4 Exercises

Do the following exercises to check your understanding of the material. The exercises with (☞) have a hint at the end of Lecture Notes.

Part I

Exercise 1.4.1. Let $x = (1, 2, 3)^t$, $y = (y_1, y_2, y_3)^t$, $z = (4, 2, 1)^t$. (Here t is the transposition operation meaning that the row-vector is converted to a column vector and vice versa.) Compute $2x$, $3y$, $x + 2y - 3z$.

Exercise 1.4.2. Which of the following sets (with natural addition and multiplication by a scalar) are vector spaces. Justify your answer.

(a) The set of all continuous functions on the interval $[0, 1]$;

(b) The set of all non-negative functions on the interval $[0, 1]$;

(c) The set of all polynomials of degree exactly n;

(d) The set of all symmetric $n \times n$ matrices, i.e., the set of matrices $A = [a_{jk}]_{j,k=1}^n$ such that $a_{ij} = a_{ji}$.

Exercise 1.4.3. True or false:

(a) Every vector space contains a zero vector;

(b) A vector space can have more than one zero vector;

(c) An $m \times n$ matrix has m rows and n columns;

(d) If f and g are polynomials of degree n, then $f + g$ is also a polynomial of degree n;

(e) If f and g are polynomials of degree at most n, then $f + g$ is also a polynomial of degree at most n.

More exercises from Treil: 2.1, 2.2, 2.3, 2.4

Exercise 1.4.4. ☞

Let a system of vectors $v_1, v_2, \dots, v_r$ be linearly independent but not generating. Show that it is possible to find a vector v_{r+1} such that the system $v_1, v_2, \dots, v_r, v_{r+1}$ is linearly independent.

Exercise 1.4.5. ☞

Is it possible that vectors v_1, v_2, v_3 are linearly dependent, but the vectors $w_1 = v_1 + v_2$, $w_2 = v_2 + v_3$ and $w_3 = v_3 + v_1$ are linearly independent?

Some proofs:

Exercise 1.4.6. Prove that a zero vector $\vec{0}$ of a vector space V is unique.

Exercise 1.4.7. Prove that the additive inverse of an element of a vector space is unique.

Exercise 1.4.8. ☞

Prove Lemma 1.1.25

1. $0 \cdot v = \vec{0}$ for all $v \in V$.
2. $\alpha \cdot \vec{0} = \vec{0}$, for all $\alpha \in F$.
3. $-v = (-1) \cdot v$, for all $v \in V$.

Part 2

From Treil:

1. 3.1(c), 3.1(d), 3.2, 3.3(b), 3.3(d), 3.4(a, b, c)
2. 5.1(a), 5.3, 5.5, 5.6, 5.7
3. 6.1, 6.2, 6.5, 6.6, 6.7, 6.8, 6.9, 6.12, 6.13
4. 7.1, 7.2, 7.5

Additional Exercises:

Exercise 1.4.9. Suppose the vector space V is the space of polynomials with real coefficients that have the degree ≤ 3. Use the basis $\{1, x, x^2, x^3\}$. In this basis, what is the matrix of the shift operator T that sends a polynomial $P(x) \to P(x+1)$?

Exercise 1.4.10. Write the matrix $\left(\left((AB)^t\right)^{-1}\right)^t$ in terms of A^{-1} and B^{-1}.

Part 3

Proof exercises:

The sequence of exercises 1.4.11 – 1.4.14 proves the existence of dimension of a vector space, that is, the theorem that every basis in a vector space has the same number of vectors. (See exercises 4.33 - 4.36 in Lipschutz - Lipson book.)

Exercise 1.4.11. Prove: Suppose two or more nonzero vectors $v_1, v_2, ..., v_m$ are linearly dependent. Then one of them is a linear combination of the **preceding** vectors.

Exercise 1.4.12. Suppose $S = \{v_1, \ldots, v_m\}$ spans a vector space V.

(a) (Addition of a vector) If $w \in V$, then $\{w, v_1, \ldots, v_m\}$ is linearly dependent and spans V.

(b) (Removal of a vector) If v_i is a linear combination of $v_1, \ldots, v_{i-1}$, then S without v_i spans V.

Exercise 1.4.13. ☞

Use two previous exercises to prove the **Exchange Lemma**: Suppose $\{v_1, v_2, \ldots v_n\}$ spans V, and suppose $\{w_1, w_2, \ldots, w_m\}$ is linearly independent. Then $m \leq n$, and V is spanned by a set of the form

$$\{w_1, w_2, \ldots, w_m, v_{i_1}, v_{i_2}, \ldots, v_{i_{n-m}}\}$$

Show that this implies that if $\{v_1, v_2, \ldots v_n\}$ spans V, then any $n+1$ or more vectors in V are linearly dependent.

Exercise 1.4.14. Using the previous exercise, prove the Dimension Theorem: Every basis of a vector spaces V has the same number of elements.

Exercise 1.4.15. Prove the following theorem: Let V be a vector space of finite dimension n. Then

(i) Any n+1 or more vectors must be linearly dependent.

(ii) Any linearly independent set $\{u_1, \ldots, u_n\}$ with n elements is a basis of V.

(iii) Any spanning set $v_1, v_2, \ldots v_n$ of V with n elements is a basis of V.

Chapter 2

Range and Nullspace of Linear Maps

Reading: Chapter 2 in Treil.

2.1 RREF: Reduced Row Echelon Form

The linear system

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, \dots, m$$

or, in matrix form, $Ax = B$. It can also be associated with "augmented"matrix $[A|B]$ obtained from A by adding an extra column to A, separated by a vertical line to remind us that it was the right-hand side. The variables $x_1, \dots, x_n$ do not explicitly appear but they implicitly associated with columns of A.

Gaussian Elimination

The elementary row operations are

(a) exchange of rows,

(b) multiplication of a row by a non-zero constant, and

(c) subtraction of a row from another row.

These operations do not change the set of solutions of the system, so they can be used to reduce the system to a simple form, when the system is easily to solve. This method is called Gaussian elimination, or row reduction.

The goal is the row echelon form or the reduced row echelon form.

In general, a matrix is in **row echelon form** if it satisfies the following two conditions:

1. All zero rows (i.e. the rows with all entries equal 0), if any, are below all non-zero entries.

2. For any non-zero row its leading entry is strictly to the right of the leading entry in the previous row. (The leading entry is the left-most non-zero entry. The leading entry in each row in echelon form is also called pivot entry, or simply pivot.)

It is in reduced echelon form if, in addition,

1. All pivot entries are equal 1;

2. All entries above the pivots are 0.

The variables which correspond to non-pivot columns are called free.

If the system is in RREF, then it is easy to solve.

Example 2.1.1. Suppose the rref form of the system is

$$[A|b] = \left[\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 5 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{array}\right]$$

Then the solution 0f the system $Ax = b$ is

$$\begin{cases} x_1 = 1 - 2x_2, \\ x_2 \text{ is free}, \\ x_3 = 2 - 5x_4, \\ x_4 \text{ is free}, \\ x_5 = 3. \end{cases}$$

In the vector format, we can write this as $x = v_0 + sv_1 + tv_2$ (with $x_2 = s$ and $x_4 = t$ being arbitrary parameters).

$$x = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ -5 \\ 1 \\ 0 \end{bmatrix}, x_2, x_4 \in F.$$

Usually, v_0 is called a particular solution of the non-homogeneous equation $Ax = b$ which is obtained by setting $x_2 = x_4 = 0$ and reading off the values of the pivot variables from the right hand side, and v_0, v_1 are two independent solutions of the homogeneous equation $Ax = 0$. (These solutions are obtained by setting $(x_2 = 1, x_4 = 0)$ or $(x_2 = 0, x_4 = 1)$, assuming right-hand side equals zero, and solving for the remaining variables. This is easy for the equation in rref form.)

2.2 LU factorization

From the second interpretation of the matrix-matrix product, we know that we can manipulate columns of matrix A by multiplying A on the right by a matrix B. Similarly, we can manipulate rows of A by multiplying it on the **left** by a suitable matrix C.

In particular, elementary row transformations can be realized by multiplying matrices on the left by elementary matrices. For example, subtraction of the twice the row 1 from the row 2 can be realized by multiplying on the left by the following matrix

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -2 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & & & & \end{bmatrix}$$

These row manipulations are all that we need to perform the Gaussian elimination. If we apply the Gaussian elimination to a square matrix A, at the end we obtain an upper diagonal matrix U.

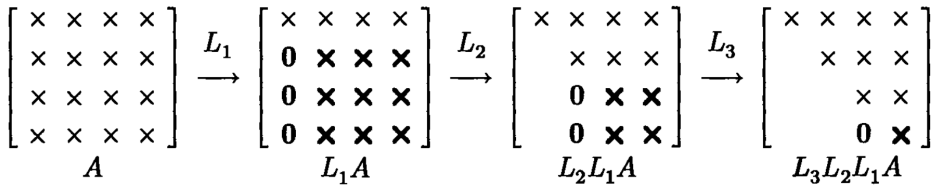

Figure 2.1: Schematics of Gaussian elimination

All the matrices we applied on the left were lower diagonal with 1 on the main diagonal, and the product of such matrices, $\widehat{L} = L_{n-1} \dots L_1$, is also lower diagonal with ones on the main diagonal. We get a formula $\widehat{L}A = U$, where U is upper diagonal.

It is easy to get $\widehat{L}$ by applying the row transformations to the extended matrix: $(A|I_n)$, where I_n is the $n \times n$ identity matrix. Then at the end of the row reduction process, we will get matrix $(U|\widehat{L})$.

The inverse of the matrix $\widehat{L}$ is also lower-diagonal. This can be checked by undoing the transformations one-by-one. For example the inverse of the matrix L_1 above is the matrix

$$L_1^{-1} = \begin{bmatrix} 1 & 0 & 0 & \dots \\ 2 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & 0 \end{bmatrix}$$

And $(\widehat{L})^{-1} = L_1^{-1} \dots L_{n-1}^{-1}$.

So, finally, we get a decomposition

$$A = LU,$$

where $L = (\widehat{L})^{-1}$ is lower-diagonal with ones on the main diagonal and U is upper-diagonal.

This is called the LU decomposition of matrix A. It turns that if this factorization exists then it is unique.

The matrix U represents the system in a **row echelon form**.

2.3 Rank and nullity

The null-space of matrix A is the null-space of the corresponding transformation L_A.

Definition 2.3.1. The **null-space** of $m \times n$ matrix A is the set of vectors $x \in \mathbb{R}^n$ such that $Ax = 0$. It is denoted Null(A) or $\ker(A)$.

In simple terms, if we have m equations in n variables represented by matrix A, then $\ker(A)$ is the space of solutions of the homogeneous system $Ax = 0$.

The dimension of the nullspace is called **nullity** of A.

The row reduction of A to $rref(A)$ gives us an easy algorithm to calculate the basis in the space of solutions of $Ax = 0$. Indeed, the elementary row transformations do not change the null space of the system. (This is why the reduction to the row echelon form is used to solve systems of linear equations.) And for the rref system we can easily write every solution of the homogeneous system as

$$x = s_1 v_1 + s_2 v_2 + \dots + s_d v_d,$$

where s_1, s_2, ..., s_d are the values of the free variables and v_1, ..., v_2 are some non-zero vectors. Moreover, this representation is unique. This implies that $v_1, \ldots, v_d$ **form a basis of the null-space and that the nullity equals to** d, the number of free variables in the rref form.

Example 2.3.2. Let, as in the previous example,

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Then the nullity is 2, the number of free variables, the free variables being x_2 and x_5, and a basis of the null-space is given by

$$\mathcal{B} = \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -5 \\ 1 \\ 0 \end{bmatrix} \right\}$$

(The vectors are obtain from setting $x_2 = 1, x_5 = 0$ in one case and $x_2 = 0, x_5 = 1$ in the other case, and solving the equations.)

Now let us turn to the rank.

Definition 2.3.3. The **range** of an $m \times n$ matrix A, denoted range(A), is the set of vectors in $\mathbb{R}^m$, that can be expressed as Ax for some x.

(It is also frequently denoted Im(A).)

The product Ax can be interpreted as a linear combination of the columns of A, so range(A) is the set of all linear combinations of columns of matrix A. Hence, range(A) is the linear space spanned by columns of A. For this reason, it is also called the **column space** of A.

The **(column) rank** of a matrix A is the dimension of range(A) (or its column space). The rank measures the size of the space of those b which can be represented as $b = Ax$.

The elementary row operations change the column space but they do not change its dimension. If several columns are dependent/independent, then they remain dependent/independent after an elementary transformation used in the reduction process.

Formally, this is a consequence of the fact that the matrix multiplication has the distributive property. If columns $v_1, \ldots, v_k$ are linearly dependent: $c_1 v_1 + \ldots + c_k v_k = 0$, and an elementary row transformation is given by

the left multiplication by matrix L, then the images of these rows are also linearly dependent:

$$c_1(Lv_1) + \ldots + c_k(Lv_k) = 0.$$

One can also go in the opposite direction because every elementary row operation can be undone by another elementary row operation.

Moreover, if certain columns form a basis of range(A), then the transformed columns form a basis of range$(\tilde{A})$, where $\tilde{A}$ is the transformed matrix.

By observing the rref matrix $\tilde{A}$ we see that the basis of its column space is given by the pivot columns. Therefore the basis of the column space A is given by the columns of A that correspond to the pivot columns. In particular, the rank of the matrix equals to the number of pivots.

Example 2.3.4.

$$A = \begin{bmatrix} 1 & 3 & 0 & -1 & 2 \\ 2 & 6 & 1 & -1 & 7 \\ 1 & 3 & 1 & 0 & 5 \end{bmatrix},$$

We can reduce A to the following matrix

$$\operatorname{rref}(A) = \begin{bmatrix} 1 & 3 & 0 & -1 & 2 \\ 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

then we have two pivot variables (corresponding to the entries shown in red) and three free variables (corresponding to blue entries), hence the rank and the nullity of the matrix are 2 and 3.

A basis of the column space range(A) is given by the first and the third columns of the original matrix A: $(1, 2, 1)^t$ and $(0, 1, 1)^t$. These columns correspond to pivot columns in $\operatorname{rref} A$.

The null space of A is the same as the null-space of rref(A). A basis is given by vectors $(-3, 1, 0, 0, 0)$, $(1, 0, -1, 1, 0)$, and $(-2, 0, -3, 0, 1)$, – we simply set one free variable to 1 and all others to 0 and determine the value of all other variables. So to get the last vector we set $x_2 = 0$, $x_4 = 0$ and $x_5 = 1$; from the equation corresponding to the second row we get $x_3 = -3$, and from the equation corresponding to the first row we get $x_1 = -2$.

Since the rank equals to the number of pivot columns and the nullity to the number of free variables, we find that for an $m \times n$ matrix A,

$$\operatorname{nullity}(A) + \operatorname{rank}(A) = n.$$

In terms of linear maps, we obtain one of the fundamental theorems of linear algebra:

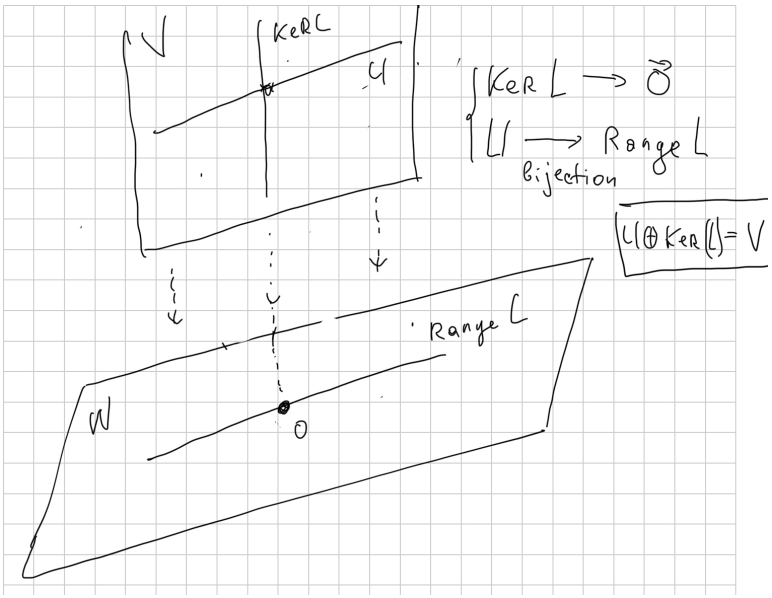

Figure 2.2: An illustration for the rank-nullity theorem

Theorem 2.3.5 (Rank-nullity Theorem)**.** *Let $L : V \to W$ be a linear map of vector spaces. Then,*

$$\dim \ker(L) + \dim \operatorname{range}(L) = \dim V.$$

That is, the sum of dimensions of the range and the kernel of a linear map are equal to the dimension of the source V. If a transformation $L : V \to W$ has trivial kernel ($\mathrm{Ker}L = \{\vec{0}\}$), then the dimensions of the domain V and of the range $\operatorname{range}(L)$ coincide. If the kernel is non-trivial, then the transformation "kills" $\dim \ker(L)$ dimensions, so $\dim \operatorname{range}(L) = \dim V - \dim \ker(L)$.

Sketch of another proof of Theorem 2.3.5.

1. Choose a basis $\{v_1, \ldots, v_s\}$ for subspace $\ker(L)$.
2. Complement it to a basis $\{v_1, \ldots, v_s, w_1, \ldots w_r\}$ in V. (One needs to prove that this is possible.)
3. Claim: $\{L(w_1), \ldots, L(w_r)\}$ form a basis in $\operatorname{range}(L)$.

   Indeed, $\{L(w_1), \ldots, L(w_r)\}$ is a generating set. If $w \in \operatorname{range}(L)$, then

$$\begin{aligned} w = L(v) &= L(c_1 v_1 + \ldots + c_s v_s + d_1 w_1 + \ldots + d_r w_r) \\ &= d_1 L(w_1) + \ldots + d_r L(w_r). \end{aligned}$$

Also, $\{L(w_1), \ldots, L(w_r)\}$ is linearly independent. Indeed, suppose

$$x_1 L(w_1) + \ldots + x_r L(w_r) = 0.$$

Then $x_1 w_1 + \ldots + x_r w_r \in \ker(L)$, so we have

$$x_1 w_1 + \ldots + x_r w_r = y_1 v_1 + \ldots + y_s v_s,$$
$$x_1 w_1 + \ldots + x_r w_r - y_1 v_1 - \ldots - y_s v_s = 0.$$

Since $\{v_1, \ldots, v_s, w_1, \ldots w_r\}$ is a basis, hence all coefficients must be zero, in particular $x_1 = \ldots = x_r = 0$.

Together this shows that $\dim \ker(L) + \dim \text{range}(L) = s + r = \dim V$ □

Example 2.3.6. Let V_n be the space of polynomials of degree $\leq n$ and let L be the differentiation map, $L : f(x) \to f'(x)$. Then, the null-space of L is the space of solutions to the equation $f'(x) = 0$, which is the space of all constants. Hence, $\dim \ker(L) = 1$. The range of L is $V_{n-1} \subset V_n$, that is, the subspace of polynomials of degree $\leq n - 1$. We have $\dim \text{range}(L) = \dim V_{n-1} = n$, and therefore,

$$\dim \ker(L) + \dim \text{range}(L) = 1 + n = \dim(V_n),$$

in agreement with the claim in Theorem 2.3.5.

Since the number of pivots cannot exceed the number of rows or the number of columns, it is clear that $\text{rank}(A) \leq \min\{n, m\}$.

Definition 2.3.7. For an $m \times n$ matrix A, if $\text{rank}(A) = \min\{n, m\}$, we say that matrix A is of **full rank**.

What does this mean that a matrix has full rank?

Theorem 2.3.8. *Let $A \in F_{m \times n}$ with $r = \text{rank}(A)$.*

(a) $L_A : F^n \to F^m$ is injective if and only if $r = n$,

(b) L_A is surjective if and only if $r = m$.

(c) L_A is bijective if and only if $m = r = n$.

Proof. (a) By Theorem 1.2.16 L_A is injective if and only if $\ker(L_A) = \{\vec{0}\}$, so if and only if $\dim \ker(L_A) = 0$. By Theorem 2.3.5, this happens if and only if $\text{rank}(L_A) = n$. (b) By Theorem 1.2.16, L_A is surjective if and only if $\text{range}(L_A) = F^m$. Since the dimension of $\text{range}(L_A) = r$, hence L_A is surjective if and only if $r = m$.

(c) This statement follows immediately from (a) and (b). □

In other words, a matrix A is full rank if and only if the linear map L_A is either injective or surjective or both.

If $m = n$ (the matrix A is square), and the matrix A has full rank, we see from the previous result that L_A is a bijection of $\mathbb{R}^n$ on $\mathbb{R}^n$ and so there exists an inverse transformation.This implies the following result

Theorem 2.3.9. *A square $n \times n$ matrix A has full-rank if and only if there exists an $n \times n$ inverse matrix A^{-1} with the properties $AA^{-1} = A^{-1}A = I_n$, where I_n is the $n \times n$* ***identity matrix.***

It is useful also to note that L_A cannot be invertible if $m \neq n$. Either injectivity or surjectivity is violated.

Row rank = column rank

We can also define the **row space** as the linear space spanned by the rows of the matrix A. Then the row rank is its dimension. It is clear that the row space space of A is isomorphic to the column space of A^t, so row rank $= \text{rank}(A^t)$.

One can easily check that elementary row operations do not change the row space.[1] Indeed, let $x = c_1 r_1 + \ldots c_k r_m$, where $r_1, \ldots r_m$ are rows. The rows of a transformed matrix can be easily written as the linear combinations of the rows of the original matrix. For example, if we subtract twice the row 1 from the row 2 then the new rows will be $r_1' = r_1$, $r_2' = r_2 - 2r_1$, $r_3' = r_3$, and so on. Then we can write x in terms of new rows. In our example this is $x = c_1 r_1' + c_2(r_2' + 2r_1') + \ldots + c_k r_k'$. So it is clear that x is in the row space of the transformed matrix LA.

It follows the elementary row operations preserve not only the column rank but also the row rank. So the row rank of A equals the row rank of rref(A) . But for this form, the non-zero rows are linearly independent, therefore the row rank equals the number of non-zero rows, so it equals the number of pivot variables! This shows that the following fundamental result holds:

$$\text{column rank}(A) = \text{ row rank}(A).$$

Theorem 2.3.10. *For every matrix A,*

$$\text{rank}(A) = \text{rank}(A^t).$$

[1] They do not change the nullspace and they do not change the row space but they do change the columns space, although they preserve the dimension of the column space!

Perhaps the following explanation is helpful. By definition $\operatorname{rank}(A^t)$ counts the number of independent rows, that is, the number of independent equations that generate all equations. So, if we have n variables, we can eliminate $\operatorname{rank}(A^t)$ of them by solving this equations. All other variables are free. This shows that $n - \operatorname{rank}(A^t) = \dim(Ker(A))$. By applying the rank-nullity theorem we find that $\operatorname{rank}(A^t) = \operatorname{rank}(A)$.

One consequence is the following "duality" theorem:

Theorem 2.3.11. *Let A be an $m \times n$ matrix. Then the equation $Ax = b$ has a solution for every $b \in F^m$ if and only if the **dual** equation $A^t x = 0$ has a unique (only the trivial) solution.*

In other words, the linear transformation $L_A : F^n \to F^m$ is surjective if and only if the transformation $L_{A^t} : F^m \to F^n$ is injective.

Proof. $Ax = b$ has a solution for every b if and only if $\operatorname{rank}(A) = m$, and $A^t x = 0$ has a unique solution if and only if $\dim \ker(A^t) = 0$, which we can rewrite as $m - \operatorname{rank}(A^t) = 0$, which is equivalent to $\operatorname{rank}(A) = m$ by the previous theorem. □

2.4 Inverse matrices

We have two interpretations for multiplication by matrix A. Correspondingly, there are two interpretations for the multiplication by the inverse matrix A^{-1}.

1. If multiplication by a square matrix A is interpreted as a linear transformation $x \to y = Ax$ from $V = \mathbb{R}^n$ to $W = \mathbb{R}^n$, then multiplication by A^{-1} is simply the inverse transformation $y \to x$ from $W \to V$.

2. If the multiplication is understood as taking a linear combination of columns of A with coefficients from x, then $A^{-1}y$ is the vector of coefficients of the expansion of y in **the basis of columns of** A. In other words, multiplication by A^{-1} can be understood as the change of basis operation. (We are given y in a standard basis, and we have a new basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. We write the coordinates of each $\boldsymbol{v}_i$ in the standard basis as i-th column of matrix A. Then the coordinates of y in the new basis are given by the entries of the vector $A^{-1}y$.)

Note, however, that in this calculation of coefficients of a vector in the basis of columns of A we assumed that A is $n \times n$ and the range of A equals

the target space R^n. We will talk more about this later when A is $m \times n$, $m > n$ and so the range can be smaller than $\mathbb{R}^m$.

It is also worthwhile to mention two useful properties of the inverse operation:

1. $(AB)^{-1} = B^{-1}A^{-1}$.
2. $(A^t)^{-1} = (A^{-1})^t$.

To see that the first property holds, note that AB is the composition of the linear maps A and B in which B acts first, and A second. We can invert this composition by doing A^{-1} first and B^{-1} second. This corresponds to the product $B^{-1}A^{-1}$.

For the second property, let $B := A^{-1}$. Then, we have $AB = I$. By taking the transposition on both sides and using one of the properties of tthe transposition operation, we get $B^tA^t = I$. This means that $B^t = (A^t)^{-1}$.

Calculation of the inverse matrix: Gauss-Jordan method

In order to calculate the inverse matrix A^{-1} we need to calculate $L_A^{-1}(e_j)$ for the standard basis e_j, $j = 1, \ldots, n$. In other words, we need to solve n linear systems $Av_j = e_j$. We can do it by applying the Gaussian elimination to the extended matrix $(A|I_n)$. After the row reduction we will get the matrix $(I_n|A^{-1})$.

More generally if we need to solve equation $AX = B$ where A is $n \times n$ invertible matrix and B is an $n \times r$ matrix, then we can apply row reduction to $(A|B)$ and the output will give us $(I_n|A^-1B)$.

Example 2.4.1. Calculate the inverse of the following matrix

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}$$

Answer:

$$A^{-1} = \begin{bmatrix} \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ \frac{4}{8} & -\frac{3}{8} & -\frac{2}{8} \\ -1 & 1 & 1 \end{bmatrix}$$

2.5 Change of basis

Suppose $L : V \to W$ is a linear transformation and we are given bases in V, $\{v_1, \ldots, v_n\}$ and in W, $\{w_1, \ldots, w_m\}$. How do we calculate the matrix of L in these bases?

As we seen, the columns of this matrix are vectors $L(v_1), \ldots, L(v_n)$ written in the basis $\{w_1, \ldots, w_m\}$. In this section, we will see two examples of this calculation.

Example 2.5.1 (Change of basis formula for $F^n \to F^n$). Consider the situation when $L = L_A : F^n \to F^n$. In other words, in the standard basis $\{e_1, \ldots, e_n\}$ in F^n the linear transformation is given by the formula $X \to AX$. In particular, the matrix of this transformation is A. Suppose, now that we consider a new basis in F^n that consists of vectors $\{b_1, \ldots b_n\}$. Let B be a matrix with columns b_i, $Col_i(B) = b_i$. Our task is to calculate the matrix of L in this new basis.

Following the plan above, we apply L to each basis vector b_i. We can write the result as a matrix AB in the sense that the columns of AB are the images of basis vectors, $L(b_i)$,

$$Col_i(AB) = ACol_i(B) = L(b_i).$$

However, the entries of these columns are the coordinates in the standard basis $\{e_i\}$, while we are interested in coordinates in the new basis $\{b_i\}$.

So, the next step is to find the coefficients of each column of AB in the basis $\{b_1, \ldots b_n\}$. By what we have seen in Section 2.4, these can be done by multiplying the matrix AB on the left by matrix B^{-1}.

Hence, the matrix of transformation $L = L_A$ in the new basis is

$$\tilde{A} = B^{-1}AB. \tag{2.1}$$

The matrices $\tilde{A}$ that can be represented in this form for an invertible matrix B are called **similar** to A. In algebraic terms $\tilde{A}$ is **conjugate** to A in the multiplicative group of the ring F_n^n, and the map $A \to \tilde{A} = B^{-1}AB$ is an **automorphism** of the ring F_n^n.

We can think about the family of similar matrices as the representations of the same linear transformation L in different bases. In particular, if we want to study the properties of the linear transformation which does not depend on the choice of the basis, we need to find the properties of matrices which does not change if we apply the similarity transformation.

Change of Basis – more general

In general we have the change of coordinates matrix $[I]_{\mathcal{BA}} = [I]_{\mathcal{B}\leftarrow\mathcal{A}}$, which represents change of coordinates from vectors written in basis $\mathcal{A}$ to coordinate written in basis $\mathcal{B}$. If $[v]_{\mathcal{A}}$ is the column of coordinates of vector $v \in V$

in basis $\mathcal{A}$ and $[v]_{\mathcal{B}}$ is the column of coordinates of this vector in basis $\mathcal{B}$, then we should have:

$$[v]_{\mathcal{B}} = [I_{\mathcal{B}\mathcal{A}}[v]_{\mathcal{A}}$$

For example, as we have seen above, if $V = F^n$ and $\mathcal{S}$ is the standard basis in F_n, which consists of vectors $\{e_1, e_2, \ldots, e_n\}$ then we have

$$[I]_{\mathcal{B}\mathcal{S}} = [I]_{\mathcal{B}\leftarrow\mathcal{S}} = B^{-1} \text{ and } [I]_{\mathcal{S}\mathcal{B}} = [I]_{\mathcal{S}\leftarrow\mathcal{B}} = B,$$

where B is the matrix whose columns are the elements of basis $\mathcal{B}$ written in the standard basis.

In this way it is easy to understand the change of basis formula (2.1) in Example 2.5.1. In order to find the matrix for linear transformation L_A in a new basis $\mathcal{B}$, we transform coordinates from basis $\mathcal{B}$ to basis $\mathcal{S}$, apply the transformation L_A in the standard basis – that is multiply by A – and then return back to the basis $\mathcal{B}$. So, the new matrix is

$$\tilde{A} = [I]_{\mathcal{B}\mathcal{S}} A [I]_{\mathcal{S}\mathcal{B}} = B^{-1}AB.$$

Here is a couple of examples to illustrate the ideas in a general setting.

Example 2.5.2. In the space of polynomials of degree at most 1 we have bases $\mathcal{A} = \{1, 1+t\}$, and $\mathcal{B} = \{1+2t, 1-2t\}$, and we want to find the change of coordinate matrix $[I]_{\mathcal{B}\mathcal{A}}$. In principle, we need to write a matrix in which the columns are the coordinates of the basis vectors of $\mathcal{A}$ written in the basis of $\mathcal{B}$. Alternatively, we can do change of coordinates in two steps and we can write:

$$[I]_{\mathcal{B}\mathcal{A}} = [I]_{\mathcal{B}\mathcal{S}} [I]_{\mathcal{S}\mathcal{A}},$$

where $\mathcal{S}$ is the standard basis $\{1, t\}$. Then we can immediately write:

$$\begin{aligned}[I]_{\mathcal{B}\mathcal{A}} &= \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \\ &= \frac{1}{4} \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.\end{aligned}$$

Example 2.5.3. Let $L : \mathbb{R}^2_2 \to \mathbb{R}^2$ be given by

$$L\Big(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \Big) = \begin{bmatrix} a - b + 2c - 3d \\ -a + 3b + 5c + d \end{bmatrix}$$

Let $\mathcal{S}$ be the standard basis of $\mathbb{R}_2^2$, i.e.,

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

and let $\mathcal{T}$ be the standard basis of $\mathbb{R}^2$. Let other ordered bases be

$$\mathcal{S}' = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right\} \text{ and } \mathcal{T}' = \left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \end{bmatrix} \right\}.$$

(a) Find the matrix A, representing L from $\mathcal{S}$ to $\mathcal{T}$.

(b) Find the matrix $\tilde{A}$, representing L from $\mathcal{S}'$ to $\mathcal{T}'$.

For (a), we simply apply L to the elements of basis $\mathcal{S}$ and read off the columns of matrix A:

$$A = \begin{bmatrix} 1 & -1 & 2 & -3 \\ -1 & 3 & 5 & 1 \end{bmatrix}.$$

For (b), what one can do is to find the coordinate change matrises $[I]_{\mathcal{S}\leftarrow\mathcal{S}'}$ and $[I]_{\mathcal{T}'\leftarrow\mathcal{T}}$ then calculate $[I]_{\mathcal{T}'\leftarrow\mathcal{T}} A [I]_{\mathcal{S}\leftarrow\mathcal{S}'}$

It might be a bit more straightforward to calculate $A[I]_{\mathcal{S}\leftarrow\mathcal{S}'}$ directly by applying the linear transformation to the basis elements of $\mathcal{S}'$. We find:

$$L(\mathcal{S}') = \left\{ \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 8 \end{bmatrix}, \begin{bmatrix} -1 \\ 6 \end{bmatrix}, \begin{bmatrix} -3 \\ 3 \end{bmatrix} \right\}$$

We need to apply the transformation $[I]_{\mathcal{T}'\leftarrow\mathcal{T}} = B^{-1}$ to each of the vectors in this set. Here B is the coordinate change matrix $[I]_{\mathcal{T}\leftarrow\mathcal{T}'}$,

$$B = \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$$

It is convenient to do all calculations simultaneously, by bringing the extended matrix

$$\left[\begin{array}{cc|cccc} 3 & 5 & -2 & 1 & -1 & -3 \\ 1 & 2 & 0 & 8 & 6 & 3 \end{array}\right]$$

to the rref.

$$\left[\begin{array}{cc|cccc} 3 & 5 & -2 & 1 & -1 & -3 \\ 1 & 2 & 0 & 8 & 6 & 3 \end{array}\right] \to \left[\begin{array}{cc|cccc} 1 & 2 & 0 & 8 & 6 & 3 \\ 3 & 5 & -2 & 1 & -1 & -3 \end{array}\right]$$

$$\to \left[\begin{array}{cc|cccc} 1 & 2 & 0 & 8 & 6 & 3 \\ 0 & -1 & -2 & -23 & -19 & -12 \end{array}\right] \to \left[\begin{array}{cc|cccc} 1 & 2 & 0 & 8 & 6 & 3 \\ 0 & 1 & 2 & 23 & 19 & 12 \end{array}\right]$$

$$\to \left[\begin{array}{cc|cccc} 1 & 0 & -4 & -38 & -32 & -21 \\ 0 & 1 & 2 & 23 & 19 & 12 \end{array}\right]$$

So we conclude that

$$\tilde{A} = \begin{bmatrix} -4 & -38 & -32 & -21 \\ 2 & 23 & 19 & 12 \end{bmatrix}$$

2.6 Rank-1 matrices

In previous section, we paid a special attention to the matrices of full rank. What about the matrices of small rank? One important case occurs when we have a matrix of rank one. In this case, the dimension of the range is 1, and by the fundamental Theorem 2.3.5 the dimension of the null space is $n - 1$.

Since the dimension of column space is 1, it means that all columns are proportional to a single column $(b_1, \ldots, b_m)^t$. So, the matrix can be written as follows:

$$A = \begin{bmatrix} a_1b_1 & a_2b_1 & \ldots & a_nb_1 \\ a_1b_2 & a_2b_2 & \ldots & a_nb_2 \\ \ldots & & & \\ a_1b_m & a_2b_m & \ldots & a_nb_m \end{bmatrix} = \boldsymbol{ba}^t,$$

where $\boldsymbol{b} = (b_1, \ldots, b_m)^t$ and $\boldsymbol{a} = (a_1, \ldots, a_n)^t$ are two column vectors. Hence, we can conclude that **every matrix of rank 1 is an outer product of two (non-zero) vectors**.

The range of L_A is spanned the vector b and a vector $v = (x_1, \ldots, x_n)$ is in the nullspace of L_A iff $a^t v = a_1x_1 + \ldots a_nx_n = 0$, in other words, if the vector v is orthogonal (i.e. perpendicular) to vector a_n.

One important case is when $a = b$ and b has unit length (i.e. $b^t b = 1$. Then this linear maps b to itself and all vectors perpendicular to b to zero vector. So, geometrically this linear transformation is the projection on the line spanned by $\vec{b}$ along the plane perpendicular to vector $\vec{b}$.

Can you see what will be $I - 2bb^t$, if b has unit length?

2.7 Exercises

Part 1. Basic questions.

Exercise 2.7.1. Let $v_1 = [1, 2, 3]^t$, $v_2 = [0, 1, 3]^t$, $v_3 = [1, 0, 1]^t$. Find the coordinates of the vector $x = e_1 = [1, 0, 0]^t$ in the basis $\{v_1, v_2, v_3\}$.

Exercise 2.7.2. Let

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 6 \\ 1 & 2 & 3 \end{bmatrix}$$

For what value of b, the system

$$Ax = \begin{bmatrix} 1 \\ 4 \\ b \end{bmatrix}$$

has a solution? Write the general solution of the system in the vector form for this value of b. Find a basis of the null-space (i.e., kernel) of A. Find a basis for the range of matrix A.

Exercise 2.7.3. Determine, if the vectors

$$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

are linearly independent or not. Do these four vectors span $\mathbb{R}^4$? (In other words, is it a generating system?) What about $\mathbb{C}^4$?

Exercise 2.7.4. A 54×37 matrix has rank 31. What are dimensions of all 4 fundamental subspaces?

Exercise 2.7.5. Consider the system of vectors

$$[1, 2, 1, 1]^t, [0, 1, 3, 1]^t, [0, 3, 2, 0]^t, [0, 1, 0, 0]^t.$$

a) Prove that it is a basis in $\mathbb{F}^4$. Try to do minimal amount of computations.

b) Find the change of coordinate matrix that changes the coordinates in this basis to the standard coordinates in $\mathbb{F}^4$ (i.e. to the coordinates in the standard basis $e_1, \dots, e_4$).

Exercise 2.7.6 (2.5.1 in Treil). True or false:

(a) Every vector space that is generated by a finite set has a basis;

(b) Every vector space has a (finite) basis;

(c) A vector space cannot have more than one basis;

(d) If a vector space has a finite basis, then the number of vectors in every basis is the same.

(e) The dimension of V_n (the vector space of polynomials of degree $\leq n$) is n;

(f) The dimension on $M_{m\times n}$ is $m+n$;

(g) If vectors $v_1, v_2, \ldots, v_n$ generate (span) the vector space V , then every vector in V can be written as a linear combination of vector $v_1, v_2, \ldots, v_n$ in only one way;

(h) Every subspace of a finite-dimensional space is finite-dimensional;

(i) If V is a vector space having dimension n, then V has exactly one subspace of dimension 0 and exactly one subspace of dimension n.

Exercise 2.7.7 (Exerices 2.6.1 in Treil). True or false

(a) Any system of linear equations has at least one solution;

(b) Any system of linear equations has at most one solution;

(c) Any homogeneous system of linear equations has at least one solution;

(d) Any system of n linear equations in n unknowns has at least one solution;

(e) Any system of n linear equations in n unknowns has at most one solution;

(f) If the homogeneous system corresponding to a given system of a linear equations has a solution, then the given system has a solution;

(g) If the coefficient matrix of a homogeneous system of n linear equations in n unknowns is invertible, then the system has no non-zero solution;

(h) The solution set of any system of m equations in n unknowns is a subspace in $\mathbb{R}^n$;

(i) The solution set of any homogeneous system of m equations in n unknowns is a subspace in $\mathbb{R}^n$.

Exercise 2.7.8 (Exercise 2.7.1 in Treil). True or false:

(a) The rank of a matrix is equal to the number of its non-zero columns;

(b) The $m \times n$ zero matrix is the only $m \times n$ matrix having rank 0;

(c) Elementary row operations preserve rank;

(d) Elementary column operations do not necessarily preserve rank;

(e) The rank of a matrix is equal to the maximum number of linearly independent columns in the matrix;

(f) The rank of a matrix is equal to the maximum number of linearly independent rows in the matrix;

(g) The rank of an $n \times n$ matrix is at most n;

(h) An $n \times n$ matrix having rank n is invertible.

Exercise 2.7.9 (Exercise 2.8.1 in Treil). True or false

(a) Every change of coordinate matrix is square;

(b) Every change of coordinate matrix is invertible;

(c) The matrices A and B are called similar if $B = Q^t AQ$ for some matrix Q;

(d) The matrices A and B are called similar if $B = Q^{-1} AQ$ for some matrix Q;

(e) Similar matrices do not need to be square.

Part 2.

Exercise 2.7.10. Find a 2×3 system (2 equations with 3 unknowns) such that its general solution has a form

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad s \in \mathbb{R}$$

Exercise 2.7.11. Find the inverse of the matrix

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & 7 & 3 \\ 2 & 3 & 4 \end{bmatrix}$$

Show all steps.

Exercise 2.7.12. ☞ Ch. 11

Prove that if V is a vector space having dimension n, then a system of vector $v_1, v_2, \ldots, v_n$ is linearly independent if and only if it spans V.

Exercise 2.7.13. Let vectors u, v, w be a basis in V. Show that $u + v + w$, $v + w$, w is also a basis in V.

Exercise 2.7.14. ☞ Ch. 11

Suppose that vectors $v_1, \ldots, v_n$ form a basis for a real vector space V. Which are also bases?

(a) $\{v_1 + v_2, v_2 + v_3, \ldots, v_{n-1} + v_n, v_n\}$;

(b) $\{v_1 + v_2, v_2 + v_3, \ldots, v_{n-1} + v_n, v_n + v_1\}$;

(c) $\{v_1 - v_n, v_2 + v_{n-1}, \ldots, v_n + (-1)^n v_1\}$.

Exercise 2.7.15. Let

$$U = \{\boldsymbol{x} \in \mathbb{R}^5 : x_1 + x_3 + x_4 = 0, 2x_1 + 2x_2 + x_5 = 0\}$$
$$W = \{\boldsymbol{x} \in \mathbb{R}^5 : x_1 + x_5 = 0, x_2 = x_3 = x_4\}.$$

Find bases for U and W containing a basis for $U \cap W$ as a subset. Give a basis for $U + W$ and show that

$$U + W = \{x \in \mathbb{R}^5 : x_1 + 2x_2 + x_5 = x_3 + x_4\}.$$

Exercise 2.7.16. ☞ Ch. 11

Prove or disprove: If the columns of a square $(n \times n)$ matrix A are linearly independent, so are the columns of $A^2 = AA$.

Exercise 2.7.17. Find the change of coordinates matrix that changes the coordinates in the basis $\{1, 1 + t\}$ in $\mathbb{P}_1[t]$ to the coordinates in the basis $\{1-t, 2t\}$.

Exercise 2.7.18. Let T be the linear operator in $\mathbb{F}^2$ defined (in the standard coordinates) by

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3x + y \\ x-2y \end{bmatrix}$$

Find the matrix of T in the standard basis and in the basis $\{[1, 1]^t, [1, 2]^t\}$.

Part 3.

Additional Exercises.

Easy questions.

Exercise 2.7.19. Do the polynomials $x^3 + 2x$, $x^2 + x + 1$, $x^3 + 5$ generate (span) $P_3[x]$? Justify your answer.

Exercise 2.7.20. Can 5 vectors in $\mathbb{F}^4$ be linearly independent? Justify your answer.

Exercise 2.7.21. Write a matrix with the required property, or explain why no such matrix exists:

a) Column space contains $[1, 0, 0]^t, [0, 0, 1]^t$, row space contains $[1, 1]^t, [1, 2]^t$,

b) Column space is spanned by $[1, 1, 1]^t$, nullspace is spanned by $[1, 2, 3]^t$,

c) Column space is R^4, row space is R^3.

Exercise 2.7.22. If A has the same four fundamental subspaces as B, does $A = B$?

Miscellaneous questions.

Exercise 2.7.23. Let B be a 4 x 4 matrix to which we apply the following operations:

1. double column 1,
2. halve row 3,
3. add row 3 to row 1,
4. interchange columns 1 and 4,
5. subtract row 2 from each of the other rows,
6. replace column 4 by column 3,
7. delete column 1 (so that the column dimension is reduced by 1).

(a) Write the result as a product of eight matrices.

(b) Write it again as a product ABC (same B) of three matrices.

Exercise 2.7.24. Compute rank and find bases of all four fundamental subspaces for the matrix

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

Exercise 2.7.25. Prove that if $A : X \to Y$ and V is a subspace of X then $\dim AV \leq \operatorname{rank} A$. ($AV$ here means the subspace V transformed by the

linear map A, i.e. any vector in AV can be represented as Av, $v \in V$). Deduce from here that for any matrices A and B such that AB is well-defined, $\operatorname{rank}(AB) \le \operatorname{rank}(A)$.

Exercise 2.7.26. Prove that if $A : X \to Y$ and V is a subspace of X then $\dim AV \le \dim V$. Deduce from here that $\operatorname{rank}(AB) \le \operatorname{rank} B$.

Exercise 2.7.27. Prove that if the product AB of two $n \times n$ matrices is invertible, then both A and B are invertible. (Use 2 previous problems, don't use determinants.)

Exercise 2.7.28. Let $L : V \to V$ be a linear map.

(a) Give an example of $L : \mathbb{R}^2 \to \mathbb{R}^2$, such that $\ker L = \operatorname{range} A$.

(b) Show that if $L^2 = L$ then $V = \ker L \oplus \operatorname{range} L$.

(c) Does your proof works for infinite dimensional V? Is the result still true?

Exercise 2.7.29. ☞ Ch. 11

Show that if the equation $Ax = 0$ has unique solution (i.e., if the echelon form of A has pivot in every column), then A is left invertible.

Exercise 2.7.30. (This is an exercise for formula (1.1).)

Let $V = \mathbb{R}^5$ and let U be the subspace of V spanned by the vectors

$$\begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ 2 \\ 1 \\ -2 \end{bmatrix},$$

and W the subspace of V spanned by the vectors

$$\begin{bmatrix} 3 \\ 2 \\ -3 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -4 \\ -1 \\ -2 \\ 1 \end{bmatrix}.$$

Determine the dimension of $U \cap W$.

Exercise 2.7.31. Let $f_1, \dots f_8$ be a set of functions defined on the interval $[1, 8]$ with the property that for any numbers $d_1, \dots, d_8$, there exists a set of

coefficients $c_l, \ldots, c_8$ such that

$$\sum_{j=1}^{8} c_j f_j(i) = d_i, \quad i = 1, \ldots, 8.$$

(a) Show that $d_1, \ldots, d_8$ determine $c_1, \ldots, c_8$ uniquely.

(b) Let A be the 8×8 matrix representing the linear mapping from data $d_1, \ldots, d_8$ to coefficients $c_1, \ldots, c_8$. What is the i, j entry of A^{-1}?

Exercise 2.7.32. ☞ Ch. 11

Prove, that if A and B are similar matrices then $\mathrm{Tr}A = \mathrm{Tr}B$.

Exercise 2.7.33. Are the matrices $\begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}$ and $\begin{bmatrix} 0 & 2 \\ 4 & 2 \end{bmatrix}$ similar? Justify.

Exercise 2.7.34. Let $L : \mathbb{R}^3 \to R^3$ be the linear map given in the standard basis by the matrix

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

a) Find the matrix representing L relative to the basis $[1, 1, 1]^t$, $[1, 1, 0]^t$, $[1, 0, 0]^t$ for both the domain and the range.

b) Write down bases for the domain and range with respect to which the matrix of L is the identity.

Exercise 2.7.35. Let Y and Z be subspaces of the finite dimensional vector spaces V and W, respectively. Let $\mathcal{L}(V, W)$ denote the vector space of all linear maps from V to W. Show that $R = \{T \in \mathcal{L}(V, W) : T(Y) \subset Z\}$ is a subspace of $\mathcal{L}(V, W)$. What is the dimension of R?

Exercise 2.7.36. Let T, U, V, W be vector spaces over $\mathbb{F}$, $\mathcal{L}(U, V)$ and $\mathcal{L}(T, W)$ be vector spaces of linear maps from U to V and from T to W, respectively. Let $\alpha : T \to U$ and $\beta : V \to W$ be fixed linear maps. Show that the mapping $\Phi : \mathcal{L}(U, V) \to \mathcal{L}(T, W)$ which sends θ to $\beta \circ \theta \circ \alpha$ is linear. If the spaces are finite-dimensional and α and β have rank r and s, respectively, find the rank of Φ.

Exercise 2.7.37. Let u and v are two vectors in $\mathbb{R}^n$. The matrix $A = I + uv^t$ is known as a **rank-one perturbation of the identity**. Show that if A is nonsingular (that is, if it has an inverse), then its inverse has the form $A^{-1} = I + \alpha uv^t$ for some scalar α and give an expression for α. For what u and v is A singular? If it is singular, what is $\ker(A)$?

Chapter 3

Determinants

Reading for this Chapter

- Treil: Chapter 3.

3.1 Definitions

Consider an $n \times n$ matrix A with columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. We might be interested in the volume of the parallelepiped (or, in simpler terms, the box) spanned by vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$.

However, one difficulty is that this function is somewhat complicated. if we call this function $\mathrm{vol}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$, then the equality

$$\mathrm{vol}(\boldsymbol{a}_1 + \boldsymbol{a}_1', \ldots, \boldsymbol{a}_n) = \mathrm{vol}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) + \mathrm{vol}(\boldsymbol{a}_1', \ldots, \boldsymbol{a}_n) \qquad (3.1)$$

sometimes holds and sometimes not: for example, volume on the left is zero if $\boldsymbol{a}_1' = -\boldsymbol{a}_1$ and the volumes on the right are (almost always) positive.

For this and other reasons, it is useful to define the signed volume of the parallelepiped. The absolute value of the signed volume equals the regular volume and its sign is determined by the **orientation** of the system of vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. We will not define the orientation rigorously but only note that it can be either positive or negative, and the orientation preserved by rotations but changes sign after a reflection.

We denote this signed volume by $\mathrm{Vol}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$. In particular $\mathrm{Vol}(v_1, v_2) = -\mathrm{Vol}(v_2, v_1)$ and $\mathrm{Vol}(-v_1, v_2) = -\mathrm{Vol}(v_1, v_2)$.

The geometric definition of the determinant of matrix A is that

$$\det(A) = \mathrm{Vol}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n).$$

This definition is a bit unsatisfactory since it depends on the definitions of volume and orientation and so it is not purely algebraic. It also does not explain how to compute the determinant.

The second definition is axiomatic. The determinant is a function that maps matrix $A = [a_1, \ldots, a_n] \in F_n^n$ to field F and satisfies the following axioms:

1. For all numbers $c \in F$,

$$\det[c\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n] = c \det[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n], \tag{3.2}$$

   (More generally, this identity should hold for all c from the field over which we define the matrices.)

2.
$$\det[\boldsymbol{a}_1 + \boldsymbol{a}_1', \ldots, \boldsymbol{a}_n] = \det[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n] + \det[\boldsymbol{a}_1', \ldots, \boldsymbol{a}_n], \tag{3.3}$$

3. For every $1 \leq i < j \leq n$, we have

$$\det[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_i, \ldots, \boldsymbol{a}_j, \ldots, \boldsymbol{a}_n] = -\det[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_j, \ldots, \boldsymbol{a}_i, \ldots, \boldsymbol{a}_n]. \tag{3.4}$$

4. For the basis vectors $\boldsymbol{e}_1 = (1, 0, 0 \ldots, 0)$, $\boldsymbol{e}_2 = (0, 1, 0, \ldots, 0)$, ..., $\boldsymbol{e}_n = (0, 0, 0, \ldots 1)$, we have

$$\det[\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n] = 1 \tag{3.5}$$

This is a nice definition but it is not clear why this function exists. (The signed volume satisfies these axioms but our goal is to avoid using the signed volume in the definition.)

We are going to show the existence of the determinant by writing the function $\det(A)$ explicitly in terms of the entries of A and checking that it satisfies the axioms. Unfortunately, this definition is somewhat cumbersome.

It can perhaps be guessed by considering examples in 2-dimensional and 3-dimensional space. For example, if we have two vectors $v_1 = (a, b)$ and $v_2 = (c, d)$ then one can show that the area of the corresponding parallelogram is $ad - bc$. One can also develop a formula for the volume of 3-dimensional parallelepiped.

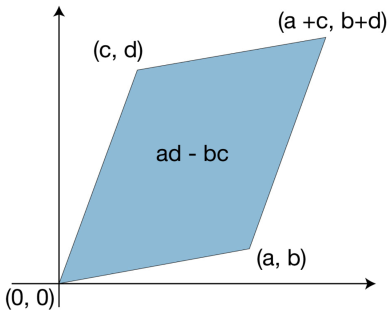

Figure 3.1: Area of a parallelogram in $\mathbb{R}^2$

Alternatively, we can note that the determinant $\det[v_1, \ldots, v_n]$ is a multi-linear function in v (these are properties (3.2) and (3.3) and this function has an additional special property to change sign when we exchange two arguments. This means that it is enough to understand what are the values of the determinant when $v_1, \ldots, v_n$ are basis vectors and then extend the definition by multi-linearity. So, as a first step, we need to understand what are $\det(e_{i_1}, e_{i_2}, \ldots, e_{i_n})$, where all e_{i_k} are basis vectors. Now, if any of these vectors repeats, then the value of the determinant must be zero if we want property (3.4) to be true. (Otherwise we could exchange the repeating vectors and would get a contradiction.) If all e_{i_k} are different then we have to figure out the determinants of the matrices like in this example:

$$\det(e_4, e_3, e_1, e_2) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{3.6}$$

Note that here we have a **permutation** of the basis vectors: e_1 was placed in 3-rd position, e_2 was placed in 4-th position and so on. By exchanging the columns we can undo this permutations and bring the matrix to the identity matrix, which should have determinant 1 by property (3.5). So, using the property (3.4), we find that

$$\det(e_{i_1}, e_{i_2}, \ldots, e_{i_n}) = (-1)^{l(i_1, i_2, \ldots i_n)},$$

where $l(i_1, i_2, \ldots i_n)$ is the number of times we need to exchange the columns to get to the identity matrix.

For an example, let us do this process for two-by-two matrices. First

$$\det(e_1, e_2) = \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 1,$$
$$\det(e_2, e_1) = \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -1.$$

Then

$$\begin{aligned}\det\begin{bmatrix}a & b\\ c & d\end{bmatrix} &= \det(ae_1+be_2, ce_1+de_2)\\ &= a\det(e_1, ce_1+de_2)+b\det(e_2, ce_1+de_2)\\ &= ac\det(e_1,e_1)+ad\det(e_1,e_2)+bc\det(e_2,e_1)+bd\det(e_2,e_2)\\ &= ad-bc.\end{aligned}$$

In general, for arbitrary n, we obtain the following constructive definition of the determinant.

Definition 3.1.1. Let $A=(a_{ij})$ be an $n\times n$ matrix, then

$$\det(A)=\sum_{\pi\in S_n}\varepsilon(\pi)a_{1\pi(1)}a_{2\pi(2)}\dots a_{n\pi(n)}. \tag{3.7}$$

Here the sum is over all permutations of the set $\{1,2,\dots,n\}$ and $\varepsilon(\pi)$ is defined below.

A permutation is the bijective mapping of this set to itself. For example, we can define a permutation of the set $\{1,2,3,4\}$ by setting $\pi(1)=3,\pi(2)=4,\pi(3)=2,\pi(4)=1$. This permutation can also be written in two-line notation:

$$\pi=\begin{matrix}1 & 2 & 3 & 4\\ 3 & 4 & 2 & 1\end{matrix}$$

or simply in one-line notation 3421 (since the first line is always the same). For each permutation, we can define its length $l(\pi)$ as the minimal number of switches of two elements which is needed to bring it to the identity permutation. (In combinatorics such a switch is called a **transposition**.) For example for our transformation $\pi=3421$, we can undo it as follows:

$$3421\xrightarrow{31}1423\xrightarrow{42}1243\xrightarrow{43}1234,$$

so the length of this permutation is three.

Then we define the function $\varepsilon(\pi):=(-1)^{l(\pi)}$, and now our formula (3.7) is well-defined.

Exercise 3.1.2. Let an inversion in a permutation π be a pair $i<j$, such that $\pi(i)>\pi(j)$. Let $inv(\pi)$ be the number of inversions in permutation π. Let (k,l) denote the permutation that switches k and l and leaves all other

elements intact, and let $\tilde{\pi} = (k, l) \circ \pi$ be the composition of (k, l) and π. Then

$$(-1)^{inv(\tilde{\pi})} = -(-1)^{inv(\pi)}$$

[Hint: prove this for (elementary) transpositions $(k, k+1)$, and then show that any transposition (k, l) can be written as a composition of an **odd** number of the transpositions of this elementary type.]

Since the identity permutation has no inversions, this exercise shows that

$$\varepsilon(\pi) = (-1)^{inv(\pi)}.$$

It also shows that the following key property of $\varepsilon(\pi)$ holds. If k is the number of elements in **any** sequence of two-element switches that brings π to the identity transformation (not necessarily the shortest sequence), then $\varepsilon(\pi) = (-1)^k$.

To illustrate formula (3.7), for a 2×2 matrix A we have only two permutations 12 and 21 with lengths 0 and 1 respectively, and the formula for the determinant is

$$\det(A) = a_{11}a_{22} - a_{12}a_{21},$$

with the first term corresponding to the identity permutation 12 and the second to the permutation 21.

If $n = 3$, we have one permutation of length 0: 123, three permutations of length 1: 213, 132, and 321, and two permutations of length 2: 231 and 312. So the formula for the determinant in this case is

$$\begin{aligned}\det(A) = {} & a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ & - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}.\end{aligned}$$

The number of terms in these formulas grows very fast so they are not useful for the actual computation of the determinants except for small matrices.

We give the following theorem without proof, although it should be sufficiently convincing after the motivation that we gave above.

Theorem 3.1.3. *The determinant function* $\det : F_n^n \to F$ *as defined in (3.7) is the unique function that satisfies properties (3.2) – (3.5).*

In particularly, the determinant which we defined previously axiomatically actually exists.

The terms in the definition of the determinant can also be re-organized to give the recursive formula for the determinant in terms of the determinants of sub-matrices.

This is called the **cofactor expansion** of the determinant.

Let A be an $n \times n$ matrix with entries a_{ij}. Let $A^{(ij)}$ be a matrix which is obtained by removing row i and column j. Then the **cofactor**

$$C_{ij} := (-1)^{i+j} \det\left(A^{(ij)}\right).$$

The **cofactor expansion** along the row i is the formula

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij}.$$

For example, for the first row, we have the expansion:

$$\begin{aligned} \det(A) &= a_{11}C_{11} + a_{12}C_{12} + \ldots + a_{1n}C_{1n} \\ &= a_{11} \det\left(A^{(11)}\right) - a_{12} \det\left(A^{(12)}\right) + \ldots + (-1)^{n+1} a_{1n} \det\left(A^{(1n)}\right) \end{aligned}$$

We have also an analogous expansion along column j:

$$\det(A) = \sum_{i=1}^{n} a_{ij} C_{ij}.$$

(So altogether, there are $2n$ different expansions, n along the rows and n along the columns.)

This result can be proved from the basic definition (3.7). In a sense, this is simply a way to organize formula (3.7) as a recursive calculation. We omit the proof.

The cofactor expansion can be used to calculate the determinant recursively but for large matrices this is usually **much slower than by reducing the matrix to the upper-diagonal form**, the method which we describe a bit later.

3.2 Properties of the determinant

One case in which it is easy to calculate the determinant from the definition (3.7) or from a cofactor expansion is the case in which the matrix is either lower-diagonal or upper-diagonal. In this case, it is easy to see that the only non-zero term in the sum in formula (3.7) is the term corresponding to the identity permutation $\pi = 12 \ldots n$. This leads to the following theorem.

Theorem 3.2.1. *If a square matrix A is upper-diagonal, or lower-diagonal, then*

$$\det A = a_{11}a_{22}\dots a_{nn}.$$

Second, this definition allows us to prove an important theorem.

Theorem 3.2.2. *For every square matrix A, we have:*

$$\det(A^t) = \det(A).$$

Proof. For the transposed matrix A^t, we have

$$\begin{aligned}\det(A^t) &= \sum_{\pi\in S_n} \varepsilon(\pi) a_{\pi(1)1}a_{\pi(2)2}\dots a_{\pi(n)n} \\ &= \sum_{\pi\in S_n} \varepsilon(\pi) a_{1\pi^{-1}(1)}a_{2\pi^{-1}(2)}\dots a_{n\pi^{-1}(n)},\end{aligned}$$

where π^{-1} is the inverse permutation to the permutation π. For example, if $\pi = \begin{bmatrix}1 & 2 & 3\\ 3 & 1 & 2\end{bmatrix}$, then $\pi^{-1} = \begin{bmatrix}3 & 1 & 2\\ 1 & 2 & 3\end{bmatrix} = \begin{bmatrix}1 & 2 & 3\\ 2 & 3 & 1\end{bmatrix}$.

It turns out that the length of the inverse permutation π^{-1} equals the length of permutation π. (If $\pi = \tau_l \circ \dots \circ \tau_2 \circ \tau_1$, where τ_i are transpositions, then $\pi^{-1} = \tau_1 \circ \tau_2 \circ \dots \circ \tau_l$)

Hence $\varepsilon(\pi^{-1}) = \varepsilon(\pi)$ and we can continue the formula above as:

$$\det(A^t) = \sum_{\pi\in S_n} \varepsilon(\pi^{-1}) a_{1\pi^{-1}(1)}a_{2\pi^{-1}(2)}\dots a_{n\pi^{-1}(n)}.$$

However if π in this sum are all possible permutations of the set $\{1,\dots,n\}$, then π^{-1} also run over all possible permutations of this set. So it is actually the same sum as in definition of $\det(A)$ and we conclude that $\det(A^t) = \det(A)$. □

In particular, this implies that properties (3.2) - (3.4) hold not only for matrix A with **columns** $a_1,\dots,a_n$ but also for matrix with **rows** $a_1,\dots,a_n$.

Here are useful consequences of these basic properties:

Theorem 3.2.3.

1. *If one row in the matrix is a multiple of another row then the determinant equals 0.*
2. *If we add a multiple of row a_i to any other row a_j then the determinant will not change.*

Proof. For the proof of the first property, without loss of generality let the second row be a multiple of the first row, then

$$\det([a_1; ca_1; \ldots]) = c\det([a_1; a_1; \ldots]) = -c\det([a_1; a_1; \ldots]),$$

where in the last equality we exchanged rows 1 and 2 and used the property (3.4). This implies that the determinant is zero. (We write $\det([a_1; a_2; \ldots; a_n])$ for the determinant of a matrix with rows given by vectors $a_1, a_2, \ldots, a_n$.

For the proof of the second property, again without loss of generality, assume that we added a multiple of the first row to the second row. Then we have:

$$\begin{aligned}\det([a_1; a_2 + ca_1; \ldots]) &= \det([a_1; a_2; \ldots]) + \det([a_1; ca_1; \ldots]) \\ &= \det([a_1; a_2; \ldots]),\end{aligned}$$

which is what we wanted to prove. (The first equality uses property (3.3) and the second one uses the property that we just proved.) □

In particular, this theorem means that if we do Gaussian elimination on matrix A (by adding the multiples of rows above to the rows below but without multiplying the rows by a constant, and without exchanging the rows), then the determinant of the matrix will not change and eventually we will be left with an upper-diagonal matrix U that have the same determinant as the original matrix. Hence, by Theorem 3.2.1 the determinant of A equals the product of the diagonal elements of U.

If we had to exchange the rows, then a more general formula applies:

$$\det(A) = (-1)^r u_{11}u_{22}\ldots u_{nn}, \tag{3.8}$$

where r is the number of times we exchanged the rows, and u_{11}, …, u_{nn} are the pivots, that is the diagonal elements of U.

This property gives an effective method to calculate the determinants.

Example 3.2.4. Calculate the determinant of

$$A = \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$$

in two different ways, by using the cofactor expansion and the Gaussian elimination.

By using the cofactor expansion over the first row, we get:

$$\begin{aligned}\det(A) &= 1 \times \det\begin{bmatrix} 8 & -9 \\ 7 & 0 \end{bmatrix} - (-4) \times \det\begin{bmatrix} -2 & -9 \\ -1 & 0 \end{bmatrix} + 2 \times \det\begin{bmatrix} -2 & 8 \\ -1 & 7 \end{bmatrix} \\ &= 63 - 36 - 12 = 15.\end{aligned}$$

By raw reduction, we find

$$\begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ 0 & 3 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & -5 \end{bmatrix},$$

where we used one exchange of rows. Consequently,

$$\det(A) = (-1)^1 \times 1 \times 3 \times (-5) = 15.$$

For large matrices, the method that uses Gaussian elimination is much more effective than the cofactor expansion method.

Example 3.2.5 (The Vandermonde determinant). One important determinant which pop-ups in many parts of mathematics is the Vandermonde determinant:

$$\det(W) = \det\begin{bmatrix} x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \\ x_1^{n-2} & x_2^{n-2} & \dots & x_n^{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ x_1 & x_2 & \dots & x_n \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

By the definition, this should be a polynomial of variables $x_1, x_2, \dots, x_n$ and that every term in this polynomial has the same total degree. It is clear that the determinant equal to zero if $x_i = x_j$ for some $i \neq j$. So, the polynomial should be divisible by all of the differences $x_i - x_j$. By checking the total degree, it follows that the polynomial is equal to the product of these differences up to the constant term and the by looking on a specific term like $x_1^{n-1} x_2^{n-2} \dots x_{n-1}$ one can find this constant. This results in the following formula:

$$\det(W) = \prod_{i<j} (x_i - x_j).$$

The formula 3.8 for determinant in terms of pivots implies the following important property of the determinants.

Theorem 3.2.6. *A square matrix A is invertible if and only if* $\det(A) \neq 0$.

Proof. The matrix A is invertible if and only if it has full rank, hence, if and only if after row reduction all the variables u_{ii} are valid pivots, that is, $u_{ii} \neq 0$, hence, by formula (3.8) if and only if $\det(A) \neq 0$. □

Another important result is that determinant multiplicative.

Theorem 3.2.7. *Let A and B be two $n \times n$ matrices, then*

$$\det AB = \det A \det B.$$

A sketch of the proof. The argument has two cases. The first case is when A is non-invertible. Then AB is also non-invertible (check it!) and we are done by Theorem 3.2.6.

If A is invertible then $\det(A) \neq 0$, and we define

$$f(B) = \frac{\det(AB)}{\det A}.$$

Then we can check that $f(B)$ satisfies all the axioms. For example, let $b_1, \ldots, b_n$ denote the columns of B, and $b'_1, b_2, \ldots, b_n$ be the columns of a matrix B' Then

$$\begin{aligned} f(b_1 + b'_1, b_2, \ldots, b_n) &= \frac{\det(A(b_1 + b'_1), Ab_2, \ldots, Ab_n)}{\det A} \\ &= \frac{\det(Ab_1, Ab_2, \ldots, Ab_n) + \det(Ab'_1, Ab_2, \ldots, Ab_n)}{\det A} \\ &= f(b_1, b_2, \ldots, b_n) + f(b'_1, b_2, \ldots, b_n). \end{aligned}$$

and the axiom (3.3) is satisfied. For the normalization axiom (3.5), we have $f(I_n) = \frac{\det(AI)}{\det(A)} = 1$. Hence, by Theorem 3.1.3, $f(B) = \det(B)$ and we obtain the identity $\det(AB) = \det A \det B$. □

(For another proof that uses elementary transformation matrices, see Section 3.3.5 in Treil.)

Corollary 3.2.8. *For a non-singular square matrix A, we have:*

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

Corollary 3.2.9. *Suppose $n \times n$ matrix V has columns $v_1, \ldots, v_n$, and let A be another $n \times n$ matrix. Then*

$$\operatorname{Vol}(Av_1, Av_2, \ldots, Av_n) = \det(A) \operatorname{Vol}(v_1, v_2, \ldots v_n).$$

Proof. Since $Av_1, Av_2, \ldots, Av_n$ are columns of the matrix AV, we have

$$\begin{aligned}\mathrm{Vol}(Av_1, Av_2, \ldots, Av_n) &= \det(AV) = \det(A)\det(V) \\ &= \det(A)\,\mathrm{Vol}(v_1, v_2, \ldots v_n).\end{aligned}$$

□

In other words, suppose we have a box with the sides given by vectors $v_1, v_2, \ldots, v_n$ and suppose this box has the oriented volume V, and we apply a linear transformation A to this box. Then this box will be mapped to a box with the oriented volume $\det(A)V$. This gives another interpretation of the determinant: it is a scale factor by which a linear transformation A extends the volume elements.

3.3 Inverse matrix and Cramer formula

A formula for inverse matrix.

Recall that cofactors are defined as

$$C_{ij} := (-1)^{i+j} \det\left(A^{(ij)}\right).$$

We can think about them as entries of a matrix C. Recall also that the (row) **cofactor expansion** along the row i is the formula

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij}.$$

Another use of cofactors is that they provide us with a formula for the inverse matrix.

Theorem 3.3.1. *Let A be a non-singular $n \times n$ matrix. Then*

$$A^{-1} = \frac{1}{\det(A)} C^t,$$

meaning that

$$(A^{-1})_{ij} = \frac{1}{\det(A)} C_{ji}.$$

Sketch of the proof: We need to prove that

$$AC^t = \det(A) I,$$

that is, that

$$\sum_{j=1}^{n} a_{ij} C_{kj} = \det(A)\delta_{ik}.$$

For $k = i$ this is simply the cofactor expansion, while for $k \neq i$, the left-hand side is the cofactor expansion for the determinant of the matrix which is obtained from matrix A by replacing the row k with the row i (and keeping all other rows intact). However, this new matrix has two identical rows and therefore its determinant is 0. □

The formula is important theoretically. However, in numerical computations, the inverse is easier to find by using the Gaussian elimination method.

Example 3.3.2. The inversion formula for 2×2 matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Example 3.3.3. If a matrix A has integer entries and $\det A = 1$, then its inverse is also an integer matrix.

A related result is Cramer's formula for the solution of linear equations.

Theorem 3.3.4. *For an invertible matrix A, the k-th entry of the solution of the equation $Ax = b$ is given by the formula*

$$x_k = \frac{\det B_k}{\det A},$$

where the matrix B_k is obtained from A by replacing column number k of A by the vector b.

Sketch of the proof: The solution is

$$x = A^{-1}b = \frac{1}{\det A} C^t b,$$

so

$$x_k = \frac{1}{\det A} \sum_{i=1}^{n} C_{ik} b_i,$$

and one can identify the sum as the cofactor expansion for the determinant of the matrix B_k along the column k. □

Again, from the practical point of view, this formula is not very useful for calculations. However, from the theoretical viewpoint, it means that the solution of **any** system of equations can be written in terms of determinants.

3.4 Block matrices and advanced properties of determinant

The theory of determinants has some beautiful identities. Here are several of them, most of which we give without proof.

Determinant of a block-diagonal matrix. Suppose A and D are square $k \times k$ and $l \times l$ matrices respectively, and let B is a $k \times l$ matrix. Then we can form a block matrix $\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$ which is a square $(k+l) \times (k+l)$ matrix. (Here 0 denotes a $l \times k$ matrix of zeros.) Then

$$\det \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = \det(A) \det(D).$$

Schur's identity is a generalization of this formula. Suppose A is square and invertible and B, C and D are such that the matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is square, then

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A) \det(D - CA^{-1}B)$$

Silvester's determinantal identity. Let A and B be $m \times n$ and $n \times m$ matrices, respectively. Then

$$\det(I_m + AB) = \det(I_n + BA)$$

The Cauchy - Binet formula. The Cauchy-Binet formula allows one to calculate the determinant of AB if A and B are not square. So, it is a generalization of the product formula for the determinant. Suppose A is $m \times n$ and B is $n \times m$ and assume that $m \leq n$ (otherwise, it is easy to show that $\det(AB) = 0$. Then one has the formula:

$$\det(AB) = \sum_S \det(A(:, S)) \det(B(S, :))$$

where the sum is over all m-element subsets S of the set $\{1, \ldots, n\}$, $A(:, S)$ is an $m \times m$ matrix whose columns are the columns of A at indices from S, and $B(S, :)$ is an $m \times m$ matrix whose rows are the rows of A at indices from S.

3.5 Exercises

Part 1.

Basics.

Exercise 3.5.1 (Exercise 3.7.1 in Treil).

1. Determinant is only defined for square matrices.
2. If two rows or columns of A are identical, then $\det A = 0$.
3. If B is the matrix obtained from A by interchanging two rows (or columns), then $\det B = \det A$.
4. If B is the matrix obtained from A by multiplying a row (column) of A by a scalar α, then $\det B = \det A$.
5. If B is the matrix obtained from A by adding a multiple of a row to some other row, then $\det B = \det A$.
6. The determinant of a triangular matrix is the product of its diagonal entries.
7. $\det(A^t) = -det(A)$.
8. $\det(AB) = \det(A)\det(B)$.
9. A matrix A is invertible if and only if $\det A = 0$.
10. If A is an invertible matrix, then $\det(A^{-1}) = 1/\det(A)$.

Exercise 3.5.2 (Exercise 3.7.2 in Treil). Let A be an $n \times n$ matrix. How are $\det(3A)$, $\det(-A)$ and $\det(A^2)$ related to $\det A$.

Exercise 3.5.3. Use a determinant to identify all values of t and k such that the following matrix is singular (i.e., noninvertible). Assume that t and k must be real numbers.

$$A = \begin{bmatrix} 0 & 1 & t \\ -3 & 10 & 0 \\ 0 & 5 & k \end{bmatrix}$$

Exercise 3.5.4. Let $A = [\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}]$ be a 4×4 matrix whose determinant is equal to 2. What is the determinant of $B = [\boldsymbol{d}, \boldsymbol{b}, 3\boldsymbol{c}, \boldsymbol{a} + \boldsymbol{b}]$? Explain.

Part 2.

Exercise 3.5.5. By applying row operations to produce an upper triangular U, compute the following determinants:

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

2.
$$A = \begin{bmatrix} 1 & t & t^2 & t^3 \\ t & 1 & t & t^2 \\ t^2 & t & 1 & t \\ t^3 & t^2 & t & 1 \end{bmatrix}$$

Exercise 3.5.6. True or false, with reason if true and counterexample if false:

1. If A and B are identical except that $b_{11} = 2a_{11}$, then $\det(B) = 2\det(A)$.
2. The determinant is the product of the pivots in the rref(A).
3. If A is invertible and B is singular, then $A + B$ is invertible.
4. If A is invertible and B is singular, then AB is singular.
5. The determinant of $AB - BA$ is zero.

Part 3.

Additional Exercises.

Exercise 3.5.7. Find the determinant of the $n \times n$ matrix $A = I + J$ where I is the identity matrix and J is the matrix with all entries equal to 1. (In other words, the diagonal entries of A equal 2 and off-diagonal entries are all equal 1.)

Can you find the determinant of the $n \times n$ matrix $A = tI + J$, that has all diagonal entries equal to $t + 1$ and all off-diagonal entries equal to 1?

Exercise 3.5.8 (3.5.3 in Treil). For the $n \times n$ matrix

$$A = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & a_0 \\ -1 & 0 & 0 & \dots & 0 & a_1 \\ 0 & -1 & 0 & \dots & 0 & a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & a_{n-2} \\ 0 & 0 & 0 & \dots & -1 & a_{n-1} \end{bmatrix}$$

compute $\det(a + tI_n)$, where I_n is an $n \times n$ identity matrix. You should get a nice expression involving $a_0, a_1, \dots, a_{n-1}$ and t. Row expansion and induction is probably the best way to go.

Exercise 3.5.9. Let D_n be the determinant of the $n \times n$ tri-diagonal matrix

$$\begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 1 & -1 & & 0 \\ 0 & 1 & 1 & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Using cofactor expansion show that $D_n = D_{n-1} + D_{n-2}$. This yields that the sequence D_n is the Fibonacci sequence $1, 2, 3, 5, 8, 13, 21, \dots$.

Chapter 4

Eigenvalues and Eigenvectors

Reading for this Chapter: Treil, Chapter 4.

4.1 Definitions

The main goal of the theory of eigenvalues and eigenvectors is to determine a basis in which the matrix of a transformation has the simplest possible form.

For this theory, some knowledge of complex numbers is unavoidable. Some basics are summarized in appendix.

Definition 4.1.1. Let A be a $n \times n$ matrix with entries in F. The number λ is an **eigenvalue** of A if there exists a non-zero vector $v \in F^n$ such that $Av = \lambda v$. Any such vector is called an **eigenvector** of an A that belongs to eigenvalue λ.

The set of all eigenvalues of a matrix A is called the **spectrum** of the matrix A.

The set of all eigenvectors belonging to eigenvalue λ (together with the zero vector) is a linear space which is called an **eigenspace**. We denote it by E_λ. The dimension of this eigenspace is called the **geometric multiplicity** of λ, $m_g(\lambda) = \dim E_\lambda$.

It is easy to see that two equivalent conditions for λ to be an eigenvalue are

1. $\dim \ker(A - \lambda I) > 0$,
2. $\det(A - \lambda I) = 0$.

Also note that the eigenspace $E_\lambda = \ker(A - \lambda I)$ and so it is easy to calculate once we know that λ is an eigenvalue. In particular, it is easy to calculate the geometric multiplicity $m_g(\lambda)$ as the dimension of this space.

Then in principle one can find all eigenvalues λ by noting that $p_A(t) = \det(A - tI)$ is a polynomial of the n-th degree in t, and $\lambda \in F$ is an eigenvalue if and only if it is a root of this polynomial. Hence, we only need to determine all the roots of this polynomial in the field F. In practice, this is a difficult task if n is large, and typically one uses other methods to find the eigenvalues, which rely on modern computer algorithms.

Definition 4.1.2. The polynomial $p_A(t) = \det(A - tI)$ is called the **characteristic polynomial** of the matrix A.

(Sometimes we will use an alternative definition $p_A(t) = \det(tI - A)$ which might differ by a sign. This definition has the benefit that then the characteristic polynomial is always monic, that is, the coefficient before the monomial of the largest degree, t^n, equals 1.)

Note that if λ is a root of $p_A(t)$, then by a theorem from algebra, $p_A(t)$ is divisible by $(t - \lambda)$. Moreover, we can write

$$p_A(t) = (t - \lambda)^m f(t), \quad m \geq 1$$

where $f(t)$ is a polynomial such that $f(\lambda) \neq 0$. Then m in this representation is called the **algebraic multiplicity** of λ, $m_a(\lambda)$. We will see later that $m_g(\lambda) \leq m_a(\lambda)$. They are typically the same. If for some λ, $m_g(\lambda) < m_a(\lambda)$, the matrix is called **defective**.

Example 4.1.3. Here is an example that the geometric and algebraic multiplicities of an eigenvalue can be different. Consider matrices

$$A = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 1 & \\ & 2 & 1 \\ & & 2 \end{bmatrix}$$

The characteristic polynomial for both matrices is $p(z) = (z - 2)^3$, so the only eigenvalue is $\lambda = 2$ and it has the algebraic multiplicity 3 for both matrices. However it is easy to check that the eigenspace of $\lambda = 2$ is the whole space $\mathbb{R}^3$ in case of matrix A, and the line spanned by the vector $e_1 = (1, 0, 0)$ in case of matrix B.

The algebraic multiplicities are also easy to calculate. It can be found either by repeated division of the polynomial $p_A(t)$ by $t - \lambda$ or by using the following result.

Exercise 4.1.4. Suppose λ is an eigenvalue of A. Show that $m_a(\lambda)$ equals to the smallest $k \geq 1$ such that

$$\frac{d^k}{dt^k} p_A(t)\Big|_{t=\lambda} \neq 0.$$

Example 4.1.5 (Rotation matrix). What are eigenvalues of matrix

$$R_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix},$$

where $\theta \in [0, 2\pi)$

We have

$$p_{R_\theta}(t) = \det \begin{bmatrix} t - \cos\theta & \sin\theta \\ -\sin\theta & t - \cos\theta \end{bmatrix} = (t - \cos\theta)^2 + \sin\theta^2 = t^2 - 2(\cos\theta)t + 1.$$

The discriminant of this quadratic polynomial is $D = \cos^2\theta - 1 \leq 0$ and so it has real roots only if $\theta = 0 or \pi$, in which case they are either both 1 or both -1, respectively. Hence in the case $\theta = 0$ the eigenvalue is 1 with algebraic and geometric multiplicities 2, and in the case $\theta = \pi$, the eigenvalue is -1, with $m_g(-1) = m_a(-1) = 2$.

For other values of θ, there are no real roots, and we conclude that this matrix has no real eigenvalues over the field $\mathbb{R}$. On the other hand, the characteristic polynomial for this matrix always has roots over field $\mathbb{C}$,

$$\lambda_{1,2} = \cos\theta \pm i\sqrt{1 - \cos^2\theta} = \cos\theta \pm i\sin\theta = e^{\pm i\theta},$$

where $i = \sqrt{-1}$ denotes the imaginary unit. In particular, the characteristic polynomial factorizes as

$$p_{R_\theta}(t) = (t - e^{i\theta})(t + e^{i\theta}),$$

and we see that the both algebraic and geometric multiplicities are 1.

Note that this is the first time when we find that the field F matters.

Example 4.1.6. Find the characteristic polynomial, eigenvalues, their multiplicities, and eigenvectors of the following matrices, considered over field $\mathbb{C}$

$$\begin{bmatrix} 1 & 2 \\ 8 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}.$$

Answers: For the first example: $p(t) = (t-1)^2 - 16$, $\lambda_{1,2} = -3, 5$, $x_1 = [1, 2]^t$, $x_2 = [1, -2]^t$.

For the second example : $p(t) = (t-1)^2 + 4$, $\lambda_{1,2} = 1 \pm 2i$, $x_1 = [1, i]^t$, $x_2 = [1, -i]^t$.

4.2 Similarity and diagonalization

Suppose that a linear transformation L has matrix A in the standard basis. Now we consider another basis $v_1, \ldots, v_n$ and $B = [v_1, \ldots, v_n]$ be a matrix whose columns are $v_1, \ldots, v_n$.

We know that the linear transformation L has the matrix $B^{-1}AB$ in the new basis. What happens with eigenvalues and eigenspaces?

Theorem 4.2.1. *Suppose B is invertible, then A and $B^{-1}AB$ have the same characteristic polynomial, eigenvalues, and algebraic and geometric multiplicities of the eigenvalues.*

In particular, we find that all of these quantities are properties of the linear transformation represented by A rather than of the matrix itself. They remain the same in every basis.

Proof. First we show that the characteristic polynomials are the same, by using properties of the determinant:

$$\begin{aligned} p_{B^{-1}AB}(t) &= \det\left(zI - B^{-1}AB\right) = \det\left(B^{-1}(tI - A)B\right) \\ &= \det\left(B^{-1}\right)\det(tI - A)\det\left(B\right) \\ &= \det(tI - A) = p_A(t). \end{aligned}$$

The equality of the characteristic polynomials implies that the eigenvalues and its algebraic multiplicities are the same for A and $B^{-1}AB$.

In order to show that the geometric multiplicities agree, it is easy to check that if E_λ is an eigenspace for A, then $B^{-1}E_\lambda$ is an eigenspace for $B^{-1}AB$ corresponding to eigenvalue λ, and conversely. In addition, these subspaces are bijectively mapped on each other by the linear transformation corresponding to B^{-1}, so they are isomorphic and have the same dimension. □

This result has an important consequence.

Theorem 4.2.2. *The algebraic multiplicity of an eigenvalue λ is at least as great as its geometric multiplicity, $m_a(\lambda) \geq m_g(\lambda)$.*

Proof. (See also exercises 4.1.7 - 4.1.9 in Treil's book.)

Let k be the geometric multiplicity of λ for matrix A, and let $\{v_1, \ldots, v_k\}$ be a basis for E_λ. Complete this basis to the full basis in F^n and let B the matrix with the columns given by vectors $v_1, \ldots, v_n$.

Then, one has

$$\tilde{A} = B^{-1}AB = \begin{bmatrix} \lambda I_k & * \\ 0_{(n-k)\times k} & D \end{bmatrix},$$

(Check this!) Then, by using the properties of determinant for block matrices, one calculates the characteristic polynomial

$$\begin{aligned} p_{\tilde{A}} = \det(zI_n - \tilde{A}) &= \det(zI_k - \lambda I_k)\det(zI_{n-k} - D) \\ &= (z-\lambda)^k p_D(\lambda). \end{aligned}$$

Therefore the algebraic multiplicity of λ as eigenvalue of $\tilde{A}$ is at least k. Since $\tilde{A}$ is similar to A, it has the same algebraic multiplicity for λ as A, and so $m_g(\lambda) = k \leq m_a(\lambda)$. □

Diagonalization

The importance of eigenvectors and eigenvalues stems from the following observation. Suppose that we have a basis $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of $\mathbb{C}^m$ that consists from eigenvectors of matrix A. Then we have

$$\begin{aligned} A\begin{bmatrix} | & | & \ldots & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_n \\ | & | & \ldots & | \end{bmatrix} &= \begin{bmatrix} | & | & \ldots & | \\ \lambda_1\boldsymbol{x}_1 & \lambda_2\boldsymbol{x}_2 & \ldots & \lambda_n\boldsymbol{x}_n \\ | & | & \ldots & | \end{bmatrix} \\ &= \begin{bmatrix} | & | & \ldots & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_n \\ | & | & \ldots & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}, \end{aligned}$$

or

$$AX = X\Lambda,$$

where X is the matrix with columns x_1, ..., x_n, and Λ is the diagonal matrix with diagonal entries equal to eigenvalues $\lambda_1, \ldots, \lambda_n$. Since $x_1, \ldots, x_n$ is a basis, the matrix X is full rank and therefore, it is invertible. So,

$$A = X\Lambda X^{-1},$$

This factorization of matrix A is called the **eigenvalue diagonalization** of matrix A.

Example 4.2.3. For matrices from Example 4.1.6, we have:

$$\begin{bmatrix} 1 & 2 \\ 8 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix}^{-1},$$

$$\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} 1+2i & 0 \\ 0 & 1-2i \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^{-1}.$$

Intuitively, in the basis of eigenvectors, the linear transformation A looks as a stretch in the directions given by eigenvectors by factors given by the eigenvalues.

Of course, a rotation is difficult to imagine like a stretch transformation. Indeed, the eigenvalue diagonalization of a rotation matrix is impossible over the real numbers, since there are no real eigenvectors. However, it is possible over the complex numbers.

However, even over complex numbers, the diagonalization is not always possible. This bad situation occurs if there are not enough eigenvectors to form a basis. So, our next task is to study when the diagonalization is possible.

Theorem 4.2.4. *An $n \times n$ matrix A is diagonalizable if and only if it is non-defective, that is, if for every eigenvalue $\lambda \in \sigma(A)$, $m_g(\lambda) = m_a(\lambda)$.*

Sketch of the proof of Theorem 4.2.4. If matrix A is diagonalizable then $A = X\Lambda X^{-1}$, so it is similar to a diagonal matrix Λ and hence has same eigenvalues with same multiplicities. It is easy to check that a diagonal matrix is non-defective, so Λ is non-defective and the same holds for A.

In the converse direction, assume that the matrix A is non-defective. Then the dimension of each eigenspace equals to the algebraic multiplicity of the corresponding eigenvalue. Hence the sum of the dimensions of these eigenspaces equals n. If we choose a basis in each of these eigenspaces, and combine the bases together then we obtain the set of n linearly independent eigenvectors [This is the place where a more accurate argument is needed.] If these m independent eigenvectors are formed into the columns of a matrix X, then X is nonsingular and we have $A = X\Lambda X^{-1}$. □

Lemma 4.2.5. *Let $\lambda_1, \ldots, \lambda_k$ be some distinct eigenvalues of A and $u_i \in E_{\lambda_i}$ for every $i = 1, \ldots, k$. Then $u_1 + \ldots + u_k = \vec{0}$ implies that $u_1 = u_2 = \ldots = u_k = \vec{0}$.*

Remark. We will see later that the claim of this lemma means that the sum $E_{\lambda_1} + \ldots + E_{\lambda_k}$ is direct.

Proof. The claim is obviously holds for $k = 1$. Suppose that we can find $k \geq 2$ for which the claim of the lemma is false. Take the smallest of these k. Then we can find k eigenvalues of A and $u_i \in E_{\lambda_i}$ such that $u_1 + \ldots + u_k = 0$ and **all** of $u_i \neq 0$. (Otherwise we could remove the zero u_i and get a smaller k.) Then $Au_1 + \ldots Au_k = 0$ which means that $\lambda_1 u_1 + \ldots + \lambda_k u_k = 0$. Multiply the first equality by λ_1 and subtract it from the second. We get

$$(\lambda_2 - \lambda_1)u_2 + \ldots + (\lambda_k - \lambda_1)u_k = 0.$$

However the vectors $\tilde{u}_i = (\lambda_i - \lambda_1)u_i \in E_{\lambda_i}$ and they are all non-zero. Hence, the claim of the lemma fails for $k - 1$ vectors in contradiction to the assumption that k was the smallest number of such vectors. This contradiction proves the lemma. □

Exercise 4.2.6. By using Lemma 4.2.5, show that if $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of A and $\mathcal{B}_i$ is a basis in E_{λ_i} then the vectors in $\bigcup_{i=1}^{k} \mathcal{B}_i$ are linearly independent. Show that this result completes the argument in the proof of Theorem 4.2.4.

The condition in the theorem can be checked by calculating the multiplicities for each eigenvalue. Indeed, $m_g(\lambda) = \dim \ker(A - \lambda I)$ and $m_a(\lambda)$ can be calculated from the characteristic polynomial. Verification that the matrix is non-defective usually takes some time. A simpler sufficient condition for diagonalizability is that all eigenvalues are distinct.

Theorem 4.2.7. *If $n \times n$ matrix A has n distinct eigenvalues then A is diagonalizable.*

Proof. This theorem is a direct consequence of Theorem 4.2.4, since the condition implies that $m_a(\lambda) = 1$ for each λ and we know that $1 \leq m_g(\lambda) \leq m_a(\lambda)$ by Theorem 4.2.2. □

What happens if a matrix is non-diagonalizable? We will discuss it in the next section. Briefly, in this case it is possible to show that there is a matrix X such that $X^{-1}AX$ has a Jordan form. In this form the matrix is block-diagonal and every block has the form

$$B = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ \ldots & \ldots & \ldots & \ldots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}$$

(The block can be 1×1 with only λ inside it.)

4.3 The determinant and trace of A and eigenvalues

Theorem 4.3.1. *The determinant and the trace of a matrix A are equal to the product and the sum of the eigenvalues of A, respectively, where the eigenvalues are enter the product and sum with their algebraic multiplicities.*

Proof. For the determinant we evaluate the characteristic polynomial at $t = 0$,

$$p_A(0) = \prod_{\lambda_i \in \sigma(A)} (\lambda_i - 0)^{m_a(\lambda_i)} = \det(A - 0 \times I) = \det(A).$$

For the trace, use the combinatorial definition of the determinant to see that

$$p_A(t) \det(tI - A) = \prod_{i=1}^{n} (t - a_{ii}) + g(t),$$

where the degree of polynomial $g(t)$ is $\leq n-2$ (verify it by using the cofactor expansion along the first column!). Then we have

$$p_A(t) = t^n - \Big(\sum_{i=1}^{n} a_{ii} \Big) t^{n-1} + \ldots,$$

where the remainder has degree $\leq n - 2$.

On the other hand, expanding

$$p_A(t) = \prod_{i=1}^{n} (t - \lambda_i) = t^n - \Big(\sum_{i=1}^{n} \lambda_i \Big) + \ldots,$$

we find that $\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii} = \operatorname{tr} A$.

□

4.4 Functions of matrices

If a matrix is not defective, then we have enough linearly independent eigenvectors $x_1, \ldots, x_n$ to build a full-rank square matrix $X = [x_1, x_2, \ldots, x_n]$. Then we have the diagonalization formula:

$$A = X \Lambda X^{-1},$$

where Λ is the diagonal matrix with the eigenvalues on the main diagonal.

This formula can be useful to compute functions of matrix A. For example if we want to calculate the k-th power of matrix A, A^k, then this formula gives us:

$$A^k = X\Lambda^k X^{-1} = X \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \\ 0 & 0 & \dots & \lambda_n^k \end{bmatrix} X^{-1}$$

Similarly, if we have a polynomial $p(z) = \sum_{k=0}^{K} c_k z^k$, then

$$p(A) := \sum_{k=0}^{K} c_k A^k = X p(\Lambda) X^{-1} = X \begin{bmatrix} p(\lambda_1) & 0 & \dots & 0 \\ 0 & p(\lambda_2) & \dots & 0 \\ 0 & 0 & \dots & p(\lambda_n) \end{bmatrix} X^{-1}$$

More generally, this formula is valid for convergent power series and for all functions that can be written as convergent power series. For example,

$$e^A := \sum_{k=0}^{K} \frac{1}{k!} A^k = X e^{\Lambda} X^{-1} = X \begin{bmatrix} e_1^{\lambda} & 0 & \dots & 0 \\ 0 & e_2^{\lambda} & \dots & 0 \\ 0 & 0 & \dots & e_n^{\lambda} \end{bmatrix} X^{-1}$$

Even more generally, this construction can be extended to all functions that can be approximated by polynomials, so eventually we get a very wide class of all measurable functions.

Example 4.4.1. Let

$$A = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$$

Find A^{2022} by diagonalizing A.

The characteristic polynomial is $p(z) = (z-4)(z-2) - 3 = z^2 - 6z + 5 = (z-1)(z-5)$. So, the eigenvalues are 1 and 5 and the corresponding eigenvectors are $[1, -1]^t$ and $[3, 1]^t$. So the diagonalization is

$$A = X \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} X^{-1},$$

where

$$X = \begin{bmatrix} 3 & 1 \\ 1 & -1 \end{bmatrix} \text{ and } X^{-1} = -\frac{1}{4} \begin{bmatrix} -1 & -1 \\ -1 & 3 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & -3 \end{bmatrix}$$

It follows that

$$\begin{aligned} A^{2022} &= \frac{1}{4}\begin{bmatrix} 3 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 5^{2022} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -3 \end{bmatrix} \\ &= \frac{1}{4}\begin{bmatrix} 3 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 5^{2022} & 5^{2022} \\ 1 & -3 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 3\times 5^{2022}+1 & 3\times 5^{2022}-3 \\ 5^{2022}-1 & 5^{2022}+3 \end{bmatrix} \\ &\approx 5^{2022}\frac{1}{4}\begin{bmatrix} 3 & 3 \\ 1 & 1 \end{bmatrix}. \end{aligned}$$

Note that we got a matrix with the columns which are very close to a multiple of the eigenvector of the largest eigenvalue 5. This observations generalizes to other matrices (under some condition) and can be used to find the largest eigenvalue of a matrix. (Or rather, the eigenvalue with the largest absolute value.)

4.5 Applications

4.5.1 Difference equations

Reading: Section 5.3 in Strang's book

A one-dimensional difference equation has the form

$$x_n = c_1 x_{n-1} + c_2 x_{n-2} + \ldots + c_k x_{n-k}$$

Here x_n is a sequence of numbers. We are given some initial conditions $x_{k-1}, x_{k-2} \ldots x_0$ or $x_0, x_{-1}, \ldots, x_{-(k-1)}$ and look to find what is the behavior of x_n for large n.

This equation can be written as the matrix equation if we introduce k-vectors $x^{(n)} = [x_n, x_{n-1}, \ldots, x_{n-k+1}]^*$ and matrix

$$A = \begin{bmatrix} c_1 & c_2 & \ldots & c_k \\ 1 & 0 & 0\ldots 0 & 0 \\ 0 & 1 & 0\ldots 0 & 0 \\ 0 & 0 & 1\ldots 0 & 0 \\ 0 & 0 & 0\ldots 1 & 0 \end{bmatrix}$$

Then we can write the difference equation in the form

$$x^{(n)} = Ax^{(n-1)}. \tag{4.1}$$

The solution of this equation is $x^{(s)} = A^s x^{(0)}$. Hence if we want to know the behavior of the sequence x_n for large n we need to know the behavior of powers of the matrix A^s.

If we can diagonalize the matrix A then we have

$$\begin{aligned} A &= X\Lambda X^{-1}, \\ A^s &= X\Lambda^s X^{-1} \end{aligned} \tag{4.2}$$

If we know both Λ and the matrix of eigenvectors X we can write an explicit formula for $x^{(n)}$. In fact, we often don't need to calculate the matrix X because formula (4.2) implies that we can write the solution as

$$x_n = \sum_{i=1}^{k} a_i \lambda_i^n, \tag{4.3}$$

where λ_i are eigenvalues of matrix A and a_i are some coefficients which can be calculated from the initial conditions.

In addition, this formula often allows us to find the asymptotic behavior of x_n. Suppose λ_1 is an eigenvalue of A that has the largest absolute value: $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_k|$. If in addition, we assume that $a_1 \neq 0$, then we have $x_n \sim a_1 \lambda_1^n$. In particular, if $|\lambda_1| < 1$ then the sequence declines to zero, and if $|\lambda_1| > 1$ then the sequence grows unboundedly.

Remark: It can be proved that the characteristic polynomial of A is

$$p_A(z) = z^k - c_1 z^{k-1} - \ldots - c_{k-1} z - c_k, \tag{4.4}$$

so the eigenvalues of A are roots of this polynomial, and our method is found to be equivalent to a popular method of solving difference equations. Namely, solve the characteristic equation (4.4) and then find the coefficients in (4.3) from the initial conditions.

Many other dynamic problems in biology, engineering and physics can be cast in the form (4.1) with $x^{(k)}$ that describe the state of a system at time k, and A that describe the evolution of the state. In this case, the stability of the system depends on the size of the eigenvalue with the largest absolute value.

Example 4.5.1 (Fibonacci numbers). A classic example for this concept is the Fibonacci numbers, which are defined by the relation:

$$F_n = F_{n-1} + F_{n-2}.$$

and the initial condition $F_1 = F_2 = 1$. Then we can define vector $\boldsymbol{f}_n = (F_{n+1}, F_n)^t$, with $\boldsymbol{f}_0 = (1, 0)^t$, and

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}^{-1},$$

where $\lambda_1 = (1 + \sqrt{5})/2$ and $\lambda_2 = (1 - \sqrt{5})/2$ are eigenvalues of matrix A. Then,

$$A^n = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}^{-1},$$

One can use this formula to calculate F_n. Alternatively, we can note that this formula implies that $F_n = a\lambda_1^n + b\lambda_2^n$, where a and b are some coefficients that do not depend on n. We can find a and b from equations $F_0 = a + b$ and $F_1 = a\lambda_1 + b\lambda_2$. The advantage of this method is that we do not need to calculate the matrix of eigenvectors X and the inverse matrix X^{-1}.

After some calculation, we can get:

$$F_n = \frac{1}{\sqrt{5}}\Big(\lambda_1^n - \lambda_2^n\Big).$$

Since $|\lambda_1| > |\lambda_2|$ we find that

$$F_n \sim \frac{1}{\sqrt{5}}\Big(\frac{1 + \sqrt{5}}{2}\Big)^n$$

For example, $F_{30} = 832,040$ and the right hand side is $832,040 + 2.4063 \times 10^{-7}$.

4.5.2 Linear differential equations

See Strang 5.4.

If we have a system of linear differential equations $x'(t) = Ax(t)$, where x is a vector with k components, when diagonalization of A decouples this system of equation. We get the formula: $x'(t) = S\Lambda S^{-1}x(t)$, where Λ is the diagonal matrix with eigenvalues of A on the main diagonal.

Let $u(t) = S^{-1}x(t)$. Each equation in the resulting system has the form $u_i'(t) = \lambda_i u_i(t)$ so its solution is $u_i(t) = e^{\lambda_i t}u_i(0)$. In vector form: $u(t) = e^{\Lambda t}u(0)$.

Therefore, the solution of the original system is

$$x(t) = Se^{\Lambda t}S^{-1}x(0) = e^{At}x(0).$$

The differential equations of higher order can be solved by a similar approach by first converting them to a system of equations.

For example, if we have a system:

$$y'' = c_1y' + c_2y,$$

then we can convert it to a system by setting $x_1(t) = y(t)$ and $x_2(t) = y'(t)$. Then we have

$$\begin{aligned} x_1'(t) &= x_2(t) \\ x_2'(t) &= c_2x_1(t) + c_1x_2(t), \end{aligned}$$

or in matrix form

$$x'(t) = \begin{bmatrix} 0 & 1 \\ c_2 & c_1 \end{bmatrix} x(t)$$

It turns out that diagonalization of the matrix A in this case equivalent to the standard method of solving these equations: find the roots of the polynomial $z^2 - c_1z - c_2 = 0$. If the roots are λ_1 and λ_2 then the general solution is

$$y(t) = a_1e^{\lambda_1 t} + a_2e^{\lambda_2 t},$$

and the coefficients a_1 and a_2 can be found by fitting the initial conditions $y(0)$ and $y(1)$.

4.6 Exercises

Basics

Exercise 4.6.1 (4.1.1 in Treil). True or false:

a) Every linear operator in an n-dimensional vector space has n distinct eigenvalues;

b) If a matrix has one eigenvector, it has infinitely many eigenvectors;

c) There exists a square real matrix with no real eigenvalues;

d) There exists a square real matrix with no (complex) eigenvectors;

e) Similar matrices always have the same eigenvalues;

f) Similar matrices always have the same eigenvectors;

g) A non-zero sum of two eigenvectors of a matrix A is always an eigenvector;

h) A non-zero sum of two eigenvectors of a matrix A corresponding to the same eigenvalue λ is always an eigenvector.

Exercise 4.6.2 (4.2.1 in Treil).

Let A be $n \times n$ matrix. True or false:

a) A^t has the same eigenvalues as A.

b) A^t has the same eigenvectors as A.

c) If A is diagonalizable, then so is A^t.

Justify your conclusions.

Exercise 4.6.3. Consider the 2×2 matrices $A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}$, $A = \begin{bmatrix} 1 & 5 \\ -2 & 3 \end{bmatrix}$. Do the following calculations by hand. (The second matrix involves calculations with complex numbers.)

a. Calculate the eigenvalues of A.

b. If possible, construct matrices P and C such that $A = PCP^{-1}$, where C is diagonal.

Exercise 4.6.4 (4.1.2 in Treil). Find characteristic polynomial, eigenvalues and eigenvectors of the following matrix:

$$\begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

Write this matrix as $A = PCP^{-1}$ with a diagonal C, if possible.

Exercise 4.6.5 (4.2.7 in Treil). Diagonalize the matrix

$$\begin{bmatrix} 2 & 0 & 6 \\ 0 & 2 & 4 \\ 0 & 0 & 4 \end{bmatrix}.$$

Exercise 4.6.6 (4.1.4 in Treil). Compute characteristic polynomials and eigenvalues of the following matrix:

$$\begin{bmatrix} 4 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 2 & 4 & e & 0 \\ 1 & 3 & 1 & 1 \end{bmatrix}$$

Do not expand the characteristic polynomials, leave them as products.

Exercise 4.6.7. a. If $A^2 = I$, what are possible eigenvalues of A?

b. If this A is 2×2 and not I or $-I$, find its trace and determinant.

c. If the first row of this matrix is $(3, -1)$, what is the second row?

Exercise 4.6.8. (a) A 2×2 matrix A satisfies $\mathrm{tr}(A^2) = 5$ and $\mathrm{tr}(A) = 3$ (where $\mathrm{tr}(X)$ denotes the trace of X). Find $\det(A)$.

(b) A 2×2 matrix A has two proportional columns and $\mathrm{tr}(A) = 5$. Find $\mathrm{tr}(A^2)$.

(c) A 2×2 matrix A has $\det(A) = 5$ and positive integer eigenvalues. What is the trace of A?

Additional:

Exercise 4.6.9. For each of the following statements, prove that it is true or give an example to show it is false. Throughout, A is a complex $m \times m$ matrix unless otherwise indicated.

a. If λ is an eigenvalue of A and $\mu \in \mathbb{C}$, then $\lambda - \mu$ is an eigenvalue of $A - \mu I$.

b. If A is real and λ is an eigenvalue of A, then so is $-\lambda$.

c. If A is real and λ is an eigenvalue of A, then so is $\overline{\lambda}$.

d. If λ is an eigenvalue of A and A is non-singular, then λ^{-1} is an eigenvalue of A^{-1}.

e. If all the eigenvalues of A are zero, then $A = 0$.

f. If A is diagonalizable and all its eigenvalues are equal, then A is diagonal.

g. If A is invertible and diagonalizable, then A^{-1} is diagonalizable.

h. Matrices A and A^t have the same eigenvalues.

Proofs:

Exercise 4.6.10 (4.1.6. in Treil). An operator (i.e., a linear map) A is called **nilpotent** if $A^k = 0$ for some k. Prove that if A is nilpotent, then $\sigma(A) = \{0\}$ (i.e. that 0 is the only eigenvalue of A). Prove that a non-zero nilpotent A cannot be diagonalizable.

Exercise 4.6.11. ☞

Suppose A and B are square and A is invertible. Prove that AB and BA have the same eigenvalues.

Exercises 4.1.7 - 4.1.9 in Treil. (These are the details for the proof of the result that the algebraic multiplicity $\geq$ the geometric multiplicity for every eigenvalue.)

Applications:

Exercise 4.6.12. Suppose each "Gibonacci" number G_{k+2} is the average of the two previous numbers G_{k+1} and G_k. Then $G_{k+2} = \frac{1}{2}(G_{k+1} + G_k)$. In matrix form this can be written as

$$\begin{bmatrix} G_{k+2} \\ G_{k+1} \end{bmatrix} = A \begin{bmatrix} G_{k+1} \\ G_k \end{bmatrix}.$$

a. Find the eigenvalues and eigenvectors of A.

b. Find the limit of the matrices A^n as $n \to \infty$.

c. If $G_0 = 0$ and $G_1 = 1$, which number do the Gibonacci numbers approach?

4.7 Appendix: Complex numbers

Complex number is a pair of real numbers (x, y). So, it is essentially a vector in $\mathbb{R}^2$. The addition of complex numbers is the addition of vectors. However, the wonderful fact is that there is also a multiplication operation. This operation has no analogue for vectors in $\mathbb{R}^n$ for general n.

It is easier to remember this operation, if we write complex numbers as $z = x + iy$, in which case the product of two complex number is defined as $z_1 z_2 := x_1 x_2 - y_1 y_2 + i(x_1 y_2 + x_2 y_1)$. This is the same as if we thought about i as a special kind of number with property $i^2 = -1$.

It turns out that this operation is associative and commutative and satisfies the distributive law with respect to the addition.

Formally, we converted the linear space $\mathbb{R}^2$ to a commutative algebra over $\mathbb{R}$.

One important new operation is that of conjugation: $\overline{x+iy} = x - iy$. Note that $z\overline{z} = x^2 + y^2 = \|z\|^2$. In the context of complex numbers, $\|z\|$ is called the **absolute value**, or the **modulus** of z.

Since we have multiplication and addition, we can define polynomials and power series using complex numbers. The convergence for series is defined using the norm in $\mathbb{R}^2$. In particular, we can define the exponential function e^z as the convergent series $\sum_{k=0}^{\infty} z^k/k! = 1 + z + z^2/2! + \ldots$. This function preserves the important property of the standard exponential function:

$$e^{z_1+z_2} = e^{z_1}e^{z_2}.$$

In particular, $e^{x+iy} = e^x e^{iy}$. In addition, directly from definition of e^z, we can obtain

$$e^{iy} = \cos y + i \sin y.$$

Therefore,

$$e^{x+iy} = e^x(\cos y + i \sin y).$$

This formula allows us to give a geometric meaning to the product operation. For this we need to represent the vector $z = (x, y)$ in polar coordinates as $(r\cos\alpha, r\sin\alpha)$. Here $r = \sqrt{x^2+y^2} = \|z\|$, and α is called the **argument** of z.

Then, $z = x + iy = e^{\log r + i\alpha}$. If we have another complex number $z' = x' + y' = e^{\log r' + i\alpha'}$ then the addition formula for the exponential gives us:

$$zz' = e^{\log r + \log r' + i(\alpha+\alpha')} = rr'e^{i(\alpha+\alpha')}.$$

In other words, when we multiply z and z', their absolute values are multiplied, and their arguments are added.

The fundamental theorem of algebra says that every equation of degree n (with coefficient in real or complex numbers) has at least one solution in complex numbers. This can be strengthened to the statement that it has exactly n solutions if we count the solutions with multiplicities. The proof of this remarkable theorem is quite non-trivial.

Chapter 5

Jordan Canonical Form

Reading:
Axler: Chapter 8
Treil: Chapter 9

5.1 Invariant subspaces

Definition 5.1.1. Let $L : V \to V$ be a linear map. A subspace $E \subset V$ is called an **invariant subspace** of L (an L-invariant subspace) if $LE \subset E$, that is, if $Lv \in E$ for all $v \in E$.

We often fix a basis and identify the linear transformation L with its matrix A. Then we say that E is an A-invariant subspace.

It is easy to check that if E is an L-invariant subspace, then E is an invariant subspace for every polynomial $p(L)$.

Example 5.1.2. Of course V is L-invariant. Then any sum of L-invariant subspaces is L invariant. Any intersection of L-invariant subspaces is L-invariant.

Example 5.1.3. Suppose L has an eigenvalue λ with eigenspace E_λ. Then E_λ is an L-invariant subspace.

We define a restriction of L to its invariant subspace E as $L|_E$. If we have $V = E_1 \oplus \ldots \oplus E_t$, and all E_i are L-invariant, then we can choose a basis $\mathcal{B}_i$ in each of them and the union of these bases gives a basis $\mathcal{B}$ of V by Theorem 1.1.30. If A_i is a matrix of $L|_{E_i}$ in basis $\mathcal{B}_i$, then the matrix of

L in basis $\mathcal{B}$ has the following block-diagonal form.

$$A = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A_t \end{bmatrix}$$

In this case we sometimes write $[A] = [A_1] \oplus \ldots \oplus [A_t]$ or $A = \operatorname{diag}(A_1, \ldots, A_t)$.

Example 5.1.4. If the matrix is diagonalizable, then we can put the matrix in diagonal form and than we have $[A] = [A_1] \oplus \ldots \oplus [A_k]$ where k is the number of distinct eigenvalues and every matrix A_i is a diagonal matrix

$$A_i = \begin{bmatrix} \lambda_i & 0 & \dots & 0 \\ 0 & \lambda_i & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_i \end{bmatrix},$$

where A_i is $d_i \times d_i$ and d_i is the geometric multiplicity of λ_i, that is, the dimension of the corresponding eigenspace. In particular, each $[A_i] = [\lambda_i] \oplus \ldots \oplus [\lambda_i]$, where the sum have d_i summands.

In the non-diagonalizable case, V is larger than the sum of eigenspaces so we have to look for other invariant spaces, and in this case the matrices A_i may be more complicated.

Our first task is to look for smallest invariant spaces, so as to make sizes of the blocks A_i as small as possible, and the second task is to make the matrices A_i as simple as possible. In the next section we address the first task.

5.2 Generalized eigenspaces

The main reason why we are not always can diagonalize a matrix is that induction does not work. Even though an eigenspace E_λ is always an A-invariant subspace, we cannot find a complementary A-invariant subspace W so that $\mathbb{R}^n = E_\lambda \oplus W$. (If we could, we would restrict to this smaller subspace W and be done by induction.)This problem boils down to the following issue:

We know from the rank nullity theorem that if $L : V \to V$, then

$$\dim \ker L + \dim \operatorname{range} L = \dim V,$$

so it is tempting to guess that $V = \ker L \oplus \operatorname{range} L$. This would be excellent because $\operatorname{range} L$ is L-invariant. [why?] However, in general this is not true.

Exercise 5.2.1. Let L has matrix

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

Show that $V \neq \ker L \oplus \operatorname{range} L$.

However, a weaker statement is valid.

Theorem 5.2.2. *Let $L : V \to V$, and suppose, for some $k \geq 1$,*

$$\ker L^{k+1} = \ker L^k,$$

then

$$V = \ker L^k \oplus \operatorname{range} L^k.$$

Exercise 5.2.3. Show that

$$\ker L \subset \ker L^2 \subset \ker L^3 \subset \dots$$

Exercise 5.2.4. Show that if $\ker L^{k+1} = \ker L^k$, then $\ker L^{m+1} = \ker L^m = \ker L^k$ for all $m \geq k$.

Exercise 5.2.5. Let $\dim V = n$. Show that $\ker L^n = \ker L^{n+1}$. In particular, the condition of the theorem is always satisfied for some $k \leq n$.

(If you have difficulties, see Axler, Propositions 8A1 - 3.)

Proof. First we show that the intersection is empty. Suppose $v \in \ker L^k \cap \operatorname{range}(L^k)$, then we have $L^k v = 0$ and $v = L^k u$ for some u, which means $L^{2k} u = 0$. Since $2k \geq k$ and so $\ker L^{2k} = \ker L^k$, we have $L^k u = 0$. But $v = L^k u$ so $v = 0$.

Hence $\ker(L^k) + \operatorname{range}(L^k) \subset V$ is a direct sum and its dimension is $\dim \ker(L^k) + \dim \operatorname{range}(L^k) = \dim V$ by the property of direct sums and the rank-nullity theorem. This implies that

$$\ker(L^k) \oplus \operatorname{range}(L^k) = V.$$

□

This theorem motivates the definition of generalized eigenspaces.

Definition 5.2.6. Let $L : V \to V$ be a linear transformation. A non-zero vector v is called a **generalized eigenvector** of L for an eigenvalue λ if $(L - \lambda I)^k = \vec{0}$ for some $k \geq 1$.

Note that $(L - \lambda I)^k v = 0$ for a non-zero v only can happen if $(L - \lambda I)^k$ is singular, which can only happen if $L - \lambda I$ is singular, so λ indeed must be an eigenvalue.

The collection of all generalized eigenvectors belonging to the eigenvalue λ (together with the zero vector) is called the **generalized eigenspace** $\tilde{E}_\lambda$. It is indeed a vector subspace of V and moreover, it is an L-invariant subspace. Indeed, if $v \in \tilde{E}_\lambda$, then $(L - \lambda I)^k v = 0$, and $(L - \lambda I)^k Lv = L(L - \lambda I)^k v = 0$, so $Lv \in \tilde{E}_\lambda$.

Let $\sigma(L)$ denote the set of all eigenvalues of L. Our goal is to prove the following result.

Theorem 5.2.7. *Let $L : V \to V$ be a linear transformation over field $\mathbb{C}$ and $\sigma(L) = \{\lambda_1, \ldots, \lambda_r\}$ denote the set of all eigenvalues of L. Then*

$$V = \bigoplus_{i=1}^{r} \tilde{E}_{\lambda_i}.$$

It is important here that the matrix is over the field of complex numbers $\mathbb{C}$.

In order to prove this theorem, we need some auxiliary results. The crucial advantage of the generalized eigenvectors over the regular eigenvector is that there are sufficiently many of them to span the whole space.

Lemma 5.2.8. *Let $L : V \to V$ be a linear transformation over $\mathbb{C}$. There is a basis of V that consists of generalized eigenvectors of L.*

Proof. The proof is by induction over the dimension of the space n. For $n = 1$ the result is true since any non-zero vector is an eigenvector. Suppose that $n \geq 2$ and that L has an eigenvalue λ (which holds because the field is algebraically closed so the characteristic polynomial has a root). Then, by Lemma 5.2.2 and Ex. 5.2.5, we have:

$$V = \ker(L - \lambda I)^n \oplus \operatorname{range}(L - \lambda I)^n$$

Note that we have a basis of generalized eigenvectors in $\ker(L-\lambda I)^n$ because this space consists of the generalized eigenvectors corresponding to λ, so any basis will work.

For range$(L-\lambda I)^n$ note that this is an L-invariant subspace. Indeed, if $v \in \text{range}(L-\lambda I)^n$, then by definition $v = (L-\lambda I)^n u$. Then $Lv = (L-\lambda I)^n(Lu)$ so Lv is also in the range of $(L-\lambda I)^n$. It follows that we can restrict L to range$(L-\lambda I)^n$. In addition $\dim \text{range}(L-\lambda I)^n = n - \dim \ker(L-\lambda I)^n < n$ by the rank-nullity theorem and the definition of the eigenvalue. Hence, we can apply the induction hypothesis and conclude that there is a basis of range$(L-\lambda I)^n$ that consists of the generalized eigenvectors of L. Combining the bases of $\ker(L-\lambda I)^n$ and range$(L-\lambda I)^n$ and applying Theorem 1.1.30, we get the claim of the lemma. □

Lemma 5.2.9. *If $\lambda \neq \mu$ are eigenvalues of L, then $\tilde{E}_\lambda \cap \tilde{E}_\mu = \{\vec{0}\}$.*

Proof. Suppose a non-zero $v \in \tilde{E}_\lambda \cap \tilde{E}_\mu$ and $(L-\lambda I)^k v = 0$, $(L-\mu I)^m v = 0$. Assume that m is the smallest $m \geq 1$ with this property. Write

$$\begin{aligned} 0 = (L-\lambda I)^k v &= (L-\mu I + (\mu-\lambda)I)^k v \\ &= \sum_{i=0}^{k} \binom{k}{i} (\mu-\lambda)^{k-i} (L-\mu I)^i v. \end{aligned}$$

By applying $(L-\mu I)^{m-1}$ on both sides, we find that

$$0 = (\mu-\lambda)(L-\mu I)^{m-1} v.$$

Since $(L-\mu I)^{m-1} v \neq 0$ by assumption, we find that $\mu = \lambda$, as desired. □

Lemma 5.2.10. *Let $v_1, \ldots, v_m$ are generalized eigenvectors corresponding to eigenvalues $\lambda_1, \ldots, \lambda_m$ which are all distinct. Then $v_1, \ldots, v_m$ are linearly independent.*

Proof. The result is true for $m = 1$. Suppose it is false for some $m \geq 2$ and choose the smallest m, for which it is false. Then we have non-zero $a_1, \ldots, a_m$ such that

$$a_1 v_1 + \ldots + a_m v_m = 0.$$

(They are non-zero by minimality of m.) Let $(L-\lambda_m I)^k v_m = 0$, and apply $(L-\lambda_m I)^k$ to the equation above. Then we get

$$\begin{aligned} a_1 (L-\lambda_m I)^k v_1 + \ldots + a_{m-1}(L-\lambda_m I)^k v_{m-1} = 0, \\ a_1 w_1 + \ldots + a_{m-1} w_{m-1} = 0, \end{aligned}$$

where $w_i := (L - \lambda_m I)^k v_i$. None of w_i is zero because of Lemma 5.2.9 and therefore w_i is a generalized eigenvector corresponding to eigenvalue λ_i. Indeed, if $(L - \lambda_i I)^{k_i} v_i = 0$, then

$$(L - \lambda_i I)^{k_i} w_i = (L - \lambda_i I)^{k_i} (A - \lambda_m I)^k v_i = (L - \lambda_m I)^k (L - \lambda_i I)^{k_i} v_i = 0.$$

However, then we have a contradiction with minimality of m. □

Now we can prove Theorem 5.2.7.

Proof. First we want to show that the sum $E_{\lambda_1} + \ldots + E_{\lambda_r}$ is direct. Indeed, suppose that $u_1 + \ldots + u_r = 0$ for $u_i \in E_{\lambda_i}$. Then all $u_i = 0$ by Lemma 5.2.10. This shows that the sum is direct.

Now we want to show that every vector in V belongs to this sum. This follows from Lemma 5.2.8. □

5.3 Jordan canonical form

Theorem 5.2.7 shows that we can find a basis in which the matrix for L has the block-diagonal form

$$A = \begin{bmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & A_r \end{bmatrix} \tag{5.1}$$

where A_i is the matrix of the restricted linear map $L|_{\tilde{E}_{\lambda_i}}$. We know that $(L - \lambda_i I)^n v = 0$ for every $v \in \tilde{E}_{\lambda_i}$, hence $A_i - \lambda I$ is a nilpotent matrix, $(A_i - \lambda_i I)^n = 0$.

There is a well developed theory of nilpotent matrices (and the corresponding nilpotent operators), which we prefer not to develop for the considerations of time. We only formulate its final conclusion.

Definition 5.3.1. Let $L : V \to V$ be a linear map over $\mathbb{C}$. A basis of V is called a **Jordan basis** for L if in this basis the matrix of L has a block-diagonal form (5.1), in which each A_i is an upper triangular matrix of the form

$$A_i = \begin{bmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 \\ 0 & 0 & \lambda_i & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & & \lambda_i \end{bmatrix}$$

Matrices A_i are often called **Jordan blocks** or **Jordan cells**. If we denote A_i by $J(\lambda_i, k_i)$ where d_i is the size of the square matrix A_i, then we can write A as a direct sum of $J(\lambda_i, d_i)$.

Theorem 5.3.2 (The existence of Jordan form)**.** *Let $L : V \to V$ be a linear transformation over $\mathbb{C}$. Then there is a basis of V which is a Jordan basis for L.*

We skip the proof of this theorem. See, however, Example 5.3.8, which gives some idea on how one can calculate the Jordan basis.

Also note that the elements of the Jordan basis are generalized eigenvectors so that the decomposition in Jordan blocks agrees with the decomposition in the blocks of $L\|_{E_{\lambda_i}}$ that we derived in Theorem 5.2.7.

Example 5.3.3. Here is an example of a linear transformation with the matrix in a Jordan basis.

$$
\begin{aligned}
A &= \begin{bmatrix}
3 & 0 & & & & & & & & & \\
0 & 3 & 1 & 0 & & & & & & & \\
 & & 3 & 1 & & & & & & & \\
 & & & 3 & 0 & & & & & & \\
 & & & 0 & 2 & 1 & & & & & \\
 & & & & & 2 & 1 & & & & \\
 & & & & & & 2 & & & & \\
 & & & & & & 0 & 2 & 1 & & \\
 & & & & & & & & 2 & & \\
 & & & & & & & & 0 & 2 & 1 \\
 & & & & & & & & & & 2
\end{bmatrix} \\
&= \begin{bmatrix}
J(3,1) & & & & & & & & & & \\
 & & J(3,3) & & & & & & & & \\
 & & & & & J(2,4) & & & & & \\
 & & & & & & & J(2,2) & & & \\
 & & & & & & & & & & J(2,2)
\end{bmatrix} \\
&= J(3,1) \oplus J(3,3) \oplus J(2,4) \oplus J(2,2) \oplus J(2,2).
\end{aligned}
$$

The sizes of the Jordan cells $J(\lambda_i, d_i)$ have to satisfy certain restrictions. The following is a list of properties.

(i) For a given λ_i, the sum of d_i equals $m_a(\lambda_i)$.
(Since the sum of these d_i is the dimension of the generalized eigenspace,

$\dim \tilde{\mathbb{E}}_{\lambda_i}$, this property says $m_a(\lambda_i) = \dim \tilde{\mathbb{E}}_{\lambda_i}$. This dimension equals $\dim \ker(L - \lambda_i I)^n$ [why?], so this property gives us an ability to calculate the algebraic multiplicity without calculating the characteristic polynomial).

In the example, we have $m_a(3) = 1 + 3 = 4$ and $m_a(2) = 4 + 2 + 2 = 8$.

(ii) The number of cells with a given λ_i equals $m_g(\lambda_i)$. [Why?] In the example $m_g(3) = 2$, $m_g(2) = 3$.

(iii) If $s_i = \max\{d_1, \dots d_m\}$ (where $d_1, \dots, d_m$ are the dimensions of the Jordan blocks for a given eigenvalue λ_i), then the polynomial

$$q(t) = (t - \lambda_1)^{s_1} \dots (t - \lambda_r)^{s_r},$$

has the property that $q(L) = 0$.

In our example $s_1 = 3$ and $s_2 = 4$, so the polynomial $q(t) = (t-3)^3(t-2)^4$.

Exercise 5.3.4. Prove these properties.

In fact, the polynomial defined in (iii) is the **minimal** non-zero polynomial $p(t)$ with the property $p(L) = 0$: every other polynomial with this property must be divisible by $q(t)$. It can be proved that this polynomial is unique if we require that the coefficient before the highest power is 1 and we can make the following definition.

Definition 5.3.5. $q(t) = (t - \lambda_1)^{s_1} \dots (t - \lambda_r)^{s_r}$ is called the **minimal polynomial** of L.

Note that $q(t)$ divides the characteristic polynomial $p_A(t)$ and in particular $p_A(L) = 0$. This is called the **Cayley–Hamilton theorem**.

Now suppose an $n \times n$ matrix A has just one eigenvalue λ. What are possible Jordan decompositions of A up to a permutation of blocks? How does the answer change if we know that the minimal polynomial for A is $(t - \lambda)^m$? It turns out that the answer can be written in terms of integer partitions of n. The requirement on the degree of the minimal polynomial corresponds to the requirement that each part of the partition does not exceed m and at least one of them equals m.

Example 5.3.6. Suppose a 5×5 matrix A has the only eigenvalue $\lambda = 2$, and the minimal polynomial is $(t - 2)^3$. In this case the size of the Jordan

blocks in the Jordan form for A cannot exceed $m = 3$ and at least one of the blocks must have this size. The only possibly Jordan blocks are

$$J(2,1) = [2], J(2,2) = B = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \text{ and } J(2,3) = C = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

Hence the only allowed partitions of 5 are $3+2$ and $3+1+1$, and they correspond to the Jordan forms $\text{diag}(C, B)$ and $\text{diag}(C, 2, 2)$.

The geometric multiplicities of $\lambda = 2$ for these forms are $m_g(2) = 2$ (two blocks) for the first form and $m_g(2) = 3$ (three blocks) for the second form. The algebraic multiplicity is of course $m_a(2) = 5$ in both cases.

Given a linear operator L or its matrix A, how do we compute its Jordan canonical corm and the basis for this form? There are some algorithms for doing this, however, we will only consider some simple examples.

Example 5.3.7. Let

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

Find the characteristic and minimal polynomials for this matrix and its Jordan form.

Clearly the characteristic polynomial is $p_A(t) = (t-1)^3$. Then we can calculate the $\ker(A - I)$:

$$\ker(A - I) = \ker \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\rangle.$$

So $m_g(1) = 1$ and there is only one Jordan cell, which must be

$$J(1,3) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

This means that the minimal polynomial is $q(t) = (t-1)^3$. (Here the degree 3 is the dimension in $J(1,3)$.)

Example 5.3.8. Let

$$A = \begin{bmatrix} 3 & 1 & -2 \\ -1 & 0 & 5 \\ -1 & -1 & 4 \end{bmatrix}.$$

Find the Jordan canonical form and Jordan basis for A.

The characteristic polynomial can be calculated as

$$p_A(t) = -(t-3)(t-2)^2.$$

Since the multiplicity of $t-3$ is 1, it corresponds to the Jordan block $J(3,1)$. For $\lambda = 2$ we can check that $\dim\ker(A-2I) = 1$. Hence the geometric multiplicity of $\lambda = 2$ is 1 and there is only one Jordan cell. We can infer that this Jordan block must be $J(2,2)$. Hence the Jordan canonical form is

$$\tilde{A} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

The first vector in the Jordan basis is an eigenvector of A corresponding to $\lambda = 3$. For example, we can take

$$v_1 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}.$$

The third vector is an eigenvector corresponding to $\lambda = 2$ and we can take

$$v_3 = \begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix}.$$

Finally, to find the second vector of the Jordan basis, v_2, we note that it satisfies equation $Av_2 = 2v_2 + v_3$. Hence we only need to solve the system $(A-2I)v_2 = v_3$. A solution is

$$v_2 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}.$$

Here a consequence of Theorem 5.3.2 which is useful in applications.

Corollary 5.3.9. *Any operator $L : V \to V$ over $\mathbb{C}$ can be represented as $L = D + N$, where D is diagonalizable (i.e. diagonal in some basis) and N is nilpotent ($N^m = 0$ for some m), and $DN = ND$.*

The first part follows directly by the existence of Jordan form. The commutativity of D and N can be checked at every $\tilde{E}_\lambda$ separately, because these are L-invariant subspaces, and there $D = \lambda I$ and therefore commute with any operator.

Without any proof we mention that this result is very helpful when one calculate the functions of L for operators L which are not diagonalizable. In this case one only needs to write the Taylor series for $f(D+N)$ around D and note that these series are finite because $N^k = 0$ for $k \geq m$:

$$f(D+N) = \sum_{k=0}^{m-1} f^{(k)}(D)\frac{N^k}{k!},$$

where $f^{(k)}$ is the k-th derivative of matrix and it is easy to calculate $f^{(k)}(D)$ because D is a diagonal matrix.

For example, if $f(x) = e^x$, then we have:

$$e^{D+N} = e^D \sum_{k=0}^{m-1} \frac{N^k}{k!}.$$

5.4 Exercises

The exercises with (☞) have a hint at the end of Lecture Notes.

Exercise 5.4.1. True or false:

(a) Eigenvectors of a linear operator T are also generalized eigenvectors of T.

(b) It is possible for a generalized eigenvector of a linear operator T to correspond to a scalar that is not an eigenvalue of T.

(c) Let T be a linear operator on an n-dimensional vector space over $\mathbb{C}$. Then, for any eigenvalue λ_i of T, $\tilde{E}_{\lambda_i} = \ker(T - \lambda I)^n$.

Proofs:

Exercise 5.4.2 (8A.1 in Axler). Suppose $T \in \mathcal{L}(V)$, i.e. T is a linear map from V to V. Prove that if $\dim \ker T^4 = 8$ and $\dim \ker T^6 = 9$, then $\dim \ker T = 9$ for all integers ≥ 5.

Exercise 5.4.3 (8A.2 in Axler). Suppose $T \in \mathcal{L}(V)$, m is a positive integer, $v \in V$ and $T^{m-1}v \neq 0$ but $T^m v = 0$. Prove that $\{v, Tv, T^2v, \ldots, T^{m-1}v\}$ is linearly independent.

Exercise 5.4.4 (8A.3 in Axler). Suppose $T \in \mathcal{L}(V)$. Prove that

$$V = \ker T \oplus \operatorname{range} T \iff \ker T^2 = \ker T.$$

Exercise 5.4.5. True or False:

1. Any linear operator on a finite-dimensional vector space has a Jordan canonical form.

2. Any linear operator on a finite-dimensional vector space over $\mathbb{C}$ has a Jordan canonical form.

3. Let T be a linear operator on a finite-dimensional vector space over $\mathbb{C}$, and let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. If, for each i, $\mathcal{B}_i$, is a basis for $\tilde{E}_{\lambda_i}$, then $\mathcal{B}_1 \cup \ldots \cup \mathcal{B}_k$ is a Jordan basis for T.

Exercise 5.4.6 (8B.12 in Axler). Give an example of an operator on $\mathbb{C}^4$, whose characteristic polynomial equals $(z-1)(z-5)^3$ and whose minimal polynomial equals $(z-1)(z-5)^2$

Exercise 5.4.7 (8B.13 in Axler). Give an example of an operator on $\mathbb{C}^4$ whose characteristic and minimal polynomials are both equal $z(z-1)^2(z-3)$.

Exercise 5.4.8. Let T be the linear operator on $P_2[t]$ (polynomials of degree ≤ 2) defined by $T(g(x)) = -g(x) - g'(x)$. Find a Jordan canonical form of T and a Jordan basis for T.

Exercise 5.4.9. Find a Jordan canonical form for

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

Proofs:

Exercise 5.4.10. [8B.4 in Axler] (☞)

Suppose $T \in \mathcal{L}V$, $n = \dim(V) \geq 2$, and $\ker T^{n-2} \neq \ker T^{n-1}$. Prove that T has at most two distinct eigenvalues.

Exercise 5.4.11. [8B.5 in Axler](☞)

Suppose $T \in \mathcal{L}V$, $n = \dim(V) \geq 2$, and 3 and 8 are eigenvalues of T. Prove that $V = (\ker T^{n-2}) \oplus (\operatorname{range} T^{n-2})$

Chapter 6

Inner Product Spaces

Reading: Treil, Chapter 5.

6.1 Inner products

In geometry we have the concept of orthogonality of lines and planes and we have the concepts of length. These two concepts are related through the Pythagorean theorem. Linear algebra captures these concepts by defining the inner product between vectors. As length of a vector can be larger or smaller, we need an order on the numbers which we use. This forces us to restrict attention to vector spaces over the field of complex numbers and or over its subfields such as $\mathbb{Q}$ or $\mathbb{R}$. [1]

First, let us start with two motivating example. Then, we will go to the general definition of the inner product.

Definition 6.1.1. A dot product of $x, y \in V = \mathbb{R}^n$ is a function $V \times V \to \mathbb{R}$ defined by

$$x \cdot y := x_1 y_1 + \dots x_n y_n = x^t y = y^t x \tag{6.1}$$

The dot product of the vector x with itself is the generalization of the square of the length in $\mathbb{R}^3$. The Euclidean norm of x is defined as the square root of $x \cdot x$:

$$\|x\| = (x \cdot x)^{1/2}.$$

[1] It is possible to define analogues of dot product $\sum_{i=1}^n x_i y_i$ over finite or p-adic fields, however this product might fail the requirement $x \cdot x \neq 0$ for all $x \neq 0$. For this reason, the questions of the existence of "orthogonal" bases of a subspace, the projections, and so on, become more difficult.

For $V = \mathbb{C}^n$, one has to modify this definition to make sure that the dot product of the vector with itself is positive. Mathematicians typically use the following modification.

$$\langle x, y\rangle = x_1\overline{y}_1 + \ldots + x_n\overline{y}_n = y^*x, \tag{6.2}$$

while physicists prefer $\langle x, y\rangle = \overline{x}_1 y_1 + \ldots + \overline{x}_n y_n = x^*y$. Here $\overline{z}$ means complex conjugation of numbers in $\mathbb{C}$ and x^* is $\overline{x}^t$, that is, a vector obtained by applying complex conjugation to every entry of x and then transposing the vector.

In these notes we follow the mathematicians' convention.

We can define more general inner products on vector spaces over $\mathbb{R}$ or $\mathbb{C}$. For concreteness, consider the vector spaces over $\mathbb{C}$.

Definition 6.1.2. An **inner product** on a complex vector space V is a function $V \times V \to \mathbb{C}$, $(u, v) \to \langle u, v\rangle$ with the following properties:

(i) $\langle v, v\rangle \geq 0$ for all $v \in V$.

(ii) $\langle v, v\rangle = 0$ if and only if $v = 0$.

(iii) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$.

(iv) $\langle \alpha u, v\rangle = \alpha\langle u, v\rangle$ for all $\alpha \in \mathbb{C}$ and all $u, v \in V$.

(v) $\langle u, v\rangle = \overline{\langle v, u\rangle}$.

For the real vector spaces definition is similar except that the function values are in $\mathbb{R}$ not in $\mathbb{C}$, the scalar $\alpha \in \mathbb{R}$ and in the last property the complex conjugation is not needed.

Here are some examples.

Example 6.1.3. The definitions in (6.1) and (6.2) give valid inner product on $\mathbb{R}^n$ and $\mathbb{C}^n$, respectively. This inner product is usually called the standard or **Euclidean inner product**.

Example 6.1.4. Let $a_1, \ldots, a_n$ be positive real numbers. Then

$$\langle u, v\rangle = a_1u_1\overline{v}_1 + \ldots a_nu_n\overline{v}_n$$

define an inner product on $\mathbb{C}^n$.

Example 6.1.5. Let $V = \mathbb{C}^n_m$ be the vector space of $m \times n$ complex matrices. Then

$$\langle A, B\rangle = \mathrm{Tr}(B^*A) = \sum_{i,j} A_{ij}\overline{B}_{ij}$$

is an inner product on V which is called the **Frobenius inner product**.

Example 6.1.6. Let V be the vector space of continuous real-valued functions on $[-1, 1]$. Then

$$\langle f, g\rangle = \int_{-1}^{1} fg\, dx$$

is an inner product.

Example 6.1.7. Let $P[t]$ be the vector space of polynomials over $\mathbb{R}$. Then,

$$\langle p, q\rangle = p(0)q(0) + \int_{-1}^{1} p'(x)q'(x)\, dx$$

is an inner product.

Example 6.1.8. Another inner product on $P[t]$ can be defined by

$$\langle p, q\rangle = \int_{0}^{\infty} p(x)q(x)e^{-x}\, dx.$$

6.2 Orthogonality and vector norms

We want to define a concept of length for vectors in a vector space. This can be done by defining a norm of a vector. Mathematically, a **norm** is a non-negative function on a linear space, which has the properties:

(i) (positivity) $\|u\| \geq 0$ for all $u \in V$.

(ii) $\|u\| = 0$ implies that $u = 0$.

(iii) (homogeneity): $\|cv\| = |c|\|v\|$, for all $c \in \mathbb{C}$,

(iv) (triangle inequality): $\|u + v\| \leq \|u\| + \|v\|$.

We want to show that we can define a vector norm of by using a given inner product:

$$\|u\| = \langle u, u\rangle.$$

In particular if we use the Euclidean inner product, the norm is called the Euclidean norm:

$$\|u\| := \sqrt{|u_1|^2 + \ldots + |u_n|^2} = \sqrt{u^* u}.$$

We want to prove that the norm defined using any inner product is indeed a vector norm. It easy to check the first three axioms. The triangle inequality is more involved. In order to prove it recall the definition of orthogonality.

Definition 6.2.1. Two vectors $u, v \in V$ are called orthogonal if $\langle u, v\rangle = 0$.

We write $u \perp v$ to denote that u and v are orthogonal.

Theorem 6.2.2 (Pythagorean Theorem)**.** *If two vectors u, v are orthogonal then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.*

Exercise 6.2.3. Prove this theorem.

Lemma 6.2.4 (Orthogonal Decomposition)**.** *Suppose $u, v \in V$ with $v \neq 0$. Let*

$$c = \frac{\langle u, v\rangle}{\|v\|^2} \text{ and } w = u - cv.$$

Then $u = cv + w$ and $w \perp v$.

Exercise 6.2.5. Prove this lemma.

Theorem 6.2.6 (Cauchy-Schwarz inequality)**.** *Let $u, v \in V$ and $\|u\|^2 := \langle u, v\rangle$. Then*

$$|\langle u, v\rangle| \leq \|u\|\|v\|$$

This inequality is an equality if and only if one of u, v is a scalar multiple of the other.

Proof. If $v = 0$, both sided of the inequality are 0. Suppose $v \neq o$, and write the orthogonal decomposition

$$u = \frac{\langle u, v\rangle}{\|v\|^2} v + w,$$

with $w \perp v$. By Pythagorean theorem,

$$\|u\|^2 = \|\frac{\langle u, v\rangle}{\|v\|^2} v\|^2 + \|w\|^2 \geq \frac{|\langle u, v\rangle|^2}{\|v\|^2},$$

which is equivalent to the Cauchy-Schwarz inequality. The equality holds only if $\|w\|^2 = 0$, that is, $w = 0$, which holds only if u is proportional to v. □

Finally we are ready to prove that $\|u\|$ satisfies the triangle inequality.

Theorem 6.2.7. *Suppose $u, v \in V$. Then*

$$\|u + v\| \leq \|u\| + \|v\|.$$

The equality holds if and only if one of u, v is a non-negative real multiple of the other.

Proof. We have

$$\begin{aligned}\|u+v\|^2 &= \|u\|^2 + \|v\|^2 + 2\mathrm{Re}\,\langle u, v\rangle \\ &\leq \|u\|^2 + \|v\|^2 + 2|\langle u, v\rangle| \\ &\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| = (\|u\| + \|v\|)^2,\end{aligned}$$

and the inequality is proved. To have the equality, we need to have the equaliyt in Cauchy-Schwarz, so one of u, v is a multiple of the other. Say $u = cv$. Then, we must also have $\mathrm{Re}\,\langle u, v\rangle = |\langle u, v\rangle|$, which implies $\mathrm{Re}\, c = |c|$ which can only occur if c is a non-negative real number. This completes the proof. □

It follows that the function $\|u\|$ defined as $\|u\| := \langle u, u\rangle^{1/2}$ for an inner product $\langle \cdot, \cdot \rangle$ is indeed a vector norm.

$$\begin{aligned}\|x\|_1 &= \sum_{i=1}^{m} |x_i|, \\ \|x\|_2 &= \left(\sum_{i=1}^{m} |x_i|^2\right)^{1/2} = \sqrt{x^* x}, \\ \|x\|_\infty &= \max_{1 \leq i \leq m} |x_i|, \\ \|x\|_p &= \left(\sum_{i=1}^{m} |x_i|^p\right)^{1/p} \quad (1 \leq p < \infty).\end{aligned}$$

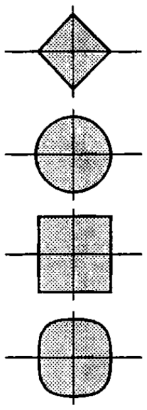

Figure 6.1: Unit balls for different vector norms

Other vector norms

It is useful to know that there are other norms besides the Euclidean norm. For example, a p-norm is defined for every $p \geq 1$ as follows. If $x \in \mathbb{R}^n$, then

$$\|x\|_p = \Big(\sum_{i=1}^{n} |x_i|^p \Big)^{1/p}.$$

This is an exercise that this function is indeed a norm.[2]

If we look at $p \to \infty$ then we get a so-called supremum norm:

$$\|x\|_\infty = \sup_i |x_i|.$$

In this notation, the Euclidean norm can be called 2-norm since it corresponds to the case $p = 2$. So, more proper notation for this norm would be $\|v\|_2$. However, we will usually use this norm rather than any other p-norm and so we will skip this subscript.

The great advantage of the 2-norm (i.e., the Euclidean norm) is that it equals the square root of the inner product of the vector with itself. Because of this, it enjoys some properties which are not true for other norms. For example, if we want to find out what is the point in a linear subspace with the smallest distance from a given point, where the distance is measured using the 2-norm, then we can use the orthogonal projection operator (which we discuss later). In contrast, if we measure distance not in the usual 2-norm but in a different norm, then this would not be true anymore and it would be more difficult to find this point.

On the other hand, the p-norms for $p \neq 2$ are sometimes used in modern statistics, so you should know about them. For example, the **lasso regression** uses the 1-norm of vectors.

The analogue of Cauchy-Schwarz inequality for more general norms on $\mathbb{R}^n$ is the Holder inequality:

$$|x^t y| \leq \|x\|_p \|y\|_q,$$

where $p^{-1} + q^{-1} = 1$.

[2] In contrast, one can check that if $p < 1$, then $\|x\|_p$ is not a norm. This is a couple of additional exercises. First is to check that if $\|\cdot\|$ is a norm, then this implies that the unit ball $B = \{x : \|x\| \leq 1\}$ must be convex. And the second is to check that if $p < 1$, then the unit ball is not convex.

6.3 More on orthogonality

6.3.1 Orthogonal systems

Definition 6.3.1. A set of vectors $u_1, \ldots, u_n$ is called an **orthogonal system** if they are all non-zero and they are pairwise orthogonal: $u_i \perp u_j$ for all $i \neq j$. It is called an **orthonormal system** if it is an orthogonal system and each of these vectors have length 1.

One useful fact about systems of orthogonal vectors is that we can use them to decompose an arbitrary vector in orthogonal components. First of all, we have the following result.

Theorem 6.3.2. *The vectors of an orthogonal system are linearly independent.*

Proof. Suppose they are dependent. The we can write, after reordering these vectors,

$$v_1 = \sum_{i=2}^{n} c_i v_i,$$

where at least one of c_i is not zero. Say, $c_i \neq 0$. Then by taking the inner product of the equality above with v_i, we get $\langle v_1, v_i \rangle = c_i \|v_i\|^2 \neq 0$, and vectors v_1 and v_i are not orthogonal, in contradiction to the assumption. □

We also have a generalization of the decomposition formula proved above. Specifically, we decompose an arbitrary vector v as a linear combination of the vectors in the orthonormal system and a "residual", which is orthogonal to every vector in this system.

Theorem 6.3.3. *Let $\{u_1, \ldots u_n\}$ is an orthonormal set of vectors in $\mathbb{F}^m$, where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$ and $m \geq n$. Then for every vector $v \in \mathbb{R}^m$, there exists a unique decomposition:*

$$v = r + \sum_{i=1}^{n} c_i u_i,$$

in which vector r is orthogonal to each of vectors u_i. The coefficients can be computed as $c_i = \langle v, u_i \rangle$.

Remarks:

(a) If $n = m$, then $\{u_1, \dots u_m\}$ is a basis (every linearly independent system of m vectors in $\mathbb{R}^m$ is a basis), so $r = 0$. The significance of the theorem in this case is that it allows us to compute the coefficients in the expansion $v = \sum_{i=1}^{m} c_i u_i$ using the inner product: $c_i = \langle v, u_i \rangle$.

(b) If $n < m$, then $\sum_{i=1}^{n} c_i u_i$ is the **orthogonal projection** of v on the subspace $U = \langle u_1, \dots, u_n \rangle$ and it can be shown that $\|r\|$ is the shortest distance from v to the subspace U: $\|r\| = \min_{\boldsymbol{x} \in U} \|v - \boldsymbol{x}\|$.

(c) If $\{u_1, \dots u_n\}$ is not orthonormal but simply orthogonal system, then the conclusion of the theorem stays the same but the coefficients c_i are calculated using a different formula:

$$c_i = \frac{\langle v, u_i \rangle}{\langle u_i, u_i \rangle} = \frac{\langle v, u_i \rangle}{\|u_i\|^2}.$$

Proof. The existence will be proved if we show that

$$r = v - \sum_{i=1}^{n} \langle v, u_i \rangle u_i$$

is orthogonal to each of vectors u_j. By multiplying with u_j, we get

$$\langle r, u_j \rangle = \langle v, u_j \rangle - \sum_{i=1}^{n} \langle v, u_i \rangle \langle u_i, u_j \rangle = \langle v, u_j \rangle - \langle v, u_j \rangle \langle u_j, u_j \rangle = 0,$$

which is the required property.

For uniqueness, we note that if we have two different decompositions like that, then we can subtract them. As a result we would have that either $r = r'$, and u_i are linearly dependent, or $r \neq r'$ and the orthogonal set $r - r', u_1, \dots, u_n$ is linearly dependent. Both are not possible by Theorem 6.3.2.

□

Example 6.3.4 (Discrete Fourier Transform (DFT)). Consider the vector space $\mathbb{C}^T$ and vectors $\boldsymbol{u}_0, \dots, \boldsymbol{u}_{T-1}$ defined as

$$\boldsymbol{u}_k = \frac{1}{\sqrt{T}} \Big[\exp\Big(\frac{2\pi i \cdot k \cdot 0}{T}\Big), \exp\Big(\frac{2\pi i \cdot k \cdot 1}{T}\Big), \dots, \exp\Big(\frac{2\pi i \cdot k \cdot (T-1)}{T}\Big) \Big].$$

(This is the function $\chi_k(t) = \exp(2\pi i \frac{kt}{T})$ evaluated at $t = 0, 1, \dots, T-1$).

Then, the vectors $\boldsymbol{u}_k$ form a orthonormal system with respect to the usual Euclidean inner product, – this is a good exercise, – and so every vector $\boldsymbol{x} = [x_0, \dots x_{T-1}]$ can be written as

$$\boldsymbol{x} = \sum_{k=0}^{T-1} c_k \boldsymbol{u}_k$$

or

$$x_t = \frac{1}{\sqrt{T}} \sum_{k=0}^{T-1} c_k \exp\Big(\frac{2\pi i \cdot k \cdot t}{T}\Big) \tag{6.3}$$

where

$$c_k = \langle \boldsymbol{x}, \boldsymbol{u}_k \rangle = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} x_t \exp\Big(-\frac{2\pi i \cdot k \cdot t}{T}\Big), \tag{6.4}$$

for $k = 0, \dots, T-1$. The vector $\hat{\boldsymbol{x}} := \boldsymbol{c} = [c_0, \dots, c_{T-1}]$ is called the Discrete Fourier transform of vector $\boldsymbol{x}$. (Sometimes $\frac{1}{\sqrt{T}}$ is omitted in the definition of DFT. This is compensated by using $\frac{1}{T}$ instead of $\frac{1}{\sqrt{T}}$ in formula (6.3).)

If we put vectors $\boldsymbol{u}_k$ as columns in matrix $F = [u_0, \dots, u_{T-1}]$, then we see that the Discrete Fourier Transform can be written in a matrix form:

$$\hat{\boldsymbol{x}} = F^* \boldsymbol{x}.$$

In order to get some intuition how the matrix F looks like, consider $T = 2$ and $T = 4$. For $T = 2$,

$$F = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

and for $T = 4$, it is

$$F = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & 1 \\ 1 & i & -1 & -i \end{bmatrix},$$

6.3.2 Unitary and orthogonal matrices

Now assume that we use the Euclidean inner product in $\mathbb{R}^n$ or $\mathbb{C}^n$, $\langle u, v \rangle = v^* u$.

Definition 6.3.5. A matrix in $\mathbb{C}_{n\times n}$ called **unitary** if the set of its column vectors is orthonormal. If, in addition, it is in $\mathbb{R}_{n\times n}$ then it is called **orthogonal**.

Definition 6.3.6. A matrix in $\mathbb{C}_{m\times n}$ is called an **isometry**, it its columns are orthonormal.

(The difference from the previous definition is that we do not require the matrix to be square.)

Alternatively, we can define isometry Q as an $m \times n$ matrix for which $Q^*Q = I_n$. If, in addition, the matrix is square, then it is a unitary matrix. If, moreover, it is real, then it is an orthogonal matrix.

For **square** matrices A and B, the identity $AB = I$ implies that $BA = I$. (This is a good exercise.) So, if Q is orthogonal then we must also have $QQ^* = I$. (However, if $m > n$ and Q is an $m \times n$ isometry matrix then $Q^*Q = I$, but it can happen that $QQ^* \neq I$.)

An important property of a linear transformation that corresponds to an isometry Q is that it preserves lengths of vectors.

$$\|Qv\|^2 = (Qv)^*Qv = v^*Q^*Qv = v^*v = \|v\|^2.$$

(This is the reason why the matrix Q is called an isometry.)

Example 6.3.7. The rotation matrix R_θ is orthogonal.

Example 6.3.8. The Discrete Fourier Transform matrix F is unitary.

6.3.3 Orthonormal bases

Theorem 6.3.3 implies that the columns of the $n \times n$ orthogonal matrix Q form a basis in Q (since in this case the maximal number of linearly independent vectors is n), and the coefficients of a vector v in this basis can be computed very conveniently as $c = Q^*v$.

Definition 6.3.9. An **orthonormal basis** of a vector space V is a basis that consists of orthonormal vectors.

Orthonormal bases have a nice property that the Euclidean inner product in any orthonormal basis $\mathcal{B}$ can be computed by using the simple formula $\langle v, u\rangle = [u]^*[v]$, where $[u]$ and $[v]$ are coordinates of vectors u and v in this basis. That is, the inner product is computed in absolutely the same way as in the standard coordinates in $\mathbb{C}^n$ (or $\mathbb{R}^n$).

Exercise 6.3.10. Let $v_1, v_2, \ldots, v_n$ be an orthonormal basis in V.

(a) Prove that for any $\boldsymbol{x} = \sum_{k=1}^n \alpha_k v_k$, $\boldsymbol{y} = \sum_{k=1}^n \beta_k v_k$,

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \sum_{k=1}^{n} \alpha_k \overline{\beta}_k.$$

(b) Deduce from this **Parseval's identity**:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{k=1}^{n} \langle \boldsymbol{x}, v_k \rangle \overline{\langle \boldsymbol{y}, v_k \rangle} = \sum_{k=1}^{n} \langle \boldsymbol{x}, v_k \rangle \langle v_k, \boldsymbol{y} \rangle$$

Example 6.3.11. If $\hat{\boldsymbol{x}} = \boldsymbol{c}$ is the DFT of $\boldsymbol{x}$ then Parceval's identity says that

$$\sum_{t=0}^{T-1} |x_t|^2 = \sum_{k=0}^{T-1} |c_k|^2.$$

Finally, we have the change of coordinates formula for linear operators. If we change the standard basis to a new orthonormal basis, then the matrix of the linear transformation in the new basis is

$$\tilde{A} = QAQ^{-1} = QAQ^*.$$

6.3.4 Orthogonal complements

Two linear subspaces V and W are orthogonal to each other ($V \perp W$) if **every (non-zero) vector in V is orthogonal to every (non-zero) vector in W**.

Note that the **intersection of two orthogonal subspaces is always zero** (i.e., the trivial subspace). Indeed, if v belongs to two subspaces simultaneously, then $v \perp v$ and $\|v\|^2 = v^*v = 0$, which implies that $v = 0$.

Let $V \subset \mathbb{R}^m$ be a linear subspace. Then its **orthogonal complement** of V in $\mathbb{R}^m$, denoted $V^\perp$, is the largest linear subspace in $\mathbb{R}^m$ orthogonal to V. Alternatively, it is the set of all vectors u that are orthogonal to V. Formally:

$$V^\perp = \{u \in \mathbb{R}^m : u^*v = 0 \text{ for all } v \in V\}.$$

Theorem 6.3.12. *For any $V \subset \mathbb{R}^m$,*

$$V + V^\perp = \mathbb{R}^m,$$

and the sum is direct.

Proof. Let $\langle u_1, \ldots, u_r \rangle$ be an orthonormal basis of V. Take an arbitrary $\boldsymbol{x} \in \mathbb{R}_m$. By Theorem 6.3.3, $\boldsymbol{x} = r + \sum c_i u_i$, where $r \perp u_i$ for every $i = 1, \ldots, r$. Then $\sum c_i u_i \in V$ and $r \in V^\perp$ and we showed that $\mathbb{R}^m = V + V^\perp$.

Since $V \cap V^\perp = \{\vec{0}\}$, the sum is direct. □

(This proof has a non-clear step since we assumed that we can always find an orthonormal basis of W. Later we will see how to construct this basis by the Gram-Schmidt orthogonalization process.)

Theorem 6.3.13. *If $V \in \mathbb{R}^m$ and $\dim(V) = k$ then $\dim(V^\perp) = m - k$.*

Proof. By previous theorem, $R^m = V \oplus V^\perp$, and therefore $m = \dim V + \dim V^\perp$ by our results about direct sums of subspaces. □

Corollary 6.3.14. *If $V^\perp$ is orthogonal complement to V in $\mathbb{R}^m$, then V is an orthogonal complement to $V^\perp$ in $\mathbb{R}^m$.*

Proof. All vectors in V are orthogonal to all vectors in $V^\perp$ [why?]. So $V \subset (V^\perp)^\perp$. We need to show that this is in fact an equality. We have $\dim((V^\perp)^\perp) = m - \dim V^\perp = m - (m - \dim V) = \dim V$ and we use the fact that if one linear space is a subspace of another one and they have the same dimension then they must coincide. □

Since $\mathbb{R}^m = V \oplus V^\perp$, therefore we can construct the basis of R^m by taking the union of the bases of V and $V^\perp$. In particular, every vector u in $\mathbb{R}^n$ can be represented in a unique way as $v + w$ where $v \in V$ and $w \in V^\perp$.

Now, here is an important example of orthogonal complements.

Theorem 6.3.15. *For an $m \times n$ matrix $A \in \mathbb{R}_{m\times n}$*

1. *The nullspace of A is the orthogonal complement of the row space of A (i.e., the range of A^t):*

$$(\ker A)^\perp = \operatorname{range}(A^t) \text{ and } (\operatorname{range} A^t)^\perp = \ker A.$$

2. *The range of A is the orthogonal complement of the left nullspace of A:*

$$(\operatorname{range} A)^\perp = \ker A^t \text{ and } (\ker A^t)^\perp = \operatorname{range} A.$$

As a corollary, we find that

$$\mathbb{R}^n = \ker A \oplus \operatorname{range}(A^t),$$
$$\mathbb{R}^m = \operatorname{range} A \oplus \ker(A^t)$$

Proof of Theorem 6.3.15. It is enough to prove one of these claims, since the proof of the other follows by considering the transpose of matrix A. Let us prove the first one.

If $u \in \operatorname{range} A^t$ then $u = A^t z$ for some vector $z \in R^m$. We aim to show that $u \in (\ker A)^\perp$.

Let $v \in \ker A$, then we can calculate the inner product

$$\langle v, u \rangle = u^t v = (A^t z)^t v = z^t A v = 0,$$

where the last step holds because $Av = 0$.

Hence $\operatorname{range} A^t \subset \ker A$. We need to show that this is in fact an equality.

The dimension of $\operatorname{range} A^t$ is $\operatorname{rank}(A)$, the dimension of $\ker A$ is $n - \operatorname{rank}(A)$ and the dimension of $(\ker A)^\perp = n - \dim \ker A = \operatorname{rank} A$. This implies that $\operatorname{range} A^t = \ker A$. □

In particular, it gives a method to calculate the orthogonal complement to a subspace spanned by vectors $\boldsymbol{c}_1, \ldots \boldsymbol{c}_n$. Write the matrix C^t with rows given by $\boldsymbol{c}_1^t, \ldots \boldsymbol{c}_n^t$, and calculate its nullspace (that is, the basis of the nullspace).

Example 6.3.16 (Example of calculation). Find a vector in the orthogonal complement to the column space of matrix

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 2 & 4 \\ 2 & 2 \end{bmatrix}.$$

This simply means calculating $\ker A^t$:

$$A^t = \begin{bmatrix} 1 & 0 & 2 & 2 \\ 1 & 1 & 4 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & 2 & 0 \end{bmatrix}$$

The free variables are x_3, x_4 and we have:

$$\operatorname{range}(A)^\perp = \ker(A^t) = \left\langle \begin{bmatrix} -2 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\rangle$$

Note also that once we know the basis vectors of $(\operatorname{range} A)^\perp$, we can easily test if a vector v belongs to $\operatorname{range} A$ by calculating the inner products of v with these basis vectors and checking if they are equal to 0.

6.4 Adjoint transformations

So far, we used notation A^* as a shortcut notation for $\overline{A}^t$. In this section we give a different definition, which might differ if the inner product is not standard.

Note that if a matrix A represents the map $L : \mathbb{C}^n \to \mathbb{C}^n$, and $\mathbb{C}^n$ has the Euclidean inner product, then matrix $A^* = \overline{A}^*$ has the property $\langle Au, v\rangle = \langle u, A^*v\rangle$ for all $u, v \in \mathbb{C}^n$.

We use this identity as the definition of the **adjoint** operator and adjoint matrix.

Definition 6.4.1. Let V is a vector space with inner product $\langle \cdot, \cdot \rangle$, and $L : V \to V$ is a linear operator, then the **adjoint operator** $L^* : V \to V$ is defined by the property that for all $u, v \in V$,

$$\langle Lu, v\rangle = \langle u, L^*v\rangle.$$

It turns out that the adjoint operator exists and unique.

If L has matrix A in a basis $\mathcal{B}$, then the matrix of L^* is the adjoint matrix A^*. In particular, if $\mathcal{B} = \{q_1, \ldots, q_n\}$ is an orthonormal basis in V, and the matrix of the transformation L is A, then the matrix of transformation A^* is as we defined it before: $A^* = \overline{A^t}$.

The benefit of the concept of the adjoint transformation is that we know that Theorems like Theorem 6.3.15 hold as statements about the range and kernel of the adjoint transformations. For example,

$$\ker L = (\operatorname{range} L^*)^\perp.$$

The operators (i.e., linear maps) that have the property $A^* = A$ are called **self-adjoint**. For the vector space $\mathbb{C}^n$ with Euclidean inner product this boils down to the property $\overline{A}^* = A$, and these matrices are called Hermitian. For $\mathbb{R}^n$ with Euclidean inner product, this simply means that the matrix A is symmetric, $A^t = A$.

Another class of operators are **normal operators**, which are defined by the property $A^*A = AA^*$. In particular, Hermitian operators are symmetric.

We will see later in Sections 8.2 and 8.3 that self-adjoint and normal operators have very remarkable properties with respect to their eigenvalues and eigenvectors.

Example 6.4.2. Suppose that the inner product in $\mathbb{R}^n$ is given by the formula $\langle x, y\rangle = \sum_{i=1}^n d_i x_i y_i$. In the matrix form it can be written as $\langle x, y\rangle = y^t D x$, where D is a diagonal matrix with the entries $d_1, \ldots, d_n$ on the main

diagonal. What is the adjoint of matrix A with respect to this inner product? What are the self-adjoint matrices?

Using $D^t = D$, we can write

$$\begin{aligned}\langle Au, v\rangle &= v^t DAu = v^t DAD^{-1}Du = \left[D^{-1}A^t Dv\right]^t Du \\ &= \langle u, D^{-1}A^t Dv\rangle.\end{aligned}$$

Hence, the adjoint of A for this real inner product is $A^* = D^{-1}A^t D$. The matrix is self-adjoint if

$$D^{-1}A^t D = A, \text{ that is, } A^t D = DA.$$

For infinite-dimensional inner-product spaces, the adjoint operators do not always exist and one needs to impose more restrictions on the operator or on the space to ensure that the adjoint operator exists.

Example 6.4.3. Consider the vector space V of real polynomials in variable x, which have the property $f(0) = f(1) = 0$ and define the inner product as $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$. Consider the differentiation operator $\mathcal{D} : f(x) \to f'(x)$. What is its adjoint?

Using integration by parts, we write

$$\begin{aligned}\langle \mathcal{D}f, g\rangle = \int_0^1 f'(x)g(x)\,dx = \Big|_0^1 f(x)g(x) - \int_0^1 f(x)g'(x)\,dx \\ = \langle f, -\mathcal{D}g\rangle.\end{aligned}$$

It follows that $\mathcal{D}^* = -\mathcal{D}$. (That is, $\mathcal{D}$ is anti-selfadjoint.)

For a vector space of complex polynomials with the property $f(0) = f(1) = 0$ and with inner product $\langle f, g\rangle = \int_0^1 f(x)\overline{g}(x)\,dx$, the operator $i\mathcal{D}$ is self-adjoint. (Here i is the imaginary unit.)

6.5 Exercises

The exercises with (☞) have a hint at the end of Lecture Notes.

Exercise 6.5.1. (Ex. 5.3.7 in Treil) True or false: if W is a subspace of V, then $\dim W + \dim(W^\perp) = \dim V$? Justify.

Exercise 6.5.2. Find all vectors that are perpendicular to (1, 4, 4, 1) and (2, 9, 8, 2).

Exercise 6.5.3. (☞)

In the vector space $V = \mathbb{R}^5$, consider the subspace U spanned by the vectors

$$\begin{bmatrix} 2 \\ 2 \\ 1 \\ 7 \\ -3 \end{bmatrix}, \begin{bmatrix} -4 \\ 1 \\ -12 \\ 6 \\ -4 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 3 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 3 \\ 1 \\ 2 \end{bmatrix}, \text{ and } \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

(a) Compute $\dim U$.

(b) Which of the vectors

$$\begin{bmatrix} 10 \\ 0 \\ 5 \\ -3 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 8 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 4 \\ 0 \\ 0 \end{bmatrix}, \text{ and } \begin{bmatrix} 5 \\ 0 \\ 5 \\ 0 \\ 2 \end{bmatrix}$$

belong to U?

Exercise 6.5.4. (Ex. 5.3.5 in Treil) Find the orthogonal projection of a vector $(1, 1, 1, 1)^t$ onto the subspace spanned by the vectors $v_1 = (1, 3, 1, 1)^t$ and $v_2 = (2, -1, 1, 0)^t$ (note that $v_1 \perp v_2$).

Exercise 6.5.5. (Ex. 5.3.6 in Treil) ☞

Find the distance from a vector $v = [1, 2, 3, 4]^t$ to the subspace spanned by the vectors $v_1 = [1, -1, 1, 0]^t$ and $v_2 = [1, 2, 1, 1]^t$ (note that $v_1 \perp v_2$). One way is to do this is by projecting v on the subspace spanned by v_1 and v_2.

Can you find the distance without actually computing the projection? That would simplify the calculations.

Exercise 6.5.6. Find a matrix A, which is in reduced echelon form, and satisfies $\dim(\text{range}(A^t)^\perp) = 4$, $\dim(\text{range}(A)^\perp) = 1$.

Exercise 6.5.7. (Ex. 5.2.3(c) in Treil) Do Exercise 6.3.10 + the following:

Assume that $v_1, v_2, \ldots, v_n$ is only an orthogonal basis in $\mathbb{R}^n$, not an orthonormal one. Can you write down Parseval's identity in this case?

Exercise 6.5.8. Let A be a **real symmetric** matrix. An **eigenvector** of matrix A is a non-zero vector x such that $Ax = \lambda x$ for some number λ which is called the **eigenvalue** corresponding to the eigenvector x.

Prove that if x and y are eigenvectors corresponding to distinct real eigenvalues λ_1 and λ_2, then x and y are orthogonal.

(This is one of the remarkable properties of self-adjoint matrices.)

Chapter 7

More about Inner Product Spaces

Reading for this Chapter: Treil, Chapter 5.

7.1 Gram-Schmidt orthogonalization

In some cases we are given a basis $(v_1, v_2, \ldots)$ of a linear space V and we want to construct a orthonormal basis $(q_1, q_2, \ldots, q_n)$. The reason for this is that in this basis it is easy to measure distances and perform orthogonal projections.

More generally, we are given an increasing sequence of spaces (a **flag**)

$$V_1 \subset V_2 \subset \ldots \subset V_n,$$

where $V_k = span(v_1, \ldots, v_k)$, and we want to construct an orthonormal system of vectors $q_1, \ldots, q_n$ so that $V_k = span(q_1, \ldots q_n)$. This can be easily done by the process that is called the **Gram-Schmidt orthogonalization**.

The process is recursive. At step 1, we take vector v_1 and normalize it to have the unit length:

$$q_1 = \frac{1}{r_{11}} v_1,$$

where $r_{11} = \|v_1\|$.

At step k we take vector v_k and subtract its projection on the subspace V_{k-1}. This is easy to do because we already know $\{q_1, \ldots, q_{k-1}\}$, which form an orthonormal basis of V_{k-1}. After this, we normalize the resulting vector so that it has the unit length.

So,

$$u_k = v_k - \langle v_k, q_1\rangle q_1 - \ldots - \langle v_k, q_{k-1}\rangle q_{k-1},$$
$$q_k = \frac{1}{r_{kk}} u_k,$$

where $r_{kk} = \|u_k\|$. (Note also that $r_{kk} = \langle u_k, q_k\rangle = \langle v_k, q_k\rangle$.)

The process will continue without interruption, provided that the inclusions $V_{k-1} \subset V_k$ are strict, which is the same as that the matrix A with columns $v_1, \ldots, v_n$ has full rank.

The formulas above can also be written differently, as

$$\begin{aligned} v_1 &= r_{11} q_1, \\ v_2 &= r_{12} q_1 + r_{22} q_2, \\ v_3 &= r_{13} q_1 + r_{23} q_2 + r_{33} q_3, \\ &\ldots \\ v_n &= r_{1n} q_1 + r_{2n} q_2 + \ldots + r_{nn} q_n, \end{aligned}$$

where $r_{ij} = \langle v_j, q_i\rangle$ when $i \le j$.

In a matrix form it can be written as

$$A = \widehat{Q}\widehat{R},$$

where A is an $m \times n$ matrix, $\widehat{Q} = [q_1, \ldots, q_n]$ is an $m \times n$ matrix with orthonormal columns and R is an upper-diagonal $n \times n$ matrix with positive diagonal elements.

$$\widehat{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \ldots & r_{1n} \\ 0 & r_{22} & r_{23} & \ldots & r_{2n} \\ 0 & 0 & r_{33} & \ldots & r_{3n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & r_{nn} \end{bmatrix}$$

This factorization is called the **reduced QR factorization** and the above argument shows that if matrix A has full rank, then this factorization exists and is unique. By extending matrix $\widehat{Q}$ to an orthogonal $m \times m$ matrix Q, and $\widehat{R}$ to an upper-diagonal $m \times n$ matrix R one can obtain the full QR factorization, although this factorization is not unique.

When doing the Gram-Schmidt by hand, it is useful to postpone the normalization steps of the algorithm. In this case we are looking for an

orthogonal basis $\{w_1, \ldots, w_n\}$, and the algorithm goes as follows. First, set $w_1 = v_1$, and then for $k = 2, \ldots, n$.

$$w_k = v_k - \frac{\langle v_k, w_1 \rangle}{\|w_1\|^2} w_1 - \ldots - \frac{\langle v_k, w_{k-1} \rangle}{\|w_{k-1}\|^2} w_{k-1}.$$

Finally, if we want to get an **orthonormal** basis, we perform normalization:

$$q_k = w_k / \|w_k\|.$$

Example 7.1.1. Let $\mathcal{B} = \{[1,0,1,0]^t, [1,1,1,1]^t, [1,2,3,4]^t\}$ be a basis of subspace $V \in \mathbb{R}^4$. Perform the Gram-Schmidt orthogonalization on $\mathcal{B}$ and find an orthonormal basis for V.

We have $w_1 = v_1$,

$$w_2 = v_2 - \frac{\langle v_2, w_1 \rangle}{\|w_1\|^2} w_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

$$w_3 = v_3 - \frac{\langle v_3, w_1 \rangle}{\|w_1\|^2} w_1 - \frac{\langle v_3, w_2 \rangle}{\|w_2\|^2} w_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} - \frac{4}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \frac{6}{2} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}.$$

$\{w_1, w_2, w_3\}$ is an orthogonal basis. To get an orthonormal basis $\{q_1, q_2, q_3\}$, we perform normalization, $q_i = w_i / \|w_i\|$ and get:

$$q_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, q_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, q_3 = \frac{1}{2} \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}.$$

Example 7.1.2. Here is an example of a QR factorization,

$$\begin{aligned} A &= \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} & \sqrt{2} \\ 0 & 1/\sqrt{2} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix} = QR \end{aligned}$$

The columns of Q are obtained as in a previous example, and the entries $r_{i,j}$ of R are computed as $r_{ij} = \langle v_j, q_i \rangle$ when $i \leq j$. (That is, the column j of R are coefficients in the expansion of v_j in the basis $\{q_1, \ldots, q_n\}$ but for $i > j$ the coefficients are zero by construction.)

The Gram-Schmidt orthogonalization is a general process, which can be applied not only to vectors in $\mathbb{R}^m$ but also to functions in a vector space of functions. One only needs to define the inner product of two functions. For example if $f(x)$ and $g(x)$ are two real-valued functions, then we can define the inner product as an integral, provided that the integral is convergent.

Here are examples of the inner products,

$$\langle f, g\rangle := \int_{-\infty}^{\infty} f(x)g(x)\,dx, \text{ or} \tag{7.1}$$

$$\langle f, g\rangle := \int_{-1}^{1} f(x)g(x)\,dx, \text{ or} \tag{7.2}$$

$$\langle f, g\rangle := \int_{-\infty}^{\infty} e^{-x^2} f(x)g(x)\,dx, \text{ or} \tag{7.3}$$

$$\dots$$

Two functions are called orthogonal if their scalar product is zero, and the norm of a function is defined naturally as $\|f\| = \sqrt{\langle f, f\rangle}$.

The Gram-Schmidt orthogonalization can also be applied to a system of functions $f_1(x), \dots, f_n(x)$ and results in a system of orthonormal functions $q_1(x), \dots, q_n(x)$.

For example, many famous families of polynomials can be obtained in this way by applying orthogonalization procedure to polynomials $1, x, x^2, x^3, \dots$ with respect to various scalar products.

The Legendre polynomials are orthonormal with respect to scalar product 7.2, Hermite's polynomials are orthonormal with respect to scalar product 7.3, etc.

This is important for the problems when one approximates functions by other functions.

7.2 Orthogonal projections

If we have a system of n orthonormal vectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_k$ that span a subspace W in an inner-product space V, then we can project every vector $\boldsymbol{x} \in V$ on W by calculating the projected vector

$$P(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{q}_1\rangle \boldsymbol{q}_1 + \dots + \langle \boldsymbol{x}, \boldsymbol{q}_k\rangle \boldsymbol{q}_n, \tag{7.4}$$

and the residual vector

$$\boldsymbol{r} = \boldsymbol{x} - P(\boldsymbol{x}),$$

and we know that $\boldsymbol{r}$ is orthogonal to every vector $\boldsymbol{q}_i$.

So effectively, we decomposed vector $\boldsymbol{x}$ as a sum of two vectors:

$$x = P_W(\boldsymbol{x}) + r,$$

where $P_W(\boldsymbol{x}) \in W$ and $r \in W^\perp$. In particular r is complementary projection of $\boldsymbol{x}$ on $W^\perp$, $r = P_{W^\perp}(v)$. (We use the word "projection" somewhat loosely here. We will soon define it precisely.)

Here P is a linear map. What is its matrix?

If we have the space $V = \mathbb{R}^n$ or $V = \mathbb{C}^n$ with the Euclidean inner product then we know that $\langle \boldsymbol{x}, \boldsymbol{q}_i \rangle = \boldsymbol{q}_i^* \boldsymbol{x}$, and we can write formula (7.4) in a matrix form by using matrix $Q = [\boldsymbol{q}_1, \dots, \boldsymbol{q}_n]$. It easy to see that

$$P(\boldsymbol{x}) = QQ^* \boldsymbol{x},$$

Example 7.2.1. If $\boldsymbol{q}_i$ is the first k coordinate vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$, then the matrix is

$$\begin{bmatrix} I_k & 0 \\ 0 & 0_{n-k} \end{bmatrix}$$

In general, the matrix QQ^* is an Hermitian $m \times m$ matrix with the following property.

$$(QQ^*)^2 = Q(Q^*Q)Q^* = QQ^*,$$

because $Q^*Q = I_n$ by orthonormality of vectors $\boldsymbol{q}_i$. Intuitively it says that projecting the same vector twice on the same subspace does not change the results.

A matrix A is **Hermitian** if $A^* = A$. For real matrices it is the same as being symmetric.

The residual can be written as

$$\boldsymbol{r} = (I - QQ^*)\boldsymbol{x},$$

and it is easy to see that $(I - QQ^*)$ is also a Hermitian matrix that has the property that

$$(I - QQ^*)^2 = I - QQ^*$$

Example 7.2.2. Suppose v is a column vector that has unit length. Then matrix $P = vv^t$ is a matrix of an orthogonal projection. It is called rank-one projection. Geometrically, $Px = v(v^*x)$ is the projection of the vector x on the line L that has the direction vector v.

One particular case is when $v = \frac{1}{\sqrt{n}}[1, 1, \dots, 1]^*$. In this case $vv^* = \frac{1}{n}J$, where J is the $n \times n$ matrix consisting of all 1s.

Example 7.2.3 (Sum of rank-one projections). If a matrix Q has column vectors $q_1, q_2, \ldots, q_n$ which form an orthonormal set, then $P = QQ^*$ is the matrix of the orthogonal projection on the linear space spanned by these vectors. It is sometimes useful to write $P = QQ^*$ as a sum of rank-one projectors from the previous example.

$$P = QQ^* = \sum_{i=1}^{n} q_i q_i^* \tag{7.5}$$

This formula is equivalent to formula (7.4) [check it!]

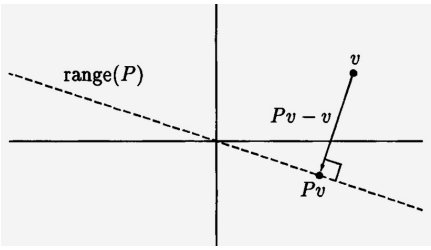

Figure 7.1: Ortogonal Projector

Let us talk about properties of orthogonal projection in a more formal way.

Definition 7.2.4. An **orthogonal projector** on subspace $W \in V$ is a linear transformation P_W that satisfies two properties: $P_W v \in W$ for every $v \in V$ and $v - P_W v \perp W$ for every vector $v \in V$.

The most useful property of orthogonal projectors is that $P_W v$ is the point in W that minimizes the distance of v from W.

Theorem 7.2.5. *The orthogonal projection $w = P_W v$ minimizes the distance from v to W, i.e. for all $x \in W$, $\|v - w\| \leq \|v - x\|$. Moreover, if for some $x \in W$, $\|v - w\| = \|v - x\|$, then $x = w$.*

Proof. Write

$$v - x = v - w + (w - x).$$

Since $v - w \in W^{\perp}$ and $w - x \in W$ hence they are orthogonal and we can apply the Pythagorean theorem:

$$\|v - x\|^2 = \|v - w\|^2 + \|w - x\|^2.$$

This implies both claims of the theorem. □

Here are some other simple properties of orthogonal projections.

Exercise 7.2.6. Let P be an orthogonal projector on $W \subset V$. Show that

(a) If $w \in W$, then $Pw = w$.

(b) If $v \in W^\perp$, then $Pv = 0$.

(c) $P^2 = P$.

(d) $I - P$ is an orthogonal projector on $W^\perp$.

Solution:

(a) $Pw - w \in W \cap W^\perp = \{\vec{0}\}$.

(b) If $v = 0$, the statement is true. Let a non-zero $v \in W^\perp$ and let $w = Pv$. Then, $w \in W$ and $v - w \in W^\perp$ by assumption. Hence $\langle v - w, w\rangle = -\|w\|^2 = 0$. So $w = 0$.

(c) Hint: write $v \in V$ as $v = w + w^\perp$, where $w \in W$, $w^\perp \in W^\perp$, and check the property.

(d) Verify the definition.

Sometimes it is needed to check if a given matrix is a matrix of an orthogonal projection.

Theorem 7.2.7. *Let $V = \mathbb{R}^n$ or $V = \mathbb{C}_n$ with the Euclidean inner product. Suppose $P : V \to V$ is a linear map such that $P^2 = P$ and $P^* = P$. Then P is the orthogonal projector on $W = \operatorname{range} P$.*

(Here we identify P with its matrix in the standard basis and P^* is the linear transformation with the matrix P^*.)

Proof. Clearly, $Pv \in W = \operatorname{range} P$ for every $v \in V$ because of the definition of the range. Then, let $w = v - Pv = (I - P)v$. We want to show that $w \in W^\perp$. Let $u \in W$, then we have $u = Px$ for some $x \in V$ and

$$\langle w, u\rangle = \langle (I - P)v, Px\rangle = (Px)^*(I - P)v = x^*P^*(I - P)v.$$

Since we assumed $P^* = P$ therefore $P^*(I - P) = P(I - P) = 0$. This shows that $(I - P)v \perp W$ and finishes the proof. □

A matrix P that has the property $P^2 = P$ but $P^* \neq P$ is called an **oblique projector**.

7.3 Projection and linear regression

7.3.1 Basic formula

Suppose a subspace $W \subset V$ spanned by vectors $a_1, a_2, \ldots, a_k$ which are not necessarily orthogonal. How can we calculate the orthogonal projection on W?

One way is obvious: apply the Gram-Schmidt orthogonalization to the set $\{a_1, a_2, \ldots, a_k\}$. It will give an orthonormal set $\{q_1, \ldots, q_k\}$ and then we can calculate the projection of x on V as QQ^*x, where $Q = [q_1, \ldots, q_k]$. There is an alternative way that avoids orthogonalization.

Theorem 7.3.1. *Let $V = \mathbb{R}^n$ or $V = \mathbb{C}^n$. Suppose $W = \langle\{a_1, a_2, \ldots, a_k\}\rangle \subset \mathbb{R}^n$, and let A be an $n \times k$ matrix with columns a_i, $i = 1, \ldots, k$, $k \leq n$. Suppose A is full rank:* $\operatorname{rank}(A) = k$. *Then*

$$P = A(A^*A)^{-1}A^* \tag{7.6}$$

is the orthogonal projector on W.

Exercise 7.3.2. ☞
Show that $\ker(A^*A) = \ker A$.

Exercise 7.3.3. Suppose $\operatorname{rank} A = k$, show that $(A^*A)^{-1}A^*$ is a surjection on $\mathbb{R}^k$ and P is a surjection on range(A).

Proof. By Exercise 7.3.2 and the assumption about the full rank, $\dim\ker(A^*A) = \dim\ker A = 0$. Hence A^*A is invertible [why?] and P is well defined. Then by direct checking, $P^2 = P$ and $P^* = P$. Hence by Theorem 7.2.7, P is an orthogonal projector on range P, which equals range A by Exercise 7.3.3. □

7.3.2 Relation to statistics

In statistics we often need to solve the following problem:

$$y_i = \beta_1 x_{i1} + \ldots \beta_n x_{in} + \varepsilon_i, \tag{7.7}$$

where $i = 1, \ldots, m$ labels observations, y_i is the value of the variable that we want to explain in observation i, and $x_{i1}, \ldots, x_{in}$ are the values of n "explanatory" variables in observation i. (They often called "features" in machine learning.) In statistics, a linear regression is usually has a constant term. Here we treat the constant term on the equal basis with other coefficients. For example, we can think about the vector $(x_{11}, \ldots, x_{1n})$ as the vector in which all components are equal to 1.

The numbers ε_i are "error terms".

Another view on this problem is that we simply trying to solve an overdetermined system of equations, where the number of equations m exceeds the number of variables n. In this case, there is no exact solution and we trying to minimize the norm of the vector of the residual terms ε_i.

Let us introduce $m \times 1$ vector $y = [y_1, \ldots, y_m]$, an $m \times n$ matrix X with entries x_{ij}, the $n \times 1$ vector of coefficients $\beta = [\beta_1, \ldots, \beta_n]$, and $m \times 1$ vector of errors $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_n]$.

Then we can re-write equation (7.7) as

$$y = X\beta + \varepsilon,$$

Our task is to minimize the norm of vector ε, which we can write as

$$(y - X\beta)^*(y - X\beta) \to \min$$

We can write the first order conditions as

$$\frac{\partial}{\partial \beta}(y - X\beta)^*(y - X\beta) = \vec{0}.$$

Here, $\frac{\partial}{\partial \beta} f(\beta)$ is the vector of partial derivatives $\frac{\partial}{\partial \beta_j} f(\beta)$. One can check directly that this leads to equations:

$$X^*(y - X\beta) = 0,$$

or

$$X^*X\beta = X^*y. \tag{7.8}$$

(Indeed

$$\frac{\partial}{\partial \beta_j} \sum_i (y_i - \sum_k X_{ik}\beta_k)^2 = -2 \sum_i X_{ij}(y_i - \sum_k X_{ik}\beta_k),$$

and this is equivalent to equation (7.8).)

In the traditional statistics, $m > n$, the number of observations exceeds the number of explanatory variables. For this reason the rank of a typical X equals n, so it is a full rank. It follows that X^*X is invertible and we can solve equation (7.8) as

$$\beta = (X^*X)^{-1}X^*y \tag{7.9}$$

The equations in (7.8) are called **normal equations** and the matrix

$$X^+ = (X^*X)^{-1}X^*$$

is sometimes called the **pseudoinverse** of matrix X.

In statistical applications we are also interested in estimated true values of y_i, when the noise ε_i is filtered out. So we define the fitted values of y as $\widehat{y} = X\beta$. Then

$$\widehat{y} = X(X^*X)^{-1}X^*y.$$

This is the linear combination of explanatory random variables which minimizes the norm of the error term $e = y - X\beta$.

From the point of view of linear algebra, $\widehat{y}$ is the orthogonal projection of vector y on the linear space spanned by the vectors of the explanatory variables $x^{(1)}$, ..., $x^{(n)}$, where $x^{(j)} = [x_{1j}, \dots x_{mj}]^t$. The matrix of the projection is

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

Example 7.3.4. Let

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

What is the orthogonal projection of $v = [1, 2, 3]^t$ on range A?

What is the matrix of the orthogonal projector P onto range(A)

Solution: Method 1: We apply Gram-Schmidt and find $w_1 = a_1 = [1, 0, 1]^t$,

$$w_2 = a_2 - \frac{\langle a_2, w_1\rangle}{\langle w_1, w_1\rangle} w_1 = a_2 = [0, 1, 0]^t.$$

(The columns of A are already orthogonal.) So the projection is

$$Pv = \frac{\langle v, w_1\rangle}{\langle w_1, w_1\rangle} w_1 + \frac{\langle v, w_2\rangle}{\langle w_2, w_2\rangle} w_2 = \frac{4}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \frac{2}{1}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}.$$

In order to get the projection matrix, we normalize the basis

$$q_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \text{ and } q_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Then

$$P = QQ^* = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

Method 2. Linear Regression formula. We calculate

$$A^*v = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$

$$A^*A = \begin{bmatrix} 2, & 0 \\ 0, & 1 \end{bmatrix}$$

$$\begin{bmatrix} 2, & 0 & | & 4 \\ 0, & 1 & | & 2 \end{bmatrix} \to \begin{bmatrix} 1, & 0 & | & 2 \\ 0, & 1 & | & 2 \end{bmatrix} \text{ so } (A^*A)^{-1}A^*v = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$A(A^*A)^{-1}A^*v = \begin{bmatrix} 1, & 0 \\ 0, & 1 \\ 1, & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}.$$

For the projection matrix, we can calculate

$$A^*A = \begin{bmatrix} 2, & 0 \\ 0, & 1 \end{bmatrix}, (A^*A)^{-1} = \begin{bmatrix} 1/2, & 0 \\ 0, & 1 \end{bmatrix}, (A^*A)^{-1}A^* = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix},$$

$$P = A(A^*A)^{-1}A^* = \begin{bmatrix} 1, & 0 \\ 0, & 1 \\ 1, & 0 \end{bmatrix} \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}.$$

For the following examples, see Treil's book.

Example 7.3.5. Projection on a plane.

Example 7.3.6. Curve fitting.

There are some generalizations of the linear regression when one puts different weights on different observations. In this case we can introduce a weighted 2-norm:

$$\|x\|_{2,w} = \sqrt{\sum_{i=1}^{n} w_i x_i^2},$$

where w_i are some positive weights. In this case one wants to minimize

$$\|y - X\beta\|^2_{2,w},$$

and this leads to the formulas

$$\widehat{\beta} = (X^*WX)^{-1}X^*Wy,$$
$$\widehat{y} = X(X^*WX)^{-1}X^*Wy,$$

where W is the diagonal matrix with weights w_i on the main diagonal.

The prediction matrix $X(X^*WX)^{-1}X^*W$ is a projector, but it is not symmetric so this projection is not orthogonal. However, it turns out that it can be thought as orthogonal if we define the orthogonality differently, namely if we say that two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal if $\sum_{i=1}^n w_i u_i v_i = 0$.

7.4 Exercises

The exercises with (☞) have a hint at the end of Lecture Notes.

Exercise 7.4.1. (Ex. 5.3.1 in Treil) Apply Gram–Schmidt orthogonalization to the system of vectors $[1, 2, -2]^t$, $[1, -1, 4]^t$, $[2, 1, 1]^t$.

Exercise 7.4.2. Apply Gram–Schmidt orthogonalization to the system of vectors $v_1 = [1, 0, 1]^t$ and $v_2 = [1, 1, -1]^t$. Using the results, write the matrix of the orthogonal projection onto 2-dimensional subspace spanned by these vectors.

Then calculate this matrix using v_1 and v_2 without orthogonalization and the projection formula (7.6).

Exercise 7.4.3. Find a basis of the orthogonal complement to the subspace spanned by the vectors $[1, 2, 3]^t$, $[1, 3, 4]^t$.

Exercise 7.4.4. ☞

Find the distance from a vector $[2, 3, 1]^t$ to the subspace spanned by the vectors $[1, 2, 3]^t$, $[1, 3, 4]^t$. Note I am only asking to find the distance to the subspace, not the orthogonal projection.

Exercise 7.4.5. (Ex. 5.3.8 in Treil)

Let P be the orthogonal projection onto a subspace E of an inner product space V , $\dim V = n$, $\dim E = r$. Find the eigenvalues and the eigenvectors (eigenspaces). Find the algebraic and geometric multiplicities of each eigenvalue. What is the trace of P?

Exercise 7.4.6 (Legendre's polynomials). Let an inner product on the space of polynomials be defined by $\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Apply the Gram-Schmidt orthogonalization to the system $1, t, t^2, t^3$. Do not normalize the polynomials during the Gram-Schmidt process.

Legendre's polynomials are particular case of the so-called orthogonal polynomials, which play an important role in many branches of mathematics.

Exercise 7.4.7. If P is an orthogonal projector, then the matrix $I-2P$ is orthogonal. Prove this algebraically, and try to give a geometric interpretation for the transformation represented by matrix $I-2P$.

Exercise 7.4.8. ☞

Let E be the $m \times m$ matrix (operator) that extracts the even part of an m-vector: $Ex = (x + Fx)/2$, where F is the $m \times m$ matrix (operator) that flips $(x_1, \ldots, x_m)^t$ to $(x_m, \ldots, x_1)^t$. Is E an orthogonal projector, an oblique projector, or not a projector at all? What are the entries of this matrix?

Exercise 7.4.9. Given an $m \times n$ matrix A with $m \geq n$, show that A^*A is invertible if and only if A has full rank. As a part of this problem, do Exercise 7.3.2.

Exercise 7.4.10. (Ex 5.4.1. in Treil) Find the least square solution of the system

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Exercise 7.4.11. (Ex. 5.4.2. in Treil) Find the matrix of the orthogonal projection P onto the column space of

$$\begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -2 & 4 \end{bmatrix}$$

Use two methods: Gram–Schmidt orthogonalization and Least Squares formula for the projection. Compare the results.

Chapter 8

Orthogonal Diagonalization

Reading for this Chapter: Treil, Chapter 6.

We showed previously that every matrix in $\mathbb{C}_{n\times n}$ can be brought to a Jordan Canonical Form by an appropriate choice of the basis. If the geometric multiplicities of eigenvalues equal to their algebraic multiplicities then the matrix can be diagonalized. For example, this happens if all eigenvalues have algebraic multiplicity 1.

What if we allow only an orthonormal bases? In this case, we cannot hope to diagonalize an arbitrary matrix. However, we can bring every matrix to the upper-triangular form by appropriate choice of the basis. This result is called Schur's factorization.

Importantly, for some classes matrices the diagonalization is always possible. This happens, for example, for Hermitian matrices, that is, for matrices that satisfy the condition $A^* = A$. (In the real case, these are symmetric matrices, $A^t = A$.)

8.1 Schur's factorization

A **Schur** factorization of a matrix A is a factorization $A = QTQ^*$, where Q is unitary and T is upper-triangular.

Theorem 8.1.1. *Every matrix $A \in \mathbb{C}_{n\times n}$ has a Schur factorization.*

Remarks:

(a) If matrix A is real and all its eigenvalues are real then it is seen from the proof that we can choose Q and T to be real in this factorization.

(b) Two matrices A and B are called **unitarily equivalent** if $A = UBU^*$ for a unitary matrix U. So, the theorem says that every square matrix is unitarily equivalent to an upper-triangular matrix.

Proof. The proof is by induction on the dimension n of A. The case $n = 1$ is obvious. Suppose $n \geq 2$. Every matrix A has at least one eigenvalue λ by one of our previous results. Let $\boldsymbol{u}_1$ be a unit eigenvector belonging to λ and set it as a first column of a unitary matrix $U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$. Then, we can check that

$$U^*AU = \begin{bmatrix} \lambda & w^* \\ 0 & B \end{bmatrix},$$

for some $w \in \mathbb{C}^{n-1}$ and $B \in \mathbb{C}_{n\times n}$. (Indeed, this simply says that in the new basis the first column of transformed matrix is $[\lambda, 0, \ldots, 0]$ and this follows from $Au_1 = \lambda u_1$.)

By inductive hypothesis, there exists a Schur factorization VTV^* of B. Then, we can set

$$Q = U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix},$$

and check that

$$Q^*AQ = \begin{bmatrix} \lambda & w^*V \\ 0 & T \end{bmatrix},$$

which is the desired Schur factorization. □

8.2 Spectral theorem for Hermitian matrices

One of the most important properties of Hermitian matrices is that they are diagonalizable and moreover, they admit a unitary diagonalization. This is often called the Spectral Theorem.

Theorem 8.2.1 (Spectral theorem for Hermitian matrices)**.** *If $A^* = A$, then A admits unitary diagonalization:*

$$A = Q\Lambda Q^*, \tag{8.1}$$

where Q is unitary and Λ is diagonal with real entries.

Remark: if A is real then by using the remark about the Schur diagonalization, the proof below shows that Q can also be chosen real.

Note that if (8.1) holds then A is Hermitian. So, in fact we obtained another definition of Hermitian (or self-adjoint) matrix. This is a matrix of an operator which in a certain orthonormal basis has the diagonal matrix with real diagonal entries.

Proof. By Schur's factorization, we can write $A = QTQ^*$, and since $A^* = A$, we have $T^* = T$, which implies that T is diagonal and its diagonal entries are real.

□

This proof is very short but it has a disadvantage that it hides the role of self-adjointness and it is not easy to generalize to infinite-dimensional space. More conceptually, one can proceed along the following lines.

Exercise 8.2.2. Suppose λ be an eigenvalue of self-adjoint operator A. Prove that λ is real.

Exercise 8.2.3. Show that $(E_\lambda)^\perp$ is A-invariant.

Then, we have a direct sum of A-invariant sub-spaces: $\mathbb{R}^n = E_\lambda \oplus (E_\lambda)^\perp$ and therefore we can study A on these two sub-spaces separately. We can choose an orthogonal basis in E_λ and in this basis $A\big|_{E_\lambda}$ is diagonal, $= \lambda I$. In addition, the subspace $(E_\lambda)^\perp$ has dimension smaller than n and we can apply induction and conclude that there is a basis of $(E_\lambda)^\perp$, in which $A\big|_{(E_\lambda)^\perp}$ is diagonal. Then we can take the union of these bases, and see that A is diagonal in this basis. This complete the alternative proof of Theorem 8.2.1.

The arguments in the proof above implicitly showed that $E_\lambda \perp E_\mu$ if eigenvalues λ and μ are distinct. Let us show this explicitly.

Theorem 8.2.4. *If $\lambda_1 \neq \lambda_2$ are two eigenvalues of Hermitian matrix A with eigenvectors u_1 and u_2, respectively, then $u_1 \perp u_2$.*

Proof. By the previous theorem the eigenvalues are real and

$$\begin{aligned} \langle u_2, Au_1\rangle = \langle Au_2, u_1\rangle, \text{ so} \\ \lambda_1\langle u_2, u_1\rangle = \lambda_2\langle u_2, u_1\rangle \end{aligned}$$

and since $\lambda_1 \neq \lambda_2$, we must have $\langle u_2, u_1\rangle = 0$. □

From Theorem 8.2.1, we have that A is diagonalizable and therefore

$$\mathbb{R}^n = E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_r},$$

where $\lambda_1, \ldots, \lambda_r$ are distinct eigenvalues of A and E_{λ_i} are corresponding eigenspaces. Theorem 8.2.4 shows that the $E_{\lambda_i} \perp E_{\lambda_j}$ for $i \neq j$.

The formula (8.1) is often written in the following form:

$$A = \sum_{i=1}^{n} \lambda_i q_i q_i^*,$$

where λ_i are eigenvalues of A (counted with multiplicities) and $\{q_i\}$ is an orthonormal basis of eigenvectors.

It can also be written as

$$A = \sum_{i=1}^{r} \lambda_i P_{E_{\lambda_i}}, \tag{8.2}$$

where λ_i runs now over all distinct eigenvalues of A and $P_{E_{\lambda_i}}$ is the orthogonal projection on eigenspace E_{λ_i}. Note that eigenspaces E_{λ_i} are orthogonal to each other and $\mathbb{R}^n = E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_r}$. In terms of orthogonal projections $P_{E_{\lambda_i}}$ this means that $P_{E_{\lambda_1}} + \ldots + P_{E_{\lambda_r}} = I_n$ and $P_{E_{\lambda_i}} P_{E_{\lambda_j}} = 0$ if $i \neq j$.

Formula (8.2) elucidate the origin of the name "Spectral Theorem". It is inspired by the parallel drawn between eigenvalues and the frequencies found within the light spectrum. Just as frequencies are fundamental components of the light spectrum, eigenvalues serve a similar foundational role in the theorem. Furthermore, the matrix is expressed as a sum of weighted projections, mirroring the way white light is decomposed into a spectrum of pure frequency light. This analogy underscores the theorem's capacity to break down complex matrices into more interpretable and fundamental elements, akin to the dispersion of white light through a prism.

8.3 Spectral theorem for normal operators

Some other classes of matrices also admit unitary diagonalization.

Definition 8.3.1. An operator (matrix) A is called normal, if $A^*A = AA^*$.

One example of normal matrices which are not Hermitian are unitary matrices.

Theorem 8.3.2. *If A is normal, then there are a unitary matrix U and a diagonal matrix D, so that*

$$A = UDU^*.$$

Note that in this case we do not claim that the diagonal terms of D are real. For example, for unitary matrices A the elements of D (i.e., the eigenvalues of A) belong to the unit circle in the complex plain. (Can you prove it?)

This theorem gives a characterization of normal matrices. These are matrices of an operator, which have a diagonal matrix in an orthonormal basis. (However, in difference from self-adjoint operators, the diagonal matrix is not required to be real.)

Proof. The claim of the theorem also follows from Schur's factorization. One can prove it by induction on n that for an upper-triangular matrix T, $T^*T = TT^*$ can hold only if T is diagonal. For details see Theorem 6.2.4 in Treil's textbook. □

8.4 Simultaneous diagonalization

Definition 8.4.1. Two Hermitian matrices A and B are called **simultaneously diagonalizable** if we can find a unitary matrix U such that

$$A = U\Lambda_A U^*,$$
$$B = U\Lambda_B U^*,$$

where Λ_A and Λ_B are the diagonal matrices with eigenvalues of A and B, respectively, on the main diagonal.

Theorem 8.4.2. *Hermitian matrices A and B are simultaneously diagonalizable if and only they commute, that is, if $AB = BA$.*

Proof. One direction is clear. If A and B are simultaneously diagonalizable, then by direct verification they commute. (Check it.)

Now assume that $AB = BA$. We proceed by induction. The case when the size of matrix $n = 1$ is clear. So suppose $n \geq 2$. We claim that A and B have a common eigenvector. Let x be an eigenvector of A with eigenvalue λ, $x \in E_\lambda(A)$. Then for every integer $k \geq 1$,

$$AB^k x = B^k Ax = \lambda B^k x,$$

so $B^k x \in E_\lambda(A)$. Note that $W = \operatorname{span}\{x, Bx, B^2x, \ldots\}$ is a finite dimensional subspace of $\mathbb{R}^n$ which is B-invariant. Hence $B\big|_W$ has a unit length eigenvector w in $W \subset E_\lambda(A)$. It follows that w is an eigenvector of both A and B. Consider subspace $\operatorname{span}\{w\}^\perp$. Since A and B are Hermitian, this

subspace is A and B invariant (check it!) and we can apply the induction hypothesis to A and B restricted to $\text{span}\{w\}^\perp$. As a result, we find an orthonormal basis in $\text{span}\{w\}^\perp$, in which the restriction of operators A and B are diagonal. By adding w to this basis, we find the orthonormal basis of $\mathbb{R}^n$ in which A and B are diagonal. □

This theorem can be generalized. If Hermitian operators $A_1, \ldots, A_m$ pairwise commute ($A_iA_j = A_jA_i$ for all i, j) then they can be simultaneously diagonalized by a unitary matrix U. The proof is based on the following exercise.

Exercise 8.4.3. Suppose operators $A_1, \ldots, A_m$ pairwise commute. By using induction, prove that they have a common eigenvector.

8.5 Positive definite matrices

Definition 8.5.1. A self-adjoint operator $A : V \to V$ is called **positive definite** if $\langle Av, v\rangle > 0$ for every $v \neq 0$. It is called **positive semidefinite** if $\langle Av, v\rangle \geq 0$ for every $v \in V$.

The notation for positive definite and positive semidefinite matrices A are $A > 0$ and $A \geq 0$, respectively.

One reason to study positive definite matrices is that we can define a new inner product on $\mathbb{R}^n$ by the formula $\langle u, v\rangle_A = \langle Au, v\rangle$ and in fact these are all possible inner products. In addition, positive semidefinite matrices often arise in practice as covariance matrices of random vectors. They also arise in mathematical analysis. (If the gradient of a multivariate function $f(x_1, \ldots, x_n)$ is zero $\nabla f(\vec{x}) = 0$ and the Hessian matrix H with entries $H_{ij} = \partial_{x_i}\partial_{x_j} f(\vec{x})$ is positive definite, then the matrix has a strict minimum at $\vec{x}$.)

Exercise 8.5.2. Let $A : \mathbb{R}^n \to R^m$. Then operator A^*A is positive semidefinite.

Theorem 8.5.3. *Let $A = A^*$. Then,*

1. *$A > 0$ if and only if all eigenvalues of A are positive.*
2. *$A \geq 0$ if and only if all eigenvalues of A are non-negative.*

Proof. The idea is to choose the orthogonal basis in which the matrix of A is diagonal, with eigenvalues on the main diagonal. In this basis the claim is clear.

More formally, by the Spectral Theorem, $A = U\Lambda U^*$, where U is orthogonal. Then $\langle Av, v\rangle = \langle U\Lambda U^* v, v\rangle = \langle \Lambda U^* v, U^* v\rangle$. Let $u = U^* v = [u_1, \dots u_n]^*$. Note that $\|u\| = \|v\|$, so $u \neq 0$ if and only if $v \neq 0$. Then $\langle Av, v\rangle = \sum_{i=1}^n \lambda_i |u_i|^2$. This is > 0 for all $u \neq 0$ if and only if all $\lambda_i > 0$. And it is ≥ 0 for all u if and only if all $\lambda_i \geq 0$. □

Corollary 8.5.4. *Let $A = A^* \geq 0$ be a positive semidefinite operator. There exists a unique positive semidefinite operator B such that $B^2 = A$.*

Proof. If the spectral decomposition for A is $A = U\Lambda U^*$, where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, then by the previous theorem all $\lambda_i \geq 0$ and we can define $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ and then we can take $B = U\Lambda^{1/2}U^*$. For uniqueness see the textbook. □

Definition 8.5.5. Let A is an operator from $\mathbb{R}^n$ to $\mathbb{R}^m$. The **modulus** of A is the operator $|A| : \mathbb{R}^n \to \mathbb{R}^n$ defined as $|A| := \sqrt{A^* A}$.

The modulus of A is useful because it is a positive-semidefinite operator that captures the size of operator A. In particular it is possible to prove that every operator from R^n to R^n has a **polar decomposition**:

$$A = U|A|,$$

where U is a unitary operator. For a proof see Thm 6.3.5 in Treils's book.

8.6 Exercises

Exercise 8.6.1. (Ex. 6.2.1 in Treil) True or false:

a) Every unitary operator $U : X \to X$ is normal.

b) A matrix is unitary if and only if it is invertible.

c) If two matrices are unitarily equivalent, then they are also similar.

d) The sum of self-adjoint operators is self-adjoint.

e) The adjoint of a unitary operator is unitary.

f) The adjoint of a normal operator is normal.

g) If all eigenvalues of a linear operator are 1, then the operator must be unitary or orthogonal.

h) If all eigenvalues of a normal operator are 1, then the operator is identity.

i) A linear operator may preserve norm but not the inner product.

Exercise 8.6.2. Let

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

A has exactly two distinct eigenvalues, which are -2, and 1.

If possible, construct matrices P and D such that $A = PDP^t$, P is a matrix with orthonormal columns, and D is a diagonal matrix.

Exercise 8.6.3. (Ex. 6.2.2 in Treil) True or false: The sum of normal operators is normal? Justify your conclusion.

Exercise 8.6.4. (Ex. 6.2.3 in Treil) Show that an operator unitarily equivalent to a diagonal one is normal.

Exercise 8.6.5. (Ex. 6.2.4 in Treil) Orthogonally diagonalize the matrix,

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

Find all square roots of A, i.e. find all matrices B such that $B^2 = A$.

Exercise 8.6.6. (Ex. 6.2.4 in Treil) True or false: any self-adjoint matrix has a self-adjoint square root. Justify.

Exercise 8.6.7. (Ex. 6.2.7 in Treil) True or false:

a) A product of two self-adjoint matrices is self-adjoint.

b) If A is self-adjoint, then A^k is self-adjoint.

Exercise 8.6.8. Do Exercise 8.5.2: Let $A : \mathbb{R}^n \to R^m$. Then operator A^*A is positive semidefinite.

Exercise 8.6.9. ☞

Suppose A and B are positive semidefinite symmetric matrices. Show that $AB + I$ is invertible.

Chapter 9

Singular Value Decomposition (SVD)

Reading for this Chapter

9.1 Matrix norms

Matrices form a linear space so we can talk about norms of matrices. Since matrices also have some additional structure: for example, they act on vectors, – there are some additional issues for matrix norms.

The two most popular matrix norms are **Frobenius** and **operator** norms. The **Frobenius norm** is defined as follows:

$$\|A\|_F := \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^2} = \sqrt{\mathrm{Tr}(A^*A)},$$

where Tr is the trace: $\mathrm{Tr}M = \sum_{i=1}^{n} M_{ii}$.

It is easy to see that the Frobenius norm of A is simply the norm of the long vector formed by stacking all column vectors of A together. The benefit of this norm is that it is essentially our familiar vector norm, in particular, there is an associated inner product: $\langle A, B\rangle = \mathrm{Tr}(B^*A)$. One of the big advantages of the Frobenius norm is that it is easy to calculate. Another useful norm, or rather a family of norms, is called the **operator norm** and it is defined by the following formula:

$$\|A\| := \sup_{v\neq 0}\frac{\|Av\|}{\|v\|} = \sup_{v:\|v\|=1}\|Av\|. \tag{9.1}$$

The operator norm depends on which vectors norms we choose to use to measure $\|v\|$ and $\|Av\|$. The most frequent situation is when both are usual Euclidean norms, that is, ℓ^2 vector norms: $\|v\| = (\sum v_j^2)^{1/2}$. Sometimes, to make this clear, the operator norm can be denoted $\|A\|_{(2,2)}$ or $\|A\|_2$. Below, if we say "operator norm" without qualifier we mean the 2-norm $\|A\|_2$.

From (9.1), it is clear that the operator norm equals the maximum increase in the length of a vector which is achieved by the linear transformation that have matrix A. Obviously, this is a useful quantity but it is more difficult to calculate.

Example 9.1.1. Let D is an $m \times n$ diagonal matrix with diagonal elements $d_1 \geq d_2 \geq \ldots \geq d_n \geq 0$. (We assume $m \geq n$.) What are the Frobenius and the operator norms of this matrix?

So far we talked about matrix norms as functions on matrices that satisfy the axioms of vector norms. However, sometimes additional requirements are imposed on matrix norms, which are related to such operations on matrices such as taking the adjoint (or transposition) and the multiplication. In particular, it is usually required that

$$\|A^*\| = \|A\|,$$

and

$$\|AB\| \leq \|A\|\|B\|.$$

Both the operator norm and the Frobenius norm satisfy these properties. For the operator norm it is essentially by definition and for the Frobenius norm it is an exercise based on the Cauchy-Schwarz inequality. (See Trefethen-Bau textbook, p.23, for a derivation.)

Another important property of these two norms is that they are invariant relative to unitary transformations.

Theorem 9.1.2. *For every $m \times n$ matrix A and every unitary $m \times m$ matrix Q, we have*

$$\|QA\|_2 = \|A\|_2, \text{ and}$$
$$\|QA\|_F = \|A\|_F.$$

9.2 Definition and existence of SVD

Recall that the motivation for the eigenvalue decomposition of a square matrix A is to find a basis $\{v_i\}_{i=1}^n$, in which the linear transformation A has

the simplest possible form: $A : v_i \to \lambda_i v_i$. When we can find such decomposition, it gives us an enormous insight in the properties of A. However, there are some problems with this approach.

1. Matrix A must be square, that is, the linear transformation must be an endomorphism: it maps the linear space to itself.
2. In many cases we encounter complex eigenvalues and eigenvectors
3. The basis of eigenvectors is not always orthogonal, which means that it is not easy to measure distances in this basis.
4. The eigenvalue decomposition does not always exist and we need to use the Jordan matrices instead of diagonal matrices.

These problems disappear for symmetric matrices but it is a big restriction.

The SVD decomposition is a different approach to the study of properties of linear transformations. Suppose $A : \mathbb{R}^n \to \mathbb{R}^m$. Then we look for two orthonormal bases $\{v_i\}_{i=1}^n$ and $\{u_i\}_{i=1}^m$ such that

$$Av_i = \sigma_i u_i, \text{ for all } i \leq \min\{n, m\}, \tag{9.2}$$

where $\sigma_i \geq 0$ are some real non-negative numbers. If $n > m$ we also require that $Av_i = 0$ for all $i > \min\{n, m\}$. (In fact this requirements is satisfied automatically.)

In matrix form, this is equivalent to the following definition.

Definition 9.2.1. A singular value decomposition (SVD) of an $m \times n$ matrix A is the following product

$$A = U\Sigma V^*, \tag{9.3}$$

where U is an $m \times m$ unitary matrix, V^* is an $n \times n$ unitary matrix and Σ is an $m \times n$ diagonal matrix with real **non-negative** entries. That is, if $i \neq j$ then $\Sigma_{ij} = 0$, otherwise $\Sigma_{ii} \geq 0$.

The diagonal elements of the matrix Σ are called singular values and denoted σ_i. For a real matrix A all elements in matrices U and V can be chosen to be real (so in particular, U and V are orthogonal matrices). By convention, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$. This is always can be achieved by re-arranging columns of U and V.

The columns of matrices V and U are the orthonormal bases $\{v_i\}_{i=1}^n$ and $\{u_i\}_{i=1}^m$ that we introduced above and it is easy to see that property (9.2) holds. Moreover, we also have the following property:

$$A^* u_i = \sigma_i v_i, \text{ for all } i \leq \min\{n, m\}. \tag{9.4}$$

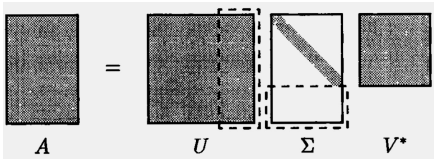

Figure 9.1: Full SVD decomposition, $m > n$

Intuitively, for $m \geq n$, if A represent a linear transformation, then we can write it as a rotation in $\mathbb{R}^n$, represented by V^*, followed by a map Σ that stretches the result and imbeds it isometrically to $\mathbb{R}^m$, and completed by another rotation in $\mathbb{R}^m$, represented by U.

In particular, this interpretation suggests that a unit sphere in $\mathbb{R}^n$ will be mapped to an ellipsoid in $\mathbb{R}^m$ and the half-lengths of the ellipsoid's principal axes will be equal to the singular values $\sigma_i := \Sigma_{ii}$.

The decomposition is clearly not unique if $m > n$. In the picture, the portion of the matrix U selected by dashed lines will be multiplied by zeros in the matrix Σ. Therefore, this portion can be chosen arbitrarily. Intuitively, we can rotate the orthogonal complement to the range of the map A in arbitrary way.

If we want to remove this source of non-uniqueness, then it is useful to define a **reduced singular value decomposition**. Assume that $m \geq n$ and that A is full rank, so that its range space has dimension n. Then the reduced SVD is

$$A = \widehat{U}\widehat{\Sigma}V^*, \tag{9.5}$$

where $\widehat{U}$ is an $m \times n$ matrix that has an orthonormal set of columns. Matrix $\widehat{\Sigma}$ is a square $n \times n$ diagonal matrix. And matrix V^* is the same as in full SVD, that is, it is an $n \times n$ unitary matrix.

In the reduced SVD, $\widehat{U}$ is not is not square (if $m \neq n$) and therefore it is not unitary. However, $\widehat{U}^*\widehat{U} = I_n$. Intutitively, the matrix $\widehat{U}$ is an isometric embedding of $\mathbb{R}_n$ in $\mathbb{R}_m$. Its columns give an orthonormal basis in the image of this embedding.

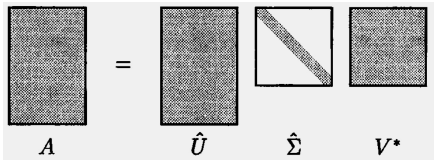

Figure 9.2: Reduced SVD decomposition, $m > n$

The reduced SVD is still not unique. However, this non-uniqueness is mild. It is up to permutation of certain columns and rows in these matrices and up to multiplication of columns and rows by ± 1. It can be almost fixed by requiring that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$ and that the first elements in columns of U and rows of V^* are positive. In exceptional cases when some σ_i are equal, some additional effort may be needed to get the uniqueness, however, this rarely happens in practice.

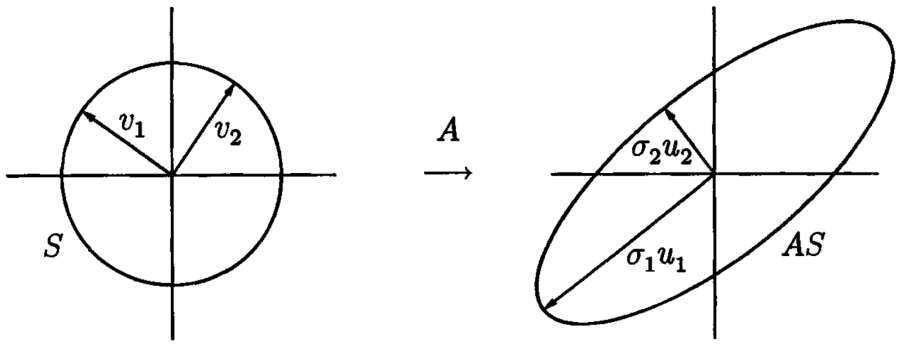

Figure 9.3: SVD decomposition of a 2×2 matrix

Geometric meaning of matrices $\widehat{U}, \widehat{\Sigma}, V$

If $v_1, v_2, \ldots v_n$ are the columns of V, $u_1, u_2, \ldots u_n$ are the columns of $\widehat{U}$, and $\sigma_1, \ldots, \sigma_n$ are the diagonal entries of $\widehat{\Sigma}$, then matrix A sends $v_i \to \sigma_i u_i$. See illustration in Figure 9.3.

The existence of the SVD decomposition

Here is Theorem 4.1 from Trefethen - Bau.

Theorem 9.2.2. *Every matrix A has a singular value decomposition (9.5). Furthermore the singular values σ_i are uniquely determined. If A is square and the σ_i are distinct then the corresponding column vectors in U and V are uniquely determined up to a multiplication by a scalar that have absolute value* 1.

For the complete proof, see the Trefethen-Bau book. Here is a sketch of the proof of the existence claim for $m \geq n$.

Proof of the existence claim. For concreteness, let us work with real matrices.

We will use induction on the size of the matrix and leave the base of the induction (the case when $n = 1$) as an exercise.

By a compactness argument, the supremum in the definition of the matrix norm (9.1) is attained on a vector v_1, and so there exist vectors u_1 and v_1 such that $u_1 = Av_1/\|A\|$, $|v_1| = 1$, $|u_1| = 1$. (In addition it can be proved that for a real matrix A, the maximizing vector v_1 can be chosen to be real.)

Let us define $\sigma_1 = \|A\|$, so we have $u_1 = \sigma_1 A v_1$. Complete the vectors u_1 and v_1 to a pair of orthonormal bases $\{u_i\}$ and $\{v_j\}$ in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Let U_1 and V_1 be the matrices with columns u_i and v_i, respectively.

Then from $Av_1 = u_1$ we have that

$$U_1^* A V_1 = S = \begin{bmatrix} \sigma_1 & w^* \\ 0 & B \end{bmatrix}.$$

We claim that in fact if the norm of A is attained on v_1, then the vector w must be zero.

Indeed, S is obtained from A by a multiplication by two orthogonal matrices on both sides, so it has the same norm as A, that is, $\|S\| = \sigma_1$. Then, we notice that the first element of the vector

$$S \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} = \begin{bmatrix} \sigma_1 & w^* \\ 0 & B \end{bmatrix} \begin{bmatrix} \sigma_1 \\ w \end{bmatrix}$$

is $\sigma_1^2 + w^* w$. Hence

$$\left\| S \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\| \geq \sigma_1^2 + w^* w = \left(\sigma_1^2 + w^* w\right)^{1/2} \left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|$$

So $\|S\| \geq \left(\sigma_1^2 + w^* w\right)^{1/2}$, so it must be that $w = 0$.

Also note that $\|B\| \leq \|A\|$.

However, then we can apply the induction hypothesis to the matrix B and notice that it can be written as $B = U_2 \Sigma_2 V_2^*$.

This leads to the decomposition

$$A = U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix}^* V_1^*,$$

which gives an SVD for matrix A. □

Note that in fact, the proof gave us more than the existence of the SVD. It also showed that the largest singular value σ_1 equals to the maximum of $\|Ax\|$ subject to the constraint $\|x\| = 1$, and that the maximum is achieved at the right singular vector v_1.

Moreover, by analyzing the proof, we see that $\sigma_2 = \max \|Ax\|$ subject to $\|x\| = 1$ and an additional constraint $x \perp v_1$ and that this maximum is achieved at v_2. We can continue this and find that $\sigma_k = \max \|Ax\|$ subject to the constraints that $\|x\| = 1$ and that $x \perp \operatorname{span}(v_1, \ldots, v_{k-1})$.

9.3 Relation to eigenvalue decomposition

Previously, we learned that many matrices can be diagonalized and represented in the form $A = X\Lambda X^{-1}$ where Λ is the diagonal matrix of eigenvalues and X is the matrix, whose columns are eigenvectors.

For general matrices, the connection between eigenvalues and singular values is not straightforward. There is a bunch of inequalities between the singular values and absolute values of eigenvalues.

The eigenvalue diagonalization is very useful when matrix A is symmetric (or Hermitian in the complex case). In this case, all eigenvalues are real and one can choose eigenvectors in such a way that they form an orthonormal set, so that matrix X is orthogonal. This is very close to the SVD decomposition and the difference is that some eigenvalues may happen to be negative, while all singular values must be non-negative.

Theorem 9.3.1. *If A is a symmetric $n \times n$ matrix, then the singular values of A are the absolute values of the eigenvalues of A, $\sigma_i = |\lambda_i|$, for $i = 1, \ldots, n$.*

Proof. In the case of symmetric (or Hermitian) matrices, we have the eigenvalue decomposition:

$$A = Q\Lambda Q^*,$$

where Λ and Q are diagonal and orthogonal (or unitary) matrices, respectively. We can easily convert it to the SVD decompositions by multiplying some of the columns by -1,

$$A = Q|\Lambda|\text{sign}(\Lambda)Q^*,$$

where $|\Lambda|$ is the diagonal matrix with $|\lambda_i|$ on the main diagonal and $\text{sign}(\Lambda)$ is the diagonal matrix with the diagonal entries $\text{sign}(\lambda_i)$. We can also choose the ordering of λ_i in such a way that its absolute values decrease: $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$. This decomposition shows that $\sigma_i = |\lambda_i|$. □

The proof also shows that in the SVD decomposition of a symmetric or an Hermitian matrix, $U = Q$ and V equals Q with some of the columns multiplied by -1.

For more general matrices, we can still use the eigenvalue diagonalization to calculate the SVD.

Theorem 9.3.2. *Let A be an $m \times n$ matrix and $t = \min\{m, n\}$. Matrices A^*A and AA^* have the same sets of t eigenvalues with the largest absolute value, and the t singular values of A are the square roots of these eigenvalues.*

Proof. Let the (full) singular value decomposition for A be

$$A = U\Sigma V^*,$$

where Σ is $m \times n$ matrix and U and V are orthogonal. Then,

$$A^*A = V(\Sigma^*\Sigma)V^*,$$
$$AA^* = U(\Sigma\Sigma^*)U^*$$

where A^*A is $n \times n$ matrix and A^*A is $m \times m$. The matrices $\Sigma^*\Sigma$ and $\Sigma\Sigma^*$ are diagonal and its first t diagonal elements are $\sigma_1^2, \ldots, \sigma_t^2$.

Hence the first t eigenvalues of A^*A with the largest absolute value are the same as the first t eigenvalues of AA^* with the largest absolute value, and both sets are equal to $\{\sigma_1^2, \ldots, \sigma_t^2\}$. □

Note that the proof also shows that V corresponds to eigenvectors of matrix A^*A. And U of the full SVD can be calculated as the matrix of eigenvectors of AA^*.

In the situation when $m > n$ we are typically interested in the reduced SVD and we can observe that A maps column vectors of V to column vectors of $\widehat{U}$ (the U matrix of the reduced decomposition, except it stretches them by the singular values. Hence, we can calculate $u_i = (1/\sigma_i)Av_i$. The situation with $\sigma_i = 0$ is special. In this case one can simply take a unit vector u_i which is perpendicular to all other left-singular vectors. (See an example below for an illustration.)

Note also that this theorem gives another proof of Theorem 9.3.1, since for a real symmetric matrix A, we have $A^*A = A^2$ and the eigenvalues of A^2 are equal to the squares of eigenvalues of A. So, by Theorem 9.3.2, singular values of A are equal to $\sqrt{\lambda_i^2} = |\lambda_i|$, absolute values of eigenvalues of A.

Example 9.3.3. Find the (reduced) SVD decomposition of the matrix

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

We have

$$A^*A = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$$

Then we can calculate the eigenvalues $\lambda_1 = 18$, $\lambda_2 = 0$. (This can be done either by finding the roots of the characteristic polynomial, or by noticing

that the matrix is singular, so one of the eigenvalues must be zero, and finding the second one from the fact that the trace of the matrix is equal to the sum of the eigenvalues.) Hence the singular values are $\sigma_1 = \sqrt{18} = 3\sqrt{2}$, $\sigma_2 = 0$. The eigenvectors of A^*A are $v_1 = \frac{1}{\sqrt{2}}[1, -1]^*$ and $v_2 = \frac{1}{\sqrt{2}}[1, 1]^*$. These are right singular vectors.

We calculate the first left singular vector as

$$u_1 = \frac{1}{3\sqrt{2}} A v_1 = \frac{1}{3\sqrt{2}} \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$$

Since $\sigma_2 = 0$, we can take any unit vector perpendicular to u_1 as the second left singular vector. For example, $u_2 = \frac{1}{\sqrt{5}}[2, 1, 0]^*$ will do.

So, one possible reduced SVD of A is

$$A = \begin{bmatrix} 1/3 & 2/\sqrt{5} \\ -2/3 & 1/\sqrt{5} \\ 2/3 & 0 \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

9.4 Properties of the SVD and singular values

Theorem 9.4.1. *Let $A = U\Sigma V^*$ be the full SVD of A and let r be the number of non-zero singular values. Then*

$$\begin{aligned} \text{range}(A) &= \text{span}\{u_1, \ldots, u_r\}, \\ \text{Null}(A) &= \text{span}\{v_{r+1}, \ldots, v_n\}, \end{aligned}$$

where u_i and v_j are columns of matrices U and V respectively. In particular the rank of A equals r.

Proof. The matrices U and V are full rank orthogonal matrices. Essentially they simply rotate $\mathbb{R}^m$ and $\mathbb{R}^n$. What is important is that the $\text{range}(\Sigma) = \text{span}\{e_1, \ldots, e_r\}$ in $\mathbb{R}^m$ and $\text{Null}(\Sigma) = \text{span}\{e_{r+1}, \ldots, e_n\}$ in R^n. □

The operator and Frobenius norms of a matrix can be written in terms of its singular values.

Theorem 9.4.2. *Let $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ be non-zero singular values of matrix A. Then,*

$$\begin{aligned} \|A\|_2 &= \sigma_1, \\ \|A\|_F &= \sqrt{\sigma_1^2 + \ldots + \sigma_r^2}. \end{aligned}$$

Proof. Note that multiplication by an orthogonal (or unitary) matrix does not change the norm of a vector. This implies that $\|A\|_2 = \|\Sigma\|_2$, and it is easy to check that $\|\Sigma\|_2 = \sigma_1$. For the Frobenius norm, we calculate:

$$\begin{aligned}\|A\|_F^2 = \mathrm{Tr}(A^*A) &= \mathrm{Tr}\Big((V\Sigma^*U^*)(U\Sigma V^*)\Big) \\ &= \mathrm{Tr}\Big(V\Sigma^*\Sigma V^*\Big) = \mathrm{Tr}(\Sigma^*\Sigma),\end{aligned}$$

where the last step is by the property of the trace: $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

And the last quantity is easy to calculate:

$$\mathrm{Tr}(\Sigma^*\Sigma) = \sigma_1^2 + \ldots + \sigma_r^2.$$

□

Now let us consider the relation of eigenvalues and singular values to the determinant. For eigenvalues, we have seen that $\det(A) = \prod_{i=1}^n \lambda_i$. If matrix A has an eigenvalue decomposition, then

$$\begin{aligned}\det(A) = \det(X\Lambda X^{-1}) &= \det(X)\det(\Lambda)\det(X)^{-1} \\ &= \det(\Lambda) = \prod_{i=1}^n \lambda_i.\end{aligned}$$

In general it follows because $\det(zI - A) = (z - \lambda_1)\ldots(z - \lambda_n)$ by setting $z = 0$.

It turns out that we can also write a similar formula using the singular values, except that we lose the information about the sign of the determinant.

Theorem 9.4.3. *For an $m \times m$ matrix A,*

$$|\det(A)| = \prod_{i=1}^m \sigma_i,$$

where σ_i are singular values of the matrix A.

Proof. By using the multiplicative property of the determinant, we write:

$$\det(A) = \det(U\Sigma V^*) = \det(U)\det(\Sigma)\det(V^*)$$

Now we use the fact that the determinant of a unitary matrix has absolute value 1. (This holds because (i) $\det(U)\det(U^*) = \det(UU^*) = 1$, and (ii)

$\det(U^*) = \overline{\det(U)}$. Hence $|\det(U)|^2 = 1$, and therefore $|\det(U)| = 1$.) Therefore

$$|\det(A)| = |\det(\Sigma)| = \prod_{i=1}^{m} \sigma_i.$$

□

9.5 Low-rank approximation via SVD

The SVD is useful because it allows us to construct low-rank approximations to a matrix which are optimal both in the Frobenius and operator norms.

Given an integer $\nu \geq 1$, a **rank-ν approximation** to a matrix A in a norm $\|\cdot\|$ is a matrix B that has rank ν and minimizes the norm of the difference $A - B$.

Theorem 9.5.1. *Let an $m \times n$ matrix A has rank r, and let $A = U\Sigma V^*$ be its SVD, with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$. Then*

$$A_\nu = \sum_{j=1}^{\nu} \sigma_j u_j v_j^*$$

is a rank-ν approximation to A in the operator norm. Moreover, for $\nu < r$ the error of the approximation

$$\inf_{B:\mathrm{rank}(B)\leq\nu} \|A - B\| = \|A - A_\nu\| = \sigma_{\nu+1}.$$

(For $\nu \geq r$, $A_\nu = A$.)

Proof. Suppose that there is some matrix B with the rank $\leq \nu$, which outperform A_ν. Namely, suppose that $\|A - B\|_2 < \|A - A_\nu\|_2 = \sigma_{\nu+1}$. Since the matrix B has rank $\leq \nu$, therefore its null-space W has dimension $\geq n - \nu$. For every vector in $w \in W$, we have

$$\|Aw\| = \|(A - B)w\| < \sigma_{\nu+1}\|w\|.$$

On the other hand, if V is the linear subspace spanned by the first $\nu + 1$ singular vectors of A, then we have that for every $v \in V$,

$$\|Av\| \geq \sigma_{\nu+1}\|v\|.$$

(Exercise: Prove this statement.) Since the sum of the dimensions of W and V exceeds n, they must have a non-zero vector in common. This gives a contradiction. □

An analogous result holds also for the Frobenius norm.

9.6 Principal Component Analysis (PCA)

The singular value decomposition is often used in data analysis for dimension reduction. The basic idea that we are trying to approximate a matrix of data with a low-rank matrix.

Suppose X is the matrix of data. The rows of this matrix are data points and the columns give values of various variables (also called features) for these datapoints. For example, rows can correspond to different individuals and columns to different characteristics of the individual. For another example, rows can correspond to dates and the columns to different financial stocks while the entries are the stock returns recorded on that day.

One statistical technique to analyze data X is called the principal component analysis. It is essentially the SVD of matrix X.

If we write the reduced SVD of $m \times n$ matrix X: $X = U\Sigma V^\star$ then the j-th column of matrix V is called the j-th **principal component** and its elements are called the **loadings** of the the j-th component.

The elements of the j-th column of matrix U are connected to the "scores" of the j-th principal component for a particular observation. So for example, $\sigma_j U_{ij}$ is the score of the j-th component for i-th observation.

We can also write:

$$X = U\Sigma V^* = \sum_{k=1}^{n} \sigma_k u^{(k)} \left(v^{(k)}\right)^*, \tag{9.6}$$

where $u^{(k)}$ and $v^{(k)}$ are k-th columns of the matrices U and V, respectively.

In particular the scores of the k-th principal component can be calculated as

$$\sigma_k u^{(k)} = X v^{(k)},$$

which means that the vector of scores for k-th component is a linear combination of columns of X ("features"), with coefficients given by entries of the vector $v^{(k)}$ ("loadings" of k-th component).

In components, this can be written as

$$x_{ij} = \sum_{k=1}^{r} \sigma_k u_{ik} v_{jk}$$

where $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

Note that V is the matrix of eigenvectors of matrix X^*X which has the meaning of the empirical covariance matrix for the data. Statistically one can think about the first column V (i.e., the first eigenvector) as the coefficients of the linear combination of variables that has the largest variance (that is, for which the quadratic form v^*X^*Xv achieves its maximum, assuming that v has unit length. In other words this column gives the coefficients of the linear combination of characteristics with the largest variation across individuals.

Similar interpretations can be given for other columns of matrix V.

Often for visualization purposes only the first two principal components are used and the observation vector $x_{i1}, \dots x_{ip}$ is replaced with the scores on the first two principal components: $\sigma_1 U_{i1}$ and $\sigma_2 U_{i2}$.

We also know that the best approximation to the matrix X with rank r is given by

$$X = U\Sigma V^* = \sum_{k=1}^{r} \sigma_k u^{(k)} \left(v^{(k)}\right)^*, \tag{9.7}$$

where the sum in (9.6) is cut at $r \leq n$. This is the basis for the dimension reduction technique when X is replaced with the matrix of the scores for the first r components, that is with matrix whose columns are $\sigma_k u^{(k)}$, $k = 1, \dots, r$.

This technique is very popular. One example is the data on financial stock returns. It turns out that the empirical covariance matrix exhibit three important factors (which are principal components with large singular values).

9.7 Condition Number

Let $y = Ax$, where A is an $m \times n$ matrix. In applications it is often important to know whether small changes in input can lead to big changes in output, and it is often useful to measure the size of the changes in relative terms. Then we define a **relative condition number** at input x, as

$$\kappa(x) := \sup_{\delta x \neq 0} \left[\frac{\|A(x+\delta x) - Ax\|}{\|Ax\|} / \frac{\|\delta x\|}{\|x\|} \right] = \sup_{\delta x \neq 0} \left[\frac{\|A\delta x\|}{\|\delta x\|} \right] \frac{\|x\|}{\|Ax\|}$$

By definition of the operator norm we see that

$$\kappa(x) = \|A\| \frac{\|x\|}{\|Ax\|}.$$

Now sometimes we want a bound on the relative condition number which would be independent of the input. Let us assume that A is a **square** non-singular matrix.

Then we have

$$\frac{\|x\|}{\|Ax\|} = \frac{\|A^{-1}Ax\|}{\|Ax\|} \leq \|A^{-1}\|.$$

Hence, we have $\kappa(x) \leq \|A\|\|A^{-1}\|$.

Theorem 9.7.1. *Let A be a square non-singular matrix and consider the equation $Ax = b$. The problem of computing b, given x, has the relative condition number*

$$\kappa = \|A\|\frac{\|x\|}{\|b\|} \leq \|A\|\|A^{-1}\|,$$

with respect to perturbations of x. The problem of computing x, given b, has the relative condition number

$$\kappa' = \|A^{-1}\|\frac{\|b\|}{\|x\|} \leq \|A\|\|A^{-1}\|,$$

with respect to perturbations of b.

Proof. We proved the first part above. For the second part, note that we can re-write the problem of computing x given b as $A^{-1}b = x$, and then we can apply the first part. □

We know that $\|A\| = \sigma_1(A)$ and $\|A^{-1}\| = 1/\sigma_n$, where $\sigma_1(A)$ and $\sigma_n(A)$ are the largest and the smallest singular values of A. So we can write a bound $\kappa(x) \leq \sigma_1/\sigma_n$.

The first part of this theorem can be easily generalized to non-square matrices. Indeed, if the matrix A is $m \times n$ with $m > n$, then we can replace A^{-1} in the arguments above with the pseudo-inverse A^+, and then $k(x) \leq \|A\|\|A^+\| = \sigma_1/\sigma_n$.

The quantity σ_1/σ_n is called the **condition number** of the matrix A. It is a universal bound for the relative condition number $\kappa(x)$ which is valid for all inputs $x \neq 0$.

In fact this number also controls the sensitivity of output to perturbations in the matrix.

Theorem 9.7.2. *Let b be fixed and consider the problem of computing $x = A^{-1}b$, where A is square and nonsingular. The relative condition number of this problem with respect to perturbations in A is*

$$\kappa = \|A\|\|A^{-1}\| = \frac{\sigma_1(A)}{\sigma_n(A)}.$$

Proof. If we perturb A in the equation $Ax = b$, we find that

$$(A + \delta A)(x + \delta x) = b.$$

By using the equality $Ax = b$ and dropping the second order term $\delta A \delta x$, we find $(\delta A)x + A\delta x = 0$, or $\delta x = -A^{-1}(\delta A)x$. This implies that $\|\delta x\| \leq \|A^{-1}\|\|\delta A\|\|x\|$, which we can re-write as:

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\|\|A^{-1}\|\frac{\|\delta A\|}{\|A\|}.$$

This shows that the relative condition number is bounded from above by $\|A\|\|A^{-1}\|$.

In fact it is possible to show that the bound is achieved for some δA. We omit the proof of this fact. See the book by Trefethen and Bau for details. □

9.8 Exercises

Exercise 9.8.1. Determine the SVDs of the following matrices (by hand calculation):

$$\text{(a)} \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}, \text{(b)} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \text{(c)} \begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \text{(d)} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \text{(e)} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

(Note that the answers can be different up to some multiplication of columns of U and V by ± 1.)

Exercise 9.8.2. Determine the SVD of the following matrix (by hand calculation):

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

Exercise 9.8.3. Example 3.6 in Trefethen-Bau shows that if A is an outer product of two vectors $A = uv^t$, then $\|A\|_2 = \|u\|_2\|v\|_2$, where $\|\cdot\|_2$ denotes

both the 2-norm on vectors (the usual Euclidean norm) and the corresponding induced operator norm on matrices.

Is the same true for the Frobenius norm, that is, is $\|A\|_F = \|u\|_F\|v\|_F$? Prove it or give a counterexample.

Exercise 9.8.4. Let A be an $m \times n$ matrix, $m \geq n$. Prove that $\|A\|_F \leq \sqrt{n}\|A\|$.

Exercise 9.8.5. Suppose A is an $m \times n$ matrix and B is the $n \times m$ matrix obtained by rotating A ninety degrees clockwise on paper. Do A and B have the same singular values? Prove that the answer is yes or give a counterexample.

Chapter 10

Bilinear and Quadratic Forms

10.1 Bilinear and quadratic forms. Congruence

A bilinear form is a map that sends a pair of vectors to a number, $B : \mathbb{R}^n \times \mathbb{R}^n \to R$. This map is required to be linear in both arguments:

$$B(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1 B(x_1, y) + \alpha_2 B(x_2, y)$$
$$B(x, \alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 B(x, y_1) + \alpha_2 B(x, y_2)$$

A bilinear symmetric form has an additional property $B(x, y) = B(y, x)$.

Remark: the bilinear forms can be defined for arbitrary fields. In case of the complex numbers, another concept is also useful, the concept of Hermitian forms, when $B(\lambda x, y) = \lambda B(x, y)$, $B(x, y) = \overline{B(y, x)}$ which implies $B(x, \lambda y) = \overline{\lambda} B(x, y)$. (So, the Hermitian form is linear in the first argument and "conjugate-linear" or antilinear in the second argument.) In the following, we focus on bilinear forms.

For a bilinear form, one can define a quadratic form $Q(x) = B(x, x)$. Conversely, if we are given a quadratic form $Q(x)$ then we can define a symmetric bilinear form as $B(x, y) = [Q(x + y) - Q(x - y)]/4$.

To every bilinear form $B(x, y)$ we can associate a matrix $B_{ij} = B(e_i, e_j)$. If the bilinear form is symmetric then the matrix is also symmetric. If $x = [x_1, \ldots, x_n]^t$ in the standard basis, then the bi-linearity of the form implies that

$$B(x, y) = x^t B y.$$

We can also approach this topic from another, more elementary angle. **Quadratic form** in variables $x_1, \ldots, x_n$ is a polynomial $Q(x_1, \ldots, x_n)$ whose monomials have the degree exactly 2. That is, it is a **homogeneous** polynomial of degree 2, and we can write it as

$$Q(x_1, \ldots, x_n) = \sum_{i=1}^{n} b_{ii} x_i^2 + 2 \sum_{1 \le i < j \le n} b_{ij} x_i x_j.$$

We can write this expression using matrices:

$$Q(\boldsymbol{x}) = \boldsymbol{x}^t B \boldsymbol{x},$$

where B is a symmetric matrix with entries $B_{ij} = B_{ji} := B_{ij}$. In this section we assume that B_{ij} are real.

This matrix B is exactly the matrix of the symmetric bilinear form that corresponds to quadratic form Q.

Example 10.1.1. What is the matrix for the form: $x_1^2 + 3x_2^2 + 5x_3^2 + 4x_1x_2 - 16x_1x_3 + 7x_2x_3$?

If we change the variables $\boldsymbol{x} = R\boldsymbol{y}$, where R is an invertible matrix, then in the new variables this form will be

$$Q(\boldsymbol{y}) = \boldsymbol{y}^t R^t B R \boldsymbol{y}.$$

The transformation

$$B \to R^t B R$$

is called the **congruence transformation** on matrices. Compare this with the **similarity transformation** $B \to X^{-1} B X$.

We are interested in the properties of quadratic forms and associated matrices which do not depend on the change of variables, that is, which are invariant with respect to congruence transformations.

The main fact here is that every symmetric matrix B can be brought to the diagonal form by a suitable congruence transformation. In fact, there are many congruence transformations that accomplish this task. The most straightforward method is based on orthogonal diagonalization. Indeed, since B is a real symmetric matrix we can write it as

$$B = Q \Lambda Q^t,$$

where Q is an orthogonal matrix. Hence,

$$Q^t B Q = \Lambda,$$

and we are done.

Example 10.1.2. Consider the quadratic form $Q(x) = 2x_1^2 + 2x_1x_2 + 2x_2^2$. Bring it to the diagonal form.

There are some other methods, which are simpler computationally and involve only algebraic operations.

One of them is based on elementary row and column operations. Suppose that we reduce B by row operations to the upper-diagonal form as we did in the algorithm for LU decomposition. Note that B is symmetric and so when we perform row operations in the row reduction procedure, we can also do analogous operations on columns. As a result we will get a decomposition of the matrix B:

$$B = LDL^t,$$

where L is a lower-triangular matrix with ones on the main diagonal and D is the diagonal matrix with the pivots on the main diagonal.

Example 10.1.3. Let us find a "congruence" diagonalization of a matrix B by using the algorithm that we just described. We are looking for C such that $C^*AC = D$, where D is diagonal and C is non-singular. Let

$$B = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 5 & -4 \\ -3 & -4 & 8 \end{bmatrix}$$

Then, we can do the following sequence of row and column transformations. [We perform column operations only on the left hand side of the augmented matrix.]

$$\left[\begin{array}{ccc|ccc} 1 & 2 & -3 & 1 & 0 & 0 \\ 2 & 5 & -4 & 0 & 1 & 0 \\ -3 & -4 & 8 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_2-2R_1} \left[\begin{array}{ccc|ccc} 1 & 2 & -3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ -3 & -4 & 8 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{C_2-C_1} \left[\begin{array}{ccc|ccc} 1 & 0 & -3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ -3 & 2 & 8 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_3+3R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & -3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 2 & -1 & 3 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{C_3+3C_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 2 & -1 & 3 & 0 & 1 \end{array}\right] \xrightarrow{R_3-2R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 0 & -5 & 7 & -2 & 1 \end{array}\right]$$

$$\xrightarrow{C_3-2C_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & -5 & 7 & -2 & 1 \end{array}\right]$$

This means that

$$C^* = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 7 & -2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -5 \end{bmatrix}$$

(From the practical point of view it is enough to do the row operations, the column operations are done only to illustrate that the matrix is indeed reduced to the diagonal form.)

Note that this algorithm fails if one needs to do an exchange of rows as for example for matrix

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

(In this case one can proceed by introducing the non-zero element by a row operation. For example:

$$\begin{bmatrix} 0 & 1 & | & 1 & 0 \\ 1 & 0 & | & 0 & 1 \end{bmatrix} \xrightarrow{R_1+\frac{1}{2}R_2} \begin{bmatrix} 1/2 & 1 & | & 1 & 1/2 \\ 1 & 0 & | & 0 & 1 \end{bmatrix}$$
$$\xrightarrow{C_1+\frac{1}{2}C_2} \begin{bmatrix} 1 & 1 & | & 1 & 1/2 \\ 1 & 0 & | & 0 & 1 \end{bmatrix} \to \dots$$

but the resulting matrix C^* is no longer lower-triangular.

10.2 Positive definite forms

Let Q be a quadratic form and A a corresponding symmetric matrix: $Q(x, x) = x^t Ax$. (I have changed the notation from B to A here, sorry.) A quadratic form Q and the corresponding matrix A are called **positive definite** if $Q(\boldsymbol{x}) = \boldsymbol{x}^t A\boldsymbol{x} > 0$ for every $\boldsymbol{x} \neq 0$. It is clear that this property does not depend on the change of variables, so for invertible matrices R, matrix A is positive definite if and only if matrix $R^t AR$ is positive definite.

In applications it is often needed to check whether a matrix is positive definite. In particular, $Q(\boldsymbol{x})$ has a strict minimum at 0 if and only $Q(\boldsymbol{x})$ is positive definite. We can check whether a quadratic form is positive-definite by using one of the criteria given by the following theorem.

Theorem 10.2.1. *Suppose $Q(\boldsymbol{x}) = \boldsymbol{x}^t A\boldsymbol{x}$ where A is a real symmetric matrix. Then, each of the following tests is a necessary and sufficient condition for the form $Q(\boldsymbol{x})$ to be positive definite:*

1. *All the eigenvalues of A satisfy $\lambda_i > 0$.*

2. *A can be reduced to the upper diagonal form without row exchanges and all the pivots (without row exchanges) satisfy $d_k > 0$.*

3. *All the upper left $k \times k$ sub-matrices A_k have positive determinants.*

Proof. Matrix A is real symmetric, so we can write an orthogonal diagonalization: $A = Q\Lambda Q^t$, so A is positive-definite if and only if Λ is positive definite, and the form for Λ is $Q(x) = \lambda_1 x_1^2 + \ldots \lambda_n x_n^2$, and so it is clear that it is positive definite if and only if $\lambda_i > 0$ for every i.

Next, we are going to prove that if A is positive-definite, then (2) holds. We perform the algorithm described above. At every stage, after we perform a row operation and a corresponding row operation, the matrix remains positive definite. In particular, there can be no zero elements on the main diagonal, so a row exchange is never required. Eventually, we will get a diagonal matrix and all the diagonal elements d_k (pivots) must be positive.

Conversely, (2) implies that the matrix A is positive-definite. Indeed, condition (2) implies that we can find a decomposition of the matrix A:

$$A = LDL^t,$$

where L is a lower-triangular matrix with ones on the main diagonal and D is the diagonal matrix with the pivots on the main diagonal. Since $d_k > 0$, it follows that A is positive definite.

Finally we claim that (2) is equivalent to (3). Indeed, the row operations do not change the determinants of A_k. So if (2) holds then all of these determinants must be positive: $\det(A_k) = d_1 \ldots d_k$. Conversely, if all of these determinants are positive, then the row exchange is never required (otherwise, one of the determinants at this stage would be equal to zero) and all pivots must be positive (or one of the determinants would be negative.

□

Another useful fact is that a positive definite matrix A can be factorized as $R^t R$.

Theorem 10.2.2. *The symmetric $n \times n$ matrix A is positive definite if and only if there is a non-singular n matrix R with independent columns such that $A = R^t R$.*

Proof. Suppose $A = R^t R$, where R is non-singular. Then, $x^t Ax = (Rx)^t Rx = \|Rx\| \geq 0$. If this quantity is zero, then $Rx = 0$, hence $x = 0$ because R is non-singular.

In the other direction, we can write $A = LDL^t$, where L is lower diagonal and D is a diagonal matrix with positive entries on the diagonal. Then we can take $R = \sqrt{D}L^t$, and observe that the columns of R are linearly independent. □

The decomposition $A = R^t R$, where R is upper-diagonal is often called the **Cholesky decomposition** of a positive definite matrix.

10.3 Law of Inertia

What can be said about more general situation, when the form $Q(x)$ represented by a symmetric matrix A is not necessarily positive definite?

It turns out that in this case, it can be reduced by a suitable change of variable to a form represented by a diagonal matrix that have only ± 1 or 0 on the main diagonal. Moreover, the number of positive, negative and zero items on the main diagonal of this diagonal matrix does not depend on the particular choice of this change of variables. This statement is called Sylvester's law of inertia and it follows from the following result.

Theorem 10.3.1. *Let A be a real symmetric matrix and C be a real invertible matrix. Then, matrix $C^t AC$ has the same number of positive eigenvalues, negative eigenvalues, and zero eigenvalues as A.*

So we can define the **signature** of a quadratic form as a triple (k_p, k_n, k_0), where the k_p, k_n, and k_0 is the number of positive, negative and zero entries on the main diagonal of any of these diagonal matrices. (Sometimes $k_p - k_n$ is also called the signature of the form.)

Proof of Theorem 10.3.1. We give a sketch of a proof for a simpler situation in which A is non-singular, so we do not need to worry about zero eigenvalues.

Let $C(u)$, $u \in [0, 1]$, be a family of matrices such that $C(0) = C$, $C(1) = Q$, where Q is an orthogonal matrix, and $C(u)$ is never singular. (We will prove that it is possible to find such family of matrices below.) Then the matrix $C(u)^* AC(u)$ is never singular, so its determinant is never zero, and therefore, its eigenvalues are never zero. In addition, the eigenvalues of $C(u)^* AC(u)$ are continuous in u. (We skip the proof of this claim.) It follows that they can never change sign, when u changes from 0 to 1, and therefore, the number of positive eigenvalues of $C^* AC$ is the same as the number of positive eigenvalues of $Q^* AQ$. However, $Q^* AQ$ has the same eigenvalues as A.

In order to prove that there is a required $C(u)$ we can take Q from the QR decomposition $C = QR$. We choose the decomposition in such a way that R has positive entries on the main diagonal. Then we can write $C(u) = Q\big(uI + (1-u)R\big)$, and this matrix is always non-singular because the matrix $\big(uI + (1-u)R\big)$ is upper-diagonal and has positive entries on its diagonal.

□

10.4 Rayleigh quotient

Rayleigh quotient is an important way to characterize eigenvalues as the maximum of a quadratic form.

In order to motivate this property, note that the largest singular value $\sigma_1(A)$ equals the norm of the matrix A, which is the maximum of the quotient

$$\frac{\|Ax\|}{\|x\|}.$$

over all possible non-zero x. The corresponding left singular vector is the vector at which this maximum is achieved. The square of this expression can be rewritten as

$$\frac{\|Ax\|^2}{\|x\|^2} = \frac{(Ax, Ax)}{(x, x)} = \frac{(x, A^*Ax)}{(x, x)}.$$

Hence the square of the largest singular value is the maximum of the expression (x, A^*Ax) given that $(x, x) = 1$.

The Rayleigh quotient is a modification of this idea, which focuses on eigenvalues instead of singular values. By definition, the **Rayleigh quotient** of a vector x is the ratio:

$$R(x) = \frac{(x, Ax)}{(x, x)}.$$

Theorem 10.4.1 (Rayleigh-Ritz)**.** *If a symmetric matrix A has eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, then λ_1 and λ_n are the maximum and the minimum, respectively, of the Rayleigh quotient $R(x)$ over all $x \neq 0$.*

Proof. We need to check that

$$\lambda_n(x, x) \leq (x, Ax) \leq \lambda_1(x, x) \tag{10.1}$$

holds and that the bounds can be achieved by a suitable choice of $x \neq 0$.

The inequalities hold because $A = U\Lambda U^*$ and so

$$(x, Ax) = (U^*x, \Lambda U^*x) = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2.$$

where $y = (y_1, \ldots, y_n)^* = U^*x$. The last expression is between $\lambda_n \|y\|^2$ and $\lambda_1 \|y\|^2$ and we know that $\|y\|^2 = \|x\|^2$.

It is also clear that the bounds in the inequalities (10.1) are achieved if we set x equal to the eigenvectors corresponding to eigenvalues λ_1 and λ_n. □

For example, for the largest eigenvalue we have

$$\lambda_1 = \max_{x \neq 0} \frac{(x, Ax)}{(x, x)}.$$

Alternatively we can write:

$$\lambda_1 = \max_{x:\|x\|=1} (x, Ax),$$

and the maximum is achieved on an eigenvector of A that corresponds to the eigenvalue λ_1.

Note that if $Q(x)$ is the quadratic form associated to the symmetric matrix A, then this gives us ability to find the maximum of $Q(x)$ on the set of all vectors x that have unit length.

Similarly, for the smallest eigenvalue we have:

$$\lambda_n = \min_{x:\|x\|=1} (x, Ax),$$

and again the minimum is achieve at the eigenvector that corresponds to the smallest eigenvalue λ_n.

Corollary 10.4.2. *The diagonal entries of any symmetric matrix are between λ_1 and λ_n.*

Proof. This is a consequence of Theorem 10.4.1 because the diagonal entry $a_{ii} = R(\boldsymbol{e}_i)$, where $\boldsymbol{e}_i = (0, \ldots, 1, \ldots, 0)$ is the i-th coordinate vector. □

Characterization of the largest and smallest eigenvalues as the maximum and minimum, respectively, of a quadratic form can be extended to intermediate eigenvalues. Let V_{k-1} be the space spanned by orthonormal system of

eigenvectors $u_1, u_2, \ldots u_{k-1}$ that correspond to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{k-1}$. Then,

$$\lambda_k = \max_{x \neq 0, x \perp V_{k-1}} R(x).$$

In order to see this, note that the space $V_{k-1}^{\perp}$ orthogonal to V_{k-1} is invariant under the transformation A and spanned by the eigenvectors corresponding to the eigenvalues $\lambda_k, \ldots, \lambda_n$. Then the desired result can be obtained by restricting the linear transformation A to the linear space $V_{k-1}^{\perp}$ and applying the Rayleigh-Ritz theorem to this restriction.

For example, the second eigenvalue of A gives the following maximum:

$$\lambda_2 = \max_{x \neq 0, x \perp u_1} \frac{(x, Ax)}{(x, x)},$$

where u_1 is the first eigenvector corresponding to λ_1. Alternatively we can write this expression as

$$\lambda_2 = \max_{x: \|x\| = 1, x \perp u_1} (x, Ax).$$

The maximum is achieved at u_2, an eigenvector that corresponds to λ_2.

A useful extension of this result is the Courant-Fisher Theorem. It says that instead of explicitly choosing V_k as the span of the first k eigenvectors, one can solve a minmax problem.

Theorem 10.4.3 (Courant-Fisher)**.** *If a symmetric matrix A has eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, then for $1 \leq k \leq n$,*

$$\lambda_k = \min_{V_{k-1}} \max_{x \neq 0, x \perp V_{k-1}} R(x),$$

where the minimization is over all $k-1$ dimensional subspaces V_{k-1}, and

$$\lambda_k = \max_{V_{n-k}} \min_{x \neq 0, x \perp V_{n-k}} R(x),$$

where the maximization over all $n-k$ dimensional subspaces V_{n-k}.

The difference (and the benefit) of this theorem from our previous considerations is that it does not define V_{k-1} as the span of eigenvectors $u_1, \ldots, u_{k-1}$ but allows V_{k-1} to run over all possible $k-1$ subspaces and chooses the "worst" possible subspace. The worst here means that it leads to the smallest of the maximal Rayleigh ratios. (The second expression is similar but the maximum and minimum are exchanged in this expression.)

The Courant-Fisher Theorem allows proving several important theoretical results. One of the most useful is a theorem by Hermann Weyl. Let us write $\lambda_j(X)$ to denote the eigenvalues of an Hermitian matrix X arranged in decreasing order.

Theorem 10.4.4 (Weyl)**.** *Let A and B be two Hermitian $n \times n$ matrices. For each $k = 1, 2, \ldots, n$, we have*

$$\lambda_k(A) + \lambda_n(B) \leq \lambda_k(A + B) \leq \lambda_k(A) + \lambda_1(B)$$

Proof. For every vector x, we have

$$\lambda_1(B) \geq \frac{x^t B x}{x^t x} \geq \lambda_n(B).$$

Hence

$$\begin{aligned}
\lambda_k(A + B) &= \min_{V_{k-1}} \max_{x \perp V_{k-1}} \frac{x^t (A + B) x}{x^t x} \\
&= \min_{V_{k-1}} \max_{x \perp V_{k-1}} \left[\frac{x^t A x}{x^t x} + \frac{x^t B x}{x^t x} \right] \\
&\leq \min_{V_{k-1}} \max_{x \perp V_{k-1}} \left[\frac{x^t A x}{x^t x} + \lambda_1(B) \right] = \lambda_k(A) + \lambda_1(B).
\end{aligned}$$

(In all maximizations over x it is assumed that $x \neq 0$.) The lower bound can be established similarly. □

Corollary 10.4.5. *If matrix B is non-negative definite, then all eigenvalues of A increase when we add B:*

$$\lambda_k(A + B) \geq \lambda_k(A)$$

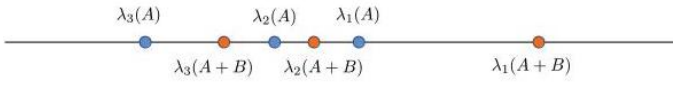

Figure 10.1

Another important and surprising result is as follows.

Theorem 10.4.6 (interlacing of eigenvalues I)**.** *If the matrix B is non-negative definite and has rank 1, then*

$$\lambda_{k+1}(A) \leq \lambda_{k+1}(A + B) \leq \lambda_k(A),$$

where $k = 0, \ldots, n - 1$, with the convention that $\lambda_0(A) = +\infty$.

In other words the rank-one perturbation of matrix A cannot move the internal eigenvalues too much. This is illustrated in Figure 10.1.

Proof. Note that $\lambda_k(A+B) \geq \lambda_k(A)$ holds by Corollary 10.4.5, so we only need to prove the other inequality.

By the eigenvalue decomposition, every non-negative definite symmetric rank 1 matrix can be written as a outer product of a vector with itself. So let $B = vv^t$. For $1 \leq k \leq n-1$ we write the following sequence of inequalities (where $x \neq 0$ always):

$$\begin{aligned}
\lambda_k(A) &= \min_{V_{k-1}} \max_{x \perp V_{k-1}} \frac{x^t(A + vv^t - vv^t)x}{x^t x} \\
&\geq \min_{V_{k-1}} \max_{x \perp V_{k-1}, x \perp v} \frac{x^t(A + B - vv^t)x}{x^t x} \\
&= \min_{V_{k-1}} \max_{x \perp (V_{k-1} \oplus \langle v \rangle)} \frac{x^t(A+B)x}{x^t x} \\
&\geq \min_{V_k} \max_{x \perp V_k} \frac{x^t A x}{x^t x} = \lambda_{k+1}(A+B).
\end{aligned}$$

The inequality in the second line of this display holds because we added a new constraint to the maximization problem. The second inequality holds because in the constraint we used arbitrary V_k instead of those V_k that required to include v.

□

A closely related result is as follows.

Theorem 10.4.7 (interlacing of eigenvalues II)**.** *Let A be a Hermitian $n \times n$ matrix, and let A' be its $(n-1) \times (n-1)$ upper-left principal submatrix.*

$$\lambda_1(A) \geq \lambda_1(A') \geq \lambda_2(A) \geq \lambda_2(A') \geq \ldots \geq \lambda_{n-1}(A') \geq \lambda_n(A).$$

(In fact the result holds for any A' which by removing k-th column and k-th row from A, where $1 \leq k \leq n$.)

Example 10.4.8. The matrix

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

has eigenvalues $\lambda_1(A) = 2+\sqrt{2}$, $\lambda_2(A) = 2$, and $\lambda_3(A) = 2-\sqrt{2}$, and matrix

$$A' = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

has eigenvalues $\lambda_1(A') = 3$ and $\lambda_2(A') = 1$, and as the theorem claims, we have:

$$2 + \sqrt{2} \geq 3 \geq 2 \geq 1 \geq 2 - \sqrt{2}.$$

It is also interesting that if the eigenvalues of $n \times n$ symmetric matrices A and B are known, and n is large, then one can calculate approximately the distribution of eigenvalues of the matrix $A + UBU^*$, where U is a random unitary matrix. This was one found recently (around 20 years ago) in research that comprised the study of random matrices and results from a field in functional analysis called free probability theory.

10.5 Exercises

Exercise 10.5.1. Let

$$A = \begin{bmatrix} 1 & -3 & 2 \\ -3 & 7 & -5 \\ 2 & -5 & 8 \end{bmatrix}$$

Find a nonsingular real matrix C, such that $D = C^*AC$ is diagonal, and find $sign(A)$, the signature of A.

Exercise 10.5.2. Determine whether each of the following quadratic forms Q is positive definite:

(a) $Q(x, y, z) = x^2 + 2y^2 - 4xz - 4yz + 7z^2$.

(b) $Q(x, y, z) = x^2 + y^2 + 2xz + 4yz + 3z^2$.

Chapter 11

Hints to Exercises

Chapter 1

- (Ex. 1.4.4) Take for v_{r+1} any vector that cannot be represented as a linear combination $\sum_{k=1}^{r} \alpha_k v_k$ and show that the system $v_1, v_2, \ldots, v_r, v_{r+1}$ is linearly independent.
- (Ex. 1.4.5) Consider $W = \text{span}\{w_1, w_2, w_3\}$. Is this subspace spanned by $\{v_1, v_2, v_3\}$? What can you say about its dimension under the assumption that v_1, v_2 and v_3 are linearly dependent?
- (Ex. 1.4.8) For 1. and 2., first show that $\vec{0}$ is the only element that satisfies equation $x + x = x$.
- (Ex. 1.4.13) Add vectors w_i one by one to the beginning of the list of vectors. After each addition, remove a suitable vector v_{j_i}. Make sure that you never need to remove a vector w_k, which have already been added.

Chapter 2

- (Ex. 2.7.12) You can do the exercise by a direct argument, using the facts from the previous chapter, in particular, that every basis must have n elements. Alternatively, you can write vectors $v_1, \ldots, v_n$ in a standard basis as columns of a matrix and argue that the claim follows from the rank-nullity theorem.
- (Ex. 2.7.14) Let $\{w_1, \ldots, w_n\}$ be new vectors. Write the matrix L that has columns $\{w_1, \ldots, w_n\}$ written in the basis $\{v_1, \ldots, v_n\}$ and check if it is full rank.

- (Ex. 2.7.16 What is the rank of A? Use Theorem 2.3.8.
- (Ex. 2.7.29) Let $L_m \dots L_1 A = B$, where $B = rref(A)$, and $L_1, \dots, Lm$ are elementary row transformations. Can you find the left inverse for B? Can you use it to write down the left inverse for A?
- (Ex. 2.7.32) Recall how $\mathrm{Tr}(XY)$ and $\mathrm{Tr}(YX)$ are related.

Chapter 4

- (Ex. 4.6.11) Are AB and BA similar?

Chapter 5

- (Ex. 5.4.10) Is 0 an eigenvalue of T? What can be said about the dimension of $\ker T^{n-1}$? What can be said about the dimension of $\tilde{E}_0$? Then, use Theorem 5.2.7.
- (Ex. 5.4.11) If 0 is not an eigenvalue, the claim is true (why?). If 0 is an eigenvalue, what can be said about the dimension of the generalized eigenspace $\tilde{E}_0$? Now note that $\ker T^k \subset \tilde{E}_0$ and use Theorem 5.2.2.

Chapter 6

- (Ex. 6.5.3) The nullspace of the relevant matrix is

$$Ker = \left\langle \begin{bmatrix} \frac{3}{5} \\ -\frac{1}{5} \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\rangle = \left\langle \begin{bmatrix} 3 \\ -1 \\ -5 \\ 0 \\ 5 \end{bmatrix} \right\rangle$$

  and the vectors in the subspace have to be orthogonal to this nullspace. (Note that the rref and the basis of the nullspace can be conveniently calculated using the **sympy** package in Python. The relevant methods are "rref" and "nullspace".)

- (Ex. 6.5.5) Here is how one can proceed without actually calculating Pv.

  We want to use a consequence of Plancherel's identity. Namely, if q_1, q_2 is orthonormal basis, then the projection is $Pv = c_1 q_1 + c_2 q_2$, and by Pythagorean theorem

$$\|v\|^2 = \|Pv\|^2 + \|v - Pv\|^2,$$

so

$$\|v - Pv\|^2 = \|v\|^2 - \|Pv\|^2 = \|v\|^2 - c_1^2 - c_2^2.$$

We need a formula for c_i. It is

$$c_i = \frac{\langle v, v_i \rangle}{\|v_i\|},$$

(check that you understand it), and the final formula is

$$\|v - Pv\|^2 = \|v\|^2 - \frac{\langle v, v_1 \rangle^2}{\|v_1\|^2} - \frac{\langle v, v_2 \rangle^2}{\|v_2\|^2}.$$

Answer: the distance is $\sqrt{\frac{170}{21}}$.

Chapter 7

- (Ex. 7.3.2)

  To prove the equality $\ker A = \ker(A^*A)$ one needs to prove two inclusions $\ker(A^*A) \subset \ker A$ and $\ker A \subset \ker(A^*A)$. One of the inclusions is simple (which one?), for the other one use the fact that

  $$\|Ax\|_2 = \langle Ax, Ax \rangle = \langle A^*Ax, x \rangle$$

- (Ex. 7.4.4) One can use results from the previous exercise.

- (Ex. 7.4.8)

  First prove that $F^2 = I$ and $F^* = F$ (by using the definition of the adjoint operator for the latter). Then check if $E = (I + F)/2$ satisfies the properties of the orthogonal projector.

  For the matrix entries E_{ij}, the Kronecker delta notation can be useful: $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

Chapter 8

- (Ex. 8.6.9) Show that AB has the same eigenvalues as $A^{1/2}BA^{1/2}$. Then show that $A^{1/2}BA^{1/2}$ is positive semidefinite. Argue that this implies that $AB + I$ is invertible.