Linear algebra – a brief review

Stilian A. Stoev
Lecture Notes for STAT 520

Fall 2009

1 Matrix algebra

1.1 Basic notions and operations

A triangular array A of scalars (here, real $\mathbb{R}$ or complex $\mathbb{C}$ numbers) is said to be a *matrix.*

Notation: $A = (a_{ij})_{m\times n}$, $m,\ n \in \mathbb{N} := \{1, 2, \ldots\}$, or

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Matrices of dimensions $m \times 1$ and $1 \times n$ are called *column* and *row vectors*, respectively. We will typically denote column and row vectors by lower case Latin letters, e.g. $a,\ b,\ x,\ y$ and other matrices by upper case Latin letters, e.g. $A,\ B,\ X,\ Y$. The scalars (or 1×1 matrices) will be frequently denoted by Greek letters $\alpha,\ \beta,\ \lambda,\ \mu$, etc. Unless stated otherwise, all scalars will be real numbers.

Operations with matrices

1. *(multiplication by a scalar)* For any $m \times n$ matrix $A = (a_{ij})_{m\times n}$ and $\forall \alpha \in \mathbb{R}$,

$$\alpha A := (\alpha a_{ij})_{m\times n}.$$

2. *(matrix multiplication)* If $A = (a_{ij})_{m\times p}$ and $B = (b_{ij})_{p\times n}$, then

$$C = AB := (c_{ij})_{m\times n}, \quad \text{where} \quad c_{ij} := \sum_{k=1}^{p} a_{ik} b_{kj}.$$

3. *(addition)* If $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{m\times n}$, then

$$C = A + B := (c_{ij})_{m\times n}, \quad \text{where} \quad c_{ij} := a_{ij} + b_{ij}.$$

4. *(transpose)* If $A = (a_{ij})_{m\times n}$, then

$$A' := (c_{ij})_{n\times m}, \quad \text{where} \quad c_{ij} := a_{ji}.$$

5. *(trace)* If $A = (a_{ij})_{m\times m}$ is a *square matrix*, then

$$\mathrm{tr}(A) := \sum_{i=1}^{m} a_{ii}.$$

The transpose operation and the trace functional have the following

PROPERTIES 1.1 *Whenever the matrix dimensions are compatible, we have:*

a. $(\alpha A)' = \alpha A'$

b. $(A')' = A$

c. $(\alpha A + \beta B)' = \alpha A' + \beta B'$

d. $(AB)' = B'A'$

e. $\mathrm{tr}(\alpha A') = \alpha \mathrm{tr}(A)$

f. $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$

g. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

h. $\mathrm{tr}(A'A) = 0$ *if and only if* $A = (0)$*, that is only if* A *is a matrix of zeros.*

Examples and more notation:

1. Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

We then have

$$AB = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 6 & 5 \end{pmatrix}.$$

The matrix B is a special case of a *transformation matrix.*

2. *(the identity matrix)* $I_m := (\delta(i-j))_{m\times m}$, where $\delta(i) = 1$, for $i = 0$ and $\delta(i) = 0$, for $i \neq 0$, i.e.

$$I_m := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Recall that for any $A = (a_{ij})_{m\times n}$, we have

$$AI_n = A \quad \text{and} \quad I_m A = A.$$

3. *(diagonal matrices)* For scalars $\lambda_i,\ 1 \leq i \leq m$

$$A = \mathrm{diag}(\lambda_1, \cdots, \lambda_m) := \left(\begin{array}{cccc} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{array} \right).$$

1.2 Determinants and inverses

Let $A = (a_{ij})_{n\times n}$ be a square matrix.

Definition 1.1 The determinant of A is a scalar, defined as follows

$$\det(A) \equiv |A| := \sum_{(i_1,\cdots,i_n)\in S_n} (-1)^{\sigma(i_1,\cdots,i_n)} a_{1i_1} \cdots a_{ni_n},$$

where the summation is over the group S_n of all permutations $(i_1, \cdots, i_n)$ of the numbers $\{1, \cdots, n\}$. Here $\sigma(i_1, \cdots, i_n)$ denotes the number of inversions in the permutation $(i_1, \cdots, i_n)$. That is, the number of pairs of indices $1 \leq j' < j'' \leq n$, such that $i_{j'} > i_{j''}$.

A useful result for calculating determinants is the:

The co–factor formula For any $1 \leq i \leq n$,

$$|A| = \sum_{j=1}^{n} a_{ij} A_{ij}, \quad A_{ij} := (-1)^{i+j} M_{ij},$$

where M_{ij}, called the *minor* of the element a_{ij} is the *determinant* of the $(n-1) \times (n-1)$ matrix obtained from A by deleting the $i-$th row and $j-$th column. The matrix A_{ij} is called the *co–factor* of the element a_{ij}.

One also has the *column* co–factor expansion:

$$|A| = \sum_{i=1}^{n} a_{ij} A_{ij}, \quad A_{ij} := (-1)^{i+j} M_{ij},$$

valid for any $j,\ 1 \leq j \leq n$.

Examples:

1.
$$\det \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) = ad - bc$$

2.
$$\det \left(\begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{array} \right) = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{12}a_{23}a_{31} - a_{31}a_{22}a_{13} - a_{21}a_{12}a_{33} - a_{11}a_{23}a_{32}.$$

PROPERTIES 1.2 *Let* $A = (a_{ij})_{n\times n} = (a_1\ a_2\ \cdots a_n)$ *be a square matrix, where* $a_j = (a_{ij})_{n\times 1}$, $1 \leq j \leq n$. *We then have:*

a. $|A'| = |A|$

b. $|\alpha A| = \alpha^n |A|$

c. *If two column (row) vectors of* A *are proportional or if one column (row) vector of* A *is zero, then the determinant is zero,* $|A| = 0$.

d. *If two column (row) vectors of* A *are interchanged, then the sign of* $|A|$ *changes, i.e.*

$$\det(a_1 \cdots a_i \cdots a_j \cdots a_n) = -\det(a_1 \cdots a_j \cdots a_i \cdots a_n).$$

e. *If a column (row) vector of* A *is multiplied by a scalar, then the determinant is multiplied by this scalar, i.e.*

$$\det(\alpha a_1 \cdots a_n) = \alpha \det(\alpha a_1 \cdots a_n)$$

f. *If a multiple of a column (row) is added to another column (row) of* A*, then the determinant does not change, i.e.*

$$\det(a_1 + \alpha a_2 \cdots a_n) = \det(a_1 \cdots a_n).$$

The next important result is known as the *product formula.*

Theorem 1.1 *For any two square matrices* $A = (a_{ij})_{n\times n}$ *and* $B = (b_{ij})_{n\times n}$*, we have*

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

For the proof, see for example, pp. 7 and 8 in Schott (2005). We conclude this brief discussion on determinants with the following fundamental result on the origin of the determinant function.

Theorem 1.2 *Let* $M_n(\mathbb{R})$ *denote the set of all square* $n \times n$ *matrices with real*[1] *entries. There exists a unique function* $f : M_n(\mathbb{R}) \to \mathbb{R}$*, which satisfies the next two conditions:*

i. f *is alternating and multi–linear column–wise, i.e.*

$$f(a_1 \cdots a_i \cdots a_j \cdots a_n) = -f(a_1 \cdots a_j \cdots a_i \cdots a_n),\ \forall 1 \leq i < j \leq n$$

and

$$f(\alpha a_1' + \beta a_1'' \cdots a_n) = \alpha f(a_1' \cdots a_n) + \beta f(a_1'' \cdots a_n).$$

ii. $f(I_n) = 1$.

Such a function f *is in fact the determinant and is given as in Definition 1.1.*

We next recall some facts about inverses of square matrices when they exist. A square matrix $A = (a_{ij})_{n\times n}$ is said to be *non–singular* if $\det(A) \neq 0$, and *singular* otherwise.

A square matrix $A = (a_{ij})_{n\times n}$ is said to have an inverse, which is a $n \times n$ matrix denoted A^{-1}, if:

$$AA^{-1} = A^{-1}A = I_n. \tag{1.1}$$

It is easy to check that the inverse is unique, if it exists. The next result is fundamental.

[1] This result is valid in a greater generality e.g. for an arbitrary field of scalars.

Theorem 1.3 *A matrix $A = (a_{ij})_{n\times n}$ has an inverse if and only if it is non–singular. In this case, the inverse is given by*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \dots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}, \tag{1.2}$$

where $A_{ij} = (-1)^{i+j} M_{ij}$ is the co–factor of the element a_{ij} of the original matrix A.

Proof: Suppose that A is non–singular. Then, one can easily verify that A^{-1} defined as in (1.2) satisfies (1.1). Indeed, for simplicity suppose that $\det(A) = 1$ and consider the matrix product $C = AA^{-1} = (c_{ij})_{n\times n}$. For any $1 \le i, j \le n$, we have

$$c_{ij} = \sum_{k=1}^{n} a_{ik} A_{jk}. \tag{1.3}$$

Notice that if $i = j$, then the last formula is simply the co–factor expansion of $\det(A) = 1$. That is, $c_{ii} = 1$, $\forall 1 \le i \le n$. Now, if $i \ne j$, then (1.3) is also a co–factor expansion of the determinant of a modification of the matrix A. Namely, a matrix obtained by replacing the the $j-$th row $(a_{j1} \ \cdots \ a_{jn})$ of A with its $i-$th row $(a_{i1} \ \cdots \ a_{in})$. Such a matrix has two identical rows and hence determinant zero. One can similarly check that $A^{-1}A = I_n$.

Now let $\det(A) = 0$ and suppose that A^{-1} exists. Then, by the product formula for determinants, we have that $0 = \det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(I_n) = 1$, which is a contradiction. Therefore, the singular matrix A has no inverse. □

Remark. Notice the *transpose* nature of the co–factors appearing in the formula for the inverse matrix (1.2). As a direct consequence of Theorem 1.3, we have:

Proposition 1.1 *Let $A = (a_{ij})_{n\times n}$ be a non–singular square matrix. Then, for any column vector $b = (b_i)_{n\times 1}$, there exists a unique column vector $x = (x_i)_{n\times 1}$ such that*

$$Ax = b. \tag{1.4}$$

Moreover, the x_i's are given by the Cramer formula:

$$x_i = \frac{\det(A_i)}{\det(A)}, \quad \textit{where} \quad A_i = (a_1 \cdots a_{i-1} \ b \ a_{i+1} \cdots a_n),$$

that is, A_i is obtained from A by replacing its $i-$th column with b.

The proof follows by noting that $x = A^{-1}Ax = A^{-1}b$ and then applying the formula for the inverse in (1.2). Notice that the matrix equation (1.4) is in fact a shorthand for the linear system of n equations and n unknowns:

$$\begin{cases} a_{11}x_1 & + \cdots & + a_{1n}x_n & = b_1 \\ \vdots & \dots & \dots & \vdots \\ a_{n1}x_1 & + \cdots & + a_{nn}x_n & = b_n \end{cases}$$

We will review the theory of general linear systems in the sequel.

PROPERTIES 1.3 *For a square and invertible (non–singular) matrix $A = (a_{ij})_{n\times n}$, we have:*

a. $(\alpha A)^{-1} = \alpha^{-1}A^{-1}$, *for any* $\alpha \in \mathbb{R},\ \alpha \neq 0$.

b. A' *is invertible and* $(A')^{-1} = (A^{-1})'$

c. A^{-1} *is invertible and* $(A^{-1})^{-1} = A$.

d. $\det(A^{-1}) = 1/\det(A)$.

e. *If* $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, *then* $A^{-1} = \mathrm{diag}(1/\lambda_1, \ldots, 1/\lambda_n)$.

f. *If* $A' = A$, *then* $A^{-1} = (A^{-1})'$.

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

1.3 Some useful applications of determinants

• *Volume of a linear transformation of a hyper–cube*

Consider the $n-$dimensional Euclidean space $\mathbb{R}^n$ of column vectors. Let $A = (a_{ij})_{n\times n}$ be a non–singular matrix and define a linear transformation $L_A : \mathbb{R}^n \to \mathbb{R}^n$ by using matrix multiplication. That is,

$$y = L_A(x),\ y \in \mathbb{R}^n, \quad \text{is such that } y = Ax,$$

namely $y_i = \sum_{j=1}^{n} a_{ij}x_j,\ \forall 1 \leq i \leq n$.

The volume of the unit cube $K = \{x \in \mathbb{R}^n :\ 0 \leq x_i \leq 1,\ \forall 0 \leq i \leq n\}$ is one. It can be shown that the volume of the image of K under L_A is equal to the absolute value of the determinant of A:

$$\mathrm{vol}(L_A(K)) = |\det(A)|.$$

This fundamental fact is essentially the reason why the Jacobian appears when changing variables in multivariate integrals.

• *Change of variables in multiple integrals*

Consider the *change of variables*:

$$\begin{cases} x_1 & = & x_1(y_1, \ldots, y_n) \\ \vdots & & \cdots \\ x_n & = & x_n(y_1, \ldots, y_n) \end{cases},$$

that is $x = x(y)$, where $x = (x_i)_{i=1}^n$ is a vector–valued function of the new *coordinates* $y \in \mathbb{R}^n$.

Assume that all components of $x = x(y)$ are continuously differentiable for all $y \in \mathbb{R}^n$. Then the *Jacobian* of the transformation $x = x(y)$ at a point $y^* \in \mathbb{R}^n$, is defined as the determinant of the matrix:

$$\frac{\partial x}{\partial y}(y^*) := \begin{pmatrix} \frac{\partial x_1(y^*)}{\partial y_1} & \frac{\partial x_1(y^*)}{\partial y_2} & \cdots & \frac{\partial x_1(y^*)}{\partial y_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial x_n(y^*)}{\partial y_1} & \frac{\partial x_n(y^*)}{\partial y_2} & \cdots & \frac{\partial x_n(y^*)}{\partial y_n} \end{pmatrix}$$

Notation: The Jacobian of $x = x(y)$ at y^* is denoted $J_x(y^*)$.

Suppose now that $x = x(y),\ y \in U \subset \mathbb{R}^n$ is continuously differentiable (smooth) for all y in the open set U. The inverse function theorem from analysis implies that if for all $y \in U$, the Jacobian $J_x(y)$ is non–zero, then the inverse $y = y(x)$ exists, it is defined for all x in the open set $V := x(U) \subset \mathbb{R}^n$ and continuously differentiable. For more details, see e.g. p. 221 in Rudin (1976).

Let ξ be a random vector with density $f_\xi(x)$, defined over the open set V. Consider the new *coordinates* y, where the map $x = x(y)$, $x : U \to V$ is invertible, onto and continuously differentiable, as above. Then, the random vector $\eta = y(\xi)$ has a density $f_\eta(y)$, defined over the open set U and such that:

$$f_\eta(y) = f_\xi(x(y))|J_x(y)|, \quad y \in U.$$

This follows from the change of variables formula for integrals:

$$\int_A f_\xi(x)dx_1 \cdots dx_n = \int_{y(A)} f_\xi(x(y))|J_x(y)|dy_1 \cdots dy_n,$$

valid for any closed $A \subset V$.

The absolute value of the Jacobian expresses the rate of change of the infinitesimal volume $dx = dx_1 \cdots dx_n$ relative to the infinitesimal volume $dy = dy_1 \cdots dy_n$ (recall that, $x = x(y)$). This, as argued above, is because the determinant function yields the volume of linear transformations of hyper–cubes.

Remark. The random vector $\eta = y(\xi)$ may have a density even if the Jacobian $J_x(y)$ is zero on a set of y's with measure zero.

1.4 Exercises

1. Verify Properties 1.1 **d.**, **g.** and **h.** by using the definitions.

2. Consider the matrix of ones $E = (a_{ij})_{m\times n},\ a_{ij} = 1,\ \forall 1 \le i \le m,\ 1 \le j \le n$. Find column vectors x and y (of dimensions $m \times 1$ and $n \times 1$, respectively) such that $E = xy'$.

3. Do the following matrix multiplications *by hand*:

   a.
   $$\begin{pmatrix} 1 \\ -1 \end{pmatrix}(1\ 2\ 3)$$

   b. $AB =?$ and $BA =?$, where
   $$A = \begin{pmatrix} 1 & 2 & 1 & 4 & -1 \\ -2 & 1 & 0 & -3 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & -2 \\ 2 & 1 \\ 1 & 0 \\ 4 & -3 \\ -1 & 1 \end{pmatrix}.$$

4. Let A be $(m \times p)$ and B be $(q \times p)$ dimensional matrices, respectively. Determine the dimensions of the matrix X in each case:

   a. $X = (BA')(BA')'$

   b. $X = (B'B)(A'A)^{-1}$.

5. Let A and B be square $n \times n$ dimensional matrices. Given that $\mathrm{tr}(A) = 2$ and $\mathrm{tr}(B) = -1$, calculate $\mathrm{tr}(AB - BA)$.

6. A permutation $\pi = (i_1, \cdots, i_n) \in S_n$ with *even* number of *inversions* is said to be *even*, otherwise it is called an *odd* permutation. The set of all permutations of the numbers $\{1, 2, \cdots, n\}$ is a group with respect to the operation composition. It is called the *symmetric group* and denoted S_n.

   Show that:

   a. Every permutation can be represented as a product (composition) of *transposition* permutations: $(ij) = (1 \cdots i \cdots j \cdots n) \mapsto (1 \cdots j \cdots i \cdots n)$. For example,

$$(132) = (32) \quad \text{and} \quad (321) = (31)(21).$$

   b. The transposition representations of even (odd) permutations involve only even (odd) number of terms.

   Hint: By replacing $(ij),\ i < j$ with the product of *odd number* of transpositions:

$$(j-1\ j) \cdots (i+1\ i+2)\ (i\ i+1)\ (i+1\ i+2) \cdots (j-1\ j),$$

   show that any permutation has a transposition representation involving only adjacent numbers. Then, use the fact that any transposition of adjacent numbers changes the number of inversions of the permutation by one to show that even (odd) permutations have even (odd) number of terms in their transposition representation.

7. By using the co–factor formulae show that:

   a.

$$\det\Big(\operatorname{diag}(\lambda_1, \cdots, \lambda_n)\Big) = \lambda_1 \cdots \lambda_n = \prod_{1 \le i \le n} \lambda_i$$

   b.

$$\det \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} = \prod_{1 \le i \le n} \lambda_i.$$

8. *(Vandermonde determinants)* Show that

$$\det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \cdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{pmatrix} = \prod_{1 \le i < j \le n} (x_i - x_j).$$

   Hint: By subtracting multiples of the first row, zero–out the first column with the exception of the diagonal element. Then, apply the co–factor formula and the properties of the determinant to reduce the 'dimension' of the problem and conclude by induction.

9. Calculate *by hand* the determinants and the inverses (if they exist) of the matrices:

$$A = \begin{pmatrix} 1 & 2 & 0 & 0 \\ -1 & 3 & 0 & 0 \\ 0 & 0 & 0.6 & -0.4 \\ 0 & 0 & 0.2 & 0.2 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 & 3 & 4 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 5 & 8 \end{pmatrix}$$

   Hint: Can you use the block–diagonal structure of the matrix A?

10. The transformations in Properties 1.2 **d.**, **e.** and **f.**, which do not change the *absolute value* of the determinant, can be realized by multiplication with a square $n \times n$ matrix.

   a. Convince yourself that left multiplication corresponds to transforming rows and right multiplication to transforming columns, respectively.

   b. Show that the product of two left (right) transformation matrices is a left (right) transformation matrix.

   c. Show that $|\det(B)| = 1$, for any transformation matrix B.

   Hint: Use the product formula for determinants and start transforming the identity matrix.

11. Let $A = (a_{ij})_{m\times n},\ m, n \in \mathbb{N}$ be an arbitrary real matrix. Show that for some transformation matrices $S = (s_{ij})_{m\times m}$ and $T = (t_{ij})_{n\times n}$, we have

$$SAT = D = \begin{pmatrix} \Lambda & (0)_{r\times(n-r)} \\ (0)_{(m-r)\times r} & (0)_{(m-r)\times(n-r)} \end{pmatrix}, \quad \text{where} \quad \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_r),$$

   and where $(0)_{p\times q}$ denotes the zero matrix of dimensions $p \times q$. Here $1 \le r \le \min\{m, n\}$.

12. Prove Proposition 1.1.

13. Let $X = \sqrt{R}\cos(\theta)$ and $Y = \sqrt{R}\sin(\theta)$, where the random variables R and θ are independent and such that $R \sim \text{Exp}(1)$ and $\theta \sim \text{Uniform}(0, 2\pi)$. Determine the density of the random vector (X, Y).

2 Vector spaces

2.1 Definition of a vector space, linear independence and basis

Definition 2.1 A non–empty set V is said to be a (linear) vector space over a field of scalars K, if:

i. V is equipped with a binary operation called addition: $\forall x, y \in V$, $\exists z \in V$, denoted $z = x + y$, such that:

- **a.** $x + y = y + x$, for all $x, y \in V$
- **b.** $x + (y + z) = (x + y) + z$, for all $x, y, z \in V$
- **c.** there exists a zero element (vector) $\vec{0} \in V$, such that $x + \vec{0} = x$, $\forall x \in V$.
- **d.** $\forall x \in V$, $\exists -x \in V$, such that $x + (-x) = \vec{0}$

ii. V is equipped with the binary operation called multiplication by scalars from K: $\forall x \in V$, $\forall \alpha \in K$, exists $z \in V$, denoted $z = \alpha x$. Moreover,

a. $\alpha(\beta x) = (\alpha\beta)x$, for all $\alpha, \beta \in K$ and $x \in V$.
b. $1x = x$, for all $x \in V$
c. $(\alpha + \beta)x = \alpha x + \beta x$, for all $\alpha, \beta \in K$ and $x \in V$.
d. $\alpha(x + y) = \alpha x + \alpha y$, for all $\alpha \in K$ and $x, y \in V$
The elements of V are called vectors.

We will only work with vector spaces over the fields of real $\mathbb{R}$ or complex $\mathbb{C}$ numbers and not mention the field of scalars, when it is $\mathbb{R}$.

Examples:

1. The set of column vectors $\mathbb{R}^n$, equipped with the usual component wise addition of vectors and multiplications by scalars is a vector space over $\mathbb{R}$.

2. The set V of matrices $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$, $\forall a, b, c \in \mathbb{R}$ is a vector space over $\mathbb{R}$ with respect to the usual matrix addition and scalar multiplication operations.

3. The set V of all real polynomials $P(x) = a_0x^n + \cdots + a_{n-1}x + a_n$ of degree not greater than n with the usual polynomial addition and scalar multiplication operations is a vector space over $\mathbb{R}$.

   Is the set W of all polynomials of degree *exactly equal to* n a vector space?

4. The set $C[0, 1]$ of all continuous functions $f : [0, 1] \to \mathbb{R}$ with the usual function addition and scalar multiplication operations is a vector space over $\mathbb{R}$.

Examples 1 to 3 involve *finite–dimensional* vector spaces and 4 involves an *infinitely–dimensional* one. We recall the concept of dimension in the sequel.

A subset $A \subset V$ of a vector space is said to be a *sub–space* if it is itself a vector space (over the same field of scalars).

Notation: $A \leq V$.

Definition 2.2 A finite set of vectors $\{v_1, \ldots, v_n\} \subset V$ in a vector space V is said to be *linearly independent* if

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = \vec{0} \quad \text{if and only if} \quad \alpha_1 = \cdots = \alpha_n = 0.$$

Otherwise, the vectors $v_1, \ldots, v_n$ are said to be *linearly dependent.*

If the vectors $v_1, \ldots, v_n$ are linearly dependent, then at least one of them can be expressed as a linear combination of the others. Recall the notion of a linear envelop or *span* of a set of vectors $\{v_1, \ldots, v_n\}$:

Notation: $\text{span}\{v_1, \ldots, v_n\} = \{v = \alpha_1 v_1 + \cdots + \alpha_n v_n,\ \forall \alpha_1, \ldots, \alpha_n \in \mathbb{R}\}$. The span of an infinite set of vectors A is the smallest vector space which contains A.

Examples:

1. $A = \{x = (x_1 \ x_2 \ 0)', \ \forall x_1, x_2 \in \mathbb{R}\} \leq \mathbb{R}^3$.

2. $C^1(\mathbb{R}) \leq C(\mathbb{R})$, and, in fact, $C^\infty(\mathbb{R}) \leq C^n(\mathbb{R}) \leq C(\mathbb{R}), \ \forall n \in \mathbb{N}$.

3. For any $v_1, \dots, v_n \in V$, we have $\mathrm{span}\{v_1, \dots, v_n\} \leq V$.

Definition 2.3 A set of vectors $S \subset V$ is called a basis for the vector space V if
a. S is linearly independent
b. $\mathrm{span}(S) = V$
The vector space V is said to be *finite–dimensional* if it has a basis of finite cardinality.

Remark. Definitions of linear independence, span and basis can be given for *infinitely–dimensional spaces.* For simplicity, we focus here only on finitely–dimensional vector spaces.

Proposition 2.1 *Let $\{v_1, \dots, v_n\}, \ n \in \mathbb{N}$ be a basis of the vector space V. Then for any $v \in V$, there exist unique set of scalars $\alpha_1, \dots, \alpha_n$, such that $v = \alpha_1 v_1 + \dots + \alpha_n v_n$.*

The proof is an exercise. The unique (ordered) set of scalars $(\alpha_1, \dots, \alpha_n)$ associated with a vector $v \in V$ are called its *coordinates* in the basis $\{v_1, \dots, v_n\}$. Since the coordinates identify uniquely an element of V, we will see below that any vector space V, over $\mathbb{R}$, for example, can be identified with the vector space of column vectors $\mathbb{R}^n$. We need however the following result which justifies the definition of the dimension of a vector space.

Theorem 2.1 *Let $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_m\}, \ m, n \in \mathbb{N}$ be sets of linearly independent vectors in the space V.*
a. *If $v_1, \dots, v_n \in \mathrm{span}\{w_1, \dots, w_m\}$, then $n \leq m$.*
b. *If both $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_m\}$ are bases of V, then, $n = m$ and for some non–singular square matrix $T = (t_{ij})_{n \times n}$,*

$$(v_1 \ \cdots \ v_n) = (w_1 \ \cdots \ w_m) T, \tag{2.1}$$

i.e. $v_j = \sum_{i=1}^m w_i t_{ij}$.

Proof: Focus on part **a.** Since $v_1, \dots, v_m \in \mathrm{span}\{w_1, \dots, w_n\}$, we have that Relation (2.1) holds for some $(m \times n)$ matrix $T = (t_{ij})_{m \times n}$. Suppose that $m < n$. One can show that for a transformation matrix $B = (b_{ij})_{n \times n}$:

$$TB = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & 0 \\ & * & \cdots & \lambda_m & 0 \end{pmatrix},$$

for some $\lambda_1, \dots, \lambda_m \in \mathbb{R}$. The right–multiplication with the transformation matrix B corresponds to interchanging columns and adding a multiple of one column to another.

We then have $vB = wTB$, where $v = (v_i)_{1 \times n}$ and $w = (w_j)_{1 \times m}$. By focusing on the last column of TB, we obtain that

$$b_{1n} v_1 + \dots + b_{nn} v_n = \vec{0}.$$

This however implies $b_{1n} = \dots = b_{nn} = 0$ because $\{v_1, \dots, v_n\}$ are linearly independent, which is impossible since $\det(B) \neq 0$ (recall that transformation matrices have non–zero determinants).

To prove part **b.** observe that in this case also $w_1, \ldots, w_m \in \text{span}\{v_1, \ldots, v_n\}$ and hence $m \le n$. This implies $n = m$ and the matrix T in (2.1) is square.

Finally, if T is singular, one of the diagonal elements λ_i's is zero and one can show that $Tb = (0)$, for some non–zero column vector $b = (b_j)_{n\times 1}$ (Why?). Hence the vector $vb = b_1v_1 + \cdots + b_nv_n = wTb$ is zero. This is impossible since the v_i's are linearly independent and hence T must be non–singular. □

Since all bases of finite–dimensional vector spaces have the same size (cardinality), we have

Definition 2.4 A finite–dimensional vector space V has dimension, denoted $\dim(V)$, equal to the cardinality of any basis.

Theorem 2.1 **b.** indicates that any two vector spaces (over a field of scalars K) with equal dimensions are *isomorphic*. The vector spaces V and W (over K) are isomorphic if there exist a bijection map (one-to-one and onto) $J : V \to W$ such that

$$J(\alpha v + \beta w) = \alpha J(v) + \beta J(w), \quad \forall \alpha, \beta \in K \ \text{ and } \ \forall v, w \in V. \tag{2.2}$$

See the exercises in Section 2.5 below.

Any vector space V over $\mathbb{R}$ of dimension $n \in \mathbb{N}$ is isomorphic to $\mathbb{R}^n$. Thus, once a basis $\{v_1, \ldots, v_n\}$ is fixed, we can work with the space of column vectors $\mathbb{R}^n$ representing the *coordinates* of vectors $v \in V$. Once the basis is changed, however, the coordinates change. The following formula relates the coordinates in two bases.

Let $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_n\}$ be two bases of the vector space V. Let also $v \in V$ has coordinates $x = (x_1 \ \cdots \ x_n)'$ in $\{v_1, \ldots, v_n\}$ and coordinates $y = (y_1 \ \cdots \ y_n)'$ in $\{w_1, \ldots, w_n\}$, respectively. In view of Relation (2.1), we have that

$$y = Tx. \tag{2.3}$$

A map $J : V \to W$ between two vector spaces is said to be *linear* if it satisfies (2.2). We study linear maps in Section 2.2 below. We conclude this section with the notions *intersection* and *direct sum* of vector spaces.

Let $V_1 \le V$ and $V_2 \le V$ be two sub–spaces of the finite–dimensional vector space V. One can easily show that $W := V_1 \cap V_2$ is a vector space and hence it is also finite–dimensional (see, Section 2.5, below). The union $V_1 \cup V_2$ is *not* a vector space unless $V_1 \subset V_2$ or $V_2 \subset V_1$. The *sum* $W = V_1 + V_2$ of the vector spaces V_1 and V_2, defined as:

$$W := V_1 + V_2 = \{v_1 + v_2 \ : \ v_i \in V_i, \ i = 1, 2\}$$

is a vector space. One can easily show that

$$\dim(V_1 + V_2) = \dim(V_1) + \dim(V_2) - \dim(V_1 \cap V_2). \tag{2.4}$$

Proposition 2.2 *Consider a finite–dimensional vector space V and let $V_1 \le V$ and $V_2 \le V$. If $V_1 \cap V_2 = \{\vec{0}\}$, then for any $w \in W = V_1 + V_2$, there exist unique $v_1 \in V_1$ and $v_2 \in V_2$, such that $w = v_1 + v_2$.*

The sub–space W in the context of Proposition 2.2 is called the *direct sum* of the sub–spaces V_1 and V_2. In this case since $V_1 \cap V_2 = \{\vec{0}\}$, in view of (2.4),

$$\dim(V_1 + V_2) = \dim(V_1) + \dim(V_2).$$

Corollary 2.1 *Consider a finite–dimensional vector space V and let $W \le V$. If $\dim(W) = V$, then $W \equiv V$. Otherwise, there exists a sub–space $\widetilde{W} \le V$, such that*

$$V = W + \widetilde{W}, \quad \textit{for some} \quad \widetilde{W} \le V, \quad \textit{with } \dim(\widetilde{W}) = \dim(V) - \dim(W).$$

In view of (2.4), $\widetilde{W} \cap W = \{\vec{0}\}$ and the space $\widetilde{W}$ is said to be a direct complement to W in V.

Remark. For $W \le V$ with $\dim(W) < \dim(V)$, there are infinitely many different direct complement spaces $\widetilde{W}$ of W in V. Intuitively, $\widetilde{W}$ may be "angled" in different ways w.r.t. W. We will see in the sequel that there is an unique *orthogonal complement* sub–space $\widetilde{W}$.

2.2 Linear transformations, change of basis

Here, we discuss briefly linear transformations, their matrix representations, range and kernel (null space). We also discuss a rank of a matrix and the fundamental theorem of linear algebra.

Consider two vector spaces V and W (say, over $\mathbb{R}$) with $\dim(V) = n$ and $\dim(W) = m$, $n, m \in \mathbb{N}$.

Definition 2.5 A map $L : V \to W$ is said to be linear (transformation), if for any scalars α and β and $v_1, v_2 \in V$, we have $L(\alpha v_1 + \beta v_2) = \alpha L(v_1) + \beta L(v_2)$.

It follows that for a linear map L, $L(\vec{0}) = \vec{0}$. Now, fix two bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ of the spaces V and W, respectively. Let $v \in V$ have coordinates $x = (x_j)_{n\times 1} \in \mathbb{R}^n$ in the basis $\{v_j\}$ and let

$$w = L(v)$$

have coordinates $y = (y_i)_{m\times 1} \in \mathbb{R}^m$ in the basis $\{w_i\}$.

Then, there exist (a unique) matrix $A = (a_{ij})_{m\times n}$,

$$y = Ax, \quad \text{where} \quad L(v_j) = \sum_{i=1}^{m} w_i a_{ij}, \tag{2.5}$$

that is, $L(v) = wA$, with $L(v) = (L(v_1)\ \cdots\ L(v_n))$ and $w = (w_1\ \cdots\ w_m)$ are the *formal row–vectors* with components elements of the vector space W.

Relation (2.5) provides the matrix representation of the linear map L in the bases $\{v_i\}$ and $\{w_j\}$. We next show how the matrix representations change when the bases change. Consider, in addition, the bases $\{v_1^*, \ldots, v_n^*\}$ and $\{w_1^*, \ldots, w_m^*\}$ of V and W, respectively. Construct the matrix $P = (p_{ij})_{n\times n}$ with columns the coordinates of v_i^* in the basis $\{v_i\}$. Observe that a vector v with coordinates $x^* = (x_i^*)_{n\times 1} \in \mathbb{R}^n$ in $\{v_i^*\}$ will then have coordinates $x = Px^*$.

Similarly, let the columns of $Q = (q_{ij})_{m\times m}$ be the coordinates of w_i^* in $\{w_i\}$ and note that $y = Qy^*$, where $y \in \mathbb{R}^n$ and $y^* \in \mathbb{R}^n$ are the coordinates of some vector w in the bases $\{w_i\}$ and $\{w_i^*\}$, respectively.

Now, formally, from $y = Ax$ we obtain

$$Qy^* = APx^* \quad \Leftrightarrow \quad y^* = Q^{-1}APx^*,$$

and hence the matrix A^* representing L in the new bases $\{v_i^*\}$ and $\{w_i^*\}$ is:

$$A^* = Q^{-1}AP. \tag{2.6}$$

Observe that by Theorem 2.1, Q^{-1} exists.

The special case when the two spaces V and W coincide and the bases $\{v_i\}$ and $\{w_i\}$ are the same is very important. We then obtain

$$A^* = P^{-1}AP.$$

Examples:

1. Let
$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$
be the matrix of a linear transformation $L : \mathbb{R}^2 \to \mathbb{R}^2$ in the *standard basis*: $v_1 = (1\ 0)'$ and $v_2 = (0\ 1)'$.

   Determine the matrix representation of the same map in the *new basis*: $v_1^* = (1\ -1)'$ and $v_2^* = (1\ 1)'$.

2. Multiplication by the matrix A, stretches by a factor of $2 = \det(A)$ and rotates the plane $\mathbb{R}^2$ by 45^o counterclockwise. Using this, one can show that
$$A^n = 2^n \begin{pmatrix} \cos(n\pi/4) & -\sin(n\pi/4) \\ \sin(n\pi/4) & \cos(n\pi/4) \end{pmatrix}.$$

For simplicity we will frequently identify linear maps with their matrices, when the bases are understood. We next introduce the important notions of a range and kernel (or null space) of a matrix $A = (a_{ij})_{m\times n}$.

Definition 2.6 **a.** The range of A, denoted $\mathcal{R}(A)$, is the span of the columns of the matrix A.
b. The kernel (or null space) of A, denoted $\mathcal{N}(A)$, is defined as $\{x \in \mathbb{R}^n \ :\ Ax = (0)\}$.
c. Range and null space are defined similarly for a linear map $L : V \to W$.

One can easily show that the range $\mathcal{R}(A)$ and the null space $\mathcal{N}(A)$ of a matrix A are *vector spaces*. The following result, known as *the fundamental theorem of linear algebra*, relates the dimensions of these spaces to the dimensions of the matrix A and its rank.

Theorem 2.2 *For any matrix $A = (a_{ij})_{m\times n}$ of real scalars, we have that:*

$$r = \dim(\mathcal{R}(A)) = \dim(\mathcal{R}(A^t)),$$

where A^t denotes the transposed matrix A. The number r is called rank of the matrix A, denoted rank(A). *In addition, we have that*

$$\dim(\mathcal{N}(A)) = n - r \quad \textit{and} \quad \dim(\mathcal{N}(A^t)) = m - r. \tag{2.7}$$

PROOF: One can easily see that for some transformation matrices $T_n = (t_{ij})_{n\times n}$ and $S_m = (s_{ij})_{m\times m}$, we have

$$S_m A T_n = D = \begin{pmatrix} \Lambda & (0)_{r\times(n-r)} \\ (0)_{(m-r)\times r} & (0)_{(m-r)\times(n-r)} \end{pmatrix}, \tag{2.8}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ and where $(0)_{p\times q}$ denotes the zero–matrix of dimensions $p \times q$. Here $1 \le r \le \min\{m, n\}$ and the entries of the diagonal matrix Λ are all non–zero (with the exception of the trivial case when $A = (0)_{m\times n}$).

The matrix A represents a linear map L of $\mathbb{R}^n$ to $\mathbb{R}^m$. Recall that transformation matrices are non–singular. Hence, we can consider bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ of $\mathbb{R}^n$ and $\mathbb{R}^m$, so that

$$(v_1 \ \ldots \ v_n) = T_n \quad \text{and} \quad (w_1 \ \ldots \ w_m) = S_m^{-1}.$$

Then, in view of (2.6), the matrix of the linear map L in these new bases is

$$(S_m^{-1})^{-1} A T_n = D$$

as in (2.8).

Notice however that $\mathcal{R}(A) = L(\mathbb{R}^n)$ and hence the dimension of the range of A does not depend on the basis. Hence $\dim(\mathcal{R}(A)) = r$.

Similarly, consider the linear map $M : \mathbb{R}^m \to \mathbb{R}^n$ with matrix A^t and let the new bases $\{v_i^*\}$ and $\{w_i^*\}$ of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, be such that the matrix of M is:

$$T_n^t A^t S_m^t = D^t.$$

As argued above we also have $\dim(\mathcal{R}(A^t)) = r$.

Observe that the value of r does not depend on the matrices S_m and T_m since it is related to the dimension of the vector spaces $L(\mathbb{R}^n)$ and $M(\mathbb{R}^m)$, respectively. These vector spaces do not depend on the choice of bases!

Now, it remains to show (2.7). Let $V_1 := \{v \in \mathbb{R}^n \ : \ L(v) = \vec{0}\}$ and observe that $\dim(V_1) = \dim(\mathcal{N}(A))$. However, when viewed in the bases $\{v_i\}$ and $\{w_i\}$, the matrix of L is D. It is clear that $\dim(\mathcal{N}(D)) = n - r$ and since D is just another representation of the same linear map L, we have that $\dim(V_1) = \dim(\mathcal{N}(D))$. The argument about A^t is similar. □

A popular definition of the rank of a matrix $A = (a_{ij})_{m\times n}$ is through *sub–matrices.* If we delete some rows and/or columns of A, we obtain a sub–matrix.

Definition 2.7 The rank of the matrix A is the dimension of the largest non–singular square sub–matrix.

This definition of $\mathrm{rank}(A)$ is equivalent to the one given in Theorem 2.2 above. Often the dimension of the range of A is referred to as a *column rank* and the dimension of the range of A^t as to *row rank.* By Theorem 2.2 it follows that column and row rank coincide.

2.3 Inner product spaces, orthogonal bases and projections

Before discussing *inner product spaces*, we recall the definition of a norm.

Definition 2.8 Let V be a vector space over the real or complex numbers. A functional assigning a non–negative real number to any $v \in V$, denoted $\|v\|$, is said to be a *norm* if:

1. *(homogeneity)* $\|\lambda v\| = |\lambda|\|v\|$, for all $v \in V$ and scalars λ.

2. *(triangle inequality)* $\|u + v\| \leq \|u\| + \|v\|$, for all $u, v \in V$.

3. $\|u\| \geq 0$, for all $u \in V$ and $\|u\| = 0$ if and only if $u = \vec{0}$.

A vector space equipped with a norm is called a normed vector space.

Examples:

1. Let V be $\mathbb{R}^n$ or $\mathbb{C}^n$, $n \in \mathbb{N}$. The following are examples of norms in the space V:

   a. *(the Euclidean norm)* $\|x\|_2 := (\sum_{i=1}^n |x_i|^2)^{1/2}$.

   b. $\|x\|_1 := \sum_{i=1}^n |x_i|$.

   c. *(the uniform norm)* $\|x\|_\infty := \max_{1 \leq i \leq n} |x_i|$.

2. Let V be the infinite dimensional vectors space of continuous, real valued functions $C[0, 1]$ with the usual function addition and scalar multiplication. The following are examples of norms:

   a. *(the uniform convergence norm)* $\|f\|_\infty := \max_{x \in [0,1]} |f(x)|$.

   b. *(the L^1–norm)* $\|f\|_1 := \int_0^1 |f(x)| dx$.

The fact that the Euclidean norm $\|x\|_2$ is a norm in $\mathbb{R}^n$ (or $\mathbb{C}^n$) is not trivial. It is most difficult to check the *triangle inequality.* For $\mathbb{R}^2$ and $\mathbb{R}^3$, this is precisely the usual triangle inequality. In general, the triangle inequality follows from the Cauchy–Buniakowski–Schwartz inequality:

$$\sum_{i=1}^n x_i y_i \leq (\sum_{i=1}^n |x_i|^2)^{1/2} (\sum_{i=1}^n |y_i|^2)^{1/2}, \tag{2.9}$$

valid for all $x_i,\ y_i \in \mathbb{C}$. We will establish (2.9) later. Now, we show that the Euclidean norm is derived naturally from an *inner product.*

Definition 2.9 Let V be a vector space over the real $\mathbb{R}$ or complex $\mathbb{C}$ numbers. A binary operation on V, assigning a scalar to any pair of vectors $u, v \in V$, denoted $\langle u, v\rangle$, is called *an inner product* if:

1. $\langle u, v\rangle = \overline{\langle v, u\rangle}$, for all $u, v \in V$, where $\overline{z}$ denotes the complex conjugate of $z \in \mathbb{C}$. Note that when the scalars are real, we simply have $\langle u, v\rangle = \langle v, u\rangle$.

2. $\langle u_1 + u_2, v\rangle = \langle u_1, v\rangle + \langle u_2, v\rangle$, for all $u_1, u_2, v \in V$.

3. $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$, for all $u \in V$, and for any scalar λ.

4. $\langle u, u\rangle \geq 0$, for all $u \in V$ and $\langle u, u\rangle = 0$, if and only if $u = \vec{0}$.

The space V, equipped with $\langle \cdot, \cdot\rangle$, is called an inner product space. A finite–dimensional inner product space over $\mathbb{R}$ is called an Euclidean space.

Examples:

1. Let $V = \mathbb{R}^n$ and for any $x = (x_i)_{n\times 1}$ and $y = (y_i)_{n\times 1}$, define

$$\langle x, y\rangle := x'y = \sum_{i=1}^{n} x_i y_i.$$

It is easy to check that $\langle x, y\rangle$ is an inner product. When the real scalars $\mathbb{R}$ are replaced by the complex $\mathbb{C}$, then the above form does *not* satisfy property **1.** of the inner product.

For $V = \mathbb{C}^n$, we have that

$$\langle x, y\rangle := x'\overline{y} = \sum_{i=1}^{n} x_i \overline{y_i}$$

defines an inner product.

2. Let $V = C[0, 1]$ be the set of continuous functions on the closed interval $[0, 1]$. Define

$$\langle f, g\rangle := \int_0^1 f(x)g(x)dx.$$

The so defined quantity $\langle f, g\rangle$ is an inner product in the vector space $C[0, 1]$.

3. Let $V = \ell_2(\mathbb{R}) = \{(x_1, x_2, \cdots) \ : \ x_i \in \mathbb{R}, \ \sum_{i=1}^{\infty} x_i^2 < \infty\}$ be the set of all infinite sequences with finite sums of squares. It is a vector space with respect to the usual component–wise addition and scalar multiplication. Define

$$\langle x, y\rangle := \sum_{i=1}^{\infty} x_i y_i, \quad \forall x = (x_i)_{i=1}^{\infty}, \ y = (y_i)_{i=1}^{\infty} \in V, \tag{2.10}$$

which is an inner product.

Remarks: Example 1 involves finite–dimensional inner product spaces whereas Examples 2 and 3 involve an infinitely dimensional ones. Example 3, moreover, is an instance of a Hilbert space, which we will study in detail later. In Example 2, V is not a Hilbert space since it is not *complete* (to be defined later).

• *The norm generated by an inner product*

One can equip any inner product space V with a natural norm, defined as follows:

$$\|x\| := \sqrt{\langle x, x\rangle}, \quad x \in V.$$

The fact that $\|x\|$ is a valid norm can be easily shown by using the the following inequality, known as the *Cauchy–Buniakowski–Schwartz inequality.* For any $x, \ y \in V$, we have

$$|\langle x, y\rangle| \le \|x\|\|y\| = \sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}. \tag{2.11}$$

We next prove the inequality (2.11) for real inner product spaces. For any $x, y \in V$ and $\lambda \in \mathbb{R}$, we have

$$g(\lambda) := \langle x + \lambda y, x + \lambda y\rangle \ge 0.$$

If $y = \vec{0}$, the inequality (2.11) is trivial. Let $y \neq \vec{0}$ and observe that for all $\lambda \in \mathbb{R}$, the quadratic polynomial

$$g(\lambda) = \lambda^2\|y\|^2 + 2\lambda\langle x, y\rangle + \|x\|^2$$

is non–negative. Thus, its discriminant is non–positive:

$$D = 4\langle x, y\rangle^2 - 4\|x\|^2\|y\|^2 \leq 0 \quad \Leftrightarrow \quad |\langle x, y\rangle| \leq \|x\|\|y\|,$$

which implies (2.11). Moreover, $|\langle x, y\rangle| \leq \|x\|\|y\|$, if and only if, $D = 0$, which implies that for $\lambda_0 = \langle x, y\rangle/\|y\| \in \mathbb{R}$, $g(\lambda_0) = 0$, i.e. $x + \lambda y = 0$ or x and y are linearly dependent.

• *Orthogonal bases. Gramm–Schmidt orthogonalization.*

Let V be an inner product space. The additional structure of the inner product provides the notion of an *angle* between two vectors and in particular an extension of the notion of *orthogonality* from Euclidean geometry.

Definition 2.10 The vectors $v_1, \ldots, v_n,\ n \geq 2$ are said to be orthogonal, if

$$\langle v_i, v_j\rangle = 0, \qquad \text{for all } i \neq j,\ 1 \leq i, j \leq n.$$

If the vectors have also unit norms: $\|v_i\| = 1,\ 1 \leq i \leq n$, then they are said to be an *orthonormal set* of vectors.

Notation: If u and v are orthogonal, then we write $u \perp v$. Further, if $u \perp v$, for all $v \in W$, then we write $u \perp W$.

One can easily show the following

Proposition 2.3 *Any set of non–zero, orthogonal vectors $\{v_1, \ldots, v_n\}$ is linearly independent.*

The proof is an exercise.

Recall that if V is finite–dimensional with $\dim(V) = n,\ n \in \mathbb{N}$, then any set of more than n vectors in V is linearly dependent. We can have however, a set of precisely n linearly independent vectors $\{v_1, \ldots, v_n\}$, which constitutes a *basis* of V.

We next answer the question: *Are there orthogonal bases of V?*

• *Gramm–Schmidt orthogonalization*

Let $\{v_1, \ldots, v_n\}$ be a basis of V. We will construct in n steps an orthogonal basis $x_1, \ldots, x_n$ of V, by iteratively "adjusting" the angles of the vectors $v_1, \ldots, v_n$.

Step 1: Set $x_1 := v_1$.

Step 2: Set $x_2 := v_2 - \lambda_{11}x_1$ and solve for the scalar λ_{11} so that $x_2 \perp x_1$. We have:

$$0 = \langle x_2, x_1\rangle = \langle v_2, x1\rangle - \lambda_{11}\langle x_1, x_1\rangle \quad \Leftrightarrow \quad \lambda_{11} = \langle v_2, x_1\rangle/\langle x_1, x_1\rangle.$$

$$\cdots$$

Step k+1: Set $x_{k+1} := v_{k+1} - \lambda_{k1}x_1 - \cdots - \lambda_{kk}x_k$ so that $x_{k+1} \perp \operatorname{span}\{x_1, \ldots, x_k\}$. We have:

$$\lambda_{k1} = \langle v_{k+1}, x_1\rangle/\langle x_1, x_1\rangle, \quad \cdots, \quad \lambda_{kk} = \langle v_{k+1}, x_k\rangle/\langle x_k, x_k\rangle.$$

$\cdots$

After the $n-$th step of the Gramm–Schmidt orthogonalization process, we have an *orthogonal* set of non–zero vectors $\{x_1, \ldots, x_n\}$, such that:

1. $\mathrm{span}\{x_1, \ldots, x_k\} = \mathrm{span}\{v_1, \ldots, v_k\}$, for all $k = 1, \ldots, n$.

2. $x_{k+i} \perp x_1, \ldots, x_k$, for all $i, 1 \leq ilen - k$ and for all $k = 1, \ldots, n$.

• *Orthogonal projection and orthogonal vector spaces*

Let W be a sub–space of a *finite–dimensional* inner product space V. Then, for every $v \in V$, we define its orthogonal projection onto W, denoted $P_W(v)$, as such an element $w_0 \in W$ that

$$v - w_0 \perp W, \quad \text{that is} \quad \langle v - w_0, w \rangle = 0, \quad \forall w \in W.$$

One can easily show that if the orthogonal projection w_0 of v onto W exists, then it is *unique.* We now show that the orthogonal projection always exists. Indeed, let $\{w_1, \ldots, w_r\}$ be a basis of W. As seen before, we can complete the basis $\{w_1, \ldots, w_r\}$ to a basis $\{w_1, \ldots, w_r, u_1, \ldots, u_{n-r}\}$ of V, where $\dim(V) = n$ and $\dim(W) = r \leq n$.

Now, the Gramm–Schmitd orthogonalization process, applied to $\{w_1, \ldots, w_r, u_1, \ldots, u_{n-r}\}$ yields an orthogonal basis $\{x_1, \ldots, x_r, x_{r+1}, \ldots, x_n\}$. In addition, we have that $\mathrm{span}\{x_1, \ldots, x_r\} = \mathrm{span}\{w_1, \ldots, w_r\} = W$.

Now, let $v \in V$ be arbitrary and let the scalars $\alpha_1, \ldots, \alpha_n$ be the coordinates of v in the basis $\{x_i\}$. We have:

$$v = \underbrace{\alpha_1 x_1 + \cdots + \alpha_r x_r} + \alpha_{r+1} x_{r+1} + \cdots + \alpha_n x_n.$$

It is then clear that $w_0 := \alpha_1 x_1 + \cdots + \alpha_r x_r$ is the orthogonal projection of v onto W.

This argument also shows that the orthogonal projection $P_W : V \to W \subset V$ is a linear operator. Indeed, the matrix representation of P_W in the basis $\{x_1, \ldots, x_r, x_{r+1}, \ldots, x_n\}$ is

$$A = \begin{pmatrix} I_r & (0)_{r\times(n-r)} \\ (0)_{(n-r)\times r} & (0)_{(n-r)\times(n-r)} \end{pmatrix}. \tag{2.12}$$

Observe that $P_W \circ P_W = P_W$ since the orthogonal projection of any $w \in W$ onto the sub–space W is in fact the same element w. This can be also seen from the structure of the matrix A, since $A^2 = A$. We discuss orthogonal and projection matrices in Section 2.4, below. We next introduce the notion of *orthogonal complement* of a sub–space.

Definition 2.11 Let $W \leq V$ be a sub–space of the inner–product space V. The set of vectors

$$W^{\perp} := \{v \in V \ : \ v \perp W\}$$

is a sub–space of V. It is called the *orthogonal complement* of W in V.

One can show that $W \perp W^{\perp}$, that is, the sub–spaces W and $W^{\perp}$ are orthogonal. This implies $W \cap W^{\perp} = \{\vec{0}\}$ and therefore, we have that V is the direct sum of the spaces W and $W^{\perp}$, denoted

$$V = W \oplus W^{\perp}.$$

In fact, if the sub–spaces $W_1, W_2, \ldots, W_k$ of V are mutually orthogonal, and $W_1 + \cdots + W_k = V$, we write:

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_k.$$

In this, case any $v \in V$, has a unique representation:

$$v = w_1 + w_2 + \cdots + w_k, \quad \text{where } w_i \in W_i,$$

and where, in fact, w_i is the orthogonal projection of v onto W_i.

We conclude this section with an important interpretation of orthogonal projections in terms of "best approximation":

Proposition 2.4 *Let $v \in V$, and $W \leq V$, where V is an inner product space. The vector $w_0 \in W$ is the orthogonal projection of v onto W, if and only if, $\|v - w_0\| \leq \|v - w\|$, for all $w \in W$.*

PROOF: Let first $w_0 = P_W(v)$. Then, for any $w \in W$, we have that $v - w_0 \perp w_0 - w$. Thus, the Pythagoras' theorem implies (draw a picture and see, Section 2.5 below), that

$$\|v - w\|^2 = \|(v - w_0) + (w_0 - w)\|^2 = \|v - w_0\|^2 + \|w_0 - w\|^2 \geq \|v - w_0\|^2.$$

Conversely, suppose that $\|v - w_0\| \leq \|v - w\|, \ \forall w \in W$. For simplicity we deal only with the case of real scalars. We want to show that $\langle v - w_0, w\rangle = 0, \ \forall w \in W$. Suppose that for some $w^* \in W$, we have $a = \langle v - w_0, w^*\rangle \neq 0$. Let $\lambda \in \mathbb{R}$ and consider the quantity

$$g(\lambda) := \|v - (w_0 + \lambda w^*)\|^2 - \|v - w_0\|^2 \geq 0, \quad \forall \lambda \in \mathbb{R}.$$

Observe that

$$g(\lambda) = -2\lambda\langle v - w_0, w^*\rangle + \lambda^2\|w^*\|^2.$$

However, $g(\lambda) \geq 0, \ \forall \lambda \in \mathbb{R}$, which is impossible. Indeed, since $w^* = \vec{0}$ and $\langle v - w_0, w^*\rangle \neq 0$, the quadratic form has two distinct roots and hence it takes negative values for some real λ. This contradiction implies $\langle v - w_0, w\rangle = 0, \ \forall w \in W$. $\Box$

Remark. We showed above that in finite–dimensional inner product spaces the orthogonal projection (of a vector onto a sub–space) always exists. This is not always the case for infinitely dimensional spaces unless they are *complete*, that is, the so–called *Hilbert spaces.*

2.4 Orthogonal projection matrices. Linear regression

A set of orthogonal non–zero vectors $\{v_1, \ldots, v_n\}$ is said to be *orthonormal* if $\|v_i\| = 1, \ \forall i$. That is, the v_i's are normalized to have unit norms.

In this section, we consider only the Euclidean spaces $\mathbb{R}^n, \ n \in \mathbb{N}$ of column vectors.

Definition 2.12 The matrix $P = (p_{ij})_{m\times n}$, with $n \leq m$ is said to be orthogonal, if its column vectors $p_j = (p_{ij})_{m\times 1}, \ j = 1, \ldots, n$ are orthonormal.

Observe that $P = (p_{ij})_{m\times n}$ is an orthogonal matrix if and only if $P^tP = I_n$.

Definition 2.13 The matrix corresponding to any (orthogonal) projection operator P_W is said to be a *projection matrix.*

Recall that, in a suitable orthogonal basis, any projection operator has the matrix representation in (2.12) above. For an a-priori fixed basis, however, the projection operator may (and typically will) have more intricate matrix. We will now derive the form on the projection matrix corresponding to an arbitrary sub–space W of $\mathbb{R}^m$.

Let the sub–space $W \le \mathbb{R}^m$ have the orthonormal basis $w_1, \ldots, w_n$, where $w_j = (w_{ji})_{m\times n}$, $j = 1, \ldots, n$ are column vectors and where $n = \dim(W)$, $n \le m$. As argued above (see (2.12)), complete the orthonormal set of column vectors $w_1, \ldots, w_n$ to the following orthonormal basis of $\mathbb{R}^m$:

$$p_1 := w_1,\ p_2 := w_2, \cdots, p_n := w_n,\ p_{n+1} := u_1, \cdots, p_m := u_{m-n}.$$

Form the matrix $P = (p_{ij})_{m\times m}$ with columns given by the so–constructed orthonormal basis.

Observe that the matrix of the projection operator P_W in the basis $\{p_i\}$ is:

$$D = \begin{pmatrix} I_m & (0) \\ (0) & (0) \end{pmatrix}.$$

However, any column vector $x = (x_i)_{m\times 1}$ has coordinates $y = (y_i)_{m\times 1}$ in the basis $\{p_i\}$, which satisfy the "change of coordinates" relation $x = Py$. Therefore, $y = P^{-1}x$ and for the matrix A of the projection operator in the standard basis of $\mathbb{R}^m$, we obtain:

$$A = PDP^{-1} = PDP^t = (W\widetilde{W}) \begin{pmatrix} I_m & (0) \\ (0) & (0) \end{pmatrix} \begin{pmatrix} W^t \\ \widetilde{W}^t \end{pmatrix}.$$

Here, we used the "block–wise" notation where $W = (w_{ij})_{m\times n}$ denotes the matrix with columns given by the w_i's and $\widetilde{W}$ is the rest of the orthogonal matrix P.

Notice now that the right–hand side of the last relation is equal to: WW^t. That is, we have shown that:

Proposition 2.5 *Consider the orthogonal projection operator a subspace W of $\mathbb{R}^m$. Then, the matrix of P_W (in the standard basis of $\mathbb{R}^m$) has the form:*

$$A = WW^t,$$

where $W = (w_{ij})_{m\times n}$ is a matrix with orthonormal columns, which span the projection sub–space $W \le \mathbb{R}^m$.

It follows from the last fact that any projection matrix A must be *symmetric.* It is also clear that $A^2 = A$ i.e. A is idempotent. These two conditions, it turns out, are sufficient for a matrix to be a projection matrix.

Proposition 2.6 *The matrix $A = (a_{ij})_{n\times n}$ is a projection matrix if and only if:*

1. *A is symmetric.*

2. *A is indempotent: $A^2 = A$.*

Proof: The "only if" part follows from Proposition 2.5. Let now A be a symmetric and idempotent matrix. Set $W = \mathcal{R}(A)$. It suffices to show that $x - Ax \perp y$, for all $y \in W$. That is, that for all x and $z \in \mathbb{R}^m$, $(Az)'(x - Ax) = 0$, or equivalently that

$$z'A'x = zA'Ax, \quad \forall x, z \in \mathbb{R},$$

which is clearly true for a symmetric and idempotent matrix A. $\square$

We next consider an important application of projection matrices to the linear regression problem.

• *Multiple linear regression through projection matrices*

We borrow the following presentation (with certain changes in notation) from Example 2.11 in Schott (2005). Consider the *multiple linear regression setup*:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon, \quad k \geq 1,$$

where y, called the dependent variable or the response is modelled as a linear function of the independent variables $x_1, \ldots, x_k$ or predictors plus unobservable noise ϵ. Typically, one observes values $Y_1, \ldots, Y_N$ of the variable y and their corresponding x_i values $X_{i1}, \ldots, X_{ik}$, $i = 1, \ldots, N$. The goal is then to find a vector of parameters $\widehat{\beta} = (\widehat{\beta}_i)_{(k+1)\times 1}$, which yields "best" approximation of the y's as a linear function of the x_i's. The classical approach is that of *least squares*, i.e. to minimize the squared error:

$$\widehat{\beta} = \mathrm{Argmin}_{\beta \in \mathbb{R}^{k+1}} \|Y - X\beta\|^2, \tag{2.13}$$

where $Y = (Y_i)_{N\times 1}$, $X = \Big((1)_{N\times 1}\ X_{ij}\Big)_{N\times(k+1)}$ and where $\beta = (\beta_0\ \beta_1\ \cdots\ \beta_k)'$.

Observe that if $Y \in \mathcal{R}(X)$, that is, if the vector Y belongs to the column space of the *matrix* X, then one can find $\widehat{\beta} \in \mathbb{R}^{k+1}$, such that $Y = X\widehat{\beta}$. This is typically not the case since $N >> k + 1 \geq \mathrm{rank}(X)$.

Notice also, in view of Proposition 2.5, that the solution of (2.13) is such that $Y_0 = X\widehat{\beta}$ is the orthogonal projection of Y onto the column space of X, i.e. onto $\mathcal{R}(X)$.

Proposition 2.7 *Let $X = (x_{ij})_{m\times n}$, $m \geq n$ be a matrix of full rank, i.e.* $\mathrm{rank}(A) = n$*. Then the square matrix $X'X$ is non–singular and the orthogonal projection operator onto the column space $\mathcal{R}(X)$ of X has matrix representation*

$$X(X'X)^{-1}X',$$

in the standard basis of $\mathbb{R}^m$.

The proof is given as an exercise. For simplicity, to avoid *multi–collinearity issues* we shall assume next that X has full rank. Then, the last result implies that

$$X(X'X)^{-1}X'Y = Y_0 = X\widehat{\beta}.$$

Now, to solve for $\widehat{\beta}$, we multiply the last relation to the left by X', first and then by the inverse matrix $(X'X)^{-1}$. This fact yields the well–known formula:

$$\widehat{\beta} = (X'X)^{-1}X'Y.$$

Notice that, in passing, we have shown that the solution to the least squares linear regression problem is *unique* in the case when X has full rank and $N > k+1$. This follows from the uniqueness of the orthogonal projection. The case when X does not have full rank is more involved.

2.5 Exercises

1. Prove Proposition 2.1 by using the definition of a basis.

2. Prove that any finite–dimensional vector space V over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$, where $n = \dim(V)$.

   Hint: Fix a basis in V and use Proposition 2.1 to construct an isomorphism.

3. Verify the formula (2.3) for change of coordinates, where T is as in (2.1).

4. **a.** Let V be a vector space (say, over $\mathbb{R}$) which is not necessarily finite–dimensional. Suppose that any collection of $n \in \mathbb{N}$ non–zero vectors in V is *linearly dependent*. Show that V is finite–dimensional and that $\dim(V) < n$.

   b. Conversely, if V is a finite–dimensional vector space and $\{v_1, \ldots, v_m\}$ is a set of linearly independent vectors, then show that $m \leq \dim(V)$.

   Hint: For part **a.**, start with $v_1 \neq \vec{0}$ and consider $V_1 = \text{span}\{v_1\}$. If $V_1 \equiv V$, then $\dim(V) = 1$. Otherwise, consider $v_2 \in V \setminus V_1$ and let $V_2 := \text{span}\{v_1, v_2\}$ and observe that v_1 and v_2 are *linearly independent*. Keep adding vectors to the linearly independent set $\{v_1, v_2, \ldots\}$.

5. Let $V \leq W$, where W is a finite–dimensional vector space. Show that V is also finite–dimensional and that $\dim(V) \leq \dim(W)$.

   Hint: Argue as in the previous exercises by iteratively constructing a set of linearly independent vectors in V. Apply Theorem 2.1 to show that the process is finite.

6. Prove Proposition 2.2.

7. Prove the formula (2.4) and Corollary 2.1.

   Hint: To prove (2.4), start with the intersection space $U := V_1 \cap V_2$ and build a basis of $V_1 + V_2$: $\{u_1, \ldots, u_r, v_1, \ldots, v_{n_1 - r}, w_1, \ldots, w_{n_2 - r}\}$, such that $u_i \in U$, $v_i \in V_1$, and $w_i \in V_2$, where $\dim(U) = r$, $\dim(V_1) = n_1$, and $\dim(V_2) = n_2$.

8. Convince yourself that the matrix A in (2.5) is unique.

9. Let $\{v_1, \ldots, v_n\}$ be a basis of the space V. Let $T = (t_{ij})_{n \times n}$ be an arbitrary non–singular matrix. Show that the vectors $\{v_1^*, \ldots, v_n^*\}$ given by

   $$(v_1^* \ \cdots \ v_n^*) = (v_1 \ \cdots \ v_n)T$$

   are linearly independent and hence form another basis of V.

   Hint: Suppose that a linear combination of the v_i^*s is zero and show that it should then be trivial.

10. Show that Definition 2.7 is equivalent to the definition of rank(A) in Theorem 2.2.

    Hint: Without loss of generality we can assume that the largest square $r \times r$ sub matrix is in the top–left corner of A (otherwise, we can just interchange some rows and/or columns). Then, by adding multiples of the first r rows and/or columns to the rest, zero–out the matrix items below and to the right of this sub–matrix. Argue that also the remaining corner $(m-r \times n-r)$ matrix is zero. As shown in the proof of Theorem 2.2, these operations do not change the row and column rank of A.

11. Show that the series in (2.10) is absolutely convergent.

12. Prove Proposition 2.3.

13. Suppose that $u \perp W$, for some subset of vectors $W \subset V$, i.e. $u \perp v$, $\forall v \in W$. Then, show that $u \perp \operatorname{span}(W)$. Consider only the case of finite–dimensional inner product spaces.

14. Convince yourself that the Gramm–Schmidt orthogonalization process works, i.e. show that on step $k+1$, $x_{k+1} \neq \vec{0}$, $x_{k+1} \perp x_1, \ldots, x_k$, and that $\operatorname{span}\{x_1, \ldots, x_{k+1}\} = \operatorname{span}\{v_1, \ldots, v_{k+1}\}$.

15. **a.** Prove that the orthogonal complement $W^\perp$ to a sub–space $W \leq V$ is itself a sub–space.

    b. Prove that the orthogonal complement $A^\perp$, of any (non–empty) sub–set $A \subset V$ (A is not necessarily a sub–space) is a sub–space of V.

16. Prove the Pythagoras' Theorem: If $v = w_1 + \cdots + w_k$, where $w_1, \ldots, w_k$ are orthogonal, then show that
$$\|v\|^2 = \|w_1\|^2 + \cdots + \|w_k\|^2.$$
    Hint: From the perspective of inner product spaces, it is easy.

17. Let $A = (a_{ij})_{m\times n}$ be a real matrix. Show that
$$\mathcal{N}(A) \oplus \mathcal{R}(A^t) = \mathbb{R}^n \quad \text{and} \quad \mathcal{N}(A^t) \oplus \mathcal{R}(A) = \mathbb{R}^m,$$
    that is, the *null space* of A is the orthogonal complement of its *row space* and the *null space* of A^t is the orthogonal complement of its *column space*.

    This provides another perspective to the *fundamental theorem of linear algebra* (see Theorem 2.2 above).

18. Consider the vectors:
$$v_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \; v_2 = \begin{pmatrix} 0 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \; v_3 = \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix}.$$
    Orthogonalize these vectors by using the Gramm–Schmidt procedure. Do the computations *by hand.*

19. Why a real matrix $P = (p_{ij})_{m\times n}$ of dimensions $m < n$ *cannot* be orthogonal?

20. Let $P = (p_{ij})_{n\times n}$ be a square orthogonal matrix.

    a. Show that P is non–singular and that $P^{-1} = P^t$.

    b. Show that $Q = P^t$ is also an orthogonal matrix.

    Hint: Consider $\det(P^tP)$ and apply the product formula for determinants.

21. Prove Proposition 2.7.

    Hint: Multiply $A := X'X$ by left– and right–transformation matrices. This does not change the rank and can transform A into a convenient from. Then, verify the conditions for a matrix to be a projection matrix and show that the matrix $B := X(X'X)^{-1}X'$ has the same range as X.

3 Matrix norms and matrix factorizations

3.1 Eigen values and eigen vectors. Symmetric matrices and quadratic forms.

Definition 3.1 Let V be a vector space and $L : V \to V$ be a linear map. The *non–zero* vector $v \in V$ is said to be an eigen vector of L if $L(v) = \lambda v$, for some scalar λ. The scalar λ is said to be the eigen value corresponding to the eigen vector v.

We typically identify the linear maps with their matrices, when the basis is fixed. Thus, a square matrix $A = (a_{ij})_{n\times n}$ is said to have a (column or right–) eigen vector $x \neq \vec{0}$ if $Ax = \lambda x$, for some scalar λ, called an eigen value.

• *Why are eigen vectors useful?*

Suppose that $x_1, \ldots, x_n$ are *linearly independent* eigen vectors of the matrix $A = (a_{ij})_{n\times n}$. Then, the linear transformation $A : \mathbb{R}^n \to \mathbb{R}^n$ in the basis $x_1, \ldots, x_n$ becomes:

$$B = \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

where λ_i is the eigen value corresponding to the eigen vector $x_i,\ i = 1, \ldots, n$.

Not all linear transformations have diagonal matrices in a suitable basis.

Example: Let

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

and observe that $x = (1\ 0)'$ is an eigen vector with eigen value 2. By examining the system of equations:

$$A\begin{pmatrix} a \\ b \end{pmatrix} = \lambda \begin{pmatrix} a \\ b \end{pmatrix},$$

we obtain that either $\lambda = 0$, or $b = 0$. But the matrix A is non–singular and hence it does not have zero as an eigen value and therefore $b = 0$. This implies that the vector $x = (1\ 0)'$ is the only (up to rescaling) eigen vector of A.

Definition 3.2 Let $L : V \to V$ be a linear map with an eigen value λ. The span $S_L(\lambda)$ of all eigen vectors corresponding to λ is said to be the *eigen space* of the eigen value λ.

In the previous example, $S_A(2) = \{(x\ 0)' \ :\ x \in \mathbb{R}\}$. Consider an arbitrary square matrix $A = (a_{ij})_{n\times n}$ and let λ be its eigen value. Then, any $x \in S_A(\lambda),\ x \neq \vec{0}$ is an eigen vector of A corresponding to eigen value λ. The eigen spaces $S_A(\lambda)$ are examples of *invariant spaces* for the matrix (or linear map) A.

•*How to find eigen values and vectors?*

Consider an arbitrary real (or complex) square matrix $A = (a_{ij})_{n\times n}$. The eigen values of A are the roots of its *characteristic polynomial*:

$$p_A(\lambda) := \det(A - \lambda I_n) = 0. \tag{3.1}$$

Observe that $p_A(\lambda)$ is a real (complex) polynomial of degree not greater than n. Therefore, it has n complex roots $\lambda_1, \ldots, \lambda_n$, some perhaps with multiplicity more than 1. When the matrix A is real, the polynomial $p_A(\lambda)$ has real coefficients and hence its non–real roots necessarily come in conjugate pairs. The multiplicity of a root of λ_* of the characteristic polynomial p_A in (3.1) is said to be the *multiplicity of the eigen value.*

The eigen values and eigen vectors have the following basic:

PROPERTIES 3.1 *Let $A = (a_{ij})_{n\times n}$ be a square matrix (of real or complex scalars).*

a. *The sets of eigen values of A' and A coincide.*

b. *The matrix A is singular if and only if A has a zero eigen value.*

c. *If A is triangular (upper or lower), then its diagonal elements are precisely its eigen values.*

d. *For any non–singular B of dimension n, the matrix $B^{-1}AB$ has the same set of eigen values as A.*

e. *If A is an orthogonal matrix, then the modulus of any of its eigen values is equal to* 1*.*

f. *If λ is an eigen value of A with multiplicity $r \le n$, then*

$$1 \le \dim S_A(\lambda) \le r.$$

g. *If λ is an eigen value of A then λ^n is an eigen value of A^n, $n \in \mathbb{N}$.*

h. *Let $\lambda_1, \ldots, \lambda_n$ be the eigen values of A (some of them may be complex). Then:*

$$\det(A) = \prod_{i=1}^{n} \lambda_i \quad \text{and} \quad \operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i.$$

Properties **a.** – **e.** follow easily by using the characteristic polynomial of A. For the proof of **f.** and **h.** see e.g. Theorem 3.3 in Schott (2005).

Real symmetric matrices play an important role in statistics. We next study the properties of their eigen values, eigen vectors, and show an important *spectral decomposition* theorem.

• *Real symmetric matrices*

Let $A = (a_{ij})_{n\times n}$ be a real symmetric matrix. If λ is an eigen value and $x \in \mathbb{R}^n$ is an eigen vector of A corresponding to λ, then

$$x'A = \lambda x.$$

That is, x is also a *left* eigen vector. We have the following important result.

Proposition 3.1 *If x and y are eigen values of the real symmetric matrix A, corresponding to different eigen values λ and μ, then x and y are orthogonal.*

The proof is an exercise.

Proposition 3.2 *All eigen values of a real symmetric matrix $A = (a_{ij})_{n\times n}$ are real.*

Proof: Let $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ are the eigen values of A. Suppose that $\lambda_1 = a + ib$, $a, b \in \mathbb{R}$ and let $z = x + iy \in \mathbb{C}^n,\ x, y \in \mathbb{R}^n$, be the eigen vector corresponding to λ_1. Since A is real, we have also that

$$\overline{\lambda}\overline{z} = A\overline{z},$$

so that $\overline{\lambda_1}$ is also an eigen value corresponding to eigen vector $\overline{z} = x - iy$. By symmetry, we also have that $z'A = \lambda_1 z'$, and hence

$$z'A\overline{z} = \lambda_1 z'\overline{z} = \overline{\lambda_1} z'\overline{z}.$$

Since $z'\overline{z} = x'x + y'y > 0$, we have that $\overline{\lambda_1} = \lambda_1$, and $\lambda_1 \in \mathbb{R}$. $\square$

The following result is fundamental. It shows that any (real) symmetric matrix is *diagonalizable*, and moreover, this can be done by an *orthogonal* transformation of $\mathbb{R}^n$.

Theorem 3.1 *For any (real) symmetric matrix $A = (a_{ij})_{n\times n}$, there exists an orthogonal (real) matrix P, such that*

$$P^{-1}AP = P^tAP = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

A short proof of this result can be found on p. 96 in Schott (2005). The following proof, however, reveals the structure of the real symmetric matrices. It uses the following

Lemma 3.1 *If a real symmetric matrix A has only zero eigen values, then it is the zero matrix.*

The proof of this lemma is given as an exercise.

Proof of Theorem 3.1: We will construct, iteratively an orthonormal basis for $\mathbb{R}^n$ which consists of eigen vectors of the matrix A. If all eigen values of A are zero, then A is the zero matrix (Lemma 3.1) and there is nothing to prove.

Let $\lambda_1, \ldots, \lambda_r,\ r \geq 1$ be the non–zero eigen values of A. We will show first that

$$A = \lambda_1 x_1 x_1' + \cdots + \lambda_r x_r x_r', \tag{3.2}$$

for some set of *orthonormal vectors* $x_1, \ldots, x_r$. This would easily imply the result.

Step 1. Consider the eigen vector $x_1 \in \mathbb{R}^n$ corresponding to λ_1 and let

$$A_1 := A - \lambda_1 x_1 x_1'.$$

Observe that the range of the matrix A_1 is orthogonal to x_1. Indeed, for any $y \in \mathbb{R}^n$, we have $y = y_1 + y_2$, where $y_1, y_2 \in \mathbb{R}^n$, $y_1 = ux_1,\ u \in \mathbb{R}$ and $x_1 \perp y_2$. Thus,

$$A_1 y = Ay_1 + Ay_2 - \lambda_1 x_1 x_1' y_1 - \lambda_1 x_1 x_1' y_2 = Ay_2,$$

but $x_1' Ay_2 = \lambda_1 x_1' y_2 = 0$, which implies $x_1 \perp A_1 y$, for all $y \in \mathbb{R}^n$.

If the range of A_1 has dimension zero, then the rank of A_1 is zero and $A = \lambda_1 x_1 x_1'$, that is, (3.2) holds.

Step 2. Notice also that A_1 is a real symmetric matrix. We apply the above argument in *Step 1* to A_1: Let $y \in \mathbb{R}^n$ be an eigen value of A_1 corresponding to a non–zero value μ. Then, since

$x_1 \perp \mathcal{R}(A_1)$, we have that $A_1 y = \mu y \neq \vec{0}$ is orthogonal to x_1 and hence $x_1 \perp y$. This implies that y is an eigen vector of A corresponding to value μ, indeed:

$$\mu y = A_1 y = Ay - \lambda_1 x_1 x_1' y = Ay.$$

Step 3. Without loss of generality, let $\lambda_2 = \mu \neq 0$, set $x_2 = y/\|y\|$ and consider the *new matrix* A_2:

$$A_2 := A - \lambda_1 x_1 x_1' - \lambda_2 x_2 x_2'.$$

As in *Step 2*, show now that $\mathcal{R}(A_2)$ is orthogonal to x_1 and x_2. Continue the iterative argument until all non–zero eigen vectors $\lambda_1, \ldots, \lambda_r$ are exhausted. Then, the matrix

$$A_r := A - \lambda_1 x_1 x_1' + \cdots + \lambda_r x_r x_r'$$

will be a symmetric matrix with only zero eigen values which by Lemma 3.1 should be the zero matrix. We have thus shown (3.2).

Now, complete the orthonormal set $x_1, \ldots, x_r$ to an orthonormal basis $x_1, \ldots, x_n$ of $\mathbb{R}^n$. Since $Ax_i = \lambda_i x_i$, for all $1 \leq i \leq r$ and $Ax_j = \vec{0}$, for all $j > r$, we have that the matrix of the linear operator $A : \mathbb{R}^n \to \mathbb{R}^n$ in the new basis $x_1, \ldots, x_n$ is:

$$P^{-1}AP = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) = \mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0).$$

Here P is the orthogonal matrix with columns, the vectors $x_1, \ldots, x_n$, respectively.

The last relation can be also shown to hold by simply writing $P^{-1}AP = P^t AP$ and using the expression (3.2) for A. $\square$

In the above proof, we showed that any real symmetric matrix A of dimension $n \times n$ can be expressed as

$$A = \lambda_1 x_1 x_1' + \cdots + \lambda_r x_r x_r', \tag{3.3}$$

where $\lambda_1, \ldots, \lambda_r,\ 1 \leq r \leq n$ are the non–zero eigen values of A, and where $x_1, \ldots, x_r$ are orthonormal eigen vectors of A. It can be shown that the rank of A is precisely r. Indeed, it is easy to see that $\mathcal{R}(A) = \mathrm{span}(x_1, \ldots, x_r)$.

The representation (3.3) is called the *spectral representation* of A. It is a special case of the *singular value decomposition* (SVD).

Symmetric matrices and quadratic forms.

We now discuss in more detail symmetric matrices from the perspective of quadratic forms.

Definition 3.3 Let $x\mathbb{R}^n$ be two column vector, where its coordinates $x_i,\ i = 1, \ldots, n$ are treated as independent variables and let $A = (a_{ij})_{n \times n}$ be a symmetric matrix. The function $Q : \mathbb{R}^n \to \mathbb{R}$, defined as:

$$Q(x) := x' Ax$$

is called a *quadratic form.* The quadratic form Q (or its corresponding matrix A) is said to be:

a. *positive definite* if $Q(x) > 0$, for all $x \neq \vec{0}$.

b. *positive semi–definite* if $Q(x) \geq 0$, for all $x \neq \vec{0}$.

c. *negative definite* if $Q(x) < 0$, for all $x \neq \vec{0}$.

d. *negative semi–definite* if $Q(x) \leq 0$, for all $x \neq \vec{0}$.

If Q (or A) are none of the above, then it is said to be *indefinite.*

Examples:

1. The matrix $A = \text{diag}(1,2)$ is positive definite, $B = \text{diag}(-1,1)$ is indefinite and $C = \text{diag}(0,2)$ is positive semi–definite.

2. Let $B = (b_{ij})_{m\times n}$ be an arbitrary real $m \times n$ matrix. Then the matrix:
$$A := B'B$$
is positive semi–definite.

3. The covariance matrix Σ of an arbitrary random vector X taking values in $\mathbb{R}^n$ is positive semi–definite. Indeed, since $\Sigma = \mathbb{E}(X-\mu)(X-\mu)'$, $\mu = \mathbb{E}X$ for all $y \in \mathbb{R}^n$, we have $y'\Sigma y = \mathbb{E}y'(X-\mu)(X-\mu)'y \geq 0$, since $y'(X-\mu)(X-\mu)'y = (y'(X-\mu))^2 \geq 0$.

By using the *spectral representation* of a real symmetric positive semi–definite matrix A, one can show the following useful result.

Corollary 3.1 *Let $A = (a_{ij})_{n\times n}$ be a real, symmetric and positive semi–definite matrix. Consider its spectral representation:*
$$A = \lambda_1 x_1 x_1' + \cdots + \lambda_n x_n x_n',$$
where $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n \geq 0$, and where x_i, $i = 1,\ldots,n$ are orthonormal, non–zero vectors in $\mathbb{R}^n$.
a. *There exists a symmetric, positive semi–definite matrix B such that*
$$A = BB.$$
The matrix B is unique and it is called the square root of the matrix A, denoted $B := A^{1/2}$.
b. *If the matrix A is non–singular, then show that*
$$A^{-1} = \lambda_1^{-1} x_1 x_1' + \cdots + \lambda_n^{-1} x_n x_n'.$$

The proof of this result is given as an exercise. We conclude the discussion on symmetric matrices with an important result. Its proof can be found on p. 163–164 in Schott (2005).

Theorem 3.2 *Let A and B be two real, symmetric matrices of dimensions $n \times n$. Then A and B commute (i.e. $AB = BA$) if and only if for some orthogonal matrix P,*
$$P^t AP = \text{diag}(\lambda_1,\ldots,\lambda_n) \quad \text{and} \quad P^t BP = \text{diag}(\mu_1,\ldots,\mu_n).$$
That is, A and B are simultaneously diagonalizable by an orthogonal change of coordinates.

• *Jordan normal forms*

We borrow the following presentation form Ch. 4.5 in Schott (2005). The $h \times h$ square complex matrix $J_h(\lambda)$:
$$J_h(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}$$

is said to be a Jordan cell or a Jordan block matrix. For $h = 1$, we set $J_h(\lambda) = \lambda$.

One can show that $J_h(\lambda)$ has only one eigen value of multiplicity h and $\dim S_{J_h(\lambda)}(\lambda) = 1$, that is, the dimension of the eigen space corresponding to λ is one.

The following result shows that any linear map $L : \mathbb{C}^n \to \mathbb{C}^n$ has a Jordan matrix in some appropriate basis.

Theorem 3.3 *Let $A = (a_{ij})_{n \to n}$ be a complex square matrix. Then, for some non–singular B of dimension n,*

$$B^{-1}AB = \mathrm{diag}\Big(J_{h_1}(\lambda_1),\ \cdots,\ J_{h_r}(\lambda_r)\Big),$$

where $h_1 + \cdots + h_r = n$ and $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$.

3.2 Matrix norms and eigen values

Let $M_n(\mathbb{R})$ (or $M_n(\mathbb{C})$) be the set of real (or complex) square $n \times n$ matrices. It is a n^2-dimensional vector space over $\mathbb{R}$ (or $\mathbb{C}$) with the usual operations of addition and scalar multiplication. So, we can equip this vector space with a norm as any other vector space.

The matrices however can be also multiplied. This extra structure is taken into account in the following definition of a *matrix norm.*

Definition 3.4 A norm $\|\cdot\|$ on the vector space $M_n(\mathbb{R})$ (or $M_n(\mathbb{C})$) of square $n \times n$ matrices is said to be a *matrix norm* if, in addition, we have:

$$\|AB\| \le \|A\|\|B\|, \quad \text{for all } A,\ B \in M_n(\mathbb{R}) \ \ (\text{or } M_n(\mathbb{C})). \tag{3.4}$$

In the sequel we work with real matrices but unless stated otherwise, the results will be valid for complex matrices, too. A first question regarding matrix norms is:

Do matrix norms exist?

We will next show that any norm $\|\cdot\|$ on the space $\mathbb{R}^n$ of column vectors generates a *matrix norm.* This matrix norm is often called an *operator norm.*

Consider a norm $\|\cdot\|$ on the space $\mathbb{R}^n$ of column vectors and for any $A \in M_n(\mathbb{R})$, define:

$$\|A\|_{op} := \sup_{x \neq \vec{0}} \frac{\|Ax\|}{\|x\|}. \tag{3.5}$$

One can easily check that (3.5) defines a *matrix norm.* The triangle inequality for the norm $\|\cdot\|$ on $\mathbb{R}^n$ implies the triangle inequality for $\|\cdot\|_{op}$ and, in fact, that $\|\cdot\|_{op}$ is a norm on the vector space $M_n(\mathbb{R})$. To check (3.4), observe that for any $x \in \mathbb{R}^n$, such that $Bx \neq \vec{0}$,

$$\|ABx\|/\|x\| = (\|A(Bx)\|/\|Bx\|)(\|Bx\|/\|x\|) \le (\sup_{y\neq\vec{0}} \|Ay\|/\|y\|)(\sup_{z\neq\vec{0}} \|Az\|/\|z\|).$$

Since $\|ABx\| = 0$, for all x, such that $Bx = \vec{0}$, in view of (3.5), the last inequality implies (3.4).

In the following examples we list several popular matrix norms.

Examples:

1.
$$\|A\|_1 := \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}|. \tag{3.6}$$

2.
$$\|A\|_\infty := \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|. \tag{3.7}$$

3. *(spectral norm)*
$$\|A\|_2 := \max_{1\le j\le n} \sqrt{\mu_j}, \tag{3.8}$$

where $\mu_j,\ j = 1, \ldots, n$ are the eigen values of the symmetric matrix $A'A$ (or of the conjugate symmetric matrix $A * A$ in the complex case).

The following notion of the *spectral radius* of a square matrix plays an important role.

Definition 3.5 Let $A = (a_{ij})_{n\times n}$ be a matrix with eigen values $\lambda_1, \ldots, \lambda_n$ (possibly complex). The spectral radius $\rho(A)$ of A is defined as:

$$\rho(A) := \max_{1\le j\le n} |\lambda_j|.$$

Note that A is the zero matrix if and only if $\rho(A)$ is zero. One can also show that

$$\rho(A) \le \|A\|, \quad \text{for all } A \in M_n(\mathbb{R}),$$

for any *matrix norm* $\|\cdot\|$. Indeed, let x be an eigen vector (possibly complex), corresponding to $|\lambda^*| := \rho(A)$. Then, let $B := x(1\ 1\ \cdots\ 1)$ be a $n\times n$ matrix with columns identical to x and observe that

$$\|AB\| = \|\lambda^* B\| = |\lambda^*|\|B\| = \rho(A)\|B\|.$$

However, $\|AB\| \le \|A\|\|B\|$, so the last relation implies $\rho(A) \le \|A\|$ since $\|B\| \ne 0$.

We next recall an important notion of *equivalence* between two different norms, defined on an arbitrary vector space V.

Definition 3.6 Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two arbitrary norms in a vector space V. They are said to be equivalent, if there exist positive constants $c, C > 0$, such that

$$c\|v\|_1 \le \|v\|_2 \le C\|v\|_1, \quad \text{for all } \ v \in V.$$

Norms provide a natural way of extending the definition of convergent sequence in a vector space. Namely, we say that $v_n \to v$, as $n \to \infty$ in the norm $\|\cdot\|$ on V, if:

$$\|v_n - v\| \to 0, \quad \text{as } n \to \infty.$$

Thus, if two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on the same vector space are equivalent, then a convergent sequence in $\|\cdot\|_1$ is convergent in $\|\cdot\|_2$ and vice versa.

Theorem 3.4 *Any two norms in finite–dimensional vector space (over $\mathbb{R}$ or $\mathbb{C}$) are equivalent.*

The proof of this theorem uses the fact that the continuous function image of a compact set is compact. It is important to note that the last result is not true for infinite–dimensional spaces (see the Exercises below).

Why are matrix norms interesting?

Example: Consider a matrix norm on $M_d(\mathbb{R})$ and let $A \in M_d(\mathbb{R})$ be an arbitrary matrix. When is the matrix $B := I_d - A$ invertible? It turns out that for any A such that $\|A\| < 1$, the matrix B is invertible. Indeed, let

$$C_n := I_n + A + A^2 + \cdots + A^n.$$

We have that $C_n \to C$, as $n \to \infty$, that is the *matrix series* $\sum_n A^n$ converges. Indeed, for any $1 \leq n < m$

$$\|C_n - C_m\| \leq \|A^n\| + \|A^{n+1}\| + \cdots + \|A^{m-1}\| \leq a^n(1 + a + \cdots) = \frac{a^n}{1-a},$$

where $a := \|A\| \, in(0,1)$. Thus the sequence $C_n,\ n \in \mathbb{N}$ is a *Cauchy sequence in a complete normed space*[2] of matrices and hence $C_n \to C,\ n \to \infty$.

Let us see that $C(I_d - A) = I_d$.

$$C_n(I - A) = I + A + \cdots + A^n - A - A^2 - \cdots - A^{n+1} = I - A^{n+1} \to I, \quad \text{as } n \to \infty.$$

However, the fact that

$$\|C(I - A) - C_n(I - A)\| = \|(C - C_n)(I - A)\| \leq \|C - C_n\|\|I - A\| \to 0, \quad \text{as } n \to \infty.$$

and the triangle inequality imply $C(I - A) = I$.

The key advantage of the extra condition (3.4) in the definition of a matrix norm as compared to the usual norms is that it implies the *continuity of the matrix multiplication* (in the matrix norm).

Proposition 3.3 *Let $A_n \to A$ and $B_n \to B$, where $A_n, B_n, A, B \in M_d(\mathbb{R})$. Then,*

$$A_n B_n \to AB, \quad \textit{as } n \to \infty.$$

The proof is an exercise. Observe that since all norms in a finite–dimensional space are equivalent. Thus, the convergence $A_n \to A$ in any matrix norm on $M_d(\mathbb{R})$ is equivalent to the convergence $a_{ij}(n) \to a_{ij}$ for all $1 \leq i, j \leq d$.

We now list some important results which relate the spectral radius of a matrix to its matrix norm. Their proofs and more details can be found in Ch. 4.8 of Schott (2005).

Theorem 3.5 *For any $\epsilon > 0$, there exists a matrix norm $\|\cdot\|$, such that*

$$\rho(A) \leq \|A\| \leq \rho(A) + \epsilon, \quad \textit{for all } A \in M_n(\mathbb{R}).$$

Proposition 3.4 *Let $A \in M_d(\mathbb{R})$. We have that $A^n \to (0),\ n \to \infty$, if and only if $\rho(A) < 1$.*

Proposition 3.5 *For any matrix norm $\|\cdot\|$ on $M_d(\mathbb{R})$,*

$$\|A^n\|^{1/n} \to \rho(A), \quad \textit{as } n \to \infty.$$

[2] do be discussed more later

3.3 An application to the multivariate normal distribution

We recall here some facts about the multivariate normal distribution and relate them to the perspective of matrices.

Definition 3.7 Let $Z_1, \dots, Z_n$ be independent standard normal random variables. A random vector X taking values in $\mathbb{R}^m$ is said to have multivariate Normal distribution if

$$X = \vec{\mu} + AZ, \quad \text{for some deterministic } \vec{\mu} \in \mathbb{R}^n, \tag{3.9}$$

for some real $m \times n$ matrix $A = (a_{ij})_{m\times n}$, where $Z = (Z_i)_{n\times 1}$.

The mean vector of X is $\mathbb{E}X = \vec{\mu}$ and the covariance matrix of X is:

$$\Sigma_X = \mathbb{E}(X - \vec{\mu})(X - \vec{\mu})' = AA'.$$

<u>Notation:</u> $X \sim \mathrm{N}(\vec{\mu}, \Sigma_X)$.

We therefore have the that $X \sim \mathrm{N}(\vec{0}, I_n)$, if and only if $X = PZ$ for some orthogonal matrix $P = (p_{ij})_{n\times n}$, where $Z \sim \mathrm{N}(\vec{0}, I_n)$.

The following is a simple consequence of the above fact:

Lemma 3.2 *Let $Z \sim \mathrm{N}(\vec{0}, I_n)$. Let $\xi = x'Z$ and $\eta = y'Z$, for some $x, y \in \mathbb{R}^n$. Then, ξ and η are independent if and only if $x \perp y$.*

Observe that X in (3.9) has density in $\mathbb{R}^m$ if and only if $m \le n$ and $\mathrm{rank}(A) = m$. That is, A is of complete rank and it has less rows than columns. Indeed, X has density, if and only if $\Sigma_X = AA'$ is non–singular. Recall that in this case the density is given by:

$$f_X(x) = \frac{1}{(2\pi)^{m/2} \det(\Sigma_X)^{1/2}} \exp\Big\{ -\frac{1}{2}(x - \vec{\mu})' \Sigma_X^{-1} (x - \mu) \Big\}.$$

By Lemma 3.2 and the fact that pair–wise independence implies independence for jointly normal random variables, we get:

Proposition 3.6 *Let $Z \sim \mathrm{N}(0, I_n)$. Let also $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{r\times n}$ be real matrices. Then, the two multivariate normal vectors $X := AZ$ and $Y = BZ$ are independent, if and only if,*

$$AB' = (BA')' = (0).$$

•A famous application

Let $Z \sim \mathrm{N}(0, I_n)$, then

$$s^2 := \frac{1}{n-1} \sum_{i=1}^n (Z_i - \overline{Z})^2 \quad \text{and} \quad \overline{Z} = \frac{1}{n} \sum_{i=1}^n Z_i$$

are independent and $s^2 \sim \chi^2(n-1)$.

Consider the matrices:

$$A = I_n - \frac{1}{n}\vec{1}\vec{1}' \quad \text{and} \quad B = \frac{1}{n}\vec{1}\vec{1}'.$$

Observe that A and B are symmetric and $n \times n$ matrices. Note moreover that

$$AB = BA = (0).$$

Thus, Proposition 3.6 implies that the vectors

$$X = AZ = (Z_1 - \overline{Z} \ \cdots \ Z_n - \overline{Z})' \quad \text{and} \quad Y = BZ = (\overline{Z} \ \cdots \ \overline{Z})',$$

are independent! Since $s^2 = X'X/(n-1)$ and $\overline{Z} = \vec{1}'Y/n$, we get that s^2 and $\overline{Z}$ are independent.

Observe that B and $A = I_n - B$ are both orthogonal projection matrices. The following proposition implies that $s^2 \sim \chi^2(n-1)$.

Proposition 3.7 *Let P be an $n \times n$ (real) orthogonal projection matrix of rank $r, 1 \leq r \leq n$. Let also $Z \sim \mathrm{N}(0, I_n)$. Then, $\xi := \|PZ\|^2 = Z'P'PZ$ is such that*

$$\xi/r \sim \chi^2(r).$$

Proof: Since P is an orthogonal projection matrix, it is symmetric and idempotent. Let $x_1, \ldots, x_n$ be an orthonormal basis, the first r vectors of which span the projection sub–space of P, i.e. $\mathcal{R}(P)$. Then, set $Q = (x_1 \ \cdots \ x_n)$ and observe that

$$P = QDQ', \quad \text{where} \ \ D = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0).$$

Then

$$\xi = Z'P'PZ = Z'PZ = Z'QDQ'Z = \widetilde{Z}_1^2 + \cdots + \widetilde{Z}_r^2,$$

where $\widetilde{Z} = Q'Z$. Since $Q'Q = I_n$, the components of the vector $\widetilde{Z}$ are independent standard normals, which implies that $\xi \sim \chi^2(r)$. $\square$

3.4 Singular value decomposition, Choleski decomposition and simulation

In this short section, we give (without proof) an important decomposition result valid for an arbitrary (real) matrix $A = (a_{ij})_{m \times n}$ of dimensions $m \times n$. The following presentation is borrowed from Schott (2005).

Theorem 3.6 *There exist orthogonal matrices $P = (p_{ij})_{n \times n}$ and $Q = (q_{ij})_{m \times m}$ of dimensions $n \times n$ and $m \times m$, respectively. Such that*

$$A = Q'DP, \quad \textit{where } D = \begin{pmatrix} \Delta & (0)_{r \times (n-r)} \\ (0)_{(m-r) \times r} & (0)_{(m-r) \times (n-r)} \end{pmatrix}, \tag{3.10}$$

and where $\Delta := \mathrm{diag}(\lambda_1^*, \ldots, \lambda_r^*)$. Here $\lambda_1^*, \ldots, \lambda_r^* > 0$, $\mathrm{rank}(A) = r$ and the values λ_i^* are the positive square–roots of the matrix $A'A$ (or, equivalently, the matrix AA'). They are called singular values of A.

For the proof, see Ch. 4.1 in Schott (2005). We now show an alternative form of the expression in (3.10).

There exist two sets of orthonormal column vectors $x_1, \ldots, x_r \in \mathbb{R}^n$ and $y_1, \ldots, y_r \in \mathbb{R}^m$, such that

$$A = \lambda_1^* y_1 x_1' + \cdots + \lambda_r^* y_r x_r'. \tag{3.11}$$

Indeed, let $\widetilde{A} = \lambda_1^* y_1 x_1' + \cdots + \lambda_r^* y_r x_r'$, where the y_is are the *columns* of Q' and the x_is are the *columns* of P'. Consider $Q\widetilde{A}P'$ and by using the orthogonality of the matrices P and Q, we get:

$$Q\widetilde{A}P' = D.$$

Since P and Q are invertible, the last relation implies $A \equiv \widetilde{A}$ and completes the proof of the representation in (3.11).

The representation in (3.11) is intuitively more appealing than the matrix relation (3.10). Indeed, it clearly shows that the matrix of the linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ in the orthonormal bases $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$, respectively, has essentially a diagonal structure.

We have already seen the representation (3.11) in the case when A is a square, symmetric matrix. In this case we have the added benefit of having $y_i \equiv x_i$. Efficient algorithms for calculating the SVD (Singular Value Decomposition) of a matrix are implemented in most statistical packages. The following is an important application of singular value decomposition to the case of simulation of a normal random vector.

Proposition 3.8 *Let $\Sigma = (\sigma_{ij})_{n\times n}$ be a positive semi–definite matrix and let $\Sigma = P'DP$ be its singular value decomposition, where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.*

Generate a vector $Z \sim \mathrm{N}(0, I_n)$ of independent standard normal components and set $X = P'\sqrt{D}PZ$, where $\sqrt{D} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$.

We then have that $X \sim \mathrm{N}(0, \Sigma)$, that is, X is a normal random vector with zero mean and covariance matrix Σ.

• *Choleski decomposition*

Any symmetric (real) positive (semi–)definite matrix S can be decomposed as $S = LL'$, where L is a *lower–triangular* matrix. Such a decomposition is called the *Choleski decomposition* of the matrix S.

As the singular value decomposition, the Choleski decomposition can be used to simulate a normal random vector with prescribed covariance matrix Σ. Indeed, let $\Sigma = LL'$ is the Choleski decomposition of Σ. Convince yourself that $X := LZ \sim \mathrm{N}(0, \Sigma)$, where $Z \sim \mathrm{N}(0, I)$.

Remark: The simulation of high–dimensional normal distributions can become computationally very demanding as the dimension grows, especially if the covariance matrix Σ has non–trivial structure. This is because large matrices are difficult to diagonalize and even store efficiently. Thus, the generic simulation methods based on SVD and Choleski decomposition above should be used with care and only if other methods are not available or when the dimension is "small".

3.5 Exercises

1. Find, *by hand*, the eigen values and eigen vectors of the matrix:

$$A = \begin{pmatrix} 4 & -\sqrt{3} \\ -\sqrt{3} & 2 \end{pmatrix}.$$

Are the eigen vectors corresponding to different eigen values orthogonal?

2. Prove Proposition 3.1.

   Hint: Consider the quantity $x'Ay$ and write it in two different ways.

3. Prove Lemma 3.1.

   Hint: Consider the symmetric matrix A^2. By using its characteristic polynomial $\det(A^2 - \lambda I_n) = \det(A - \sqrt{\lambda} I_n)\det(A + \sqrt{\lambda} I_n)$ show that it has eigen values equal to the squares of the eigen values of A. Then, use the fact that $\mathrm{tr}(A'A) = 0$ if and only if A is the zero matrix.

4. By using the spectral representation of the symmetric matrix A, show that:

   a. A is positive (negative) definite if and only if all of its eigen values are positive (negative).

   b. A is positive (negative) semi–definite if its eigen values are non–negative (non–positive).

5. Let A be a positive definite symmetric matrix. Show that

   $$\langle x, y \rangle_A := x'Ay, \quad x, y \in \mathbb{R}^n,$$

   defines an inner product in $\mathbb{R}^n$.

6. Let $\Sigma = (\sigma_{ij})_{n\times n}, \ n > 1$ be the covariance matrix of a random vector X, taking values in $\mathbb{R}^n$, where $n \geq 2$. Show that if Σ is singular, there exists a non–zero and non–random vector $y \in \mathbb{R}^n$, such that

   $$y \perp X - \mu, \quad \text{almost surely, where} \quad \mu = \mathbb{E}X.$$

   Intuition: The distribution of the vector X is concentrated on a lower dimensional hyper–plane of $\mathbb{R}^n$.

7. Let A and B be two a real, symmetric matrices of dimensions $n \times n$. Let B be such that any eigen vector of A is an eigen vector of B. Show that A and B commute.

   Hint: Show first that there exists an orthonormal basis $x_1, \ldots, x_n$ of $\mathbb{R}^n$ such that

   $$A = \lambda_1 x_1 x_1' + \cdots + \lambda_n x_n x_n' \quad \text{and} \quad B = \mu_1 x_1 x_1' + \cdots + \mu_n x_n x_n'.$$

   That is, for the orthogonal matrix P with columns, the vectors $x_i, \ i = 1, \ldots, n$:

   $$P^t A P = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \quad \text{and} \quad P^t B P = \mathrm{diag}(\mu_1, \ldots, \mu_n).$$

   Intuition: The linear operators A and B are *simultaneously diagonalizable*, more precisely, they both have diagonal matrices in a suitable *orthonormal basis*.

8. Let $x_1, \ldots, x_r \in \mathbb{R}^n$ and $y_1, \ldots, y_r \in \mathbb{R}^n$, be two sets of orthonormal vectors. Suppose that

   $$W = \mathrm{span}(x_1, \ldots, x_r) = \mathrm{span}(y_1, \ldots, y_r).$$

   Then, show that the matrices

   $$A := x_1 x_1' + \cdots + x_r x_r' \quad \text{and} \quad B := y_1 y_1' + \cdots + y_r y_r',$$

   coincide.

   Hint: Relate A and B to the orthogonal projection operator P_W onto the common space W spanned by the vectors x_is and y_is.

9. Prove Corollary 3.1.

   Hint: Consider the matrix $B^* = \sqrt{\lambda_1}x_1x_1' + \cdots + \sqrt{\lambda_n}x_nx_n'$ and show that $A = B^*B^*$. To prove uniqueness, show that any matrix B, such that $A = BB$ has the form $B = \sqrt{\lambda_1}y_1y_1' + \cdots + \sqrt{\lambda_n}y_ny_n$, for some orthonormal y_i's, which are also eigen vectors of the matrix A. Then, show that the eigen spaces of B and B^* coincide and use the previous problem.

10. Let $\|\cdot\|_{op}$ be an operator matrix norm in $M_n(\mathbb{R})$ corresponding to a norm $\|\cdot\|$.

    a. Show that
    $$\|I_n\|_{op} = 1.$$

    b. Show that $\|A\| := C\|A\|_{op}$ is a matrix norm if and only if $C > 1$. Conclude that not all matrix norms are operator norms for some norm on $\mathbb{R}^n$.

11. Given a matrix norm $\|\cdot\|$ on $M_n(\mathbb{R})$, and an arbitrary fixed non–singular matrix $B \in M_n(\mathbb{R})$, define $\|A\|_B := \|B^{-1}AB\|$. Show that $\|\cdot\|_B$ is also a matrix norm.

12. Let $\|x\|_\infty := \max_{1\le i\le n} |x_i|$, for $x = (x_i)_{n\times 1} \in \mathbb{R}^n$. Show that the operator norm $\|\cdot\|_{op}$ corresponding to $\|\cdot\|_\infty$ is as in (3.7)
    $$\|A\|_{op} \equiv \|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^n |a_{ij}|.$$
    Is the matrix norm in (3.6) an operator norm? If yes, then determine its corresponding column vector norm.

13. Consider the vector space $\ell_2 = \{x = (x_i)_{i=1}^\infty \ : \ x_i \in \mathbb{R}, \sum_{i=1}^\infty x_i^2 < \infty\}$ with the usual vector addition and scalar multiplication.

    a. Show that $\|x\|_2 := \sqrt{\sum_i x_i^2}$ and $\|x\|_\infty := \sup_i |x_i|$ are norms. Show that for any $x \in \ell_2$, we have $\sup_i |x_i| = x_{i_0}$ for some $i_0 \in \mathbb{N}$. That is, the last supremum is achieved can be written as a maximum.

    b. If $v_n \to v,\ n \to \infty$ in $\|\cdot\|_2$, then show that $v_n \to v,\ n \to \infty$ in $\|\cdot\|_\infty$. That is, a convergent sequence in $\|\cdot\|_2$ is also convergent in $\|\cdot\|_\infty$.

    c. Construct a "counterexample" of a sequence $v_n \to v,\ n \to \infty$ in $\|\cdot\|_\infty$ which does not converge in $\|\cdot\|_2$. That is, show that the norms $\|\cdot\|_2$ and $\|\cdot\|_\infty$ are not equivalent.

14. Let $B \in M_n(\mathbb{R})$ be non–singular and let $\|\cdot\|$ be a an arbitrary matrix norm in $M_n(\mathbb{R})$.

    a. Show that there exists $\epsilon > 0$, such that the matrix $B + A$ is invertible, for any $A \in M_n(\mathbb{R})$ with $\|A\| < \epsilon$.

    b. Given an arbitrary matrix $A \in M_n(\mathbb{R})$, show that $C_\delta := B + \delta A$ is invertible for some $\delta > 0$,

15. Prove Proposition 3.3.

16. Prove Proposition 3.8.

17. Let A be an $n \times n$ symmetric positive semi–definite matrix (e.g. a covariance matrix of a random vector). Suppose that all entries of the matrix A are non–negative and that $\lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \leq \lambda_n \geq 0$ are the eigen values of A. That is, the largest eigen value λ_1 has algebraic multiplicity one.

    a. Let $\vec{x}$ be an eigen vector corresponding to the largest eigen value of A. Show that $\vec{x}$ does not have entries with different signs.

    b. *(Hint to* **a.***)* Show that limit $m^{-1}A^m$ exists as $m \to \infty$ and identify its value. Can you determine a upper bound on the rate of the convergence?

References

Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw–Hill, New York, third edition edition.

Schott, J. (2005). *Matrix Analysis for Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons.