Overview of Linear Algebra 3

by **Fabrizio Donzelli**, with contributions from **Patrick Boily**

3.1 Vector Spaces 147
Practical Definition 147
Linear Combinations 149
Bases and Dimension 150
Vector Subspaces 151
Spanning Sets 153
Dot Product 153
Cross Product in $\mathbb{R}^3$ 154
3.2 Linear Transformations . . 155
3.3 Matrix Algebra 157
Matrix Operations 157
Square Matrices 159
Determinants 160
3.4 Linear Systems 162
Gauss-Jordan Elimination . 165
Linear Systems & Matrices 167
3.5 Matrix Diagonalization . . 168
Eigenvalues & Eigenvectors 168
Similar Matrices 174
Diagonalization 175
3.6 Exercises 177
Chapter References 180

This chapter contains an essential introduction to linear algebra. The goal is to provide the readers interested in statistics and/or data science with some basic mathematical tools that are at the base of the algorithms and the mathematical models of statistical analysis. Theoretical details, such as rigorous proofs and definitions, will be kept at a minimal level.

A more detailed introduction to linear algebra can be found in [2, 3, 4].

3.1 Vector Spaces

At its most fundamental level, linear algebra deals with **vector spaces** and **linear transformation** between these.

Linear transformation are represented by **matrices**; a good portion of this chapter will be therefore dedicated to matrix algebra.[1]

3.1.1 Practical Definition

While there is a formal definition of vector spaces (see [3], for instance), we will eschew it in these notes. Instead, we use a "recipe" that contains all that we will need.

In the context of linear algebra, the set $\mathbb{R}^n$ is the **n-dimensional vector space**, consisting of **n-dimensional vectors**.[2]

Here are the key defining properties of these vectors:

- a n-dimensional vector $\mathbf{v}$ is a collection of n numbers: $\mathbf{v} = \langle v_1, \cdots, v_n \rangle$, where the numbers v_k are the **components** of the vector;[3]
- vectors belonging to the same vector space can be added, while remaining a part of that vector space: the **vector sum** of $\mathbf{v} = \langle v_1, v_2, \cdots, v_n \rangle$ and $\mathbf{w} = \langle w_1, w_2, \cdots, w_n \rangle$, is

$$\mathbf{v} + \mathbf{w} = \langle v_1 + w_1, v_2 + w_2, \cdots, v_n + w_n \rangle;$$

- in vector algebra, simple numbers are **scalars** – the **multiplication of a vector by a scalar** is defined in the "obvious way": if c is a scalar, and $\mathbf{v} = \langle v_1, v_2, \cdots, v_n \rangle$ is a vector, then

$$c\mathbf{v} = \langle cv_1, cv_2, \cdots, cv_n \rangle;$$

- the zero n-dimensional vector is denoted by $\mathbf{0} = \langle 0, 0, \cdots, 0 \rangle$.

1: Note that the order in which the material covered by a first year university linear algebra course could be different than the order presented here – it is common for texts of this nature to start with linear systems before moving to vector spaces; this is not how we will approach the presentation, in no small part because the language of vectors is very useful, not only in mathematics, but also in coding. A mastery of this language makes mathematical modeling more accessible, in general.

2: This definition is not ideal since it implicitly assumes that the vector are expressed with respect to the **standard basis**

$$\begin{aligned}\mathbf{e}_1 &= \langle 1, 0, \cdots, 0 \rangle, \\ \mathbf{e}_2 &= \langle 0, 1, \cdots, 0 \rangle, \\ &\vdots \\ \mathbf{e}_n &= \langle 0, 0, \cdots, 1 \rangle.\end{aligned}$$

3: In the other chapters, we will use $(v_1, \cdots, v_n)$ when the context is clear.

Example An aircraft is flying from Ottawa to Milan. The direction and its speed are determined by three values that change over time: latitude $x(t)$, longitude $y(t)$, and altitude $z(t)$. Hence, the velocity of the aircraft is modeled using a 3-dimensional vector $\mathbf{v}(t) = \langle x'(t), y'(t), z'(t)\rangle \in \mathbb{R}^3$.

Note however that the 3 quantities $x(t)$, $y(t)$, and $z(t)$ are not truly Cartesian in nature, since longitude and latitude are described by angles. Locally, however,[4] this $\mathbb{R}^3$ model is a good approximation, assuming that the Earth is **locally flat**.

4: That is to say, as long as we do not look at long distance trajectories, say.

Example A boat is sailing in the Pacific Ocean with a velocity vector $\mathbf{v} = \langle 1, 2\rangle$. At some point the wind starts blowing with speed $\mathbf{w} = \langle 2, 4\rangle$, helping the boat to sail faster. What is the estimate of the effective velocity of the boat under the influence of the wind?

We need to add the vectors. Luckily for us, velocities add **linearly**, hence the velocity of the wind-boosted boat is

$$\mathbf{v}_{\text{tot}} = \mathbf{v} + \mathbf{w} = \langle 1, 2\rangle + \langle 2, 4\rangle = \langle 3, 6\rangle.$$

The result is only an approximation of the real situation, since in reality there are dissipation effects that may reduce the speed of the boat.[5] ■

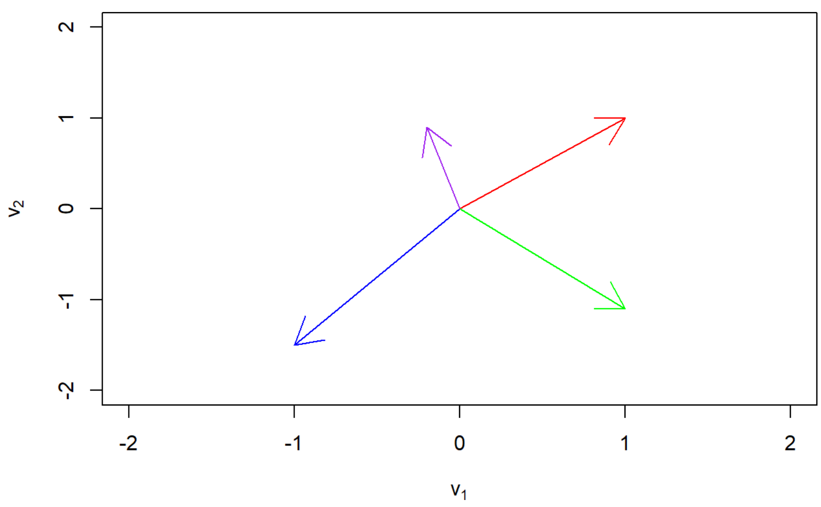
5: But that is a problem for engineers, really, and we will sidestep the challenge simply by ignoring it.

While vectors can be of arbitrary dimension, having a low-dimensional geometric picture helps strengthen vector intuition, which may be otherwise sound too abstract. In practice, vectors in $\mathbb{R}^2$, $\mathbb{R}^3$ are represented by arrows, emanating from the same origin point.

Example Here is an example of a representation of 2-dimensional vectors, which include a basic R script that produces the picture.[6]

6: Which can be improved, see Chapter 18 and [1].

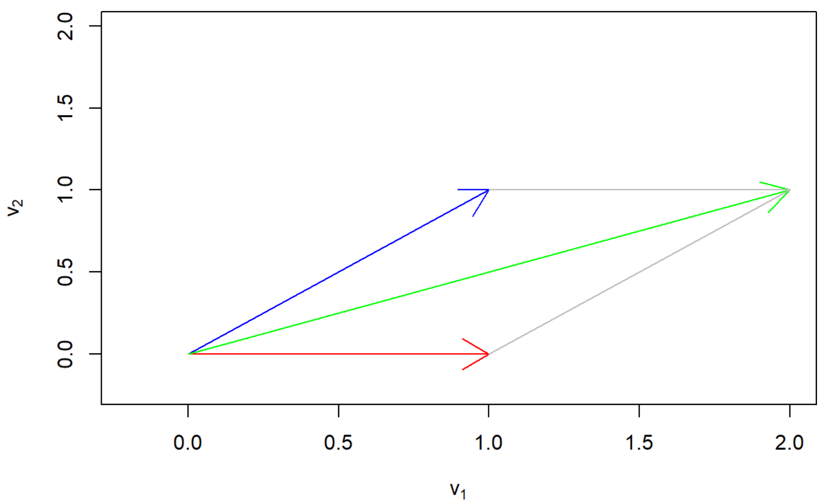
```
plot(NA,xlim=c(-2,2), ylim=c(-2,2),
    xlab = expression(list(v[1])),
    ylab=expression(list(v[2])))
arrows(0,0,1,1, col="red"); arrows(0,0,-1,-1.5, col="blue")
arrows(0,0,1,-1.1, col="green"); arrows(0,0,-0.2,0.9, col="purple")
```



In principle, arrows exist in arbitrary dimensions, but they are difficult to visualize. As we can always represent a vector as an arrow, the next rule applies no matter the dimension n.

Parallelogram rule: the sum of two vectors $\mathbf{v}$ and $\mathbf{w}$ is the diagonal of the parallelogram generated by $\mathbf{v}$ and $\mathbf{w}$, emanating from the origin.

```
plot(NA,xlim=c(-0.2,2), ylim=c(-0.22,2),
    xlab = expression(list(v[1])),
    ylab=expression(list(v[2])))
arrows(0,0,1,0, col="red")
arrows(0,0,1,1, col="blue")
arrows(0,0,2,1, col="green")
segments(1,1,2,1, col="grey")
segments(1,0,2,1, col="grey")
```



3.1.2 Linear Combinations

Given a finite collection of n-dimensional vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ and scalar coefficients $c_1, c_2, \cdots, c_k$, the vector

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

is called the **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ with coefficients $c_1, c_2, \cdots, c_k$.

Example Show that the vector $\langle 2, 3\rangle$ can be written as a linear combination of $\mathbf{e}_1 = \langle 1, 0\rangle$ and $\mathbf{e}_2 = \langle 0, 1\rangle$.

This problem can be set up and solved using an algorithm that solves a system of linear equations.[7]

7: See Section 3.4.

However, the situation at hand is a simpler matter of applying the definition of linear combination. We see that we can express

$$\langle 2, 3\rangle = \langle 2, 0\rangle + \langle 0, 3\rangle = 2\langle 1, 0\rangle + 3\langle 0, 1\rangle = 2\mathbf{e}_1 + 3\mathbf{e}_2.$$

3.1.3 Bases and Dimension

As we mentioned previously, the components of a vector are not defined in a "universal way", but they depend on the choice of a set of "reference vectors", which form a **basis**: a set of vectors which cover once and only once all possible **independent** directions of the vector space.

Let V be a vector space, and let $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ be a finite list of vectors in V. We say that the vectors are **linearly independent** if:

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0} \text{ if and only if } c_1 = c_2 = \cdots = c_k = 0.$$

Otherwise, we say that they are **linearly dependent**.

If we expand the equation above, we see that the condition of linear independence is equivalent to state that the **homogeneous linear system** (see Section 3.4)

$$\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k\end{bmatrix} \cdot \begin{bmatrix}c_1 & c_2 & \cdots & c_k\end{bmatrix}^T = \begin{bmatrix}0 & 0 & \cdots & 0\end{bmatrix}^T$$

only has the trivial solution $c_1 = c_2 = \cdots = c_k = 0$.

We can also view linear dependence is as follows. Suppose, for instance, that we have three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ related by a linear dependence relation. For example, let us assume that

$$\mathbf{v}_1 - \mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}.$$

Then we can rewrite this expression as

$$\mathbf{v}_1 = \mathbf{v}_2 + \mathbf{v}_3,$$

which provides an intuition for the idea of linear dependence: one (or more) vector in the collection can be reconstructed as a linear combination of the remaining vectors.

A **basis** of a vector space V is a collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ such that: + The vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly independent. + Every vector $v \in V$ can be expressed **in a unique way** as a linear combination of the basis element $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$.

Note that the linear combination expressed from a basis is **unique**,[8] that is the coefficients $c_1, c_2, \cdots, c_n$ of the equation

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

are uniquely determined.

8: Because the basis vectors are linearly independent.

While a vector space V has more than one basis, **all of its bases have the same cardinality**, meaning that all bases have the same number of vectors. This number n is the **dimension** of the vector space.

The vector space $\mathbb{R}^n$ is $n-$dimensional; we usually (but not always) represent vectors with respect to the **standard basis** $\{\mathbf{e}_1, \ldots, \mathbf{e}_2\}$.

The uniqueness of the expression of a vector as a linear combination of basis vectors explains why we can interpret the components of the vector as coordinates.

Example Determine if the following 4 vectors form a basis in $\mathbb{R}^4$:

$$\begin{aligned}
\mathbf{v}_1 &= \langle 1, 0, 0, 0 \rangle \\
\mathbf{v}_2 &= \langle 1, 1, 1, 1 \rangle \\
\mathbf{v}_3 &= \langle 1, 0, 1, -2 \rangle \\
\mathbf{v}_4 &= \langle 0, 1, 0, -1 \rangle
\end{aligned}$$

We need to solve the equation $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + c_4\mathbf{v}_4, = \langle 0, 0, 0, 0 \rangle$, which unwraps into:

$$\begin{aligned}
&c_1 = 0 \\
&c_1 + c_2 + c_3 + c_4 = 0 \\
&c_1 + c_3 - 2c_4 = 0 \\
&c_2 - c_4 = 0
\end{aligned}$$

Is it clear that the only solution is the trivial one $c_1 = c_2 = c_3 = c_4 = 0$? We will discuss how to demonstrate that it is indeed the only solution in Section 3.4.[9]

9: We can also verify linear independence using the properties of determinants (see Section 3.3.3).

Example Show with an example that there can be infinitely many bases for a vector space of positive dimension.

For each $\theta \in [0, 2\pi)$, the set

$$B_\theta = \{\langle \cos\theta, \sin\theta \rangle, \langle -\sin\theta, \cos\theta \rangle\}$$

is a basis of $\mathbb{R}^2$. ■

We will not discuss infinite dimensional vector spaces (that's a topic for advanced courses), but we provide one such example, for curiosity's sake.

Example The space

$$\mathbb{P}[x] = \{a_0 + a_1 x + \cdots + a_k x^k \mid a_i \in \mathbb{R}, k \in \mathbb{N}\}$$

of all polynomials in one variable x over the reals is an infinite dimensional vector space; the vectors are polynomials. For all $n \in \mathbb{N}$, the monomials $1, x, x^2, \cdots, x^n$ are linearly independent for all n, so there are infinitely many linearly independent vectors in $\mathbb{P}[x]$.[10]

10: This example is interesting not just because it deals with an infinite-dimensional vector space, but also because it shows that the notion of vector space applies beyond the intuitive geometric notion of arrows represented in vector components.

3.1.4 Vector Subspaces

The space $W = \mathbb{R}^2$ consists of vectors of the form $\langle x, y \rangle$. The space $V = \mathbb{R}^3$ consists of vectors of the form $\langle x, y, z \rangle$. We can interpret W as a smaller vector space contained in V, from which it inherits the operations of sum and multiplication by scalar.

Example Show that a linear combination of 2-dimensional vectors of the form $\langle x, y, 0\rangle$ has the same form (i.e., the third component remains zero).

This is a classic problem that looks hard the first time we learn linear algebra, but in fact the solution consists a simple check. Take two arbitrary vectors $\mathbf{v}_1 = \langle x_1, y_1, 0\rangle$ and $\mathbf{v}_2 = \langle x_2, y_2, 0\rangle$. Then, for arbitrary scalars a, b, the linear combination of them has the expression

$$a\mathbf{v}_1 + b\mathbf{v}_2 = a\langle x_1, y_1, 0\rangle + b\langle x_2, y_2, 0\rangle = \langle ax_1 + bx_2, ay_1 + by_2, 0\rangle,$$

of the form $\langle x, y, 0\rangle$, if we let $x = ax_1 + bx_2$ and $y = ay_1 + by_2$. ■

Let V be a vector space, and $W \subset V$, a subset of V: we say that W is a **vector subspace** (subspace in short) of V, denoted $W < V$, if W is a vector space itself (which inherits the operations from the bigger space V in which it is contained).

In particular, if $W < V$, and $\mathbf{v}, \mathbf{w} \in W$ and $a, b \in \mathbb{R}$, then:

- $\mathbf{0} \in W$, and
- $a\mathbf{v} + b\mathbf{w} \in W$.

Note that, by definition, V is a subspace of itself.

The result of the previous example can be recast as $\mathbb{R}^2$ being a vector subspace of $\mathbb{R}^3$.

Example Let V be a vector space. What is the "largest" subspace of V? What is the "smallest" subspace of V?

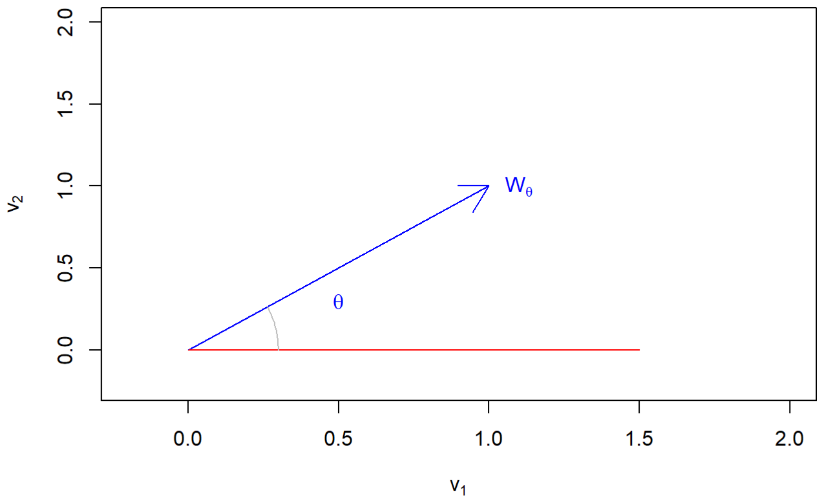
As $V \subseteq V$ is itself a subspace of V, it is also the largest subspace of V. The smallest subspace of V is the zero-dimensional vector space $\{\mathbf{0}\}$, which consists solely of the zero vector. ■

Let V be a vector space of dimension n. Then, it should be intuitive that if W is a subspace of V, then $\dim(W) \leq \dim(V)$.

The zero space from the previous example is the only zero-dimensional vector subspace; the space V itself is the only subspace of maximal dimension n. There are infintely many "intermediate dimension" (proper) subspaces as soon $\dim V \leq 2$.

Example Let $W_\theta = \{a\langle\cos\theta, \sin\theta\rangle \mid a \in \mathbb{R}\} < \mathbb{R}^2$, $\theta \in [0, 2\pi)$. For each angle value θ, the vector $\langle\cos\theta, \sin\theta\rangle$ gives a different direction, hence $W_{\theta_1} = W_{\theta_2}$ if and only if $\theta_1 = \theta_2$.

```
library(plotrix)
plot(NA,xlim=c(-0.2,2), ylim=c(-0.22,2),
     xlab = expression(list(v[1])),
     ylab=expression(list(v[2])))
arrows(0,0,1,1, col="blue")
segments(0,0,1.5,0, col="red")
draw.arc(0,0,0.3,0,0.5,col="grey")
text(0.5,0.3,expression(theta),col="blue")
text(1.1,1,expression(list(W[theta])),col="blue")
```



3.1.5 Spanning Sets

How do we "create" subspaces? As long as we do not worry too much about "clean production", we take a finite set of vectors of a given vector space V, and consider **all possible linear combination** of such vectors.

Let V a vector space and $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N \in V$. The **spanning set**

$$\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N\} = \{a_1\mathbf{v}_1 + \cdots + a_N\mathbf{v}_N \mid a_i \in \mathbb{R}\} < V.$$

Example Let V be a vector space and $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N \in V$. Then $\mathbf{v} \in \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ if and only if $\mathbf{v} = c_1\mathbf{v}_1 + \cdots c_n\mathbf{v}_N$, for some coefficients $c_1, c_2, \cdots, c_N \in \mathbb{R}$.

This is a "trivial" statement – we simply translated the condition "belonging to span" into the equation "$\mathbf{v}$ is a linear combination of the spanning vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N$".[11]

11: Being trivial, it can still cause confusion at the beginning; but it is crucial to learn how to translate math-related sentences into formulas or equations.

The problem with the definition of the spanning set of a collection of vectors is that it says nothing about the **dimension of the vector space**.

Example Let $V = \mathbb{R}^2$. We can write $V = \text{Span}\{\langle 1, 0\rangle, \langle 0, 1\rangle\}$, which makes sense since the two vectors form a basis of V. However, we can also generate the entire vector space with three vectors, so that the number of vectors is not linked to the dimension: $V = \text{Span}\{\langle 1, 0\rangle, \langle 0, 1\rangle, \langle 1, 1\rangle\}$.

3.1.6 Dot Product

The dot product of two vectors is a scalar quantity that in some sense measure how much of their components two vectors share. The **dot** (or scalar) **product** of two n-dimensional vectors $\mathbf{v} = \langle v_1, v_2, \cdots, v_n\rangle$ and $\mathbf{w} = \langle w_1, w_2, \cdots, w_n\rangle$:

$$\mathbf{v} \cdot \mathbf{w} = v_1w_1 + v_2w_2 + \cdots + v_nw_n.$$

From the dot product, we can define the Euclidean **length** (or **norm**) of a vector $\mathbf{v}$:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Two vectors $\mathbf{v}$ and $\mathbf{w}$ are **orthogonal** if and only if $\mathbf{v}\cdot\mathbf{w} = 0$. In general:

$$\mathbf{v}\cdot\mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\|\cos(\theta),$$

where θ is the angle formed by the vector $\mathbf{v}$ and $\mathbf{w}$.[12]

12: In fact, this is how we define the angle between two vectors when the geometrical interpretation is unavailable to us.

Two non-zero vectors $\mathbf{v}, \mathbf{w}$ create two angles, θ and $2\pi - \theta$: does the dot product depends on the choice between the two angles?

No, because for all angles θ we have:

$$\cos(2\pi - \theta) = \cos(\theta).$$

Example Find the (smallest) angle θ formed by the vectors $\mathbf{v} = \langle 1, 2\rangle$ and $\mathbf{w} = \langle -1, 1\rangle$.

It's a one line calculation:

$$\theta = \arccos\left(\frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|}\right) = \arccos\left(\frac{-1+2}{\sqrt{1+4}\sqrt{1+1}}\right) = \arccos\left(\frac{1}{\sqrt{10}}\right) = 1.25 \text{ radians}\ .$$

Example Let t be a real parameter Find the vectors of the form $\langle 1, t\rangle$ and with length equal to 5.

The general vector $\langle 1, t\rangle$ has length

$$\|\langle 1, t\rangle\| = \sqrt{1^2 + t^2} = \sqrt{1 + t^2}.$$

We look for the values of t such that

$$\|\langle 1, t\rangle\| = \sqrt{1 + t^2} = 5,$$

which are found by solving the **quadratic equation**:

$$\sqrt{1+t^2} = 5 \implies 1 + t^2 = 25 \implies t^2 = 24 \implies t = \pm\sqrt{24} = \pm 2\sqrt{6}.$$

As expected, there are two vectors $\langle 1, t\rangle$ of length 5: $\langle 1, \pm 2\sqrt{6}\rangle$.

3.1.7 Cross Product in $\mathbb{R}^3$

The dot product is also called scalar product, since it outputs a scalar from two given vectors. The cross (or vector) product, which will define below, produces a new vector out of two input vectors.

Given two 3-dimensional vectors $\mathbf{v} = \langle v_1, v_2, v_3, \rangle$ and $\mathbf{w} = \langle w_1, w_2, w_3, \rangle$, the **cross** (or vector) **product** formula can be symbolically represented with the help of a determinant:

$$\mathbf{v} \times \mathbf{w} = \det \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix}$$
$$= \langle v_2 w_3 - v_3 w_2, -(v_1 w_3 - v_3 w_1), v_1 w_2 - v_2 w_1 \rangle.$$

Note that we left the formula without multiplying out negative sign in front of the second entry, in order to remind the reader that the determinant is an alternating sum.[13]

13: See Section 3.3.3.

Whereas the dot product can be extended to vector space of all dimensions, the cross product is only defined on $\mathbb{R}^3$.

3.2 Linear Transformations and Matrices

A **matrix** of size $m \times n$ is a collection of $m \times n$ numbers aligned along m rows and n columns:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

We refer to matrices of size $n \times n$ as **square matrices** of size n.

Let V and W be two vector spaces (of arbitrary dimension, possibly infinite-dimensional). A **linear map** $T : V \to W$ is a function that preserves linear combinations of vectors:

$$T(a\mathbf{v} + b\mathbf{w}) = aT(\mathbf{v}) + bT(\mathbf{w}), \text{ for all } a, b \in \mathbb{R} \text{ and for all } \mathbf{v}, \mathbf{w} \in V.$$

Given a basis $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ of V and a basis $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$ of W, we can construct the **matrix elements** $t_{i,j}$ of the matrix representing the linear transformation T with respect to the given bases. In fact, there are coefficients T_{ij} such that

$$T(\mathbf{v}_i) = \sum_{j=1}^{m} t_{i,j} \mathbf{w}_j$$

We will use the **convention** that a matrix is given with respect to the **standard basis**.

A linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is represented by **matrix-vector multiplication**. We write the vectors of $\mathbb{R}^n$ and $\mathbb{R}^m$ as **column vectors**:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \in \mathbb{R}^m.$$

The vector-matrix multiplication defines the linear map $T(\mathbf{v}) = \mathbf{w}$ (relative to bases of $\mathbb{R}^n$ and $\mathbb{R}^m$):

$$\begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} t_{1,1} & \cdots & t_{1,n} \\ \vdots & \ddots & \vdots \\ t_{m,1} & \cdots & t_{m,n} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} t_{1,1}v_1 + \cdots + t_{1,n}v_n \\ \vdots \\ t_{m,1}v_1 + \cdots + t_{m,n}v_n \end{bmatrix}.$$

Linear maps can be composed in the same way as regular functions, assuming that the range of the second is in the domain of the first.

If $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $S : \mathbb{R}^m \longrightarrow \mathbb{R}^p$ are two linear maps, then the **composition** of S and T (the order is important) is the linear map $S \circ T : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ defined by

$$(S \circ T)(\mathbf{v}) = S(T(\mathbf{v})) = (ST)\mathbf{v}.$$

If the maps S and T are represented by the matrices S=

$$\begin{bmatrix} s_{1,1} & \cdots & s_{1,m} \\ \vdots & \ddots & \vdots \\ s_{p,1} & \cdots & s_{p,m} \end{bmatrix} \text{ and } T = \begin{bmatrix} t_{1,1} & \cdots & t_{1,n} \\ \vdots & \ddots & \vdots \\ t_{m,1} & \cdots & t_{m,n} \end{bmatrix},$$

then the composite map corresponds to the matrix obtained by **matrix multiplication** (or **matrix product**)

$$ST = \begin{bmatrix} s_{1,1} & \cdots & s_{1,m} \\ \vdots & \ddots & \vdots \\ s_{p,1} & \cdots & s_{p,m} \end{bmatrix} \begin{bmatrix} t_{1,1} & \cdots & t_{1,n} \\ \vdots & \ddots & \vdots \\ t_{m,1} & \cdots & t_{m,n} \end{bmatrix} = \begin{bmatrix} s_{1,1}t_{1,1} + \cdots + s_{1,m}t_{m,1} & \cdots & s_{1,1}t_{1,n} + \cdots + s_{1,m}t_{m,n} \\ \vdots & \ddots & \vdots \\ s_{p,1}t_{1,1} + \cdots + s_{p,m}t_{m,1} & \cdots & s_{p,1}t_{1,n} + \cdots + s_{p,m}t_{m,n} \end{bmatrix}$$

Note that the formula of matrix multiplication can be more easily understood using dot products:

$$(st)_{ij} = (\text{row } i \text{ of S}) \cdot (\text{column } j \text{ of T}).$$

Example For any angle value in radians, measured counterclockwise with respect to reference to the positive x-axis, the matrix

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

rotates vectors in the xy-plane around the origin by an angle θ. For instance, we can rotate the vector $\langle 1, 0\rangle$ by $\frac{\pi}{4}$ (45 degrees counterclockwise):

$$R_{\pi/4}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}.$$

The new vector has the same length as the original one, in agreement with the fact that rotations do not change lengths, and it forms an angle of 45 degrees with respect to the positive x-axis.

For a fixed θ, the rotation is a linear map:

$$R_\theta(a\mathbf{v} + b\mathbf{w}) = aR_\theta(\mathbf{v}) + bR_\theta(\mathbf{w}) \qquad \text{(prove it!)}.$$

3.3 Matrix Algebra

We have already introduced matrix multiplication as a way to define the composition of two compatible linear maps. In this section we collect the essential rules of matrix algebra. We start with operations that make sense for all matrices, and then specialize to operations that are defined only for square matrices. For convenience we report again the definition of matrix multiplication.

3.3.1 Matrix Operations

Matrix Multiplication

Formally, let $A \in \mathcal{M}_{m,n}$ (i.e., A is a $m \times n$ matrix) and $B \in \mathcal{M}_{n,p}$ (i.e., B is a $n \times p$ matrix). Then the **matrix product** of A by B is the matrix $AB \in \mathcal{M}_{m,p}$ (i.e., AB is of size $m \times p$), where the entries $(ab)_{i,j}$ are

$$(ab)_{i,j} = \text{ row } i \text{ of } A \cdot \text{ column } j \text{ of } B.$$

Unlike multiplication between scalars, the product of matrices is not generally commutative – assuming that both AB and BA exist, it is not always the case that $AB = BA$.

Example If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$, then

$$AB = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix} \neq \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix} = BA.$$

The matrix product AB is only defined when the number of columns of A is equal to the number of rows of B:

$$\underbrace{A}_{m\times n}\underbrace{B}_{n\times p} = \underbrace{AB}_{m\times p}.$$

The dot product of two n-dimensional vector can also be understood in term of matrix multiplication: if we represent $\mathbf{v}$ as a row vector, and $\mathbf{w}$ as a column vector, then

$$\mathbf{v} \cdot \mathbf{w} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

Matrix Addition

Given two matrices $A, B \in \mathcal{M}_{m,n}$, their **sum** is the matrix $A + B \in \mathcal{M}_{m,n}$ obtained by adding A and B entry-by-entry, that is

$$(a + b)_{i,j} = a_{i,j} + b_{i,j}.$$

Example If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$, then $A + B = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$.

Note that, unlike matrix multiplication, matrix addition is commutative: $A + B = B + A$ for all compatible matrices.

Multiplication by a Scalar

For any matrix $A \in \mathcal{M}_{m,n}$ and scalar $c \in \mathbb{R}$, the **scalar multiplication** of A by c is the matrix $cA \in \mathcal{M}_{m,n}$ whose entries are the entries of A scaled by the factor c, that is:

$$(ca)_{i,j} = ca_{i,j}.$$

Example If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $c = -2$, then $cA = \begin{bmatrix} -2 & -4 \\ -6 & -8 \end{bmatrix}$.

Transpose of a Matrix

The **transpose** of $A \in \mathcal{M}_{m,n}$ is the matrix $A^\top \in \mathcal{M}_{n,m}$ whose columns are the rows of A:

$$(a^\top)_{i,j} = a_{j,i}.$$

Example If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then $A^\top = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$.

The transpose is a linear operation: $(A + B)^\top = A^\top + B^\top$ for all compatible matrices A, B. However, it behaves "unexpectedly" with respect to matrix multiplication: $(AB)^\top = B^\top A^\top$ for all compatible matrices A, B.[14]

14: While this is not a proof, we see that this formula is at the very least aligned with the compatibility of matrix multiplication: if $A \in \mathcal{M}_{m,n}$ and $B \in \mathcal{M}_{n,p}$, then $AB \in \mathcal{M}_{m,p}$ and $(AB)^\top \in \mathcal{M}_{p,m}$. Since $B^\top \in \mathcal{M}_{p,n}$ and $A^\top \in \mathcal{M}_{n,m}$, we see that $B^\top A^\top$ is always defined, but that $A^\top B^\top$ is only defined when $m = p$.

Matrix Spaces

The **column space** of a matrix $A = [A_1 \mid \cdots \mid A_n] \in \mathcal{M}_{m,n}$ is the vector subspace of $\mathbb{R}^m$ spanned by the column vectors of A:

$$\text{colsp}(A) = \text{Span}\{A_1, \cdots, A_n\}.$$

The **rank** of A is the dimension of colsp(A). If we interpret A as a linear map (as discussed in Section 3.2), then colsp(A) is in fact the **image** of this map:

$$\text{Im}(A) = \{A\mathbf{v} \mid \mathbf{v} \in \mathbb{R}^n\} < \mathbb{R}^m.$$

The **nullspace** (or **kernel**) of A is the vector subspace of $\mathbb{R}^m$ that are mapped to the null vector $\mathbf{0}$ by A:

$$\text{nullsp}(A) = \ker(A) = \{\mathbf{v} \in \mathbb{R}^m \mid A\mathbf{v} = \mathbf{0}\} \subseteq \mathbb{R}^m.$$

That these two sets are indeed vector subspaces of $\mathbb{R}^m$ is clear:

- $\mathbf{0} \in \text{Im}(A), \ker(A)$ since $A\mathbf{0} = \mathbf{0}$;[15]

15: The null vector pulls double-duty here.

- if $\mathbf{v}, \mathbf{w} \in \ker(A)$, $a, b \in \mathbb{R}$, then $a\mathbf{v} + b\mathbf{w} \in \ker(A)$ since

$$A(a\mathbf{v} + b\mathbf{w}) = aA\mathbf{v} + bA\mathbf{w} = a\mathbf{0} + b\mathbf{0} = \mathbf{0};$$

- if $\mathbf{v}, \mathbf{w} \in \mathrm{Im}(A)$, $a, b \in \mathbb{R}$, then $a\mathbf{v} + b\mathbf{w} \in \mathrm{Im}(A)$ since there exists $\mathbf{u}, \mathbf{z} \in \mathbb{R}^n$ such that $A\mathbf{u} = \mathbf{v}$ and $A\mathbf{z} = \mathbf{w}$, and so

$$a\mathbf{v} + b\mathbf{w} = aA\mathbf{u} + bA\mathbf{z} = A(a\mathbf{u} + b\mathbf{z}).$$

In particular, neither of these spaces is empty since they always contain at least $\mathbf{0}$.

Rank-Nullity Theorem

Let $A \in \mathcal{M}_{m,n}$; then

$$\dim(\ker(A)) + \dim(\mathrm{Im}(A)) = m.$$

This theorem is a basic (and very useful) result of linear algebra, with counterparts in other sectors of algebra (such as group theory).

3.3.2 Square Matrices

The **identity matrix** of size n is the square matrix, denoted by $\mathbf{I}_n$:

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

The **diagonal** of a square matrix A is the list of elements A_{ii} (that is, the values along the diagonal).

A square matrix is said to be a **diagonal matrix** if the non-diagonal entries are all zero.

A square matrix A is said to be **symmetric** if $A = A^\top$. In fact, the entries are symmetric with respect to the diagonal of the matrix.

A square matrix A of size n is said to be **invertible** (or non-singular) if there exists a matrix, denoted by A^{-1}, such that $AA^{-1} = A^{-1}A = \mathbf{I}_n$. The matrix A^{-1} is called the **inverse** of A. Note that the inverse of A^{-1} is A (in other words, $(A^{-1})^{-1} = A$).

If A is invertible, then

$$(A^{-1})^\top = (A^\top)^{-1}.$$

If A and B are both invertible (and have the same size), then

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

We will discuss a way to compute the inverse of a non-singular matrix in Section 3.4.2.

3.3.3 Determinants

There is an important numerical value that can be associated to any square matrix A, its **determinant** $\det(A)$.

When we work with large-sized matrices, we rely on a computer program to compute the determinant. However, we need to know what it is and how to compute it for small size examples.

The purely algebraic definition of the determinant makes use of the language of multilinear algebra, which will not discuss here; instead, we proceed with a computational definition.

- For a scalar $a \in \mathbb{R} = \mathcal{M}_{1,1}$, $\det(a) = a$.
- For $A \in \mathcal{M}_{2,2}$,

$$\det(A) = \det \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

- For $A \in \mathcal{M}_{n,n}$, consider the first row, which consists of the elements $[a_{1,1}, a_{2,1}, \cdots, a_{1,n}]$. Let $M_{1,k}$ be the square matrix of size $n-1$ obtained by removing from A the row and column passing through $a_{1,k}$. Then the determinant of A is the **alternating sum**:

$$\det(A) = \det(M_{1,1}) - \det(M_{1,2}) + \cdots + (-1)^{n+1} \det(M_{1,n})$$

  The quantities $\det(M_{i,j})$ are called the **minors** of the matrix.

In fact, we can pick any row or any column and apply the alternating sum formula as above. However, we need to be careful about the sign in front of the minor $\det(M_{i,j})$, which is called the **cofactor** $C_{i,j}$:

$$C_{i,j} = (-1)^{i+j} \det(M_{i,j}).$$

For more details about the general formula, we refer to [4].

Properties

The determinant determines important properties of a square matrix.

- The determinant of a **diagonal** matrix is the product of its diagonal entries.
- The determinant behaves nicely when it comes to matrix multiplication and inversion (assuming A and B are both square and of the same size):

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

  and, if A is invertible, then

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

- The determinant is **invariant under transposition**:

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

- Let A be a square matrix and let $A[R_i \leftrightarrow R_j]$ (resp. $A[C_i \leftrightarrow C_j]$) be the matrix obtained by interchanging row i with row j (resp. column i with column j). Then

$$\det(A[C_i \leftrightarrow C_j]) = -\det(A)$$
$$\det(A[R_i \leftrightarrow R_j]) = -\det(A)$$

- More generally, if we perform an odd number of permutations of rows (columns), the determinant changes sign; if we perform an even number of permutations of rows (columns), the determinant stays the same.

Let A be a square matrix, of size n. Then the following conditions are equivalent:

1. $\det(A) \neq 0$;
2. A is invertible;
3. the n column vectors of A are linearly independent, hence they form a basis of $\mathbb{R}^n$;
4. the n row vectors of A are linearly independent, hence they form a basis of $\mathbb{R}^n$;
5. the rank of A is n (maximal rank);
6. the nullspace (kernel) of A consists only of the zero vector $\mathbf{0}$.

Examples Determine if the following matrices are invertible or not, without computing the inverse.

1. $A = \begin{bmatrix} 2 & 3 \\ -1 & -3 \end{bmatrix}$ is invertible, since $\det A = 2(-3) - 3(-1) = -2 \neq 0$.
2. $B = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 \\ -1 & 1 & -1 & 1 \end{bmatrix}$ is not invertible, since the first and third rows are equal (and so they are linearly dependent).
3. $C = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 2 & 1 \\ 2 & 3 & 5 & 5 \\ -1 & 1 & -1 & 1 \end{bmatrix}$ is not invertible, as we can see either by computing that $\det C = 0$, or by observing that $R_1 + R_2 = R_3$.
4. $D = \begin{bmatrix} 1 & 42 & 0.12 & 4 \\ 0 & 1 & -2 & 21 \\ 1.2 & 23 & 0.5 & 5 \\ -2.2 & 1 & 0 & -0.55 \end{bmatrix}$ is invertible as can be seen in the following `R` code.

```
D <- rbind(c(1,42,0.12,4),c(0,1,-2,21),c(1.2,23,0.5,5),c(-2.2,1,0,-0.55))
det(D)
```

```
[1] -1336.74
```

5. Suppose that A and B are square matrices of the same size, and that $\det(A) = 3$, $\det(B) = -5$; then

$$\det(A^{-1}B^3A) = \frac{1}{\det(A)}\cdot(\det(B)^3)\cdot\det(A) = (\det(B))^3 = (-5)^3 = -125.$$

There is a closed-form formula for finding the inverse of a square matrix of arbitrary size. Computing the inverse can be very time consuming, and, when the matrices are very large (thousands of entries), we typically consider numerical methods.

But it is convenient to at least remember how to find the inverse of a 2×2 matrix.[16]

16: For inversion of matrices of arbitrary size, we refer to [4]. We mention in passing that the general formula for A^{-1} contains a factor $\frac{1}{\det A}$, re-descovering the fact zero-determinant matrices can not be inverted.

For a 2×2 matrix $A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}$, say, the inverse (when it exits) is

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} a_{2,2} & -a_{1,2} \\ -a_{2,1} & a_{1,1} \end{bmatrix}$$

The formula of the inverse starts with $\frac{1}{\det A}$. If the determinant of A is non-zero, but **close to** zero, we could have issues with the finite precision arithmetic.

We will discuss a row-reduction method to compute the inverse of a general non-singular matrix in the next section.

Example Let A and B be the following matrices:

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

Solve the equation $AX = B$ for X, where $X \in \mathcal{M}_{2,2}$. We see that

$$AX = B \Rightarrow X = A^{-1}B,$$

but is A invertible? A quick check using the determinant confirms that it is since $\det(A) = 1 \cdot 1 - 2 \cdot 3 = -5 \neq 0$. Using the formula of the inverse of a 2×2 matrix we obtain:

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

Finally X, the solution of the equation, is

$$X = A^{-1}B = \begin{bmatrix} -\frac{1}{5} & \frac{2}{5} \\ \frac{3}{5} & -\frac{1}{5} \end{bmatrix} \begin{bmatrix} 0 & -2 \\ 8 & 1 \end{bmatrix} = \begin{bmatrix} \frac{16}{5} & \frac{4}{5} \\ -\frac{8}{5} & -\frac{7}{5} \end{bmatrix}$$

3.4 Linear Systems

A big motivation for developing the machinery of linear algebra is to find systematic methods for solving **systems of linear equations**, which we can call, in short, **linear systems**. A linear system in n unknowns $x_1, x_2, \cdots, x_n$ and m equations is a system of m linear equations

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= b_2 \\ &\vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= b_m \end{aligned}$$

Collecting the coefficients of the left hand side of the system into a $m \times n$ matrix, and the coefficients of the right hand side into a m dimensional column vector, we obtain the **matrix-vector form of the linear system**, $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

We say that a system of m equations and n variables has size $m \times n$.

If $\mathbf{b} = \mathbf{0}$, the system is called **homogeneous**.

Example Let A and B be the following matrices:

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

We have shown how to solve the equation $AX = B$ for X, where $X \in \mathcal{M}_{2,2}$. Expand this equation to show that is equivalent to a linear system. Write the linear system in matrix vector form $A\mathbf{x} = \mathbf{b}$.[17]

17: This will not be the same A as in the statement.

The 4 unknowns are the entries of the matrix $X = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$. Then

$$AX = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} 0 & -2 \\ 8 & 1 \end{bmatrix}.$$

Expanding the product AX gives the equation

$$AX = \begin{bmatrix} x + 2z & y + 2w \\ 3x + z & 3y + w \end{bmatrix} = \begin{bmatrix} 0 & -2 \\ 8 & 1 \end{bmatrix}.$$

Equating the 4 components gives us a system of 4 equations in 4 unknowns:

$$\begin{aligned} x + 2z &= 0 \\ y + 2w &= -2 \\ 3x + z &= 8 \\ 3y + w &= 1. \end{aligned}$$

In matrix vector form the system is of the form $A\mathbf{x} = \mathbf{b}$, where $A \in \mathcal{M}_{4,4}$, whose entries are specified in the equation below. The right-hand side is the vector of 4 constant entries, and the unknown vector has component x, y, z, w. The system is therefore

$$\begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & 1 & 0 \\ 0 & 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ 8 \\ 1 \end{bmatrix}.$$

Rearranging the entries of a matrix in order to obtain a new matrix of different size is a common procedure in coding. Programs like R or

`Python` come with predefined functions that do the rezising for us (but we need to know how they operate!)

The solution set of an arbitray (non-linear) system of equations in n variables is a region of $\mathbb{R}^n$. We learn that such regions are recognized to be objects of euclidean geometry: as we learn in pre-calculus, for example, the solutions of the equation $x^2 + y^2 = 1$ are the points of the circle of radius 1 and centre at the origin of the Cartesian plane.

For a linear system, we will expect the solution set to have some "linearity". More precisely:

- The solution set of a homogeneous linear system $A\mathbf{x} = \mathbf{0}$, in the n unknowns $\mathbf{x} = (x_1, x_2, \cdots, x_n)^\top$, is a vector subspace of $\mathbb{R}^n$.
- The solution set of a linear system $A\mathbf{x} = \mathbf{b}$ is a "vector space shifted away from the origin" of $\mathbb{R}^n$. More precisely, let $\mathbf{x}_p$ be any solution of the system (we call it a particular solution). Then any solution of the system is of the form $\mathbf{x}_0 + \mathbf{x}_p$, where $\mathbf{x}_0$ is a solution of the associated homogeneous linear system $A\mathbf{x} = \mathbf{0}$.

Example Let us illustrate the last two points with a simple example. Notice that this example is not meant to propose an algorithm to solve a linear system, but rather to explain the geometrical aspect of the solution set of a linear system. Consider the linear system consisting of one equations in two variables:

$$x + y = 2.$$

It is not homogeneous, since the left hand side coefficient of the equation is not zero. Since there are two variables but only one equation, we expect the general form of the solution of this system to have one free parameter (or free variable), that can be arbitrary chosen. If we use t as the name for the parameter, we write

$$\begin{aligned} x &= t \in \mathbb{R} \\ y &= 2 - t. \end{aligned}$$

Note in particular that with $t = 0$ we obtain the particular solution

$$\mathbf{x}_p = \begin{bmatrix} 0 \\ 2 \end{bmatrix},$$

which we will use in a second.

The associated homogeneous linear system is

$$x + y = 0.$$

The solution set of the homogeneous system is the line $y = -x$, which in parametric form becomes

$$\begin{aligned} x &= t \\ y &= -t. \end{aligned}$$

Geometrically, the general solution of the non-homogeneous system is obtained by shifting the line $y = -x$ by the vector $(0, 2)^\top$. If we let

$$\mathbf{x}_0(t) = \begin{bmatrix} t \\ -t \end{bmatrix},$$

we see that the general solution is of the form $\mathbf{x}(t) = \mathbf{x}_p + \mathbf{x}_0(t)$.

Example Which of the following equation is linear? Why is it important to identify if an equation (or a system of equation) is linear?

a) $x + y - z = 4$ b) $x^2 - y + z = 4$ c) $4x + 4y - z - 4 = 0$

The system a) and c) are linear, while the x^2 term in b) makes that one non-linear. It is important to know what are the properties of linear systems: the linear algebra algorithms, such as Gauss-Jordan elimination, do not apply to non-linear systems.

3.4.1 Gauss-Jordan Elimination

In introductory linear algebra courses we often start by learning linear systems and how to the **Gauss-Jordan elimination** algorithm. We will not discuss the details of the method in this chapter and we refer to [4] for more details.

The idea of the elimination algorithm is to transform the matrix associated with a linear system into a simpler one. The common approach is to transform the original matrix to a **row echelon form**, or even better the **row reduced echelon form**. Reading the solution of a matrix in echelon form then is quite easy.

The principles behind the Gauss-Jordan elimination are the following. We say that two linear systems are **equivalent** if they have the same solution set.

- Given a linear system, an equivalent system is obtained by adding to one equation a multiple of another one. In term of the matrix associated to the linear system, this amounts to adding to a row a multiple of another one.
- Given a linear system, an equivalent system is obtained by rescaling an equation by a non-zero factor. In term of the matrix associated to the linear system, this amounts to multiplying the row vector corresponding to the equation by a scalar.

We can therefore proceed and start eliminating as much variables as we can, trying to obtain a matrix from which reading the solution is a simple procedure. Let us see an example.

Example We solve the following 2×3 linear system:

$$\begin{aligned} x - y - 2z &= 0 \\ 3x + 2y + z &= 2 \end{aligned}$$

We start by writing the **augmented matrix**

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

which includes the right hand side of the system in the last column. We proceed with the row reduction in order to reduce the system to an equivalent one that is easier to solve.

We denote by R_k the row number k of the matrix (in this example, $k = 1, 2$). Assume $a \neq 0, b \in \mathbb{R}$; $R_k \longrightarrow aR_k + bR_j$ denotes the operation of replacing R_k with the linear combination $aR_k + bR_j$.[18] Then

18: $a \neq 0$ is crucial.

$$\begin{bmatrix} 1 & -1 & 2 & 0 \\ 3 & 2 & 1 & 2 \end{bmatrix} \xrightarrow{R_2 \to R_2 - 3R_1} \begin{bmatrix} 1 & -1 & 2 & 0 \\ 0 & 5 & -5 & 2 \end{bmatrix} \xrightarrow{R_2 \to \frac{R_2}{5}} \begin{bmatrix} 1 & -1 & 2 & 0 \\ 0 & 1 & -1 & \frac{2}{5} \end{bmatrix} \xrightarrow{R_1 \to R_1 + R_2|} \begin{bmatrix} 1 & 0 & 1 & \frac{2}{5} \\ 0 & 1 & -1 & \frac{2}{5}. \end{bmatrix}$$

The column in position j corresponds to the variable in position j.

With the help of row reduction, the original linear system has been transformed into the equivalent system:

$$\begin{aligned} x + z &= \frac{2}{5} \\ y - z &= \frac{2}{5}, \end{aligned}$$

Selecting z as a **free variable**, we re-write it as:

$$\begin{aligned} x &= \frac{2}{5} - z \\ y &= \frac{2}{5} + z, \end{aligned}$$

We see that (x, y) depends on the value of z. The solution set of the linear system is therefore one-dimensional: geometrically, it is the line parametrized by the two equations above, with z being the free parameter. Note that the line does not pass through the origin, in agreement with the fact that the system is not homogeneous. ■

The solution set of a system of homogeneous linear equations is a vector space. The dimension of this vector space coincides with the number of free variables. In particular, if there are no free variables then either the solution is **unique** or the system is **inconsistent** – it does not have solutions.

Example Find an example of a linear system with a) no solutions, and b) an example of a linear system whose solution set has 3 free variables out of a total of 5.

To find an example of a) is very easy: write an equation "$\cdots = 1$", then add another equation, obtained by changing the constant to the right hand side, "$\cdots = 2$". Let us take the following example:

$$\begin{aligned} 3x + y - z + w &= 1 \\ 3x + y - z + w &= 2. \end{aligned}$$

It should be clear that no solution can exist, since $1 \neq 2$. Proceeding with row reduction, we can see it algorithmically: we replace $R_2 \longrightarrow R_2 - R_1$ and we obtain the system:

$$\begin{aligned} 3x + y - z + w &= 1 \\ 0 &= 1, \end{aligned}$$

which is inconsistent.

As for b), we can produce an example of a matrix that gives 3 free variables, if treated as the augmented matrix of a linear system:

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

The examples of this section have shown that, if A is a matrix associated with the linear system $A\mathbf{x} = \mathbf{b}$, then:

- the rows of the matrix corresponds to the system's equations, the column to its variables;
- interchanging two rows of the matrix swaps the corresponding equations in the linear system; interchanging two columns swaps the corresponding variables.

Example The system

$$\begin{aligned} 3x - y + z &= 0 \\ x + y + z &= 3 \end{aligned}$$

corresponds to the matrix

$$\begin{bmatrix} 3 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

If we switch y and z,[19] we obtain

$$\begin{aligned} 3x + z - y &= 0 \\ x + z + y &= 3, \end{aligned}$$

which corresponds to the matrix

$$\begin{bmatrix} 3 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}.$$

19: Which should not be done unless absolutely necessary, to be honest, but nevermind that for now.

3.4.2 Linear Systems and Matrices

Row reduction can be used to invert non-singular matrices. Let $A \in \mathcal{M}_{n,n}$ be such that $\det(A) \neq 0$. Construct the augmented matrix $(A \mid \mathbf{I}_n)$ and row reduce it using only the 3 following allowable operations:

- $R_j \longrightarrow R_j + bR_k$, $j \neq k$;
- $R_j \longrightarrow aR_j$, $a \neq 0$;
- $R_j \leftrightarrow R_k$, $j \neq k$.

The process leads to

$$(A \mid \mathbf{I}_n) \xrightarrow{\text{RREF}} (\mathbf{I}_n \mid A^{-1}).$$

Example Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$. We have seen that $\det(A) = -5$ and so that A is invertible. We reduce the augmented matrix:

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 1 \end{bmatrix} \xrightarrow{R_2 \to R_2 - 3R_1} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & -5 & -3 & 1 \end{bmatrix} \xrightarrow{R_2 \to -\frac{R_2}{5}} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & \frac{3}{5} & -\frac{1}{5} \end{bmatrix} \xrightarrow{R_1 \to R_1 - 2R_2|} \begin{bmatrix} 1 & 0 & -\frac{1}{5} & \frac{2}{5} \\ 0 & 1 & \frac{3}{5} & -\frac{1}{5}; \end{bmatrix}$$

so

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

3.5 Matrix Diagonalization

Through a series of specific transformations, some matrices can be brought into diagonal form. This seemingly inconspicuous property has far-reaching consequences.

3.5.1 Eigenvalues and Eigenvectors

A matrix is **diagonal** if its non-zero entries can only be found along the diagonal.[20] Diagonal matrices are very simple: in associated linear systems, the variables involved are "decoupled", and solving the system amounts to solving a collection of linear equations in one variable. In fact, for the diagonal matrix A with diagonal entries denoted, **in order**, by $\lambda_1, \lambda_2, \dots, \lambda_n$, the linear system $A\mathbf{x} = \mathbf{b}$ is

$$\begin{aligned} \lambda_1 x_1 &= b_1 \\ \lambda_2 x_2 &= b_2 \\ &\vdots \\ \lambda_n x_n &= b_n. \end{aligned}$$

20: Note that the diagonal entries themselves could be zero.

Note that if $\lambda_j = 0$ for some index j, the system has solution only if $b_j = 0$, and the variable x_j corresponds to a subspace belonging to $\ker(A)$.

But matrices are not "absolute objects", in the sense that the values of the entries of a matrix depend on the choice of a basis of the vector space where the matrix operates as a linear map. Can we change the coordinates so that a given matrix, with respect to this new coordinate system, is diagonal?[21]

21: The answer to this question is: "not always, but we can still do partial diagonalization".

The first step in answering this question requires the introduction of **eigenvalues** and **eigenvectors**.

- Let A be a square matrix of size n. Let $\mathbf{v} \neq \mathbf{0} \in \mathbb{R}^n$. We say that $\mathbf{v}$ is an **eigenvector** of A if

$$A\mathbf{v} = \lambda \mathbf{v}$$

for some scalar $\lambda \in \mathbb{C}$. The number λ is said to be the **eigenvalue** of A associated to the eigenvector $\mathbf{v}$.
- If $\mathbf{v} \neq \mathbf{0} \in \mathbb{R}^n$ is an eigenvector of A associated with eigenvalue λ, then so is $c\mathbf{v}$, $c \neq 0$. Indeed, if $A\mathbf{v} = \lambda\mathbf{v}$, then

$$A(c\mathbf{v}) = cA\mathbf{v} = c\lambda\mathbf{v} = \lambda(c\mathbf{v}).$$

By definition, the zero vector $\mathbf{0}$ cannot be an eigenvector. Also, note that for a given eigenvector, only one eigenvalue is associated to it.[22]

22: But eigenvalues/eigenvectors can be complex, even if the matrix only has real entries.

What happens when we apply a matrix to one of its eigenvector? A eigenvector spans a one dimensional vector space (a line), and **along this line** the matrix acts like a scalar, rescaling $\mathbf{v}$ by λ.

The goal of diagonalization is to transform the matrix to a form which is as close as possible to a diagonal; the best form would be a diagonal matrix, as we can see in the next exercise.

Example Let A be a diagonal matrix. Show that the eigenvalues of A are the diagonal values. What are the eigenvectors of A?

The matrix is of the form

$$A = \begin{bmatrix} a_{1,1} & 0 & \cdots & 0 \\ 0 & a_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{n,n} \end{bmatrix}.$$

For the vector

$$\mathbf{e}_k = (0, 0, \ldots, 0, 1, 0, \ldots, 0)^\top,$$

it is easy to verify that

$$A\mathbf{e}_k = a_{k,k}\mathbf{e}_k.$$

Hence $\mathbf{e}_k$ is the eigenvector with eigenvalue $\lambda_k = a_{k,k}$. ■

An eigenvector,[23] can come from only one eigenvalue. That is in fact almost obvious. Suppose that an eigenvector $\mathbf{v}$ of a matrix A satisfies the eigenvector equation with two different eigenvalues, which we call λ and μ, which is to say that

23: Or the 1-dimensional eigenspace spanned by it.

$$A\mathbf{v} = \lambda\mathbf{v} \quad \text{and} \quad A\mathbf{v} = \mu\mathbf{v}.$$

Since the two left-hand sides of the equations above are the same, it follows that $\lambda\mathbf{v} = \mu\mathbf{v}$. Since $\mathbf{v}$, being an eigenvector, is non-zero by definition, this last equation implies that $\lambda = \mu$.

Example Can two linearly independent eigenvector have the same eigenvalue? If you believe that this is true (which it is), prove it by finding an example of a a matrix which has the same eigenvalue for more than one independent eigenvector.

The zero matrix can be used, but let us take a non-trivial example. Fix any $\lambda \neq 0 \in \mathbb{R}$ and consider the matrix

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

The eigenvalue λ is associated to two linearly independent eigenvectors, $\mathbf{i} = (1,0,0)^\top$ and $\mathbf{j} = (0,1,0)^\top$;[24] the eigenvalue 0 is associated to the eigenvector $\mathbf{k} = (0,0,1)^\top$.[25] ■

But what do eigenvectors represent, geometrically?[26]

24: These are not the only two linearly independent eigenvectors, however.

25: In particular, $\mathbf{k}$ spans $\ker(A)$.

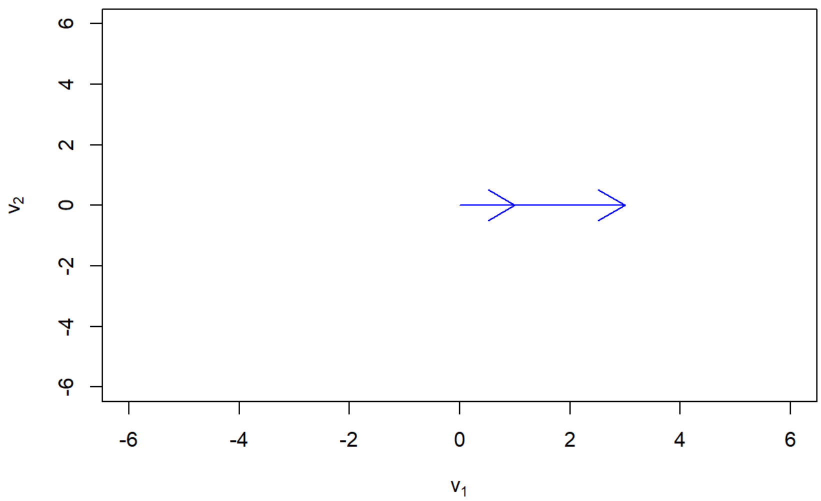
26: It is important to note that while we have illustrated the eigenconcepts with arrows in $\mathbb{R}^n$, any linear mapping of a vector space to another could have eigenvectors; in some cases eigenvectors are functions, not geometrical vectors.

Example Let

$$A = \begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix}.$$

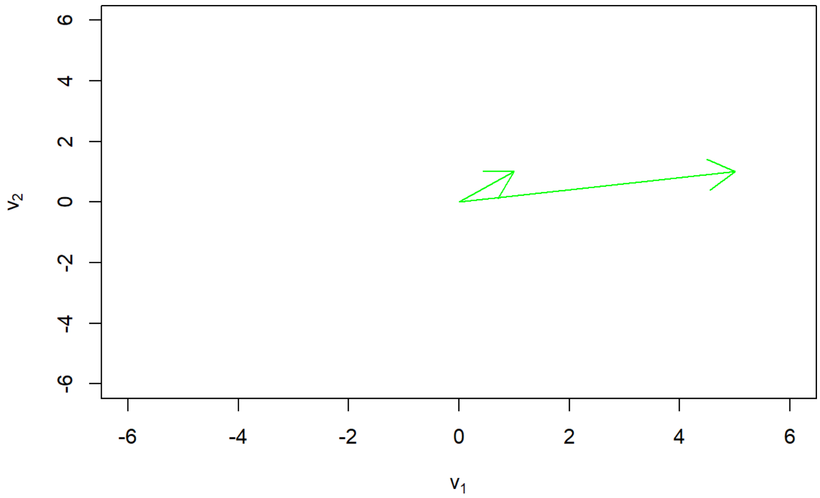
We can show that $\mathbf{v} = (1,0)^\top$ is an eigeventor of A, with eigenvalue 3, since $A\mathbf{v} = 3\mathbf{v}$. Applying A to $\mathbf{v}$ stretches it by a factor of 3, as seen below.

```
plot(NA,xlim=c(-6,6), ylim=c(-6,6),
     xlab = expression(list(v[1])),
     ylab=expression(list(v[2])))
arrows(0,0,1,0, col="blue")
arrows(0,0,3,0, col="blue")
```



But the vector $\mathbf{w} = (1,1)^\top$ is not an eigenvector of A since $A\mathbf{w} = (5,1)^\top \neq \lambda(1,1)^\top$, no matter the value of λ. Applying A to $\mathbf{w}$ does not only dilate it, it also **rotates** it.

```
plot(NA,xlim=c(-6,6), ylim=c(-6,6),
    xlab = expression(list(v[1])),
    ylab=expression(list(v[2])))
arrows(0,0,1,1, col="green")
arrows(0,0,5,1, col="green")
```



The previous examples are easy because the involved matrices are diagonal; finding the eigenvalues and eigenvectors of a general matrix will help us transform it to a form that is closer to a diagonal.

The recipe for finding the eigenvalues and eigenvector of a matrix A starts with constructing a polynomial equation, known as the **characteristic equation**, such that its roots are the eigenvalues of A.[27]

27: The characteristic equation is a direct consequence of the properties of determinant from Section 3.3.3.

Suppose that λ is an eigenvalue of A (the exact value does not matter): by definition, there is a non-zero eigenvector $\mathbf{v}$ such that $A\mathbf{v} - \lambda v = \mathbf{0}$, which can be re-written as

$$(A - \lambda \mathbf{I}_n)\mathbf{v} = \mathbf{0},$$

where $\mathbf{I}_n$ is the identity matrix with the same size as A.

The matrix $A - \lambda \mathbf{I}_n$ has therefore a non-zero nullspace, since it contains the nonzero vector $\mathbf{v}$. It follows that $A - \lambda \mathbf{I}_n$ is not invertible which means that its determinant is zero.

Hence, the eigenvalue λ is a solution of the **characteristic equation**

$$\det(A - \lambda \mathbf{I}_n) = 0.$$

The expression $\det(A - \lambda \mathbf{I}_n)$ is a polynomial in the variable λ, called the **characteristic polynomial of** A. The **degree** of the characteristic polynomial (its highest exponent in λ) is the size n of the A.

This works for all sizes n, but it is typically easier to find the eigenvalues when $2 \leq n \leq 4$, due to the insolvability of the quintic; for $n \geq 5$, we have to use numerical methods (see Chapter 4).

Example Write the characteristic polynomial of the matrix

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

and find its eigenvalues.

We need to apply the definition of the characteristic polynomial, expand the determinant, and simplify. The eigenvalues will be the roots of a quadratic equation, since A is of size 2.

$$\begin{aligned} \det(A - \lambda \mathbf{I}_2) &= \det\left(\begin{bmatrix} 1 & 4 \\ 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \\ &= \det\left(\begin{bmatrix} 1 & 4 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) \\ &= \det \begin{bmatrix} 1-\lambda & 4 \\ 1 & 2-\lambda \end{bmatrix} \\ &= (1-\lambda)(2-\lambda) - 4 = \lambda^2 - 3\lambda - 2. \end{aligned}$$

The eigenvalues of A are thus the solutions of the equation

$$\lambda^2 - 3\lambda - 2 = 0,$$

namely

$$\lambda_{1,2} = \frac{3 \pm \sqrt{17}}{2}.$$

In this example, both eigenvalues are real. ■

Let A be a square matrix, of any size, and suppose that $\mathbf{v}$ and $\mathbf{w}$ are two eigenvectors of A. Is their sum an eigenvector? What about a linear combination of them?

In general the sum is not an eigenvector. However, if $\mathbf{v}$ and $\mathbf{w}$ are associated **with the same eigenvalue** λ, then their sum is another eigenvector of A with the same eigenvalue, as the following calculations demonstrates:

$$A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w} = \lambda\mathbf{v} + \lambda\mathbf{w} = \lambda(\mathbf{v} + \mathbf{w}).$$

The sum $\mathbf{v} + \mathbf{w}$ is a linear combination; it should not be too difficult to show that a non-trivial linear combination $a\mathbf{v} + b\mathbf{w}$, $a, b \neq 0$ is not an eigenvector of A, unless $\mathbf{v}$ and $\mathbf{w}$ share their associated eigenvalue.

After we obtain the eigenvalues of A from the characteristic equation, the next step is to find the corresponding eigenvectors.

As before, we let $A \in \mathcal{M}_{n,n}$ and λ be an eigenvalue of A. The vector subspace of $\mathbb{R}^n$ spanned by all eigevenctors with this eigenvalue is **eigenspace** E_λ. The dimension E_λ, as a vector subspace of $\mathbb{R}^n$, is the **geometric multiplicity** of the associated eigenvalue λ.

The eigenspace corresponding to an eigenvalue is obtained by solving the homogeneous linear system $(A - \lambda\mathbf{I}_n)\mathbf{v} = \mathbf{0}$, where the unknown are the components of the eigenvector $\mathbf{v}$.

Example What are the eigenvectors of the matrix A from the previous example?

We already know the eigenvalues of A:

$$\lambda_{1,2} = \frac{3 \pm \sqrt{17}}{2}.$$

To find $\mathbf{v}_1$, the eigenvector of A associated to λ_1, we must solve the system

$$\begin{bmatrix} 1-\lambda_1 & 4 \\ 1 & 2-\lambda_1 \end{bmatrix} \begin{bmatrix} v_{1,1} \\ v_{1,2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Expanding the system gives two equations in the unknowns x, y (the components of the eigenvector $\mathbf{v}_1$).

$$\left(1 - \frac{3+\sqrt{17}}{2}\right) v_{1,1} + 4v_{1,2} = 0$$
$$v_{1,1} + \left(2 - \frac{3+\sqrt{17}}{2}\right) v_{1,2} = 0.$$

We expect this system to have a free variable, since the eigenspace has to be one dimensional.[28]

28: Why is that the case?

We can either get the solution *via* the Gauss-Jordan elimination algorithm or, we can solve directly by substitution since n is quite small. Proceeding with the second option, we solve both equations for $v_{1,2}$, and the second equation collapses into the first:

$$v_{1,2} = \frac{1+\sqrt{17}}{8} v_{1,1}.$$

As expected, we found a one dimensional eigenspace, parametrized by $v_{1,1}$. We can exhibit a basis for E_{λ_1} by selecting any non-zero eigenvector in this space; setting $v_{1,1} = 1$, we find

$$E_{\lambda_1} = \mathrm{Span}\{\mathbf{v}_1\} = \mathrm{Span}\left\{\begin{bmatrix} 1 \\ \frac{1+\sqrt{17}}{8}. \end{bmatrix}\right\}$$

Similar computations, which we let the reader perform, yield

$$E_{\lambda_2} = \mathrm{Span}\{\mathbf{v}_2\} = \mathrm{Span}\left\{\begin{bmatrix} 1 \\ \frac{1-\sqrt{17}}{8}. \end{bmatrix}\right\}$$

The **multiplicity** of an eigenvalue is linked to the number of times it appears as a solution of the characteristic equation. We can count properly the number of eigenvalues and eigenvector making use of this concept.

- An eigenvalue is a solution of the characteristic equation $\det(A - \lambda I)$: the multiplicity of the solution is called the **algebraic multiplicity** of the eigenvalue.
- It can be shown that the **geometric multiplicity**, i.e., the dimension of the associated eigenspace E_λ, is smaller than or equal to the **algebraic multiplicity** (defined above).

3.5.2 Similar Matrices

Eigenvectors define subspaces along which the matrix acts by scalar multiplication.[29] Once we have the eigenvectors, we apply a similarity transformation to transform our matrix to a "more diagonal one".

29: That is to say, by stretching or dilation.

Before proceeding, we need to define similarity of matrices: two square matrices A and B of the same size are said to be **similar** if there is an invertible matrix P such that

$$B = P^{-1}AP.$$

The transformation $A \longrightarrow B = P^{-1}AP$ is a **similarity transformation**.

Example Similarity is an **equivalence relation**, which means that it satisfies the 3 following properties:

1. **reflexivity** – A is similar to itself;
2. **symmetry** – A is similar to B if and only if B is similar to A;
3. **transitivity** – if A is similar to B and B is similar to C, then A is similar to C.

This exercise is more "theoretical" than our usual fare, but the proof is easy and it will help us familiarize ourselves with the algebra of matrices.

1. Let $P = \mathbf{I}$, the identity matrix of the same size of A: then

$$P^{-1}AP = \mathbf{I}^{-1}A\mathbf{I} = \mathbf{I}A\mathbf{I} = A.$$

2. Let $B = P^{-1}AP$ be the similarity relation. Then we can multiply both of its sides to the left by P and to the right by P^{-1}:

$$PBP^{-1} = PP^{-1}APP^{-1} = (PP^{-1})A(PP^{-1}) = \mathbf{I}A\mathbf{I} = A.$$

If we let $Q = P^{-1}$, we therefore obtain the similarity relation:

$$A = Q^{-1}BQ.$$

3. Let $B = P^{-1}AP$ and $C = Q^{-1}BQ$ be the hypothetical similarity relations. Substituting the second into the first yields:

$$C = Q^{-1}BQ = Q^{-1}(P^{-1}AP)Q = (Q^{-1}P^{-1})A(PQ) = (PQ)^{-1}A(PQ).$$

Hence C is similar to A.

It is important to respect the properties of matrix multiplication: for numbers (scalars), the similarity relation reduces directly to equality since $p^{-1}bp = p^{-1}pb = b$ for any number.

For matrices the similarity relation is not trivial, since the matrix product is not commutative... but it does satisfy the other "standard properties" of numbers.

In the proof of the second property above, for instance, we made use of the associative property of matrix multiplication.

3.5.3 Diagonalization

Now that we have defined the concept of similarity between matrices, we can conclude our discussion about eigenvalues and eigenvectors with the last step: the diagonalization of a matrix.

We say that a square matrix A is **diagonalizable** if it is similar to a diagonal matrix. That is, there exists an invertible matrix P such that

$$D = P^{-1}AP$$

is a **diagonal matrix**.

As discussed previously, a square matrix $A \in \mathbb{R}^n$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^n$. Thus, the matrix A is diagonalizable if and only if there exists a basis of $\mathbb{R}^n$ of eigenvectors of A, with respect to which the linear map is represented by a diagonal matrix.

The diagonal values of D are in fact the eigenvalues of A, as we will explain in detail soon.

Once the matrix is diagonal, it is "easy to use": a linear system associated to a diagonal matrix of size n, for example, is equivalent to n linear equations in one variable. The difficult part is to find the eigenvalues and eigenvectors, since we need to solve equations.[30]

30: Thankfully, we have already discussed how to do this.

Suppose that we found the matrix is diagonalizable, then what is the relation with the **eigenvalue problem**?

Let A be a **square** matrix of size n. Suppose that A is diagonalizable. Then A has n (possibly repeated) eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$. Denote by D the diagonal matrix of the eigenvalues,

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix},$$

and denote by P the matrix whose columns are the eigenvectors A (**presented in the same order as the eigenvalues!**):

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}.$$

The **diagonalization** of A is given by the **similarity transformation**:

$$D = P^{-1}AP.$$

An easy consequence of all this (which we will not prove) is that all **symmetric matrices** are diagonalizable. Moreover, if such a matrix only has real entries, then all of its eigenvalues are real.

Example Show that the equation $D = P^{-1}AP$ is equivalent to the equation $A = PDP^{-1}$

We multiply the two sides by P^{-1} from the left, P from the right:

$$P^{-1}AP = P^{-1}(PDP^{-1})P = (P^{-1}P)D(P^{-1}P) = IDI = D.$$

Example Prove that the matrix A below is diagonalizable. Diagonalize it. How are the eigenvalues related to the determinant?

$$A = \begin{bmatrix} 2 & 3 & 0.4 & 1 \\ 3 & -1.3 & 0.6 & 17 \\ 0.4 & 0.6 & 0.1 & -23 \\ 1 & 17 & -23 & 0 \end{bmatrix}$$

The matrix is symmetric, hence it is diagonalizable. We expect 4 real eigenvalues (some of which could be duplicates).

We could try to solve the problem by hand, but it would most likely be rather time-consuming. We use R to speed up the process.

```
D <- rbind(c(2,3,0.4,1),c(3,-1.3,0.6,17),
           c(0.4,0.6,0.1,-23),c(1,317,-23,0))
eigen(D)
det(D)
prod(eigen(D)$values)
```

```
eigen() decomposition
$values
[1] -77.8741054  76.0897048   2.9324699  -0.3480693

$vectors
             [,1]        [,2]        [,3]        [,4]
[1,] -0.005237695 -0.01940525  0.46752893 -0.26108024
[2,] -0.210057716 -0.20443758  0.06120210  0.07064722
[3,]  0.278113608  0.28194691  0.87637211  0.96259458
[4,]  0.937283918 -0.93719510 -0.09819847  0.01605501

[1] 6048.09

[1] 6048.09
```

The output of the first two lines of codes produces the set of eigenvectors and eigenvalues. In particular, A is transformed to the diagonal matrix D *via* the eigenvector matrix P. The third line computes the determinant of the matrix, which we see is the same as the product of the eigenvalues of A, as shown by the fourth line of code.[31]

31: This will always be the case.

Invariance of the Determinant

The value of the determinant is respected by similarity transformation: if A and B are similar matrices, then $\det(A) = \det(B)$. We can use this fact to prove that the determinant of a diagonalizable matrix is the product of its eigenvalues.

To prove the fist part, we use the property that the determinant respects the product and inverses: $\det(P^{-1}AP) = \det(P)^{-1}\det(A)\det(P) = \det(A)$. From here, the second part is clear, since for a diagonal matrix the determinant is the product of the diagonal entries.

But we must be careful: **not every square matrix is diagonalizable**!

Example For any $t \neq 0$, the matrix $T = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$ only has one eigenvector $(0,1)^{\mathsf{T}}$, with eigenvalue t. The similarity recipe presented above can thus not be applied.

While the matrix is not diagonalizable, we can still construct its **Jordan normal form**, which is a more general version of a diagonal matrix [3].[32]

32: This is a topic for an advanced linear algebra course; we will not address it here.

3.6 Exercises

1. The augmented matrix $[A|B]$ of a system has 15 rows and 18 columns. Assume $\operatorname{rank}(A) = 12$ and $\operatorname{rank}([A|B]) = 13$. Which of the following statements is necessarily true?
   a) The system is inconsistent.
   b) The system has more than one solution, expressed with one parameter.
   c) The system has more than one solution, expressed with two parameters.
   d) The system has a unique solution.
   e) The system has more than one solution, expressed with three parameters.
   f) The system has more than one solution, expressed with four parameters.
2. Find all values of b for which the following system is consistent:
$$\begin{aligned} x + y - z &= 2 \\ x + 2y + z &= 3 \\ x - 3z &= 2b - 1 \end{aligned}$$
3. Find all the values of h for which the following vectors are linearly independent:
$$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ h \end{bmatrix}.$$
4. Which of the following sets are subspaces of $\mathbb{R}^2$?
$$\begin{aligned} S &= \{(x,y) \in \mathbb{R}^2 \mid 2x - y = 1\} \\ T &= \operatorname{Span}\{(-1,1),(2,-1)\} \\ U &= \{(x,y) \in \mathbb{R}^2 \mid y = x^2\} \\ V &= \{(x,y) \in \mathbb{R}^2 \mid x - 3y = 0\} \end{aligned}$$

5. A is a 3×3 matrix. Suppose that $\det(A) = 3$. What is $\det(2A^{\top}A)$? (Hint: $A^{\top}$ is the transposed of A.)
6. Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}$. Which of the following statements is true?
   a) $AB = \begin{pmatrix} -1 & 1 & 2 \\ 5 & 4 & 5 \end{pmatrix}$
   b) $BA = \begin{pmatrix} -1 & 1 & 2 \\ 5 & 4 & 5 \end{pmatrix}$
   c) $BA = \begin{pmatrix} 3 & 3 & 4 \\ 3 & 0 & -1 \end{pmatrix}$
   d) $AB = \begin{pmatrix} 3 & 3 & 4 \\ 3 & 0 & -1 \end{pmatrix}$
   e) $BA = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$
   f) $AB = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$
7. What is the determinant of $\begin{pmatrix} 0 & 0 & 0 & 5 & 0 \\ 2 & 0 & 3 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 7 & 0 & 0 & 0 \end{pmatrix}$?
8. Let a, b, c, d, e, f be the constants and x, y be the unknowns of the system

$$\begin{aligned} ax + by &= e \\ cx + dy &= f. \end{aligned}$$

   a) What condition(s) on a, b, c, d, e, f are needed in order for the system to have a unique solution?
   b) What condition(s) on a, b, c, d, e, f are needed in order for the system to have infinitely many solutions?
   c) What condition(s) on a, b, c, d, e, f are needed in order for the system to have no solution?
9. Let $B = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}$. Find all 2×2 matrices A that satisfy $AB = BA$.
   (Hint: write $A = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$, and compute AB and BA. Then, solve the system of 4 equations in 4 unknowns that arises from $AB = BA$.)
10. Consider the matrix $A = \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}$.
    a) Find the eigenvalues of A.
    b) For each eigenvalue of A, find the corresponding eigenspace of A, and state its dimension.
11. Consider the matrix $A = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ -1 & 1 & 0 & -5 & 1 \\ 1 & -1 & 0 & 5 & 1 \\ 2 & -2 & -1 & 9 & 0 \end{pmatrix}$, whose re-

duced row echelon form is

$$\tilde{A} = \begin{bmatrix} 1 & -1 & 0 & 5 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

a) Find the column space of A? (Hint: find the columns of A that are necessary to express the column space of A.)
b) Are the columns of A linearly independent?
c) What is the dimension of the column space of A?
d) Find a basis for the nullspace of A.
e) Does the system $Ax = 0$ have a unique solution?

12. Find all values of x for which $\det \begin{bmatrix} 1 & x & x \\ -x & -2 & -x \\ x & -x & -3 \end{bmatrix} = 0$.
13. Let V be a vector space and let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$. Which of the following statements are true?

13.. If $\{\mathbf{u}, \mathbf{v}\}$ is linearly independent, so is $\{\mathbf{u}, \mathbf{v}, \mathbf{u} + \mathbf{v}\}$.
13.. If $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent, so is $\{\mathbf{u}, \mathbf{v}\}$.
13.. If $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent, so is $\{\mathbf{u}, \mathbf{v}\}$.
13.. If $\{\mathbf{u}, \mathbf{v}\}$ is linearly independent, so is $\{\mathbf{u}, \mathbf{u} + \mathbf{v}\}$.

14. Which of the following statements are true?

a) The set $\{(x, x-1, y) \in \mathbb{R}^3 \mid x, y \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^3$.
b) The set $\{p(x) \in \mathbb{P}_4 \mid p(2) = 0\}$ is a subspace of $\mathbb{P}_4$.
c) The set $\{A \in \mathcal{M}_{2,2} \mid A^2 = A\}$ is not a subspace of $\mathcal{M}_{2,2}$.

15. Let $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ be a set of linearly independent vectors. Which of the following sets of vectors are linearly dependent?

a) $\{\mathbf{u} + \mathbf{v}, \mathbf{v} + \mathbf{w}, \mathbf{w} + \mathbf{u}\}$
b) $\{\mathbf{u}, \mathbf{u} + \mathbf{z}, \mathbf{v}, \mathbf{v} + \mathbf{w}\}$
c) $\{\mathbf{u} - \mathbf{v}, \mathbf{v} - \mathbf{w}, \mathbf{w} - \mathbf{z}, \mathbf{z} - \mathbf{u}\}$
d) $\{\mathbf{u}, \mathbf{u} + \mathbf{z}, \mathbf{z}\}$

16. If $\det \begin{bmatrix} 3 & -1 & x \\ 2 & 6 & y \\ -5 & 4 & z \end{bmatrix} = ax + by + cz$, what is the value of c?
17. Let A, B, C be square $n \times n$ matrices with $\det(A) = 1$, $\det(B) = 4$ and $\det(C) = -3$. What is the value of $\det(A^2BC^\top B^{-1})$?
18. For each of the following subspaces, exhibit a basis and find the dimension.

a) $\{(x, y, z, w) | x - y + z - w = 0\}$
b) $\{A \in \mathcal{M}_{2,2} \mid A^\top = -A\}$

19. Let $A = \begin{bmatrix} 2 & -1 & 0 \\ -3 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$.

a) Find $c_A(\lambda)$, the characteristic polynomial of A.
b) Use your answer in (a) to determine the eigenvalues of A.
c) Find a basis for two of the eigenspaces of A.

20. Let U and W be subspaces of V. Define

$$U \cup W = \{\mathbf{v} \in V \mid \mathbf{v} \in U \text{ or } \mathbf{v} \in W\}$$
$$U \cap W = \{\mathbf{v} \in V \mid \mathbf{v} \in U \text{ and } \mathbf{v} \in W\}.$$

   a) Show that $U \cap W$ is a subspace of V.
   b) Is $U \cup W$ necessarily a subspace of V? Explain.

21. The *trace* of a matrix A, denoted by $\operatorname{tr}(A)$, is the sum of the elements on the diagonal of A. Thus, $\operatorname{tr}\begin{bmatrix} x & y \\ z & w \end{bmatrix} = x + w$.
   a) Show that $\operatorname{tr} : \mathcal{M}_{2,2} \to \mathbb{R}$ is linear, that is, show that

$$\operatorname{tr}\left[a\begin{bmatrix} x_1 & y_1 \\ z_1 & w_1 \end{bmatrix} + b\begin{bmatrix} x_2 & y_2 \\ z_2 & w_2 \end{bmatrix}\right] = a\operatorname{tr}\begin{bmatrix} x_1 & y_1 \\ z_1 & w_1 \end{bmatrix} + b\operatorname{tr}\begin{bmatrix} x_2 & y_2 \\ z_2 & w_2 \end{bmatrix}$$

   for all $a, b, x_i, y_i, z_i, w_i \in \mathbb{R}$.
   b) Let $x \in \mathbb{R}$. Find a matrix $A \in \mathcal{M}_{2,2}$ such that $\operatorname{tr}(A) = x$.
   c) Using the Rank-Nullity Theorem and the result from part b), can you deduce the value of $\dim(\ker(\operatorname{tr}))$?

22. Let $A = \begin{bmatrix} 2 & -1 & 0 \\ -3 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$. Find $\operatorname{rowsp}(A)$ (the space spanned by the rows of A), $\operatorname{colsp}(A)$ and $\operatorname{nullsp}(A)$.

23. Find (if possible) conditions on a, b and c such that the system

$$x + ay = 0, \quad y + bz = 0, \quad z + cx = 0.$$

   has:

   a) no solution.
   b) one solution. What is the solution in this case?
   c) infinitely many solutions. What are the solutions in this case?

24. Amongst the following vectors, which one is a linear combination of $(1, 0, 0)$ and $(0, 1, 1)$?

$$(1, 2, 3), \quad (1, 0, 1), \quad (0, 0, 1), \quad (1, 1, 1), \quad (0, 1, 0), \quad (3, 2, 1).$$

25. Let $T : \mathbb{R}^2 \to \mathbb{R}$ be a linear transformation. If $T(1, 2) = 3$ and $T(1, 0) = -1$, what is $T(1, 1)$?

26. Amongst

$$U = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}; \quad V = \{(x, y) \in \mathbb{R}^2 \mid x + y \leq 0\};$$
$$W = \{(x, y) \in \mathbb{R}^2 \mid x = 2y\},$$

   which sets are subspaces of $\mathbb{R}^2$?

Chapter References

[1] P. Boily, S. Davies, and J. Schellinck. *The Practice of Data Visualization*. Data Action Lab, 2023.

[2] M. Nevins, T. Giordano, and B. Jessup. *Vector Spaces First*. 4th. 2021.

[3] W.K. Nicholson. *Linear Algebra with Applications*, *3rd Edition*. PWS Publishing Company, 1994.

[4] G. Strang. *Introduction to Linear Algebra*. Wellesley, 2016.