Elements of Mathematics: an embarrasignly simple (but practical) introduction to algebra

Jordi Villà i Freixa (jordi.villa@upf.edu), Pau Rué

November 23, 2011

Contents

Summary. Playing around with matrices and their properties. Some examples of resolution of systems of linear equations.

1 Introduction

This is a non-exhaustive review of matrices and their properties. The practical part can be performed with the help of `octave` (http://www.octave.org). There are versions of the program for cygwin and linux.

Some on-line additional sources of information can be found at:

```
http://joshua.smcvt.edu/linalg.html
http://www.math.unl.edu/~tshores/linalgtext.html
http://archives.math.utk.edu/tutorials.html
http://www.cs.unr.edu/~bebis/MathMethods/
http://en.wikibooks.org/wiki/Linear_Algebra
http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/mathematics-2/book.html
```

For help on octave and linear algebra:

```
http://math.iu-bremen.de/oliver/teaching/iub/resources/
       octave/octave-intro/octave-intro.html
http://www2.math.uic.edu/~hanson/Octave/OctaveLinearAlgebra.html
```

Check also [1, 2, 3].

2 Sets

Any collection of objects, for example the points of a given segment, the collection of all integer numbers between 0 and 10, the students in a classroom, etc. is called a *set*. The objects inside the set (the points, the numbers and the students) are called *elements* of the set. In algebra it is common to represent sets using uppercase letters and elements using lowercase letters. The elements of a set are specified between curly brackets. For example $A = \{a, b, c, d\}$ represents a set formed of 4 elements.

A set can be specified either in an *extensive* way as in the case of $A = \{a, b, c, d\}$ or in an *intesive* way, where there is no need to specify all the elements belonging to it but only the properties they satisfy. As an example, the set, A, of all integer numbers between 0 and 10 can be specified as $A = \{x|x \in \mathbb{Z}, 0 \leq x \leq 10\}$[1].

There is a huge amount of literature descibing formally what a set is. We will just stick to the idea that a set is a collection of elements none of them equal to another. The following is a list of basic properties and definitons concerning sets:

[1] $x \in A$ is the mathematical way of representing that the element x belongs to the set A

- A set A is said to be *included* within a set B (or that A is a subset of the set B, or that B contains A) if and only if all the elements of A belong to B. In this case we will write $A \subset B$. So in a strictly mathematical way we would write

$$A \subset B \quad \text{if and only if} \quad \forall x \in A \Rightarrow x \in B$$

[2]

- Hence, two sets, A and B, are said to be *equal*, $A = B$, if and only if both conditions $A \subset B$ and $B \subset A$ are fulfilled.

Example
The usual numeric sets $\mathbb{N} = 1, 2, 3, \cdots$ (the natural numbers), $\mathbb{Z} = 0, 1, -1, 2, -2, \cdots$ (the integer numbers), $\mathbb{Q}$ (the rational numbers), $\mathbb{R}$ (the real numbers) and $\mathbb{C}$ (the complex numbers) are related in the following way:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

- There is only one unique set that contains no elements. It is called the *empty set* and it is denoted by $\emptyset$.
- In addition, the universe of a given problem is the reference set, U, that contains all the sets used in that particular problem.
- Given a universe U, and a subset A, the complement of A in U, $\bar{A}$, is the set of all elements in U that do not belong to A. Formally,

$$\bar{A} = \{x \in U | x \notin A\}.$$

- Given a universe U and two sets A and B we can define de following operations:

1. The *union* of A and B is the set having all the elements from both sets A and B and no other element.

$$A \cup B = \{x \in U | x \in A \text{ or } x \in B\}.$$

   Notice that this operation is commutative: $A \cup B = B \cup A$ and associative: $A \cup (B \cup C) = (A \cup B) \cup C$.

Example
Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{2, 3, 7, 8\}$, then

$$A \cup B = \{1, 2, 3, 4, 5, 7, 8\}.$$

2. The *intersection* of A and B is set the having all the elements common to A and B and no other element.

$$A \cap B = \{x \in U | x \in A \text{ and } x \in B\}.$$

   This operation is also commutative: $A \cap B = B \cap A$ and associative: $A \cap (B \cap C) = (A \cap B) \cap C$.

[2] $\forall$ is a mathematical symbol meaning *for all*, as in $\forall x \in A$(for all element x in the set A)

Example
Let $A = \{1, 2, a, b\}$ and $B = \{2, 3, a, c\}$, then

$$A \cap B = \{2, a\}.$$

3. The *set difference* of A and B is the set having all the elements in A that are not found in B and no other element.

$$A \setminus B = \{x \in U | x \in A \text{ and } x \notin B\}.$$

This operation is neither commutative nor associative. Notice also that we can write $A \setminus B = A \cap \bar{B}$

4. Given two elements a and b, we call an *ordered pair* the collection of these two elements in a given order. We denote the ordered pair with a being the *first coordinate* and b the *second coordinate* as in (a, b). Notice that with this definiton order matters (i.e. $(a, b) \neq (b, a)$). The *cartesian product* of A and B, $A \times B$, is then defined as the set of all ordered pairs of elements where the element in the first coordinate belongs to A, and the element in the second coordinate belongs to B.

Example
Let $A = \{1, 2, 3\}$ and $B = \{a, b\}$, then

$$A \times B = \{(1, a), (2, a), (3, a), (1, b), (2, b), (3, b)\}$$

and

$$B^2 = B \times B = \{(a, a), (a, b), (b, a), (b, b)\}.$$

In the same manner, starting from the set of real number $\mathbb{R}$, also known as the real line, we can generate the set known as the real plane

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) | x, y \in \mathbb{R}\}$$

and the real space

$$\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z) | x, y, z \in \mathbb{R}\}$$

- A *binary operation* on a set is a calculation involving two *operands* (elements of the set) and its *result* is an element of the set. Let $\star$ be an operation in a set A. We write:

$$\begin{array}{cccc} \star: & A \times A & \rightarrow & A \\ & (a, b) & \mapsto & c = a \star b \end{array}$$

which means that given two elements $a, b \in A$, the result of operating a and b is an element $c = a \star b$ which also belongs to A.

  – The property that the result of operating two elements of the set A is also an element of the set A is called the *closure property*.

Example

  * The normal sum (+) of natural numbers ($\mathbb{N}$) and real numbers ($\mathbb{R}$) is an operation that fulfills the closure property.
  * The normal subtraction (−) of natural numbers is an operation that does not fulfill the closure property ($2-4=-2 \notin \mathbb{N}$) while in the case of real numbers it is fulfilled.
  * The product of rational numbers ($\mathbb{Q}$) is an operation that fulfills the closure property.

- Another property an operation can have is *associativity*. An operation $\star$ in a set A is said to be associative if for all elements a, b and c in A it holds:

$$a \star (b \star c) = (a \star b) \star c.$$

Example
Usual sum and product in the natural, integer, rational and real numbers are associative operations. On the other hand, subtraction and division are not. Take as examples these cases: $3-(2-1) \neq (3-2)-1$ and $3/(5/2) \neq (3/5)/2$

- Notice that as the definition of operation is based on the cartesian product ($A \times A$) the order of the operands does matter in principle. An operation where the order of the operands does not matter is said to be *commutative*. Formally, an operation $\star$ is commutative if for all $a, b \in A$ it holds $a \star b = b \star a$.

Example
Again, normal addition and multiplication in the natural, integer, rational and real numbers are commutative operations while division and subtraction are not.

Another operation that is non-commutative is the product of matrices. For instance, if $M = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $N = \begin{pmatrix} 3 & 4 \\ 0 & 2 \end{pmatrix}$ then $M \cdot N = \begin{pmatrix} 3 & 8 \\ 9 & 20 \end{pmatrix}$ while $N \cdot M = \begin{pmatrix} 15 & 22 \\ 6 & 8 \end{pmatrix}$

- We say that an operation $\star$ in A has an identity (also called neutral) element, e, if there exists an element $e \in A$ such that for all elements $a \in A$ it holds

$$a \star e = e \star a = a$$

.

Example
The normal addition and multiplication in the integer, rational and real numbers have identity elements 0 and 1 respectively. Notice that $0 \notin \mathbb{N}$

- Given an operation $\star$ in A and $a \in A$, let e be the identity element of $\star$ in A. An element $b \in A$ is said to be an inverse of a if $a \star b = b \star a = e$. It can be easily shown that if this exists it is unique (no other element can be the inverse of a). We will write the inverse of a as $-a$ or a^{-1}.

Example

In the normal addition in $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ all numbers have an inverse. This is not the case for the natural numbers. For the multiplication case in $\mathbb{Q}$ and $\mathbb{R}$ all numbers but 0 have inverse.

– Given two operations $\star$ and $\circ$ in A, we say that $\star$ is *distributive* over $\circ$ if $a \star (b \circ c) = (a \star b) \circ (a \star c)$ and $(b \circ c) \star a = (b \star a) \circ (c \star a)$ for all $a, b, c \in A$

Example

Normal multiplication in the natural, integer, rational and real numbers is distributive over the normal addition.

3 Groups and fields

In this section we will introduce the notions of *group* and *field*. Both concepts are fundamental in all fields of mathematics. A group is nothing other than a set of elements together with an operation that combines any two of its elements to form a third element plus a few requirements on the operation behavior which naturally lead to the concept of subtraction. Many basic mathematical structures are groups (say $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ with the usual addition, for instance).

On the other hand, a field is a set with two operations designated as addition and multiplication with some properties that lead naturally to the operations of subtraction and division.

Definition 1. A *group* is a set, G, together with an operation $\star$

$$\begin{array}{rccc} \star: & A \times A & \to & A \\ & (a,b) & \mapsto & c = a \star b \end{array}$$

which satisfies the following axioms:

1. $a \star b \in G \quad \forall a, b \in G$ (closure)
2. $a \star (b \star c) = (a \star b) \star c \in G \quad \forall a, b, c \in G$ (associativity)
3. $\exists e \in G$ such that $a \star e = e \star a = a \ \forall a, b \in G$ (identity element)
4. $\forall a \in G \ \exists b \in G$ such that $a \star b = e$ and $b \star a = e$ (inverse element).

. We denote the group as $(G, \star)$.

As a remark, the associativity property is the one that allows us to get rid of the parentheses when summing of multiplying several numbers. That is, we usually write $a \cdot b \cdot c \cdot d$ instead of $a \cdot (b \cdot (c \cdot d))$ or $(a \cdot b) \cdot (c \cdot d)$ or $((a \cdot b) \cdot c) \cdot d$ as the multiplication is defined as a binary operation. It is correct to write it without parentheses beacuse the multiplication is associative.

Example

- $(\mathbb{Z}, +)$ is a group.
- $(\mathbb{N}, +)$ is not a group, as there is no identity element for the sum.
- $(\mathbb{Z}, -)$ is not a group, as the associativity property is not fulfilled.
- $(\mathbb{Z}, \cdot)$ is not a group, as there are no inverse elements for the elements 2, 3, 4, etc. (i.e. there is no integer x such that $2 \cdot x = 1$).
- $(\mathbb{Q}, +)$ and $(\mathbb{R}, +)$ are also groups.
- $(\mathbb{Q}, \cdot)$ and $(\mathbb{R}, \cdot)$ are not groups. The only property that is violated is the inverse element. The element 0 does not have an inverse (i.e. there is no number x such that $x \cdot 0 = 1$). If we remove 0 from these sets it is easy to see that $(\mathbb{Q} \setminus \{0\}, \cdot)$ and $(\mathbb{R} \setminus \{0\}, \cdot)$ are groups.

- The set of polynomials of degree n with coefficients in $\mathbb{Z}$ $(a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 | a_n, \cdots, a_0 \in \mathbb{Z})$ are a group with the addition of polynomials. The identity element is the constant polynomial 0. Given a polynomial $p(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ its inverse is $q(x) = -a_nx^n + (-a_{n-1})x^{n-1} + \cdots + (-a_1)x + (-a_0)$.

Notice that the commutativity property is not required in the definition of a group. There are groups that are not commutative.

Definition 2. A *field* is a set, F, with two operations, $+$ and $\cdot$, such that

1. $a + b \in F \; \forall a, b \in F$ (closure for $+$).
2. $a + (b + c) = (a + b) + c$ (associativity for $+$).
3. $\exists e_+ \in F$ such that $a + e_+ = e_+ + a = a \; \forall a \in F$ (neutral element for $+$). We will denote this element as 0.
4. $\forall a \in F \; \exists b$ such that $a + b = b + a = 0$ (inverse element for $+$). We will denote this element as $-a$.
5. $a + b = b + a \; \forall a, b \in F$ (commutativity for $+$).
6. $a \cdot b \in F \; \forall a, b \in F$ (closure for $\cdot$).
7. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ (associativity for $\cdot$).
8. $\exists e_. \in F$ such that $a \cdot e_. = e_. \cdot a = a \; \forall a \in F$ (neutral element for $\cdot$). We will denote this element as 1.
9. $\forall a \in F \setminus \{0\} \; \exists b$ such that $a \cdot b = b \cdot a = 1$ (inverse element for $\cdot$ for all elements but 0). We will denote this element as a^{-1}.
10. $a \cdot b = b \cdot a \; \forall a, b \in F$ (commutativity for $\cdot$).
11. $a \cdot (b + c) = a \cdot b + a \cdot c \; \forall a, b, c \in F$ (distributive property of $\cdot$ with respect to $+$).
12. $1 \neq 0$ (nontriviality, the neutral element for $+$ and the neutral element for $\cdot$ must be different).

We will write $(F, +, \cdot)$.

From this definition one can easily see that $(F, +)$ and $(F \setminus \{0\}, \cdot)$ are commutative groups. Recall that $\mathbb{Q}$ and $\mathbb{R}$ satisfy this. In fact, $(\mathbb{Q}, +, \cdot)$ and $(\mathbb{R}, +, \cdot)$ are fields but not $(\mathbb{N}, +, \cdot)$ and $(\mathbb{Z}, +, \cdot)$.

The complex numbers $\mathbb{C} = \{a + ib | a, b \in \mathbb{R} \text{ and } i^2 = -1\}$ with the complex addition and multiplication:

- $(a + ib) + (c + id) = (a + c) + i(b + d)$
- $(a + ib) \cdot (c + id) = (ac - bd) + i(bc + ad)$

are a field.

Not all fields have infinite elements. This might seem counterintuitive if one takes into account that addition and multiplication operations are closed. The idea behind the finite groups is that these operations are adapted in such a way that the closure property is fulfilled together with all the other required properties. For instance, let's consider the set $\mathbb{Z}_2 = \{0, 1\}$. And let's define the operations

+	0	1
0	0	1
1	1	0

and

·	0	1
0	0	0
1	0	1

. Then $(\mathbb{Z}_2, +, \cdot)$ is a field.

4 Matrices

Definition 3. Let F be a field (e.g. $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$). A *matrix* $\mathbf{A}$ with coefficients in F of order $n \times m$ $(n, m \in \mathbb{N})$ is a collection of $n \times m$ ordered elements of F.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} = \{a_{ij}\}_{1 \le i \le n, 1 \le j \le m}$$

The first index refers to the row number and the second to the column number. A $1 \times n$ matrix is called a *row vector* whereas an $m \times 1$ matrix is called *column vector*. In the case of $n = m$ the matrix is said to be square of order n.

Example

- $(3\ 4\ 2)$ is a row vector while $\begin{pmatrix} 3 \\ 5 \\ 1 \end{pmatrix}$ is a column vector.
- The matrix $\begin{pmatrix} 1.3 & 4 \\ \frac{6}{5} & 5 \\ 7 & \frac{1}{7} \end{pmatrix}$ is a 3×2 matrix in $\mathbb{Q}$ and the matrix $\begin{pmatrix} \frac{\pi}{2} & e^2 & 7 \\ 0 & \frac{3\pi}{5} & \frac{1}{e} \end{pmatrix}$ is a 2×3 matrix with coefficients in $\mathbb{R}$

We refer to the set of all $n \times m$ matrices with coefficients in F by $M_{n\times m}(F)$.

Definition 4.
- The *main diagonal* of a matrix $\mathbf{A} = (a_{ij})$ is the set of coefficients a_ij such that $i = j$.
- A *zero matrix*, $\mathbf{0}$ is a matrix all of whose elements are equal to zero.
- The *unit matrix* or *identity matrix* of order n, $\mathbf{I}_n$ is the square matrix of order n in which all the coefficients in the main diagonal are equal to one and all other elements are 0. The identity matrix of order n is written as $\mathbf{I}_n$.

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Definition 5. For any square matrix, the trace is evaluated by:

$$tr(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

with properties:

$$tr(A^{\mathrm{T}}) = tr(A)$$

$$tr(A \pm B) = tr(A) \pm tr(B)$$
$$tr(AB) = tr(BA)$$
$$tr(AB) \neq tr(A)tr(B)$$

4.1 Basic operations

4.1.1 Sum

Definition 6. Let $\mathbf{A}, \mathbf{B} \in M_{n\times m}(F)$. The *sum of matrices* $\mathbf{A}$ and $\mathbf{B}$ is defined as the matrix $\mathbf{C} \in M_{n\times m}(F)$ such that

$$c_{ij} = a_{ij} + b_{ij}\ \forall 1 \leq i \leq n, 1 \leq j \leq m.$$

We write then $\mathbf{C} = \mathbf{A} + \mathbf{B}$.

The sum of matrices is a commutative operation, hence for any two matrices $\mathbf{A}$ and $\mathbf{B}$ we have $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.

Example

$$\begin{pmatrix} 2 & 1 & 5 \\ 4 & 2 & 1 \\ 0 & 3 & 2 \end{pmatrix} + \begin{pmatrix} 2 & 5 & 0 \\ 1 & 1 & 3 \\ 2 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 6 & 5 \\ 5 & 3 & 4 \\ 2 & 3 & 2 \end{pmatrix}$$

4.1.2 Transposition

Definition 7. Let $\mathbf{A} = (a_{ij}) \in M_{n\times m}(F)$. The transpose of $\mathbf{A}$, $\mathbf{A}^\intercal \in M_{m\times n}(F)$ is the matrix of order $m \times n$

$$\mathbf{A}^\intercal = (b_{ij})_{1\leq i\leq n; 1\leq j\leq m} \text{ where } b_{i_j} = a_{ji}$$

Example

$$\begin{pmatrix} 4 & 2 & 1 \\ 0 & 3 & 2 \end{pmatrix}^\intercal = \begin{pmatrix} 4 & 0 \\ 2 & 3 \\ 1 & 2 \end{pmatrix}$$

Definition 8. Let $\mathbf{A} \in M_n(F)$ be a square matrix. $\mathbf{A}$ is said to be *symmetric* if $\mathbf{A} = zA^\intercal$ and *antisymmetric* if $\mathbf{A} = -\mathbf{A}^\intercal$. Antisymmetric matrices have 0 diagonals ($a_{ii} = -a_{ii} \Rightarrow a_{ii} = 0\ \forall i$)

4.1.3 Product

Definition 9. Let $\mathbf{A} = (a_{ij}) \in M_{n\times p}(F)$ and $\mathbf{B} = (b_{ij}) \in M_{p\times m}(F)$. The product of $\mathbf{A}$ and $\mathbf{B}$ is the matrix $\mathbf{C} = \mathbf{A}\cdot\mathbf{B} \in M_{n\times m}(F)$, where

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}.$$

Notice that the number of columns of the first (left) factor must be the same as the number of rows of the second (right) factor. The resulting matrix has the same number of rows as the left factor and the same number of columns as the right factor. Notice also that, even though square matrices can be multiplied in either order (swapping the matrices order) the product is not commutative.

Example

Let $\mathbf{A} = \begin{pmatrix} 4 & 2 & 1 \\ 0 & 3 & 2 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 7 & 3 \\ 0 & 5 \end{pmatrix}$. Then

- $\mathbf{A}\cdot\mathbf{B}$ is not defined, as the number of columns of $\mathbf{A}$ and the number of rows of $\mathbf{B}$ is not the same.
- On the other hand, $\mathbf{B}\cdot\mathbf{A} = \begin{pmatrix} 28 & 23 & 13 \\ 0 & 15 & 10 \end{pmatrix}$
- $\mathbf{B}\cdot\mathbf{B} = \begin{pmatrix} 49 & 36 \\ 0 & 25 \end{pmatrix}$ and $\mathbf{A}^\mathsf{T}\cdot\mathbf{B} = \begin{pmatrix} 28 & 12 \\ 14 & 21 \\ 7 & 13 \end{pmatrix}$
- $\mathbf{A}^\mathsf{T}\cdot\mathbf{A} = \begin{pmatrix} 16 & 8 & 4 \\ 8 & 13 & 8 \\ 4 & 8 & 5 \end{pmatrix}$ and $\mathbf{A}\cdot\mathbf{A}^\mathsf{T} = \begin{pmatrix} 21 & 8 \\ 8 & 13 \end{pmatrix}$

```
octave:1> A=[1,2;4,0;3,-2;5,1]
A =

   1   2
   4   0
   3  -2
   5   1

octave:2> B=[1,2,0;5,-1,3]
B =

   1   2   0
   5  -1   3

octave:3> C=A*B
C =
```

```
11    0    6
 4    8    0
-7    8   -6
10    9    3
```

Definition 10. Let $\mathbf{A} \in M_n(F)$ be a square matrix. A square matrix $\mathbf{B}$ is *the inverse of* $\mathbf{A}$ if $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$ and $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$. A matrix $\mathbf{A}$ is called *invertible* if it has an inverse.

If the inverse of a matrix exists, it is unique. Hence, the inverse of a matrix $\mathbf{A}$ can be written as $\mathbf{A}^{-1}$.

Example

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$$

If A is not square, then we define the pseudo-inverse A^+ as:

$$A^+ = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$$

and it can easily been shown that

$$A^+A = I$$

4.1.4 Determinant of a matrix

Definition 11. The *determinant* is an operation from the set of all square matrices of order n with coefficients in a field F to the field F. That is, for any matrix $\mathbf{A}$, $\det(\mathbf{A})$ is an element of F. The determinant can be defined in many ways. The definition which leads to the simplest way of computing determinants is by the *Laplace expansion (cofactor expansion)*. We write $\det(\mathbf{A}) = |\mathbf{A}|$

Invertible matrices are precisely those matrices with a nonzero determinant.

In the case of square matrices of order two, the determinant can be computed in the following way

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

And in the case of order 3 matrices:

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{13}a_{22}a_{31}-a_{12}a_{21}a_{33}-a_{11}a_{23}a_{32}.$$

Let $\mathbf{A} \in M_n(F)$, the Laplace expansion of $\det(\mathbf{A})$ is a way to express it as a sum of n determinants of $(n-1)\times(n-1)$ sub-matrices of $\mathbf{A}$. Let's define the i, j–*minor* of $\mathbf{A}$, $\mathbf{A}_{ij} \in M_{n-1}(F)$ as the matrix obtained by removing the ith row and jth column of **A**. Fix any index j, then the determinant of **A** can be defined recursively as

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

The same result is also true by fixing a row and summing over the columns. Notice that there are $2n$ different possible expansions, the result does not depend on the chosen column/row.

The following is a list of the most important properties of the determinants:

- If a matrix has two columns or two rows equal (or proportional) then the determinant is zero.
- If $\mathbf{A}$ is a matrix such that $\det(\mathbf{A}) = 0$ then there is a linear combination of rows (columns) of $\mathbf{A}$ equal to zero.
- If one column/rows of a matrix is a linear combination of other columns/rows then its determinant is zero.
- $\det(\mathbf{A}^\mathsf{T}) = \det(\mathbf{A})$.
- $\det(\mathbf{A} \cdot \mathbf{B}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$

Proposition 1. Let $\mathbf{A} \in M_n$. $\mathbf{A}$ is invertible if and only if $\det(A) \neq 0$.

There are different ways of computing the inverse of a matrix.

4.1.5 Rank of a matrix

Definition 12. Let $\mathbf{A} \in M_{n\times m}$. A *minor* of order r of $\mathbf{A}$ is a square submatrix of $\mathbf{A}$ of order r. That is, a submatrix of $\mathbf{A}$ obtained by removing $n - r$ rows and $m - r$ columns.

Definition 13. The *rank* of a matrix $\mathbf{A} \in M_{n\times m}(F)$ is r if any minor of order $> r$ has a determinant of zero and there exists a minor of order r with a non–zero determinant.

Example

Let $\mathbf{A} = \begin{pmatrix} 3 & 1 & 5 \\ 2 & 1 & 4 \\ 5 & 0 & 5 \end{pmatrix}$. Then, as $\det(\mathbf{A}) = 0$ but $\det \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} = 1 \neq 0$ the rank of $\mathbf{A}$ is $\mathrm{rk}(\mathbf{A}) = 2$.

```
octave:4> D=[1,2,2,3;4,0,8,-1;3,-2,6,0;5,1,10,-8]
D =

    1    2    2    3
    4    0    8   -1
    3   -2    6    0
    5    1   10   -8
```

```
octave:5> det(D)
ans = 0

octave:9> E=[1,2,3;4,0,-1;3,-2,0]
E =

   1   2   3
   4   0  -1
   3  -2   0

octave:10> det(E)
ans = -32
```

In the above example, matrix D has rank 3. It is also the maximum number of linearly independent columns or rows of D.

There several properties of matrices based on the rank:

- $rank(A^{m\times n}) \leq \min m, n$.
- $rank(A^{n\times n}) = n$ if and only if A is nonsingular (invertible).
- $rank(A^{n\times n}) = n$ if and only if $det(A) \neq 0$.
- $rank(A^{n\times n}) < n$ if and only if A is singular.

4.2 Orthogonal/orthonormal matrices

Consider the matrix A:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ & & \vdots & \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix}$$

Let's take the vectors formed by the rows (or columns) of matrix A:

$$u_1^{\mathrm{T}} = (a_{11}, a_{12}, \dots, a_{1n})$$

$$u_2^{\mathrm{T}} = (a_{21}, a_{22}, \dots, a_{2n})$$

$$\vdots$$

$$u_m^{\mathrm{T}} = (a_{m1}, a_{m2}, \dots, a_{mn})$$

Let us consider the properties:

1. $u_k u_k = 1$ or $\|u_k\| = 1$, for every k
2. $u_j u_k = 0$, for every $j \neq k$

A is orthonormal if both conditions are satisfied. A is ortogonal if only condition 2 is satisfied.

If A is orthonormal, then:

$$AA^{\mathrm{T}} = A^{\mathrm{T}}A = I$$

or what is the same,

$$A^{-1} = A^{\mathrm{T}}$$

$$\|Av\| = \|v\|$$

5 Systems of linear equations

Example

Let's consider the reaction of glucose ($C_6H_{12}O_6$) oxidation in which carbon dioxide and water are obtained. Suppose we don't know the stoichiometric coefficients of the reaction, which we will designate by the unknowns x, y, z and t as shown in:

$$xC_6H_{12}O_6 + yO_2 \longrightarrow zCO_2 + tH_2O$$

The number of atoms of each element must be the same on each side of the reaction, hence we can establish the following relations:

$$\begin{aligned} 6x &= z \\ 12x &= 2t \\ 6x + 2y &= 2z + t \end{aligned}$$

We will see that, this system is *compatible indeterminate*, meaning that it accepts infinite solutions. Setting $x = 1$ we get only one solution which is $(x, y, z, t) = (1, 6, 6, 6)$

Definition 14. A *system of linear equations* is a collection of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots a_{1m}x_m &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2m}x_m &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots a_{nm}x_m &= b_n \end{aligned}$$

where the numbers $a_{ij} \in \mathbb{R}$ are the *coefficients* and b_i are the independent or constant term. This system can be represented in the matrix form

$$\mathbf{Ax} = \mathbf{b},$$

where $\mathbf{A} = (a_{ij})_{1\le i\le m; 1\le j\le n}$ is the *matrix of the sytem*, $\mathbf{x} = (x_1, \cdots, x_m)^{\mathsf{T}}$ is the variable vector and $\mathbf{b} = (b_1, \cdots, b_n)^{\mathsf{T}}$ is the vector of independent terms. A column vector $\mathbf{s} = (s_1, \cdots, s_m)^{\mathsf{T}} \in \mathbb{R}^n$ is a *solution of the system* if substituting $\mathbf{x}$ by $\mathbf{s}$ gives a true statement

$$\mathbf{As} = \mathbf{b}.$$

That is, $\mathbf{s} = (s_1, \cdots, s_m)^{\mathsf{T}}$ is a solution of all the equations in the system.

- Not all systems have a unique solution (e.g. $2x + 4y = 0$ accepts infinite solutions).
- There are systems with no solutions (e.g. $-2x + 4y = 1,\ x - 2y = 3$ has no solutions).

Therefore, we need a criterium to decide whether a given system of linear equations has a solution or not. One of the most used criteria is the Rouché–Frobenius criterium which is based on the rank of the systems matrix $\mathbf{A}$ and the augmented matrix $\mathbf{A}|\mathbf{b}$ (which is $\mathbf{A}$ with the column vector $\mathbf{b}$ appended). It says:

- If $\text{rk}(\mathbf{A}) = \text{rk}(\mathbf{A}|\mathbf{b})$, the system is said to be *compatible* and it accepts solutions.
  - If $\text{rk}(\mathbf{A}) = m$ (the number of variables), the system is said to be *determinate* and it has one unique solution.
  - Otherwise, the system is said to be indeterminate and it has an infinite number of solutions.
- If $\text{rk}(\mathbf{A}) \neq \text{rk}(\mathbf{A}|\mathbf{b})$, the system is said to be *incompatible* and there are no solutions to it.

Once we know whether a system has solutions or not, we have to solve it (in the former case). Although there are many methods for solving systems of linear equations using computers, the standard method of resolution by hand is the *method of Gauss* or *Gaussian elimination method*. This is based on replacing equations in the system by linear combinations of other equations in such a way that the obtained system is equivalent to the original one in the sense that they share the same solutions but the new system is upper diagonal (and hence can be trivially solved). This method is based on the following result:

Theorem 1. If a system of linear equations is changed to another by one of these transformations:

1. an equation is swapped with another equation
2. an equation has both sides multiplied by a nonzero constant
3. an equation is replaced by the sum of itself and a multiple of another

then the two systems have the same set of solutions

Example

Given the system

$$\begin{pmatrix} -3 & 2 & -6 \\ 5 & 7 & -5 \\ 1 & 4 & -2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 6 \\ 6 \\ 8 \end{pmatrix}$$

We know that the system is compatible determinate and hence it only has one unique solution.

$$\mathbf{A}|\mathbf{b} = \left(\begin{array}{ccc|c} -3 & 2 & -6 & 6 \\ 5 & 7 & -5 & 6 \\ 1 & 4 & -2 & 8 \end{array}\right) \xrightarrow{\rho_1+3\rho_3} \left(\begin{array}{ccc|c} 0 & 14 & -12 & 30 \\ 5 & 7 & -5 & 6 \\ 1 & 4 & -2 & 8 \end{array}\right) \xrightarrow{\frac{1}{2}\rho_1} \left(\begin{array}{ccc|c} 0 & 7 & -6 & 15 \\ 5 & 7 & -5 & 6 \\ 1 & 4 & -2 & 8 \end{array}\right) \xrightarrow{-5\rho_3+\rho 2}$$

$$\left(\begin{array}{ccc|c} 0 & 7 & -6 & 15 \\ 0 & -13 & 5 & -34 \\ 1 & 4 & -2 & 8 \end{array}\right) \xrightarrow{2\rho_1+\rho 2} \left(\begin{array}{ccc|c} 0 & 7 & -6 & 15 \\ 0 & 1 & -7 & -4 \\ 1 & 4 & -2 & 8 \end{array}\right) \xrightarrow{\rho_1-7\rho 2} \left(\begin{array}{ccc|c} 0 & 0 & 43 & 43 \\ 0 & 1 & -7 & -4 \\ 1 & 4 & -2 & 8 \end{array}\right) \xrightarrow{\frac{1}{43}\rho_1}$$

$$\left(\begin{array}{ccc|c} 0 & 0 & 1 & 1 \\ 0 & 1 & -7 & -4 \\ 1 & 4 & -2 & 8 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} 1 & 4 & -2 & 8 \\ 0 & 1 & -7 & -4 \\ 0 & 0 & 1 & 1 \end{array}\right)$$

Hence $z = 1$, $y - 7 = -4 \Rightarrow y = 3$ and $x + 12 - 2 = 8 \Rightarrow x = -2$

Summarizing:

If A is invertible, $AX = b$ has exactly one solution:

$$x = A^{-1}b$$

The following statements are equivalent:

1. A is invertible
2. $Ax = 0$ has only the trivial solution
3. $det(A) \neq 0$
4. b is in the column space of A.

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} x_1 + \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} x_1 + \cdots \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} x_1 = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

5. $rank(A|b) = rank(A)$ and $rank(A) = n$
6. The column/row vectors of A are linearly independent
7. The column/row vectors of A span R^n

The system has *no solution* if $rank(A|b) > rank(A)$. The system has infinitely many solutions if $rank(A|b) = rank(A) < n$.

5.1 Elementary matrices and inverse

The same method Gauss-Jordan can be applied to obtain the inverse matrix. Let us define first the above transformation steps in a more precise way. Indeed, there are just three types of transformations, and all can be associated to the product for a so-caled elementary matrix:

1. Switching two rows in the matrix. For example, switching rows 2 and 3 in a given $3 \times m$ matrix A is equivalent to do

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} A = E_{23}A$$

2. Multiplying a row by a given value. For example,

$$\begin{pmatrix} c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} A = E_1(c)A$$

3. Summing up one row to the product of another by a number. This is:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} A = E_{23}(c)A$$

It is easy to see that we can build an inverse matrix making use of elementary transformations. Let A be an invertible $n \times n$ matrix. Suppose that a sequence of elementary row-operations reduces A to the identity matrix. Then the same sequence of elementary row-operations when applied to the identity matrix yields A^{-1}. To see how this is the case, let $E_1, E_2, \ldots, E_k$ be a sequence of elementary row operations such that $E_1 E_2 \cdots E_k A = I_n$. Then $E_1 E_2 \cdots E_k I_n = A^{-1}$ which, in turn, implies $A^{-1} = E_1 E_2 \cdots E_k$.

6 Vector spaces

Vector spaces are the mathematical structures most oftenly found in Bioinformatics. The real numbers $\mathbb{R}$, real plane $\mathbb{R}^2$ and real space $\mathbb{R}$ are the most common vector spaces. The idea behind a vector space is that its elements, the vectors, can be added between them and also scaled by real numbers.

Definition 15. A *vector space* over $\mathbb{R}$ consists of a set V along with two operations $+$ and $\cdot$ such that:

1. If $\vec{v}, \vec{w} \in V$ then their *vector sum* $\vec{v} + \vec{w} \in V$ and
   - $\vec{v} + \vec{w} = \vec{w} + \vec{v}$ (commutative)
   - $\vec{v} + (\vec{w} + \vec{u}) = (\vec{v} + \vec{w}) + \vec{u}$ for $\vec{u} \in V$ (associative)
   - there is a *zero vector* $\vec{0} \in V$ such that $\vec{0} + \vec{v} = \vec{v}$
   - $\forall \vec{v} \in V\ \exists \vec{w}$ such that $\vec{w} + \vec{v} = \vec{0}$ (additive inverse)
2. If $r, s \in \mathbb{R}$ (scalars) and $\vec{v}, \vec{w} \in V$, then $r\vec{v} \in V$ and
   - $(r + s)\vec{v} = r\vec{v} + s\vec{v}$
   - $r(\vec{v} + \vec{w}) = r\vec{v} + r\vec{w}$
   - $(rs)\vec{v} = r(s\vec{v})$
   - $1\vec{v} = \vec{v}$

Observe that we are using two kinds of additions, the real numbers addition and the vector addition in V

$$\underbrace{(r+s)\vec{v}}_{\text{real numbers addition}} \quad = \quad \underbrace{r\vec{v} + s\vec{v}}_{\text{vector addition}}$$

Example

- The set $\mathbb{R}^2$ is a vector space if the operations $+$ and $\cdot$ have their usual meaning:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}$$
$$r \cdot \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} r \cdot x_1 \\ r \cdot y_1 \end{pmatrix}.$$

  The zero vector of this vector space is $\vec{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. In fact $\mathbb{R}^n$ is a vector space, for any $n > 0$.

- $P = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^n | x + y + z = 0\}$ is a vector space: If $\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^n$, then for any $r \in \mathbb{R}$ $r \cdot \mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ and $r \cdot x + r \cdot y + r \cdot z = r \cdot (x + y + z) = 0$, hence $r \cdot \mathbf{v} \in P$.

- The set with only one element, the zero vector, is a vector space called the trivial vector space: $\{\vec{0}\}$.
- The set of polynomials of degree 3 or less with real coefficients, $P_3(\mathbb{R}) = \{a_0 + a_1x + a_2x^2 + a_3x^3 | a_0, a_1, a_2, a_3 \in \mathbb{R}\}$ is a vector space with the usual polynomial sum and product by constant. In fact, $P_n(\mathbb{R})$ is a vector space for any $n > 0$.
- The set of solutions of a homogeneous system of linear equations $S = \{\mathbf{v} \in \mathbb{R}^m | \mathbf{Av} = 0\}$, $\mathbf{A} \in M_{n\times m}(\mathbb{R})$ is also a vector space:

$$\mathbf{v}, \mathbf{w} \in S \Rightarrow \mathbf{A}(\mathbf{v}+\mathbf{w}) = \mathbf{Av} + \mathbf{Aw} = 0$$
$$\mathbf{v}, \in S, r \in \mathbb{R} \Rightarrow \mathbf{A}(r\mathbf{v}) = r\mathbf{Av} = 0$$

Definition 16. For any vector space V, any subset that is itself a vector space is a *subspace* of V

The linear combination of n vectors in the vector space E over $\mathbf{K}$, with $n \in \mathbf{N}$ and coefficient s $\alpha_i \in \mathbf{K}(i = 1, \dots, n)$ is defined as

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = \sum_{i=1}^{n} \alpha_i v_i$$

and we will say that $v \in E$ is a linear combination of $v_1, \dots, v_n \in E$ if there exist a set of coefficien ts $\alpha_i \in \mathbf{K}(i = 1, \dots, n)$ such that

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

Given n vectors, the subspace that is formed by all their possible linear combinations is called the subspace "g enerated" or "spanned" by them, $< v_1, \dots, v_n >$. This set of vectors, represented by $\{v_1, \dots, v_n\}$, is called "spanning set" of $< v_1, \dots, v_n >$.

Let's imagine that we want to span the vector zero from a linear combination of vectors of the set $\{v_1, \dots, v_n\}$.If this is only possibly done by the so-called "trivial solution", this is, with all α_i equal to zero, t hen we will say that $\{v_1, \dots, v_n\}$ is a set of vectors "linearly independent" or a "free set". If there exists some way to obtain 0 without all coefficients being 0, then we will say that $\{v_1, \dots, v_n\}$ i s a set of linearly dependent vectors.

A "basis" is a set of vectors that spans the subspace and at the same time is linearly independent.

This is, $B = v_1, \dots, v_n$ is a basis of the subspace V if:

- each vector of V is a linear combination of $v_1, \dots, v_n$, and
- the vectors $v_1, \dots, v_n$ are linearly independent.

If so, there will exist an ordered list of scalars such that: $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Thus, once we know the vector of the basis we know the whole subspace.

Example

Show that in $\mathbb{R}_4$, the set of vectors whose components follow:

$$x_1 + x_2 + x_3 + x_4 = 0$$

form a vector subspace with dimension 3. Find a basis.

We can answer both questions at once. We only need to solve the systems of equations describing the vector subspace. Thus, the solutions to:

$$x_1 + x_2 + x_3 + x_4 = 0$$

have the form:

$$\begin{aligned} x_1 &= -x_2 - x_3 - x_4 = -a - b - c \\ x_2 &= a \\ x_3 &= b \\ x_4 &= c \end{aligned}$$

or, equivalently:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -a-b-c \\ a \\ b \\ c \end{pmatrix} = a \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

This is to say, the vector subspace is spanned by these three vectors.

Exercise 1. *Find out if in the vector space $P_2[x]$ of the polinomials with order less than or equal to 2 over $\mathbb{R}$, the fol lowing vectors form a basis:*

$$\begin{aligned} u_1 &= 1 + 2x \\ u_2 &= -1 - 2x^2 \\ u_3 &= -2x + 2x^2 \end{aligned}$$

Exercise 2. *Let be $P_3[x]$ the vector space of the polinomials or order 3 or less with real coefficients and real variable ov er the commutative body $\mathbb{R}$. Be the set of vectors $G = \{(x^2 + x + 2), (x^3 + 3x)\}$ belonging to $P_3[x]$. Find a basis of $P_3[x]$ by completing the set G.*

Lemma 1. For any nonempty subset W of a vector space V under the inheritet operations, the following statements are equivalent

1. W is a subspace of V.
2. W is closed under linear combinations of pairs of vectors: $\forall \mathbf{v_1}, \mathbf{v_2} \in W$ and $r_1, r_2 \in \mathbb{R}$, $r_1\mathbf{v}_1 + r_2\mathbf{v}_2 \in W$.
3. W is closed under linear combinations of any number of vectors: $\forall \mathbf{v_1}, \cdots, \mathbf{v_n} \in W$ and $r_1, \cdots, r_n \in \mathbb{R}$, $r_1\mathbf{v}_1 + \cdots + r_n\mathbf{v}_n \in W$.

This last result tells us that to assess if a subset of a known vector space is also a vector space (a subspace), we don't have to check everything, just that it is closed under linear combinations.

Definition 17. The *span* (or *linar closure*) of a nonempty subset, W, of vector space V is the set of all linear combinations of vectors from W:

$$[W] = \{c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n | \mathbf{w_1}, \cdots, \mathbf{w_n} \in W, c_1, \cdots, c_n \in \mathbb{R}\}.$$

Lemma 2. In a vector space, the span of any subset is a subspace (i.e. the span is cloesed under linear combinations). The converse also holds: any subspace is the span of some set.

Example

- The span of one vector $\mathbf{v} \in V$, is: $[\{\mathbf{v}\}] = \{r \cdot \mathbf{v} | r \in \mathbb{R}\}$
- Any two linearly independent vectors span $\mathbb{R}^2$. For instance,

$$\left[\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}-1\\1\end{pmatrix}\right\}\right] = \mathbb{R}^2$$

  Any vector $\begin{pmatrix}x\\y\end{pmatrix}$ can be written as: $\frac{x+y}{2}\begin{pmatrix}1\\1\end{pmatrix} + \frac{y-x}{2}\begin{pmatrix}-1\\1\end{pmatrix}$.

Exercise 3. *Check that the set of vectors* $F = \{(x, y, z) \in \mathbb{R}^3 / x+y+z = 9\}$ *is not a vector subspace in* $\mathbb{R}^3$.

Definition 18. If in a vector space there exist a basis formed by n elements and $m > n$, then we can assure that any set of m vectors is linearly dependent. **In any finite-dimensional vector space, all of the bases have the same number of elements** The "dimension" of a vector space is the number of vectors in any of its bases. As a consequence, for the above mentioned vector space:

1. n linearly independent vectors form a basis
2. n spanning vectors form a basis
3. If V is a subspace of E, then V has a basis $(V \neq 0), dimV \leq n$, and the equality only holds if and only if $V = E$.
4. If $r < n$ and $v_1, \ldots, v_r$ are linearly independent vectors, then there exist $n - r$ vectors $v_{r+1}, \ldots, v_n$ such that $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ is a basis of E.

Example

We consider, in the vector space $\mathbb{R}^3$ over $\mathbb{R}$, two subspaces $E_1 = \langle(1, 1, 1), (1, -1, 1)\rangle$ and $E_2 = \langle(1, 2, 0), (3, 1, 3)\rangle$

- Find the set of vectors that belong to $E_1 \cap E_2$.
- Check if it is a subspace of $\mathbb{R}^3$.

- What is the dimension of subspace $\{E_1 \cap E_2\}$?

The solution is immediate if we consider a geometrical view. In $\mathbb{R}^3$, a vector subspace with dimension 2 is a plane and that two planes can intersect in a line or can be coincident. In both cases, then, we would have vector subscapes. In this way, both planes can be found in an easy way, yielding $E_1 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 / x - z = 0\}$ and $E_2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 / 2x - y - \frac{5}{3}z = 0\}$. Joining these two expressions we will see that they are L.I. and in this way we would have three unknown variables for two equations: one degree of freedom and thus we are describing a line on $\mathbb{R}^3$.

Example

Let us consider the subspace in $\mathbb{R}^4$ defined as:

$$F = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 / x_3 = 2x_1 + 3x_2; x_4 = 2x_2 - 3x_1\} \tag{1}$$

Find a basis for the subspace and complete it until obtaining a basis for $\mathbb{R}^5$.

The equations defining the subspace can be written also as:

$$\begin{aligned} x_1 &= x_1 \\ x_2 &= x_2 \\ x_3 &= 2x_1 + 3x_2 \\ x_4 &= 2x_2 - 3x_1 \end{aligned}$$

or, equivalently:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \\ 2 \\ -3 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \\ 3 \\ 2 \end{pmatrix}$$

Thus, these two vectors form a basis of the subspace, with dimension 2. To complete the set of vectors until having a basis in $\mathbb{R}^4$, we only need to choose two vectors that, along with the vectors we already have, form a L.I. set of vectors. We could try, for example, vectors $(1, 0, 0, 0)$ i $(0, 1, 0, 0)$:

$$\begin{aligned} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 2 & 3 & 0 & 0 \\ -3 & 2 & 0 & 0 \end{pmatrix} &\approx \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 3 & -2 & 0 \\ 0 & 2 & 3 & 0 \end{pmatrix} \\ &\approx \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -2 & -3 \\ 0 & 0 & 3 & 2 \end{pmatrix} \\ &\approx \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -2 & -3 \\ 0 & 0 & 0 & -13 \end{pmatrix} \end{aligned}$$

Thus, the 4 chosen vectors are L.I. and therefore form a basis in $\mathbb{R}^4$.

6.1 Basis change

The representation of a vector as a column of components depends, obviously, on the basis. For each basis the representation will be different. How do we relate these representations?

Given two different basis in a vector subspace, $B = v_1, \dots, v_n$ and $B' = w_1, \dots, w_n$, and knowing the representation of vector v according to the first of the bases $\text{Rep}_B(v)X = (x^i)_{1\leq i\leq n} \in \mathbf{K}^n$, this is:

$$v = x^1 v_1 + x^2 v_2 + \cdots + x^n v_n \quad (2)$$

To obtain the representation of vector v according to the basis B' it is enough with knowing the representation of the vectors v_i in B':

$$\begin{gathered} v_1 = a_1^1 w_1 + a_1^2 w_2 + \cdots + a_1^n w_n \\ v_2 = a_2^1 w_1 + a_2^2 w_2 + \cdots + a_2^n w_n \\ \vdots \\ v_n = a_n^1 w_1 + a_n^2 w_2 + \cdots + a_n^n w_n \end{gathered}$$

By replacing this representation in Eq. 2 we get:

$$\begin{array}{rcl} v & = & x^1(a_1^1 w_1 + a_1^2 w_2 + \cdots + a_1^n w_n) + \\ & + & x^2(a_2^1 w_1 + a_2^2 w_2 + \cdots + a_2^n w_n) + \\ & \vdots & \\ & + & x^n(a_n^1 w_1 + a_n^2 w_2 + \cdots + a_n^n w_n) \end{array}$$

rearranging:

$$\begin{array}{rcl} v & = & (x^1 a_1^1 + x^2 a_2^1 + \cdots x^n a_n^1) w_1 \\ & + & (x^1 a_1^2 + x^2 a_2^2 + \cdots x^n a_n^2) w_1 \\ & \vdots & \\ & + & (x^1 a_1^n + x^2 a_2^n + \cdots x^n a_n^n) w_1 \end{array}$$

Writing this in matrix form we see that calling P the matrix that represents the vectors of the basis B into the basis B', $Y = \text{Rep}_{B'}(v)$, this is, $v = y^1 w_1 + y^2 w_2 + \cdots + y^n w_n$, is given by:

$$Y = PX \quad (3)$$

or, what is the same, $\text{Rep}_{B'}(v) = P\text{Rep}_B(v)$. This matrix $P = (\text{Rep}_{B'}(v_1), \text{Rep}_{B'}(v_2), \dots, \text{Rep}_{B'}(v_n))$ is called "matrix of basis change". This matrix can be inverted, and P^{-1} is the matrix for changing from basis B' into B.

Example

Be $B' = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ a basis in $\mathbb{R}_3$ and $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ another basis in the same space, defined as:

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{u}_1 - \mathbf{u}_3 \\ \mathbf{v}_2 &= \mathbf{u}_1 + 2\mathbf{u}_2 + \mathbf{u}_3 \\ \mathbf{v}_3 &= \mathbf{u}_2 + 2\mathbf{u}_3 \end{aligned}$$

If $\mathbf{w}$ is a vector in $\mathbb{R}^3$with coordinates (2,1,-1) with respect to basis B, calculate the coordinates of $\mathbf{w}$ with respect to basis B'.

The above equations directly yield the transformation matrix:

$$Rep'_B(\mathbf{w}) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix} \cdot Rep_B(\mathbf{w}) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ -4 \end{pmatrix}$$

Example

Given the vector

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

expressed in the base

$$B = \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

what are its coordinates in the basis

$$B' = \left\{ \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix} \right\}$$

We only need to find the matrix for the basis transformation. This is built with the representations of the vectors of the old basis with respect to the vectors of the new basis. Thus:

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \alpha_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

from which we can obtain

$$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}$$

and

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \alpha_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \beta_2 \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

from which we can obtain

$$\begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} -1 \\ \frac{1}{2} \end{pmatrix}$$

Finally,

$$\mathbf{v}_{B'} = \begin{pmatrix} 0 & -1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \cdot \mathbf{v}_B$$

$$\mathbf{v}_{B'} = \begin{pmatrix} 0 & -1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ \frac{3}{2} \end{pmatrix}$$

6.2 The vector space $\mathcal{L}(V, W)$

If V and W are two vector spaces over the same body $\mathbf{K}$, a linear map $F : V \to W$ is a map that r espects the following linear operations:

$$\forall v, w \in V, F(v + w) = F(v) + F(w)$$

$$\forall \alpha \in \mathbf{K}, \forall v \in V, F(\alpha v) = \alpha F(v)$$

or, equivalently:

$$\forall \alpha_1, \alpha_2 \in \mathbf{K}, \forall v_1, v_2 \in V, F(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 F(v_1)\ alpha_2 F(v_2)$$

Let us consider, for example, the matrices with m rows and n columns: $A \in M_n^m(\mathbf{K})$. These matrices can be used to represent a linear map of $\mathbf{K}^n$ into $\mathbf{K}^m$:

$$F_A : \mathbf{K}^n \to \mathbf{K}^m$$

o bé,

$$F_A(X) = AX, \forall X \in \mathbf{K}^n$$

This map is linear because it follows the above conditions. If we define a second linear map G_A, analogous to F_A it is simple to prove that the space formed by all poss ible linear transforms $\mathcal{L}(V, W)$ has the structure of a vector space.

Exercise 1
Discuss if these transforms are or not linear:

$$F : \mathbb{R}^3 \to \mathbb{R}; F(X) = 2x - 3y + 4z$$

$$G : \mathbb{R}^2 \to \mathbb{R}^3; G(X) = (x + 1, 2y, x + y)$$

6.3 Rangespace and nullspace (Kernel)

Rangespace of a linear transformation is the set of images of all the vectors of V, $F(V)$:

$$\mathrm{Im}F = \{w \in W | \exists v \in V \text{ with } F(v) = w\}$$

The dimension of the rangespace is the map's *rank*, rg F = dim Im F. In any linear transformation, the rangespace of any subspace of the starting set into the arriving set is also a s ubspace. The whole rangespace of the linear transform is also a vector subspace. The *nullspace* or *kernel* of a linear map is the inverse image of the zero vector in the arriving space:

$$\mathrm{Nuc}F = \mathrm{Ker}F = \{v \in V | F(v) = 0\}$$

Both the rangespace and the kernel are vector subspaces.

A linear transformation is injective if and only if Nuc$F = 0$. If $F : V \to W$ is linear with Nuc $F = \{0\}$ and $v_1, \ldots, v_n$ are linearly independent vectors of V, then $F(v_1), \ldots, F(v_n)$ are also linearly independent. Thus, for an injective linear transform, the rangespace of V is a basis of W. Some definitions:

- *homomorphism* is equivalent to linear transform.
- *epimorphism* is a linear transform that is exhaustive in W.
- *isomorphism* is a one-to-one linear transform: both injective and exhaustive: bijective.
- *endomorphism* is a linear transform of a vector space in itself (also called sometimes operator)
- *automorphisms* are both endomorphisms and isomorphisms.

Rearranging the previous statements, is F is injective, it is also an isomorphism between V and Im F. If V is a vector space of finite dimension and $F : V \to W$ is linear, then,

$$\text{dim}V = \dim \text{Nuc}F + \dim \text{Im}F$$

$\dim \text{Nuc}F$ is sometimes caled the *nulity* of F and $\dim \text{Im}F$ its rank.

Exercise 2
Let be $F : \mathbb{R}^5 \to \mathbb{R}^3$ the linear transform defined as

$$F(X) = (x + 2y + z - 3s + 4t, 2x + 5y + 4z - 5s + 5t, x + 4y + 5z - s - 2t)$$

. Find a basis and the dimension of the rangespace of F.

Exercise 3
Let be the linear transform $F : \mathbb{R}^3 \to \mathbb{R}^3$ that has an associated matrix on the canonical bases

$$F = \begin{pmatrix} 1 & 2 & 5 \\ 3 & 5 & 13 \\ -2 & -1 & -4 \end{pmatrix}$$

Find a basis and the dimension of both the rangespace and the kernel.

Exercise 4
Find the kernel of the isoorphism $H : M_2^2 \to P_3$ as defined by:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \to (a + b + 2d) + 0x + cx^2 + cx^3$$

Example
The matrix in exercise 6.1 changes a vector from its representation in the base B to its representation in the base B'. Show that it is an automorphism. What would be the transformation matrix from B' to B? It can be shown in an analogous way to what we did in exercise 6.3. In this case the kernel is void: obvious! the only vector that when changing basis gets transformed into $\mathbf{0}$ is $\mathbf{0}$ itself!. A basis transformation is represented by a square matrix. If the kernel is empty and the matrix is

square this means the the rank for the associated matrix is 3 in the present case (check it). Or we can equivalently say that the determinant of the matrix is different than zero (check it). Applying:

$$\dim V = \dim \mathrm{Nuc} F + \dim \mathrm{Im} F$$

we see that the dimension of the origin is the same as the dimension of the image, which is at the same time the same as the whole final space: 3. Thus, we talk on an automorphism.

Using simple matrix algebra:

$$Rep'_B(\mathbf{w}) = \mathbf{A} \cdot Rep_B(\mathbf{w})$$

$$\mathbf{A}^{-1} \cdot Rep'_B(\mathbf{w}) = \underbrace{\mathbf{A}^{-1} \cdot \mathbf{A}}_{I} \cdot Rep_B(\mathbf{w})$$

$$\mathbf{A}^{-1} \cdot Rep'_B(\mathbf{w}) = Rep_B(\mathbf{w})$$

Thus, the matrix we are looking for is the inverse. This will exist, as a basis transformation is always an automorphism.

Example

Given the linear transform (homomorphism) $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by:

$$T(x_1, x_2, x_3) = (x_1 + x_2, x_2 + x_3)$$

1. find the associated matrix
2. find the kernel of the transformation
3. is it an isomorphism? is it an epimorphism?

1. the associated matrix $\mathbf{A}$ will be given in general by the images of the vectors in the starting subspace:

$$T(1,0,0) = (1,0)$$
$$T(0,1,0) = (1,1)$$
$$T(0,0,0) = (0,1)$$

   Thus:

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

   and thus if $\mathbf{v} \in \mathbb{R}^3$ and $\mathbf{v} \in \mathbb{R}^2$, we can represent the transformation by:

$$\mathbf{w} = \mathbf{A} \cdot \mathbf{v}$$

2. the kernel of the transformation will be formed by the vector subspace in the origin space having as imatge the nul vector.

$$\mathbf{0} = \mathbf{A} \cdot \mathbf{v}$$

   or:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

We solve the system:

$$0 = x + y$$
$$0 = y + z$$

getting:

$$\begin{aligned} x &= -y = -a \\ y &= a \\ z &= -y = -a \end{aligned}$$

or equivalently:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = a \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$$

thus, the dimension of the kernel is 1 and the vector (-1,1,-1) form a base of it.

3. it is not injective because the dimension of the kernel is not zero. Also, if V is a vector space with finite dimension and $F : V \to W$ is linear, then,

$$\dim V = \dim \mathrm{Nuc} F + \dim \mathrm{Im} F$$

In our case 3=1+$\dim \mathrm{Im} T$. Thus, $\dim \mathrm{Im} T = 2 = \dim W$, because W is $\mathbb{R}^2$. Thus the transformation is exhaustive.

Example

Is the application $F : \mathbb{V} \to \mathbb{W}$, to which the following matrix is associated:

$$A = \begin{pmatrix} 1 & -2 & -4 \\ -2 & 0 & 4 \\ 1 & 3 & 1 \end{pmatrix}$$

bijective?

The matrix determinant is 0. Thus, the three columns, corresponding to the images of the vectors that form the canonical basis of $\mathbb{V}$, are not linearly independent. The first two columns are L.I., for example, and thus $Dim(ImF) = 2$. As $Dim(\mathbb{V}) = 3$, then $Dim(KerF) = 1$. The transform is not injective, as the dimension of the kernel is not zero. The transform is not exhaustive as the dimension of the image is different than the dimension of $\mathbb{W}$. The transform is not bijective by any of these two reasons.

6.4 Composition and inverse

If $F : V \to W$ and $G : W \to U$ are two linear transforms, then $G \circ F : V \to U$ is linear. This composition is associative, but not commutative.

A linear transform $F : V \to W$ is invertible if there exists $G : W \to V$ linear, such that $G \circ F = id_V$ and $F \circ G = id_W$, and that we will call "inverse". Automorphisms F are one-to one linear transforms, and in this cases F^{-1} is also an automorphism.

6.5 Linear transforms and matrices

We can express linear transforms as matrices. This is to say, once we have set the basis for the starting and arrival spaces, we can stablish a one-to-one correspondence between linear transforms and matrices, which will have advantages because we know how to do matrix operations. This correspondence will be an isomorphism and the matrix corresponding to the linear transform will be formed by the images of the basis of V.

For example:

Let us consider a linear transform $h : \mathbb{R}^2 \to \mathbb{R}^3$. Let us consider that the basis of V and W are, respectively:

$$B = <\begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \end{pmatrix}>$$

$$D = <\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}>$$

The linear transform is defined by its action on the basis vectors in V:

$$\begin{pmatrix} 2 \\ 0 \end{pmatrix} \stackrel{h}{\longmapsto} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 4 \end{pmatrix} \stackrel{h}{\longmapsto} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

In order to evaluate how this linear transform affects any vector in its domain, first we need to express $h(b_1)$ and $h(b_2)$ in the basis of the rangespace:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{so} \quad \mathrm{Rep}_D(h(b_1)) = \begin{pmatrix} 0 \\ -1/2 \\ 1 \end{pmatrix}_D$$

and

$$\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - 1 \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{so} \quad \mathrm{Rep}_D(h(b_2)) = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}_D$$

Now, for each member of the starting space, we can express its image according to h in terms of the images of the basis vectors B:

$$\begin{aligned} h(v) &= h(c_1 \cdot \begin{pmatrix} 2 \\ 0 \end{pmatrix} + c_2 \cdot \begin{pmatrix} 1 \\ 4 \end{pmatrix}) \\ &= c_1 \cdot h(\begin{pmatrix} 2 \\ 0 \end{pmatrix}) + c_2 \cdot h(\begin{pmatrix} 1 \\ 4 \end{pmatrix}) \\ &= c_1 \cdot (0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}) + c_2 \cdot (1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - 1 \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}) \\ &= (0c_1 + 1c_2) \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (-\frac{1}{2}c_1 - 1c_2) \cdot \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + (1c_1 + 0c_2) \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \end{aligned}$$

Thus,

$$\text{with } \mathrm{Rep}_B(\vec{v}) = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \text{ then } \mathrm{Rep}_D(\, h(\vec{v})\,) = \begin{pmatrix} 0c_1 + 1c_2 \\ -(1/2)c_1 - 1c_2 \\ 1c_1 + 0c_2 \end{pmatrix}.$$

For example,

$$\text{with } \mathrm{Rep}_B(\begin{pmatrix} 4 \\ 8 \end{pmatrix}) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}_B \text{ then } \mathrm{Rep}_D(\, h(\begin{pmatrix} 4 \\ 8 \end{pmatrix})\,) = \begin{pmatrix} 2 \\ -5/2 \\ 1 \end{pmatrix}.$$

We can express these calculaitons in matrix form:

$$\begin{pmatrix} 0 & 1 \\ -1/2 & -1 \\ 1 & 0 \end{pmatrix}_{B,D} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}_B = \begin{pmatrix} 0c_1 + 1c_2 \\ (-1/2)c_1 - 1c_2 \\ 1c_1 + 0c_2 \end{pmatrix}_D$$

The interesting part of this expression is that the matrix representing a linear transform is generated, simply, by putting in columns the images of the vectors of the domain basis as a function of the vectors in the basis of the image.

I a more formal way: Let us suppose that V and W are vector spaces with dimensions n and m with basis B and D, and that $h\colon V \to W$ is a linear transform connecting them. If

$$\mathrm{Rep}_D(h(b_1)) = \begin{pmatrix} h_{1,1} \\ h_{2,1} \\ \vdots \\ h_{m,1} \end{pmatrix}_D \quad \ldots \; \mathrm{Rep}_D(h(b_n)) = \begin{pmatrix} h_{1,n} \\ h_{2,n} \\ \vdots \\ h_{m,n} \end{pmatrix}_D$$

then

$$\mathrm{Rep}_{B,D}(h) = \begin{pmatrix} h_{1,1} & h_{1,2} & \ldots & h_{1,n} \\ h_{2,1} & h_{2,2} & \ldots & h_{2,n} \\ & \vdots & & \\ h_{m,1} & h_{m,2} & \ldots & h_{m,n} \end{pmatrix}_{B,D}$$

is the matrix representation of the transformation.

Exercise 5

Represent the matrix of the linear transform $t_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ which transforms the vectors by rotating them clockwise any given angle θ.

Every matrix represents a linear transform.

6.6 Composition and matrix product

We already know how to change bases and we know how to represent linear transforms by means of matrices. Now we want to do the following scheme:

$$\begin{array}{ccc} V_B & \xrightarrow[H]{h} & W_D \\ \text{id}\big\downarrow & & \text{id}\big\downarrow \\ V_{\hat{B}} & \xrightarrow[\hat{H}]{h} & W_{\hat{D}} \end{array}$$

Or, what is identical in a matrix representation::

$$\hat{H} = \text{Rep}_{D,\hat{D}}(\text{id}) \cdot H \cdot \text{Rep}_{\hat{B},B}(\text{id}) \qquad (*)$$

For example, the matrix

$$T = \begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{pmatrix} = \begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix}$$

represents, with respect to $\mathcal{E}_1, \mathcal{E}_2$, the linear transformation $t\colon \mathbb{R}^2 \to \mathbb{R}^2$ that rotates the vectors $\pi/6$ radians anticlockwise. We can transform this representation with respect to $\mathcal{E}_2, \mathcal{E}_2$ to another one with respect to

$$\hat{B} = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} \rangle \qquad \hat{D} = \langle \begin{pmatrix} -1 \\ 0 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} \rangle$$

using what we just learnt:

$$\begin{array}{ccc} \mathbb{R}^2_{\mathcal{E}_2} & \xrightarrow[T]{t} & \mathbb{R}^2_{\mathcal{E}_2} \\ \text{id}\big\downarrow & & \text{id}\big\downarrow \\ \mathbb{R}^2_{\hat{B}} & \xrightarrow[\hat{T}]{t} & \mathbb{R}^2_{\hat{D}} \end{array} \qquad \hat{T} = \text{Rep}_{\mathcal{E}_2,\hat{D}}(\text{id}) \cdot T \cdot \text{Rep}_{\hat{B},\mathcal{E}_2}(\text{id})$$

$\text{Rep}_{\mathcal{E}_2,\hat{D}}(\text{id})$ can be written as the inverse of $\text{Rep}_{\hat{D},\mathcal{E}_2}(\text{id})$.

$$\begin{aligned} \text{Rep}_{\hat{B},\hat{D}}(t) &= \begin{pmatrix} -1 & 2 \\ 0 & 3 \end{pmatrix}^{-1} \begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix} \\ &= \begin{pmatrix} (5-\sqrt{3})/6 & (3+2\sqrt{3})/3 \\ (1+\sqrt{3})/6 & \sqrt{3}/3 \end{pmatrix} \end{aligned}$$

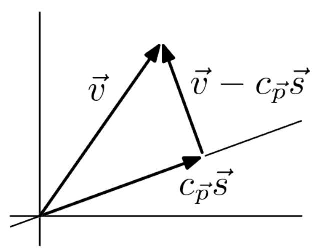
Exercise 6
Check if the effect of the new matrix is the same as the original matrix, with the new basis.

7 Projection

7.1 Orthogonal Projection Into a Line

We first consider orthogonal projection into a line. To orthogonally project a vector $\vec{v}$ into a line ℓ, darken a point on the line if someone on that line and looking straight up or down (from that person's point of view) sees $\vec{v}$.

The picture shows someone who has walked out on the line until the tip of $\vec{v}$ is straight overhead. That is, where the line is described as the span of some nonzero vector $\ell = \{c \cdot \vec{s} \mid c \in \mathbb{R}\}$, the person has walked out to find the coefficient $c_{\vec{p}}$ with the property that $\vec{v} - c_{\vec{p}} \cdot \vec{s}$ is orthogonal to $c_{\vec{p}} \cdot \vec{s}$.



We can solve for this coefficient by noting that because $\vec{v} - c_{\vec{p}}\vec{s}$ is orthogonal to a scalar multiple of $\vec{s}$ it must be orthogonal to $\vec{s}$ itself, and then the consequent fact that the dot product $(\vec{v} - c_{\vec{p}}\vec{s}) \cdot \vec{s}$ is zero gives that $c_{\vec{p}} = \vec{v} \cdot \vec{s}/\vec{s} \cdot \vec{s}$.

The orthogonal projection of $\vec{v}$ into the line spanned by a nonzero $\vec{s}$ is this vector.

$$\text{proj}_{[\vec{s}]}(\vec{v}) = \frac{\vec{v} \cdot \vec{s}}{\vec{s} \cdot \vec{s}} \cdot \vec{s} \tag{4}$$

The wording of that definition says 'spanned by $\vec{s}$' instead the more formal 'the span of the set $\{\vec{s}\}$'. This casual first phrase is common.

Exercise 7

In $\mathbb{R}^3$, the orthogonal projection of a general vector

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{5}$$

into the y-axis is

$$\frac{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix} \tag{6}$$

which matches our intuitive expectation.

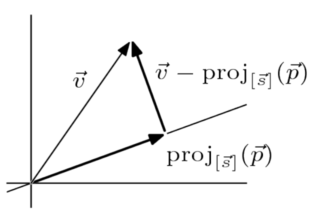
The picture above with the stick figure walking out on the line until $\vec{v}$'s tip is overhead is one way to think of the orthogonal projection of a vector into a line. We finish this subsection with two other ways.

Thus, another way to think of the picture that precedes the definition is that it shows $\vec{v}$ as decomposed into two parts, the part with the line (here, the part with the tracks, $\vec{p}$), and the part that is orthogonal to the line (shown here lying on the north-south axis). These two are "not interacting" or "independent", in the sense that the east-west car is not at all affected by the north-south part of the wind. So the orthogonal projection of $\vec{v}$ into the line spanned by $\vec{s}$ can be thought of as the part of $\vec{v}$ that lies in the direction of $\vec{s}$.

This subsection has developed a natural projection map: orthogonal projection into a line. As suggested by the examples, it is often called for in applications. The next subsection shows how the definition of orthogonal projection into a line gives us a way to calculate especially convienent bases for vector spaces, again something that is common in applications. The final subsection completely generalizes projection, orthogonal or not, into any subspace at all.

7.2 Gram-Schmidt Orthogonalization

The prior subsection suggests that projecting into the line spanned by $\vec{s}$ decomposes a vector $\vec{v}$ into two parts



$$\vec{v} = \mathrm{proj}_{[\vec{s}]}(\vec{v}) + \Big(\vec{v} - \mathrm{proj}_{[\vec{s}]}(\vec{v})\Big)$$

that are orthogonal and so are "not interacting". We will now develop that suggestion.

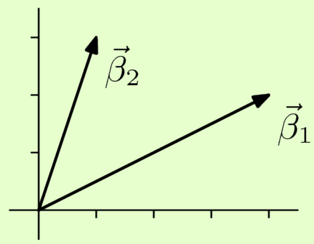
Vectors $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^n$ are mutually orthogonal when any two are orthogonal: if $i \neq j$ then the dot product $\vec{v}_i \cdot \vec{v}_j$ is zero.

If the vectors in a set $\{\vec{v}_1, \ldots, \vec{v}_k\} \subset \mathbb{R}^n$ are mutually orthogonal and nonzero then that set is linearly independent.

Exercise 8

The members $\vec{\beta}_1$ and $\vec{\beta}_2$ of this basis for $\mathbb{R}^2$ are not orthogonal.

$$B = \langle \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix} \rangle$$

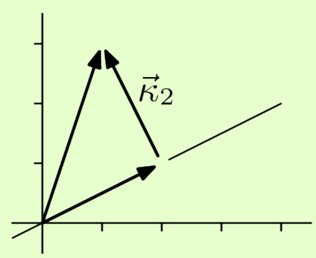


However, we can derive from B a new basis for the same space that does have mutually orthogonal members. For the first member of the new basis we simply use $\vec{\beta}_1$.

$$\vec{\kappa}_1 = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \tag{7}$$

For the second member of the new basis, we take away from $\vec{\beta}_2$ its part in the direction of $\vec{\kappa}_1$,

$$\vec{\kappa}_2 = \begin{pmatrix} 1 \\ 3 \end{pmatrix} - \text{proj}_{[\vec{\kappa}_1]}(\begin{pmatrix} 1 \\ 3 \end{pmatrix}) = \begin{pmatrix} 1 \\ 3 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$



which leaves the part, $\vec{\kappa}_2$ pictured above, of $\vec{\beta}_2$ that is orthogonal to $\vec{\kappa}_1$ (it is orthogonal by the definition of the projection into the span of $\vec{\kappa}_1$). Note that, by the corollary, $\{\vec{\kappa}_1, \vec{\kappa}_2\}$ is a basis for $\mathbb{R}^2$.

An orthogonal basis for a vector space is a basis of mutually orthogonal vectors.

Exercise 9

To turn this basis for $\mathbb{R}^3$

$$\langle \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} \rangle \tag{8}$$

into an orthogonal basis, we take the first vector as it is given.

$$\vec{\kappa}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{9}$$

We get $\vec{\kappa}_2$ by starting with the given second vector $\vec{\beta}_2$ and subtracting away the part of it in the direction of $\vec{\kappa}_1$.

$$\vec{\kappa}_2 = \begin{pmatrix}0\\2\\0\end{pmatrix} - \text{proj}_{[\vec{\kappa}_1]}(\begin{pmatrix}0\\2\\0\end{pmatrix}) = \begin{pmatrix}0\\2\\0\end{pmatrix} - \begin{pmatrix}2/3\\2/3\\2/3\end{pmatrix} = \begin{pmatrix}-2/3\\4/3\\-2/3\end{pmatrix} \tag{10}$$

Finally, we get $\vec{\kappa}_3$ by taking the third given vector and subtracting the part of it in the direction of $\vec{\kappa}_1$, and also the part of it in the direction of $\vec{\kappa}_2$.

$$\vec{\kappa}_3 = \begin{pmatrix}1\\0\\3\end{pmatrix} - \text{proj}_{[\vec{\kappa}_1]}(\begin{pmatrix}1\\0\\3\end{pmatrix}) - \text{proj}_{[\vec{\kappa}_2]}(\begin{pmatrix}1\\0\\3\end{pmatrix}) = \begin{pmatrix}-1\\0\\1\end{pmatrix} \tag{11}$$

Again the corollary gives that

$$\langle \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}-2/3\\4/3\\-2/3\end{pmatrix}, \begin{pmatrix}-1\\0\\1\end{pmatrix} \rangle \tag{12}$$

is a basis for the space.

The next result verifies that the process used in those examples works with any basis for any subspace of an $\mathbb{R}^n$ (we are restricted to $\mathbb{R}^n$ only because we have not given a definition of orthogonality for other vector spaces).

If $\langle \vec{\beta}_1, \ldots \vec{\beta}_k \rangle$ is a basis for a subspace of $\mathbb{R}^n$ then, where

$$\begin{aligned}
\vec{\kappa}_1 &= \vec{\beta}_1\\
\vec{\kappa}_2 &= \vec{\beta}_2 - \text{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_2)\\
\vec{\kappa}_3 &= \vec{\beta}_3 - \text{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_3) - \text{proj}_{[\vec{\kappa}_2]}(\vec{\beta}_3)\\
&\vdots\\
\vec{\kappa}_k &= \vec{\beta}_k - \text{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_k) - \cdots - \text{proj}_{[\vec{\kappa}_{k-1}]}(\vec{\beta}_k)
\end{aligned}$$

the $\vec{\kappa}$'s form an orthogonal basis for the same subspace.

Beyond having the vectors in the basis be orthogonal, we can do more; we can arrange for each vector to have length one by dividing each by its own length (we can normalize the lengths).

Exercise 10

Find an orthonormal basis for this subspace of $\mathbb{R}^4$.

$$\{ \begin{pmatrix}x\\y\\z\\w\end{pmatrix} \mid x - y - z + w = 0 \text{ and } x + z = 0\} \tag{13}$$

When using octave to do a Gram-Schmidt ortonormalization, the results are not always as expected, because of the different methods implemented in the program:

```
octave:24> X=[1,1,1;0,1,1;0,0,1]
X =

  1  1  1
  0  1  1
  0  0  1

octave:25> Q=orth(X)
Q =

  -0.73698  -0.59101  -0.32799
  -0.59101   0.32799   0.73698
  -0.32799   0.73698  -0.59101
```

7.3 Projection Into a Subspace

The prior subsections project a vector into a line by decomposing it into two parts: the part in the line $\text{proj}_{[\vec{s}\,]}(\vec{v}\,)$ and the rest $\vec{v} - \text{proj}_{[\vec{s}\,]}(\vec{v}\,)$. To generalize projection to arbitrary subspaces, we follow this idea.

For any direct sum $V = M \oplus N$ and any $\vec{v} \in V$, the projection of $\vec{v}$ into M along N is

$$\text{proj}_{M,N}(\vec{v}\,) = \vec{m} \tag{14}$$

where $\vec{v} = \vec{m} + \vec{n}$ with $\vec{m} \in M$, $\vec{n} \in N$.

This definition doesn't involve a sense of 'orthogonal' so we can apply it to spaces other than subspaces of an $\mathbb{R}^n$. (Definitions of orthogonality for other spaces are perfectly possible, but we haven't seen any in this book.)

Exercise 11

The space $M_{2\times2}$ of 2×2 matrices is the direct sum of these two.

$$M = \{\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \mid a, b \in \mathbb{R}\} \qquad N = \{\begin{pmatrix} 0 & 0 \\ c & d \end{pmatrix} \mid c, d \in \mathbb{R}\} \tag{15}$$

To project

$$A = \begin{pmatrix} 3 & 1 \\ 0 & 4 \end{pmatrix} \tag{16}$$

into M along N, we first fix bases for the two subspaces.

$$B_M = \langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \rangle \qquad B_N = \langle \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \rangle \tag{17}$$

The concatenation of these

$$B = B_M{}^\frown B_N = \langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \rangle \tag{18}$$

is a basis for the entire space, because the space is the direct sum, so we can use it to represent A.

$$\begin{pmatrix} 3 & 1 \\ 0 & 4 \end{pmatrix} = 3 \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + 4 \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{19}$$

Now the projection of A into M along N is found by keeping the M part of this sum and dropping the N part.

$$\text{proj}_{M,N}(\begin{pmatrix} 3 & 1 \\ 0 & 4 \end{pmatrix}) = 3 \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 0 & 0 \end{pmatrix} \tag{20}$$

Exercise 12

Both subscripts on $\text{proj}_{M,N}(\vec{v}\,)$ are significant. The first subscript M matters because the result of the projection is an $\vec{m} \in M$, and changing this subspace would change the possible results. For an example showing that the second subscript matters, fix this plane subspace of $\mathbb{R}^3$ and its basis

$$M = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \bigm| y - 2z = 0\} \qquad B_M = \langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \rangle \tag{21}$$

and compare the projections along two different subspaces.

$$N = \{k \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \bigm| k \in \mathbb{R}\} \qquad \hat{N} = \{k \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} \bigm| k \in \mathbb{R}\} \tag{22}$$

(Verification that $\mathbb{R}^3 = M \oplus N$ and $\mathbb{R}^3 = M \oplus \hat{N}$ is routine.) We will check that these projections are different by checking that they have different effects on this vector.

$$\vec{v} = \begin{pmatrix} 2 \\ 2 \\ 5 \end{pmatrix} \tag{23}$$

For the first one we find a basis for N

$$B_N = \langle \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rangle \tag{24}$$

and represent $\vec{v}$ with respect to the concatenation $B_M {}^\frown B_N$.

$$\begin{pmatrix} 2 \\ 2 \\ 5 \end{pmatrix} = 2 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} + 4 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{25}$$

The projection of $\vec{v}$ into M along N is found by keeping the M part and dropping the N part.

$$\text{proj}_{M,N}(\vec{v}\,) = 2 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} \tag{26}$$

For the other subspace $\hat{N}$, this basis is natural.

$$B_{\hat{N}} = \langle \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} \rangle \tag{27}$$

Representing $\vec{v}$ with respect to the concatenation

$$\begin{pmatrix}2\\2\\5\end{pmatrix} = 2\cdot\begin{pmatrix}1\\0\\0\end{pmatrix} + (9/5)\cdot\begin{pmatrix}0\\2\\1\end{pmatrix} - (8/5)\cdot\begin{pmatrix}0\\1\\-2\end{pmatrix} \tag{28}$$

and then keeping only the M part gives this.

$$\text{proj}_{M,\hat{N}}(\vec{v}\,) = 2\cdot\begin{pmatrix}1\\0\\0\end{pmatrix} + (9/5)\cdot\begin{pmatrix}0\\2\\1\end{pmatrix} = \begin{pmatrix}2\\18/5\\9/5\end{pmatrix} \tag{29}$$

Therefore projection along different subspaces may yield different results.

Example

The members $\vec{\beta}_1$ and $\vec{\beta}_2$ of this basis for $\mathbb{R}^2$ are not orthogonal.

$$B = \langle\begin{pmatrix}4\\2\end{pmatrix},\begin{pmatrix}1\\3\end{pmatrix}\rangle$$

Can we derive from B a new basis for the same space that does have mutually orthogonal members?

For the first member of the new basis we simply use $\vec{\beta}_1$.

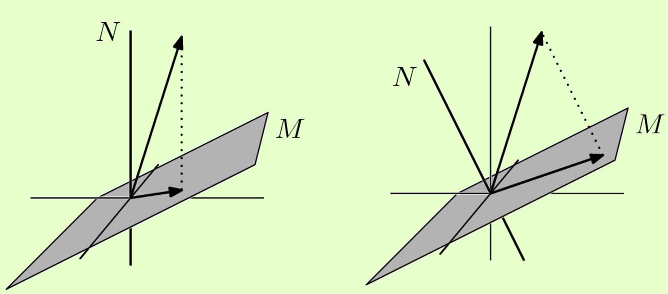
$$\vec{\kappa}_1 = \begin{pmatrix}4\\2\end{pmatrix}$$

For the second member of the new basis, we take away from $\vec{\beta}_2$ its part in the direction of $\vec{\kappa}_1$,

$$\vec{\kappa}_2 = \begin{pmatrix}1\\3\end{pmatrix} - \text{proj}_{[\vec{\kappa}_1]}(\begin{pmatrix}1\\3\end{pmatrix}) = \begin{pmatrix}1\\3\end{pmatrix} - \begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}-1\\2\end{pmatrix}$$

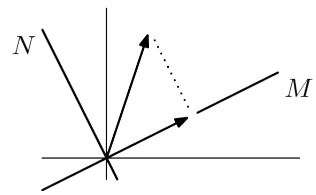
which leaves the part, $\vec{\kappa}_2$ pictured above, of $\vec{\beta}_2$ that is orthogonal to $\vec{\kappa}_1$ (it is orthogonal by the definition of the projection into the span of $\vec{\kappa}_1$). Note that $\{\vec{\kappa}_1, \vec{\kappa}_2\}$ is a basis for $\mathbb{R}^2$.

These pictures compare the two maps. Both show that the projection is indeed 'into' the plane and 'along' the line.



Notice that the projection along N is not orthogonal -there are members of the plane M that are not orthogonal to the dotted line. But the projection along $\hat{N}$ is orthogonal.

A natural question is: what is the relationship between the projection operation defined above, and the operation of orthogonal projection into a line? The second picture above suggests the answer -orthogonal projection into a line is a special case of the projection defined above; it is just projection along a subspace perpendicular to the line.



In addition to pointing out that projection along a subspace is a generalization, this scheme shows how to define orthogonal projection into any subspace of $\mathbb{R}^n$, of any dimension.

The orthogonal complement of a subspace M of $\mathbb{R}^n$ is

$$M^{\perp} = \{\vec{v} \in \mathbb{R}^n \mid \vec{v} \text{ is perpendicular to all vectors in } M\} \tag{30}$$

(read "M perp"). The orthogonal projection $\text{proj}_M(\vec{v})$ of a vector is its projection into M along $M^{\perp}$.

Exercise 13

In $\mathbb{R}^3$, to find the orthogonal complement of the plane

$$P = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid 3x + 2y - z = 0\} \tag{31}$$

we start with a basis for P.

$$B = \langle \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \rangle \tag{32}$$

Any $\vec{v}$ perpendicular to every vector in B is perpendicular to every vector in the span of B. Therefore, the subspace $P^{\perp}$ consists of the vectors that satisfy these two conditions.

$$\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \qquad \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \tag{33}$$

We can express those conditions more compactly as a linear system.

$$P^{\perp} = \{\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \mid \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\} \tag{34}$$

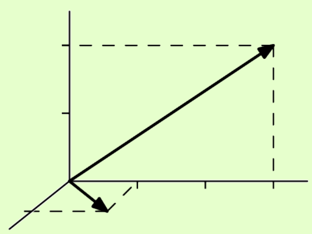
We are thus left with finding the nullspace of the map represented by the matrix, that is, with calculating the solution set of a homogeneous linear system.

$$P^{\perp} = \{ \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \mid \begin{matrix} v_1 & & +3v_3 = 0 \\ & v_2 & +2v_3 = 0 \end{matrix} \} = \{ k \begin{pmatrix} -3 \\ -2 \\ 1 \end{pmatrix} \mid k \in \mathbb{R} \} \quad (35)$$

Exercise 14

Where M is the xy-plane subspace of $\mathbb{R}^3$, what is $M^{\perp}$? A common first reaction is that $M^{\perp}$ is the yz-plane, but that's not right. Some vectors from the yz-plane are not perpendicular to every vector in the xy-plane.

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \not\perp \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix}$$



$$\theta = \arccos\left(\frac{1 \cdot 0 + 1 \cdot 3 + 0 \cdot 2}{\sqrt{2} \cdot \sqrt{13}}\right) \approx 0.94 \text{ rad}$$

Instead $M^{\perp}$ is the z-axis, since proceeding as in the prior example and taking the natural basis for the xy-plane gives this.

$$M^{\perp} = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \} = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x = 0 \text{ and } y = 0 \} \quad (36)$$

The next result justifies the second sentence.

Let M be a subspace of $\mathbb{R}^n$. The orthogonal complement of M is also a subspace. The space is the direct sum of the two $\mathbb{R}^n = M \oplus M^{\perp}$. And, for any $\vec{v} \in \mathbb{R}^n$, the vector $\vec{v} - \text{proj}_M(\vec{v})$ is perpendicular to every vector in M.

Let $\vec{v}$ be a vector in $\mathbb{R}^n$ and let M be a subspace of $\mathbb{R}^n$ with basis $\langle \vec{\beta}_1, \ldots, \vec{\beta}_k \rangle$. If A is the matrix whose columns are the $\vec{\beta}$'s then $\text{proj}_M(\vec{v}) = c_1\vec{\beta}_1 + \cdots + c_k\vec{\beta}_k$ where the coefficients c_i are the entries of the vector $(A^{\mathsf{T}}A)A^{\mathsf{T}} \cdot \vec{v}$. That is, $\text{proj}_M(\vec{v}) = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} \cdot \vec{v}$.

Exercise 15

To orthogonally project this vector into this subspace

$$\vec{v} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \qquad P = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x + z = 0 \} \quad (37)$$

first make a matrix whose columns are a basis for the subspace

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{38}$$

and then compute.

$$\begin{aligned} A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1/2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & 0 & -1/2 \\ 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \end{pmatrix} \end{aligned}$$

With the matrix, calculating the orthogonal projection of any vector into P is easy.

$$\text{proj}_P(\vec{v}) = \begin{pmatrix} 1/2 & 0 & -1/2 \\ 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} \tag{39}$$

```
octave:1> A=[0,1;1,0;0,-1]
A =

   0   1
   1   0
   0  -1

octave:2> A*inv(A'*A)*A'
ans =

   0.50000   0.00000  -0.50000
   0.00000   1.00000   0.00000
  -0.50000   0.00000   0.50000

octave:3> v=[1,-1,1]
v =

   1  -1   1

octave:4> A*inv(A'*A)*A'*v'
ans =

   0
  -1
   0
```

Example

Orthogonally project the vector $\binom{2}{3}$ into the line $y = 2x$. We first pick a direction vector for the line. For instance,

$$\vec{s} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

will do. Then the calculation is routine.

$$\frac{\begin{pmatrix} 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}}{\begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{8}{5} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 8/5 \\ 16/5 \end{pmatrix}$$

Exercise 16

Get the best linear fit to this data by means of the above procedure.

Table 1: Cancer Deaths in Oregon. Source: R. Fadeley, Journal of Environmental Health 27 (1965), pp. 883-897.

County/city	Index	Deaths
Umatilla	2.5	147
Morrow	2.6	130
Gilliam	3.4	130
Sherman	1.3	114
Wasco	1.6	138
Hood River	3.8	162
Portland	11.6	208
Columbia	6.4	178
Clatsop	8.3	210

8 Diagonalization

8.1 Diagonalizability

A transformation is diagonalizable if it has a diagonal representation with respect to the same basis for the codomain as for the domain. A diagonalizable matrix is one that is similar to a diagonal matrix: T is diagonalizable if there is a nonsingular P such that PTP^{-1} is diagonal.

Exercise 17

Exercise 18

Not every matrix is diagonalizable. The square of

$$N = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{42}$$

is the zero matrix. Thus, for any map n that N represents (with respect to the same basis for the domain as for the codomain), the composition $n \circ n$ is the zero map. This implies that no such map n can be diagonally represented (with respect to any B, B) because no power of a nonzero diagonal matrix is zero. That is, there is no diagonal matrix in N's similarity class.

That exercise shows that a diagonal form will not do for a canonical form -we cannot find a diagonal matrix in each matrix similarity class. However, the canonical form that we are developing has the property that if a matrix can be diagonalized then the diagonal matrix is the canonical representative of the similarity class. The next result characterizes which maps can be diagonalized.

A transformation t is diagonalizable if and only if there is a basis $B = \langle \vec{\beta}_1, \ldots, \vec{\beta}_n \rangle$ and scalars $\lambda_1, \ldots, \lambda_n$ such that $t(\vec{\beta}_i) = \lambda_i \vec{\beta}_i$ for each i.

Exercise 19

To diagonalize

$$T = \begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} \tag{43}$$

we take it as the representation of a transformation with respect to the standard basis $T = \text{Rep}_{\mathcal{E}_2,\mathcal{E}_2}(t)$ and we look for a basis $B = \langle \vec{\beta}_1, \vec{\beta}_2 \rangle$ such that

$$\text{Rep}_{B,B}(t) = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \tag{44}$$

that is, such that $t(\vec{\beta}_1) = \lambda_1 \vec{\beta}_1$ and $t(\vec{\beta}_2) = \lambda_2 \vec{\beta}_2$.

$$\begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} \vec{\beta}_1 = \lambda_1 \cdot \vec{\beta}_1 \qquad \begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} \vec{\beta}_2 = \lambda_2 \cdot \vec{\beta}_2 \tag{45}$$

We are looking for scalars x such that this equation

$$\begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = x \cdot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \tag{46}$$

has solutions b_1 and b_2, which are not both zero. Rewrite that as a linear system.

$$\begin{aligned} (3-x)\cdot b_1 + \quad 2\cdot b_2 &= 0 \\ (1-x)\cdot b_2 &= 0 \end{aligned} \qquad (*)$$

In the bottom equation the two numbers multiply to give zero only if at least one of them is zero so there are two possibilities, $b_2 = 0$ and $x = 1$. In the $b_2 = 0$ possibility, the first equation gives that either $b_1 = 0$ or $x = 3$. Since the case of both $b_1 = 0$ and $b_2 = 0$ is disallowed, we are left looking at the possibility of $x = 3$. With it, the first equation in $(*)$ is $0 \cdot b_1 + 2 \cdot b_2 = 0$ and so associated with 3 are vectors with a second component of zero and a first component that is free.

$$\begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ 0 \end{pmatrix} = 3 \cdot \begin{pmatrix} b_1 \\ 0 \end{pmatrix} \qquad (47)$$

That is, one solution to $(*)$ is $\lambda_1 = 3$, and we have a first basis vector.

$$\vec{\beta}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad (48)$$

In the $x = 1$ possibility, the first equation in $(*)$ is $2 \cdot b_1 + 2 \cdot b_2 = 0$, and so associated with 1 are vectors whose second component is the negative of their first component.

$$\begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ -b_1 \end{pmatrix} = 1 \cdot \begin{pmatrix} b_1 \\ -b_1 \end{pmatrix} \qquad (49)$$

Thus, another solution is $\lambda_2 = 1$ and a second basis vector is this.

$$\vec{\beta}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \qquad (50)$$

To finish, drawing the similarity diagram

$$\begin{array}{ccc} \mathbb{R}^2_{\text{w.r.t. } \mathcal{E}_2} & \xrightarrow[T]{t} & \mathbb{R}^2_{\text{w.r.t. } \mathcal{E}_2} \\ \text{id}\big\downarrow & & \text{id}\big\downarrow \\ \mathbb{R}^2_{\text{w.r.t. } B} & \xrightarrow[D]{t} & \mathbb{R}^2_{\text{w.r.t. } B} \end{array} \qquad (51)$$

and noting that the matrix $\text{Rep}_{B,\mathcal{E}_2}(\text{id})$ is easy leads to this diagonalization.

$$\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \qquad (52)$$

8.2 Eigenvalues and Eigenvectors

A transformation $t\colon V \to V$ has a scalar eigenvalue λ if there is a nonzero eigenvector $\vec{\zeta} \in V$ such that $t(\vec{\zeta}) = \lambda \cdot \vec{\zeta}$.

Exercise 20
The projection map

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \stackrel{\pi}{\longmapsto} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \qquad x, y, z \in \mathbb{C} \tag{53}$$

has an eigenvalue of 1 associated with any eigenvector of the form

$$\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \tag{54}$$

where x and y are non-0 scalars. On the other hand, 2 is not an eigenvalue of π since no non-$\vec{0}$ vector is doubled.

That example shows why the 'non-$\vec{0}$' appears in the definition. Disallowing $\vec{0}$ as an eigenvector eliminates trivial eigenvalues.

Exercise 21
The only transformation on the trivial space $\{\vec{0}\,\}$ is $\vec{0} \mapsto \vec{0}$. This map has no eigenvalues because there are no non-$\vec{0}$ vectors $\vec{v}$ mapped to a scalar multiple $\lambda \cdot \vec{v}$ of themselves.

Exercise 22
Consider the homomorphism $t\colon \mathcal{P}_1 \to \mathcal{P}_1$ given by $c_0 + c_1 x \mapsto (c_0 + c_1) + (c_0 + c_1)x$. The range of t is one-dimensional. Thus an application of t to a vector in the range will simply rescale that vector: $c + cx \mapsto (2c) + (2c)x$. That is, t has an eigenvalue of 2 associated with eigenvectors of the form $c + cx$ where $c \neq 0$.

This map also has an eigenvalue of 0 associated with eigenvectors of the form $c - cx$ where $c \neq 0$.

A square matrix T has a scalar eigenvalue λ associated with the non-$\vec{0}$ eigenvector $\vec{\zeta}$ if $T\vec{\zeta} = \lambda \cdot \vec{\zeta}$.

Although this extension from maps to matrices is obvious, there is a point that must be made. Eigenvalues of a map are also the eigenvalues of matrices representing that map, and so similar matrices have the same eigenvalues. But the eigenvectors are different -similar matrices need not have the same eigenvectors.

For instance, consider again the transformation $t\colon \mathcal{P}_1 \to \mathcal{P}_1$ given by $c_0 + c_1 x \mapsto (c_0 + c_1) + (c_0 + c_1)x$. It has an eigenvalue of 2 associated with eigenvectors of the form $c + cx$ where $c \neq 0$. If we represent t with respect to $B = \langle 1 + 1x, 1 - 1x \rangle$

$$T = \text{Rep}_{B,B}(t) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \tag{55}$$

then 2 is an eigenvalue of T, associated with these eigenvectors.

$$\{\begin{pmatrix} c_0 \\ c_1 \end{pmatrix} \mid \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 2c_0 \\ 2c_1 \end{pmatrix}\} = \{\begin{pmatrix} c_0 \\ 0 \end{pmatrix} \mid c_0 \in \mathbb{C},\, c_0 \neq 0\} \tag{56}$$

On the other hand, representing t with respect to $D = \langle 2 + 1x, 1 + 0x \rangle$ gives

$$S = \text{Rep}_{D,D}(t) = \begin{pmatrix} 3 & 1 \\ -3 & -1 \end{pmatrix} \tag{57}$$

and the eigenvectors of S associated with the eigenvalue 2 are these.

$$\{\begin{pmatrix} c_0 \\ c_1 \end{pmatrix} \mid \begin{pmatrix} 3 & 1 \\ -3 & -1 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 2c_0 \\ 2c_1 \end{pmatrix}\} = \{\begin{pmatrix} 0 \\ c_1 \end{pmatrix} \mid c_1 \in \mathbb{C},\, c_1 \neq 0\} \tag{58}$$

Thus similar matrices can have different eigenvectors.

Here is an informal description of what's happening. The underlying transformation doubles the eigenvectors $\vec{v} \mapsto 2 \cdot \vec{v}$. But when the matrix representing the transformation is $T = \text{Rep}_{B,B}(t)$ then it "assumes" that column vectors are representations with respect to B. In contrast, $S = \text{Rep}_{D,D}(t)$ "assumes" that column vectors are representations with respect to D. So the vectors that get doubled by each matrix look different.

The next example illustrates the basic tool for finding eigenvectors and eigenvalues.

Exercise 23

If

$$S = \begin{pmatrix} \pi & 1 \\ 0 & 3 \end{pmatrix} \tag{59}$$

(here π is not a projection map, it is the number $3.14\ldots$) then

$$\left| \begin{pmatrix} \pi - x & 1 \\ 0 & 3 - x \end{pmatrix} \right| = (x - \pi)(x - 3) \tag{60}$$

so S has eigenvalues of $\lambda_1 = \pi$ and $\lambda_2 = 3$. To find associated eigenvectors, first plug in λ_1 for x:

$$\begin{pmatrix} \pi - \pi & 1 \\ 0 & 3 - \pi \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Longrightarrow \quad \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix} \tag{61}$$

for a scalar $a \neq 0$, and then plug in λ_2:

$$\begin{pmatrix} \pi - 3 & 1 \\ 0 & 3 - 3 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Longrightarrow \quad \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} -b/\pi - 3 \\ b \end{pmatrix} \tag{62}$$

where $b \neq 0$.

The characteristic polynomial of a square matrix T is the determinant of the matrix $T - xI$, where x is a variable. The characteristic equation is $|T - xI| = 0$. The characteristic polynomial of a transformation t is the polynomial of any $\text{Rep}_{B,B}(t)$.

A linear transformation on a nontrivial vector space has at least one eigenvalue. Notice the familiar form of the sets of eigenvectors in the above examples.

The eigenspace of a transformation t associated with the eigenvalue λ is $V_\lambda = \{\vec{\zeta} \mid t(\vec{\zeta}) = \lambda\vec{\zeta}\} \cup \{\vec{0}\}$. The eigenspace of a matrix is defined analogously. An eigenspace is a subspace.

Exercise 24

In the above example the eigenspace associated with the eigenvalue π and the eigenspace associated with the eigenvalue 3 are these.

$$V_\pi = \{\begin{pmatrix} a \\ 0 \end{pmatrix} \mid a \in \mathbb{R}\} \qquad V_3 = \{\begin{pmatrix} -b/\pi - 3 \\ b \end{pmatrix} \mid b \in \mathbb{R}\} \tag{63}$$

Exercise 25

In the above example, these are the eigenspaces associated with the eigenvalues 0 and 2.

$$V_0 = \{\begin{pmatrix} a \\ -a \\ a \end{pmatrix} \mid a \in \mathbb{R}\}, \qquad V_2 = \{\begin{pmatrix} b \\ 0 \\ b \end{pmatrix} \mid b \in \mathbb{R}\}. \tag{64}$$

The characteristic equation is $0 = x(x-2)^2$ so in some sense 2 is an eigenvalue "twice". However there are not "twice" as many eigenvectors, in that the dimension of the eigenspace is one, not two. The next example shows a case where a number, 1, is a double root of the characteristic equation and the dimension of the associated eigenspace is two.

Exercise 26

With respect to the standard bases, this matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{65}$$

represents projection.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \stackrel{\pi}{\longmapsto} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \qquad x, y, z \in \mathbb{C} \tag{66}$$

Its eigenspace associated with the eigenvalue 0 and its eigenspace associated with the eigenvalue 1 are easy to find.

$$V_0 = \{\begin{pmatrix} 0 \\ 0 \\ c_3 \end{pmatrix} \mid c_3 \in \mathbb{C}\} \qquad V_1 = \{\begin{pmatrix} c_1 \\ c_2 \\ 0 \end{pmatrix} \mid c_1, c_2 \in \mathbb{C}\} \tag{67}$$

If two eigenvectors $\vec{v}_1$ and $\vec{v}_2$ are associated with the same eigenvalue then any linear combination of those two is also an eigenvector associated with that same eigenvalue. But, if two eigenvectors $\vec{v}_1$ and $\vec{v}_2$ are associated with different eigenvalues then the sum $\vec{v}_1 + \vec{v}_2$ need not be related to the eigenvalue of either one. In fact, just the opposite. If the eigenvalues are different then the eigenvectors are not linearly related.

For any set of distinct eigenvalues of a map or matrix, a set of associated eigenvectors, one per eigenvalue, is linearly independent.

Exercise 27

The eigenvalues of

$$\begin{pmatrix} 2 & -2 & 2 \\ 0 & 1 & 1 \\ -4 & 8 & 3 \end{pmatrix} \tag{68}$$

are distinct: $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$. A set of associated eigenvectors like

$$\{\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 9 \\ 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}\} \tag{69}$$

is linearly independent.

An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Example

What are the eigenvalues and eigenvectors of this matrix?

$$T = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & -2 \\ -1 & 2 & 3 \end{pmatrix}$$

To find the scalars x such that $T\vec{\zeta} = x\vec{\zeta}$ for non-$\vec{0}$ eigenvectors $\vec{\zeta}$, bring everything to the left-hand side

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & -2 \\ -1 & 2 & 3 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} - x \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \vec{0}$$

and factor $(T - xI)\vec{\zeta} = \vec{0}$. (Note that it says $T - xI$; the expression $T - x$ doesn't make sense because T is a matrix while x is a scalar.) This homogeneous linear system

$$\begin{pmatrix} 1-x & 2 & 1 \\ 2 & 0-x & -2 \\ -1 & 2 & 3-x \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

has a non-$\vec{0}$ solution if and only if the matrix is singular. We can determine when that happens.

$$\begin{aligned} 0 &= |T - xI| \\ &= \begin{vmatrix} 1-x & 2 & 1 \\ 2 & 0-x & -2 \\ -1 & 2 & 3-x \end{vmatrix} \\ &= x^3 - 4x^2 + 4x \\ &= x(x-2)^2 \end{aligned}$$

The eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = 2$. To find the associated eigenvectors, plug in each eigenvalue. Plugging in $\lambda_1 = 0$ gives

$$\begin{pmatrix} 1-0 & 2 & 1 \\ 2 & 0-0 & -2 \\ -1 & 2 & 3-0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad \Longrightarrow \qquad \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} a \\ -a \\ a \end{pmatrix}$$

for a scalar parameter $a \neq 0$ (a is non-0 because eigenvectors must be non-$\vec{0}$). In the same way, plugging in $\lambda_2 = 2$ gives

$$\begin{pmatrix} 1-2 & 2 & 1 \\ 2 & 0-2 & -2 \\ -1 & 2 & 3-2 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad \Longrightarrow \qquad \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} b \\ 0 \\ b \end{pmatrix}$$

with $b \neq 0$.

When using `octave`, the diagonalization has this simple form:

```
octave:5> b=[1,2,1;2,0,-2;-1,2,3]
b =

   1   2   1
   2   0  -2
  -1   2   3

octave:6> [ev,eval]=eig(b)
ev =

    5.7735e-01   -7.0711e-01    7.0711e-01
   -5.7735e-01   -2.0000e-08   -2.0000e-08
    5.7735e-01   -7.0711e-01    7.0711e-01

eval =

  0.00000  0.00000  0.00000
  0.00000  2.00000  0.00000
  0.00000  0.00000  2.00000
```

9 Singular value decomposition (SVD) and principal component analysis (PCA)

- The **Singular Values** of the square matrix A are defined as the square root of the eigenvalues of $A^{\mathsf{T}}A$.
- The **Condition Number** is the ratio of the largest to the smallest singular value.
- A matrix is **Ill Conditioned Matrix** if the condition number is too large. How large the condition number can be, before the matrix is ill conditioned, is determined by the machine precision.
- A matrix is **Singular** if the condition number is infinite. The determinant of a singular matrix is 0.
- The **Rank** of a matrix is the dimension of the range of the matrix. This corresponds to the number of non-singular values for the matrix, i.e. the number of linear independent rows of the matrix.

9.1 Spectral decomposition of a square matrix

Any real symmetric $m \times m$ matrix A has a spectral decomposition of the form

$$A = U\Lambda U^{\mathsf{T}} \tag{70}$$

where U is an orthonormal matrix (matrix of orthogonal unit vectors: $U^{\mathsf{T}}U = I$ or $\sum_k u_{ki}u_{kj} = \delta_{ij}$ and Λ is a diagonal matrix. The columns of U are the eigenvectors of matrix A and the diagonal elements of Λ are the eigenvalues. If A is positive-definite, the eigenvalues will all be positive. Multiplying with U, equation 70 can be re-written to

$$AU = U\Lambda U^{\mathsf{T}}U = U\Lambda$$

This can be written as a normal eigenvalue equation by defining the i^{th} column of U as $\mathbf{u}_i$ and the eigenvalues as $\lambda_i = \Lambda_{ii}$:

$$A\mathbf{u}_i = \lambda_i \mathbf{u}i$$

9.2 Singular Value Decomposition

A real $n \times m$ matrix B, where $n \geq m$ has the decomposition,

$$B = U\Gamma V^{\mathsf{T}} \tag{71}$$

where U is a $n \times m$ matrix with orthonormal columns ($U^{\mathsf{T}}U = I$), while V is a $n \times n$ orthonormal matrix ($V^{\mathsf{T}}V = I$), and Γ is a $m \times m$ diagonal matrix with positive or zero elements, called the singular values. From B we can construct two positive-definite symmetric matrices, BB^{T} and $B^{\mathsf{T}}B$, each of which we can decompose

$$BB^{\mathsf{T}} = U\Gamma V^{\mathsf{T}}V\Gamma U^{\mathsf{T}} = U\Gamma^2 U^{\mathsf{T}}$$

$$B^{\mathsf{T}}B = V\Gamma^2 V^{\mathsf{T}}$$

Keep in mind that $n \geq m$. We can now show that BB^{T} which is $n \times n$ and $B^{\mathsf{T}}B$ which is $m \times m$ will share m eigenvalues and the remaining $n - m$ eigenvalues of BB^{T} will be zero.

Using the decomposition above, we can identify the eigenvectors and eigenvalues for $B^{\mathsf{T}}B$ as the columns of V and the squared diagonal elements of Γ, respectively. (The latter shows that the eigenvalues of $B^{\mathsf{T}}B$ must be non-negative). Denoting one such eigenvector by $\mathbf{v}$ and the diagonal element by γ, we have

$$B^{\mathsf{T}}B\mathbf{v} = \gamma^2\mathbf{v}$$

then we can multiply on both sides with B to get

$$BB^{\mathsf{T}}B\mathbf{v} = \gamma^2 B\mathbf{v}$$

But this means that we have an eigenvector $\mathbf{u} = B\mathbf{v}$ and eigenvalue γ^2 for BB^{T} as well, since

$$(BB^{\mathsf{T}})B\mathbf{v} = \gamma^2 B\mathbf{v}$$

We have now shown that BB^{T} and $B^{\mathsf{T}}B$ share m eigenvalues.

We still need to prove that the remaining $n \times m$ eigenvalues of BB^{T} are zero. To do that let us consider an eigenvector for BB^{T}, $\mathbf{u}_\perp = \beta_\perp \mathbf{u}_\perp$ which is orthogonal to the m eigenvectors $\mathbf{u}_i$ already determined, i.e. $U^{\mathsf{T}}\mathbf{u}_\perp = 0$. Using the decomposition $BB^{\mathsf{T}} = U\Gamma^2 U^{\mathsf{T}}$, we immediately see that the eigenvalues $\beta_\perp$ must all be zero

$$BB^{\mathsf{T}}\mathbf{u}_\perp = U\Gamma^2 U^{\mathsf{T}}\mathbf{u}_\perp = 0\mathbf{u}_\perp$$

The Rank R of BB^{T} is determined by the smallest dimension of B, ($R \leq m$). This ensures that BB^{T} has at most m eigenvalues larger than zero. Note that the relation for BB^{T} corresponds to the usual spectral decomposition since the "missing" $(n-m)$ eigenvalues are zero. It is then evident that the two square matrices can be interchanged. This is a property we can advantage of when dealing with data matrices where we have many more features than examples.

Summarizing, in equation 71, matrices U and V are such that they are orthogonal. The columns of U are called **left singular values** (gene coefficients) and the rows of V^{T} are called **right singular values** (expression level vectors). To calculate the matrices U and V, one must calculate the eigenvectors and eigenvalues of BB^{T} and $B^{\mathsf{T}}B$. These multiplications of B by its transpose results in a square matrix (the number of columns is equal to the number of rows).

The columns of V are made from the eigenvectors of $B^{\mathsf{T}}B$ and the columns of U are made from the eigenvectors of BB^{T}. The eigenvalues obtained from the products of BB^{T} and $B^{\mathsf{T}}B$, when square-rooted, make up the columns of Γ in equation 71. The diagonal of Γ is said to be the singular values of the original matrix, B.

9.3 Properties of a data matrix -first and second moments

Let $\mathbf{x}$ (with components x_j, $j = 1, \ldots, n$) be a stochastic vector with probability distribution $P(\mathbf{x})$. Let $\{\mathbf{x}^\alpha \mid \alpha = 1, \ldots, m\}$ be a sample from $P(x)$. We will choose a convention for the data matrix X, where the rows denote the features $j = 1, \ldots, n$ and the columns the samples $\alpha = 1, \ldots, m$: in other words the components are $X_{j,\alpha} = x_j^\alpha$. Principal component analysis is based on the two first empirical moments of the sample data matrix. The mean vector

$$\langle \mathbf{x} \rangle \equiv \frac{1}{m} \sum_{\alpha=1}^{m} \mathbf{x}^\alpha$$

and the empirical covariance matrix,

$$C \equiv \frac{1}{m}\sum_{\alpha=1}^{m}(\mathbf{x}^{\alpha}-\langle\mathbf{x}\rangle)(\mathbf{x}^{\alpha}-\langle\mathbf{x}\rangle)^{\mathsf{T}}$$

Using the matrix formulation we can write

$$C \equiv \frac{1}{m}XX^{\mathsf{T}}$$

where we have removed the mean of the data:

$$X_{j,\alpha} := X_{j,\alpha} - \langle x_j\rangle$$

9.4 Principal component analysis (PCA)

In principal component analysis we find the directions in the data with the most variation, i.e. the eigenvectors corresponding to the largest eigenvalues of the covariance matrix, and project the data onto these directions. The motivation for doing this is that the most second order information are in these directions[3]. Each eigenvector described in section 9.2 represents a principle component, PC1 (Principle Component 1), which is defined as the eigenvector with the highest corresponding eigenvalue. The individual eigenvalues are numerically related to the variance they capture via PC's - the higher the value, the more variance they have captured. The choice of the number of directions are often guided by trial and error, but principled methods also exist. If we denote the matrix of eigenvectors sorted according to eigenvalue by $\tilde{U}$, we can then PCA transformation of the data as $Y = \tilde{U}^{\mathsf{T}}X$. The eigenvectors are called the principal components. By selecting only the first d rows of Y, we have projected the data from n down to d dimensions.

9.5 PCA by SVD

We can use SVD to perform PCA. We decompose X using SVD, i.e.

$$X = U\Gamma V^{\mathsf{T}}$$

and find that we can write the covariance matrix as

$$C = \frac{1}{n}XX^{\mathsf{T}} = \frac{1}{n}U\Gamma^2U^{\mathsf{T}}$$

In this case U is a $n\times m$ matrix. Following from the fact that SVD routine order the singular values in descending order we know that, if $n < m$, the first n columns in U corresponds to the sorted eigenvalues of C and if $m \geq n$, the first m corresponds to the sorted non-zero eigenvalues of C. The transformed data can thus be written as

$$Y = \tilde{U}^{\mathsf{T}}X = \tilde{U}^{\mathsf{T}}U\Gamma V^{\mathsf{T}}$$

where $\tilde{U}^{\mathsf{T}}U$ is a simple $n\times m$ matrix which is one on the diagonal and zero everywhere else. To conclude, we can write the transformed data in terms of the SVD decomposition of X.

[3]This also means we might discard important non-second order information by PCA

9.6 PCA by SVD in Octave

It is common in gene expression analysis or QSAR studies that we have many more features than samples, $n \ll m$. The covariance matrix itself is therefore very unpleasant to work with because it is very large and as we have proved above singular. However, using the relations above, we find that is suffices to decompose the smaller $m \times m$ matrix

$$D \equiv \frac{1}{m} X^{\mathsf{T}} X$$

Given a decomposition of D we can find the interesting non-zero principal directions and components for C, $U = XVS^{-1}$. You can instruct octave to always use the smallest matrix by using the command `[u s v] = svd(X,0)`, see also 'help svd' in octave. However, in that case we have to be careful about which matrices to use for the transformation.

Example

From the following matrix a we can obtain the eigenvalues and eigenvectors of the square of the matrix by:

```
octave:1> a=[2,4;1,3;0,0;0,0]
a =

  2  4
  1  3
  0  0
  0  0

octave:2> a*a'
ans =

  20  14   0   0
  14  10   0   0
   0   0   0   0
   0   0   0   0

octave:3> [vect,val]=eig( a*a')
vect =

  -0.00000   0.00000   0.57605  -0.81742
  -0.00000   0.00000  -0.81742  -0.57605
   1.00000   0.00000   0.00000  -0.00000
   0.00000   1.00000   0.00000  -0.00000

val =

   0.00000   0.00000   0.00000   0.00000
   0.00000   0.00000   0.00000   0.00000
   0.00000   0.00000   0.13393   0.00000
   0.00000   0.00000   0.00000  29.86607
```

which will lead to the building of the SVD for that matrix a. Now, we can do a similar job by simply performing a regular SVD in octave:

```
octave:4> [u s v] = svd(a)
u =

  -0.81742  -0.57605   0.00000   0.00000
  -0.57605   0.81742   0.00000   0.00000
   0.00000   0.00000   1.00000   0.00000
   0.00000   0.00000   0.00000   1.00000

s =

  5.46499  0.00000
  0.00000  0.36597
  0.00000  0.00000
  0.00000  0.00000

v =

  -0.40455  -0.91451
  -0.91451   0.40455
```

Exercise Using the gene expression data in the file FA.zip, perform an SVD and find the PC's.

9.7 More samples than variables

In some cases, the number of variables is smaller than the number of examples ($n < m$). In these cases, decomposition and dimension reduction might still be desirable for the $n \times m$ matrix X. Dimension change on X however also results in dimension change on U, Γ and V, who respectively get the sizes $n{\times}n$, $n{\times}m$ and $m{\times}m$. The dimension changes the svd routine in octave slow and adds unnecessary rows to the V matrix. The problem can be avoided using `[V; S; U] = svd(X'; 0); U = U'; V = V';`, in the cases where $n < m$.

9.8 Number of Principal Directions

The no of principal components to use d, is not always easy to determine. The energy fraction could be used to argue for the usage of a given number of principal components. The number of components could also be determined from the characteristics of the singular values. When the singular values stabilize, the remaining components is usually contaminated with much noise and therefore not useful.

9.9 Similar Methods for Dimensionality Reduction

There exists multiple methods that can be used for dimensionality reduction. Some of them are given in the list below.

- Singular Value Decomposition (SVD)i
- Independent Component Analysis (ICA)
- Non-negative Matrix Factorization (NMF)
- Eigen Decomposition
- Random Projection
- Factor Analysis (FA)

References

[1] JJ Faraway. Practical Regression and Anova using R. 2002. *URL http://cran. r-project. org/doc/contrib/Faraway-PRA. pdf*, 2002.

[2] A. Isaev. *Introduction to mathematical methods in bioinformatics.* Springer Verlag, 2004.

[3] J. Hefferon. *Linear Algebra. web edition, 2008.*