Chapter 1

Basic Linear Algebra

In this chapter, we will introduce the student to the basic knowledge of linear algebra, which we adapted from the textbooks from Trefethen and Bau[1] and the classical book of Arfken[2]. Then, we start our presentation of introductory linear algebra techniques by defining the solution to simple linear systems.

1.1 Linear Systems and Gauss-Elimination

Example 1.1 (From Nine Chapters on the Mathematical Art[3])**.** *The problem of the two scales: If one plum and three oranges weight a total of* 750 *g in the first scale, and two plums and one orange weight a total of* 500 *g in the other. How much does a single orange and plum weights?*

Solution: Using Fig.1.1, a) Double contents of the first scale. b) Substract from this the quantities of the second scale. Therefore 1 orange weights 200 g. c) Take orange off the second scale. Therefore plum weights 150 g.

[1] L. N. Trefethen and D. Bau, Numerical Linear Algebra, Society for Industrial and Applied Mathematics, 1997.

[2] G. Arfken, Mathematical Methods for Physics, Academic Press, 1970

[3] A textbook called "Jiuzhang Suanshu" or "Nine Chapters on the Mathematical Art" (written over some time from about 200 BC onwards, probably by a variety of authors) became an essential tool in the education of such civil service, covering hundreds of problems in practical areas such as trade, taxation, engineering and the payment of wages. Moreover, it was crucial as a guide to solving equations using a sophisticated matrix-based method that did not appear in the West until Carl Friedrich Gauss re-discovered it at the beginning of the 19th Century (and is now known as Gaussian elimination).

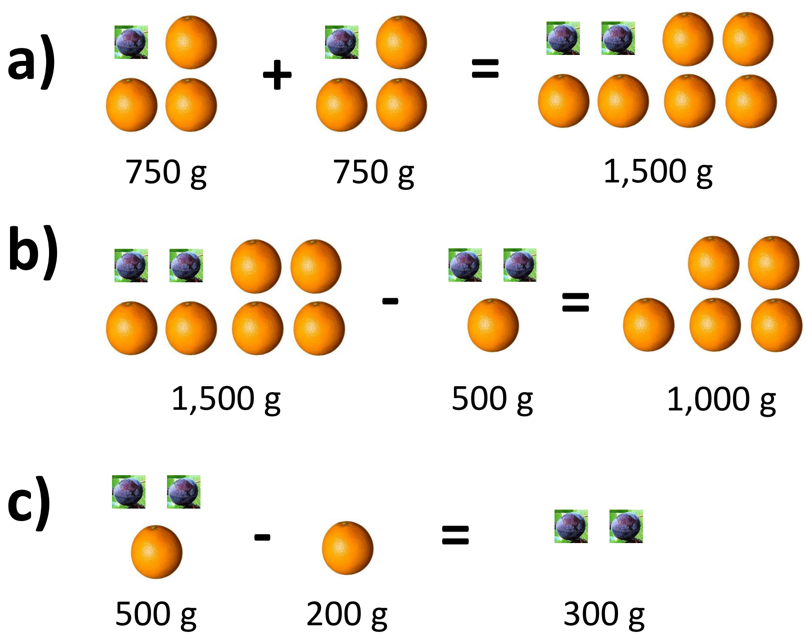

Figure 1.1: Method of solution to example 1.1.

These type of problems can be written in a more abstract form as a linear system

Example 1.2.

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

Solution: We can solve easily this system by adding both equations so $3x = 2$ then $x = 2/3$. The second equation is used to obtain $y = x - 2 = -4/3$.

Example 1.3.

$$
\begin{aligned}
2x_1 + x_2 &= 0 \\
4x_1 + 2x_2 &= 2
\end{aligned}
$$

Solution: Here we have no solution. Multiply first equation by -2 then add them so we obtain $0 = 2$.

Example 1.4.

$$
\begin{aligned}
2x_1 + x_2 &= 1 \\
4x_1 + 2x_2 &= 2
\end{aligned}
$$

Solution: Here we have infinite solutions. Multiply first equation by -2 then add them so we obtain $0 = 0$.

For a system of two variables, each equation represents the plot of a line. The intersection of two lines gives the solution of the system. In general there are three possibilities for the solution (see Fig.1.2): a) one solution, b) no solution and c) infinite solutions. In general, linear systems of two or more equations can be written in matrix form as

$$
\begin{pmatrix}
a_{11} & a_{12} & \dots & a_{1m} \\
a_{21} & a_{22} & \dots & a_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nm}
\end{pmatrix}
\begin{pmatrix}
x_1 \\ x_2 \\ \vdots \\ x_m
\end{pmatrix}
=
\begin{pmatrix}
b_1 \\ b_2 \\ \vdots \\ b_n
\end{pmatrix}. \tag{1.1}
$$

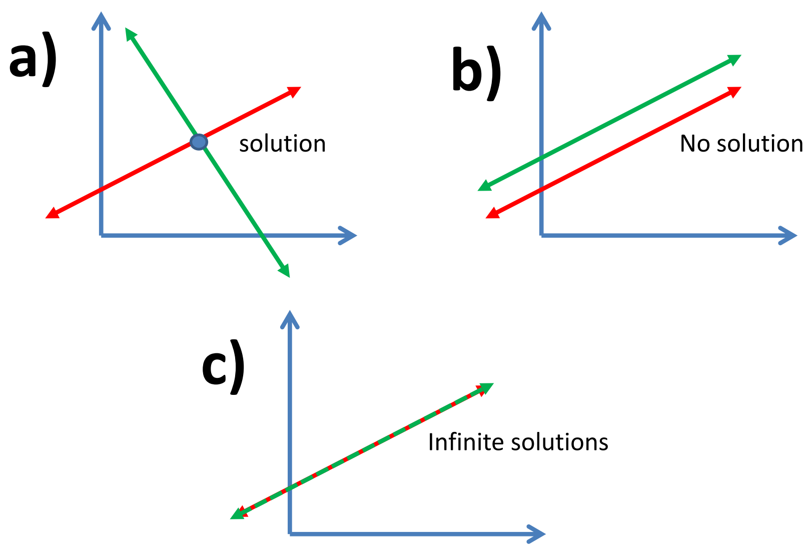

Figure 1.2: Geometrical representation of a solution to a linear system of two variables.

The process of finding the solution of a matrix system using a sequence of elementary row operations to modify the matrix until the lower left-hand corner of the matrix is filled with zeros, as much as possible. There are three types of elementary row operations:

1. Swapping two rows,
2. Multiplying a row by a non-zero number,
3. Adding a multiple of one row to another row.

Using these operations, a matrix can always be transformed into an **upper triangular matrix**, and one that is in **row echelon form**. Once all of the leading coefficients (the left-most non-zero entry in each row) are 1, and in every column containing a leading coefficient has zeros elsewhere, the matrix is said to be in **reduced row echelon form**.

Example 1.5.

$$\begin{aligned} 2x_1 + 3x_2 + x_3 &= 2 \\ 4x_1 + 2x_2 + x_3 &= 1 \\ -4x_1 + 3x_2 + 4x_3 &= -1 \end{aligned}$$

Solution: We write in matrix form

$$\begin{pmatrix} 2 & 3 & 1 \\ 4 & 2 & 1 \\ -4 & 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$$

and use the simplified form to apply Gauss elimination

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 2 \\ 4 & 2 & 1 & 1 \\ -4 & 3 & 4 & -1 \end{array}\right)$$

Then the Gauss-Elimination process is

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 2 \\ 4 & 2 & 1 & 1 \\ -4 & 3 & 4 & -1 \end{array}\right) \Rightarrow_{(2)+(3)->(3)} \left(\begin{array}{ccc|c} 2 & 3 & 1 & 2 \\ 4 & 2 & 1 & 1 \\ 0 & 5 & 5 & 0 \end{array}\right) \Rightarrow_{2(1)-(2)->(2)}$$

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 2 \\ 0 & 4 & 1 & 3 \\ 0 & 5 & 5 & 0 \end{array}\right) \Rightarrow_{5(2)-4(3)->(3)} \left(\begin{array}{ccc|c} 2 & 3 & 1 & 2 \\ 0 & 4 & 1 & 3 \\ 0 & 0 & -15 & 15 \end{array}\right)$$

Here the matrix is reduced to an upper triangular matrix or row echelon form. This form will be enough to solve the matrix system.

In addition the system can be reduced further

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 2 \\ 0 & 4 & 1 & 3 \\ 0 & 0 & -15 & 15 \end{array}\right) \Rightarrow_{-(3)/15->(3)}$$
$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 2 \\ 0 & 4 & 1 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right) \Rightarrow_{(2)-(3)->(2)} \left(\begin{array}{ccc|c} 2 & 3 & 1 & 2 \\ 0 & 4 & 0 & 4 \\ 0 & 0 & 1 & -1 \end{array}\right) \Rightarrow_{(1)-(3)->(1)}$$
$$\left(\begin{array}{ccc|c} 2 & 3 & 0 & 3 \\ 0 & 4 & 0 & 4 \\ 0 & 0 & 1 & -1 \end{array}\right) \Rightarrow_{(2)/4->(2)} \left(\begin{array}{ccc|c} 2 & 3 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right) \Rightarrow_{(1)-3(2)->(1)}$$
$$\left(\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right) \Rightarrow_{(1)/2->(1)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right)$$

This is the reduced row echelon form. From this reduced system we get that $x_1 = 0$, $x_2 = 1$ and $x_3 = -1$.

Example 1.6.

$$\begin{aligned} 2x_1 + 3x_2 + x_3 &= 1 \\ 4x_1 + 6x_2 + 2x_3 &= 2 \\ -x_1 + x_2 + 4x_3 &= 4 \end{aligned}$$

Solution: The Gauss-Elimination process is

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 1 \\ 4 & 6 & 2 & 2 \\ -1 & 1 & 4 & 4 \end{array}\right) \Rightarrow_{(2)+4(3)->(3)} \left(\begin{array}{ccc|c} 2 & 3 & 1 & 1 \\ 4 & 6 & 2 & 2 \\ 0 & 10 & 18 & 18 \end{array}\right) \Rightarrow_{2(1)-(2)->(2)}$$
$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 10 & 18 & 18 \end{array}\right) \Rightarrow_{(2)<->(3)} \left(\begin{array}{ccc|c} 2 & 3 & 1 & 1 \\ 0 & 10 & 18 & 18 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

which means that as in example 1.4 the system have infinite solutions.

Homework 1.1. *Use Gaussian-elimination to solve the following systems*

$$a)\ \begin{array}{r} x_1 + x_2 = 1 \\ x_2 = 2 \end{array}, \quad b)\ \begin{array}{r} x_1 = 6 \\ x_1 - 2x_2 = 0 \end{array}, \quad c)\ \begin{array}{r} x_1 + 2x_2 = 3 \\ 3x_1 + 4x_2 = 7 \end{array}$$

Homework 1.2. *Use Gaussian-elimination to solve the following systems*

$$a)\ \begin{array}{r} x_1 + 2x_2 = 1 \\ 2x_1 + 4x_2 = 2 \end{array}, \quad b)\ \begin{array}{r} x_1 + 2x_2 = 1 \\ 2x_1 + x_2 = -1 \end{array}, \quad c)\ \begin{array}{r} x_1 + 2x_2 = 1 \\ 3x_1 + 4x_2 = 1 \end{array}$$

Homework 1.3. *Use Gaussian-elimination to solve the following systems*

$$a)\ \begin{array}{r} 2x_1 + x_3 = 1 \\ 4x_1 + 6x_2 + 2x_3 = 2 \\ -x_1 + x_2 + 4x_3 = 4 \end{array}, \quad b)\ \begin{array}{r} 2x_1 + 3x_2 + x_3 = 6 \\ x_1 - 2x_2 + x_3 = 0 \\ -x_1 + x_2 + 4x_3 = 4 \end{array}, \quad c)\ \begin{array}{r} x_1 + 3x_2 = 7 \\ x_1 + 2x_2 + x_3 = 8 \\ x_1 + x_2 + x_3 = 6 \end{array}$$

1.2 Orthogonal Vectors

The results of the theory of linear equations can be expressed concisely by the notation of vector analysis. A system of n real (or complex) numbers is called an n-dimensional vector and is denoted by the bold face letter $\mathbf{x}$. The numbers x_i, $i = 1, ..., n$ are called the components of the vector $\mathbf{x}$. We will use the notation $\mathbf{x} \in \mathbb{R}^n$ (or specifically $\mathbb{R}^{n\times 1}$) to denote the column vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

We use the symbol $\mathbb{R}$ to denote real numbers, so we assume that the components of the vector are real. Similarly we could assume that the components of the vector are complex numbers which will not change the results described in this chapter. If all components vanish, the vector is said to be zero $\mathbf{0}$ or the *null vector*. For $n = 2$ or $n = 3$ a vector can be interpreted geometrically as a "position vector" leading from the origin to the point with the rectangular coordinates. For $n > 3$ the geometrical visualization is no longer possible but the geometrical terminology remains suitable.

The *inner product* between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ is defined as

$$(\mathbf{x} \cdot \mathbf{y}) := \mathbf{x}^T \mathbf{y} = \sum_{j=1}^{n} x_j y_j = x_1 y_1 + x_2 y_2 + ... + x_n y_n.$$

Here the upperscript "T" denote the transpose of a vector which transform the column vector into a row vector ($\mathbf{x}^T \in \mathbb{R}^{1\times m}$). The euclidian length of a vectors $\mathbf{x} \in \mathbb{R}^m$ is defined as

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}.$$

The inner product between two vectors can be express in terms of the angle θ between them

$$\mathbf{x}^T\mathbf{y} = \cos\theta\|\mathbf{x}\|_2\|\mathbf{y}\|_2. \tag{1.2}$$

MATLAB programming note 1.1.

We can define the column vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$

```
>> x = [1;2;3;4;5];
>> y = [-2;3;5;6;7];
>> z = [2;0;1;8;1];
```

We can compute the dot products $(\mathbf{x}\cdot\mathbf{y})$, $(\mathbf{y}\cdot\mathbf{z})$ as

```
>> x'*y
ans =
    78
>> y'*z
ans =
    56
```

The inner product is *bilinear*, which means that it is linear in each vector separately

$$\begin{aligned}(\mathbf{x}+\mathbf{y})^T\mathbf{z} &= \mathbf{x}^T\mathbf{z} + \mathbf{y}^T\mathbf{z},\\ \mathbf{x}^T(\mathbf{y}+\mathbf{z}) &= \mathbf{x}^T\mathbf{y} + \mathbf{x}^T\mathbf{z},\\ (\alpha\mathbf{x})^T(\beta\mathbf{y}) &= \alpha\beta\mathbf{x}^T\mathbf{y}.\end{aligned}$$

A set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\} \in \mathbb{R}^{m\times 1}$ is *linear independent* if no vector in the set is a linear combination of the others. That is,

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + ... + c_n\mathbf{x}_n = \mathbf{0}$$

only for $c_k = 0$ for every k. Finally, a set of non-zeros vectors $\{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_m\} \in \mathbb{R}^{m\times 1}$ is called a *basis* in $\mathbb{R}^m$ if for each $\mathbf{x} \in \mathbb{R}^{m\times 1}$ we have that

$$\mathbf{x} = \sum_{k=1}^{n} c_k\mathbf{q}_k,$$

where c_k are certain coefficients.

Example 1.7. *The set of vectors*

$$\mathbf{x_1} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{x_2} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{x_3} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

is linear independent.

Solution: By definition of linear independence we have to show that the solution of the system

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

is $c_1 = c_2 = c_3 = 0$. We use Gauss-Elimination, so we write the system in reduced form and obtain

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right) \Rightarrow_{(1)-(2)->(2)}$$
$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right) \Rightarrow_{(2)<->(3)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

which proves that the vectors are linearly independent.

Example 1.8. *The set of vectors*

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 2 \\ 4 \\ 4 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

is not linear independent.

Solution: We write the system in reduced form and obtain

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 1 & 4 & 2 & 0 \\ 0 & 4 & 2 & 0 \end{array}\right) \Rightarrow_{(1)-(2)->(2)}$$
$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & -2 & -1 & 0 \\ 0 & 4 & 2 & 0 \end{array}\right) \Rightarrow_{2(2)+(3)->(3)} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & -2 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

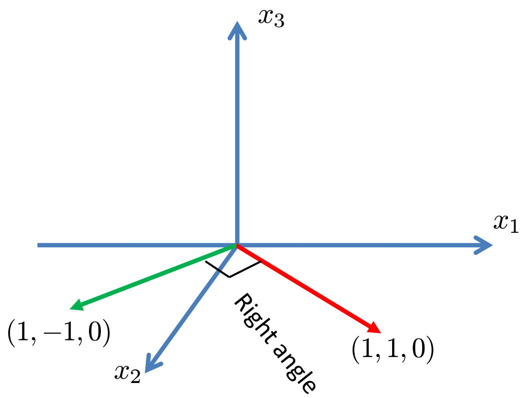

Figure 1.3: Geometrical of orthogonality between two vectors.

From the row echelon form we get $c_3 = -2c_2$, $c_1 = 0$. So the coefficients can be written as

$$\mathbf{c} = \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 \\ 2 \\ -4 \end{pmatrix}, \quad \text{etc.}$$

A pair of vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ are *orthogonal* if $(\mathbf{x} \cdot \mathbf{y}) = 0$. This means that they lie at right angles to each other in $\mathbb{R}^m$. A set $\mathcal{S}$ of non-zeros vectors $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\} \in \mathbb{R}^m$ is orthogonal if each of its elements are pairwise orthogonal. The set $\mathcal{S}$ is *orthonormal* if is orthogonal and in addition $\|\mathbf{x}_k\|_2 = 1$ for every k.

Example 1.9. *The set of vectors*

$$\mathbf{x_1} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{x_2} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \mathbf{x_3} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

is orthogonal.

Solution: It is easy to check that $\mathbf{x}_1^T\mathbf{x}_2 = \mathbf{x}_2^T\mathbf{x}_3 = \mathbf{x}_1^T\mathbf{x}_3 = 0$. Also notice that $\mathbf{x}_1^T\mathbf{x}_1 = 2$, $\mathbf{x}_2^T\mathbf{x}_2 = 2$ and $\mathbf{x}_3^T\mathbf{x}_3 = 4$.

Theorem 1.1. *An orthogonal set of vectors* $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\} \in \mathbb{R}^{m\times 1}$ $(m \geq n)$ *is linear independent.*

Proof. Let

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + ... + c_n\mathbf{x}_n = \mathbf{0}.$$

Using the orthogonality, we take the inner product for $\mathbf{x}_1$ to obtain

$$0 = c_1\mathbf{x}_1^T\mathbf{x}_1 + c_2\mathbf{x}_1^T\mathbf{x}_2 + ... + c_n\mathbf{x}_1^T\mathbf{x}_n = c_1\|\mathbf{x}_1\|_2^2,$$

which yields that $c_1 = 0$. Notice that we can do the same process for all vectors, which yields that $c_1 = ... = c_n = 0$. This proves that the vectors are linear independent. $\square$

The most important idea from the concepts of inner products and orthogonality is that inner products can be used to decompose arbitrarily vectors into orthogonal components. For example, suppose that $\mathbf{q}_1, \mathbf{q}_2$ are two orthonormal vectors and let $\mathbf{v}$ be an arbitrarily vector. The inner products $(\mathbf{q}_1 \cdot \mathbf{v})$ or $(\mathbf{q}_2 \cdot \mathbf{v})$ are scalars. Utilizing these scalars as coordinates of expansion, we find that the vector

$$\mathbf{r} = \mathbf{v} - (\mathbf{q}_1 \cdot \mathbf{v})\,\mathbf{q}_1 - (\mathbf{q}_2 \cdot \mathbf{v})\,\mathbf{q}_2$$

is orthogonal to $\mathbf{q}_1, \mathbf{q}_2$. This can be easily verified by calculating

$$\begin{aligned}
\mathbf{q}_1^T\mathbf{r} &= \mathbf{q}_1^T\mathbf{v} - (\mathbf{q}_1 \cdot \mathbf{v})\,\mathbf{q}_1^T\mathbf{q}_1 - (\mathbf{q}_2 \cdot \mathbf{v})\,\mathbf{q}_1^T\mathbf{q}_2 = (\mathbf{q}_1 \cdot \mathbf{v}) - (\mathbf{q}_1 \cdot \mathbf{v}) = 0,\\
\mathbf{q}_2^T\mathbf{r} &= \mathbf{q}_2^T\mathbf{v} - (\mathbf{q}_1 \cdot \mathbf{v})\,\mathbf{q}_2^T\mathbf{q}_1 - (\mathbf{q}_2 \cdot \mathbf{v})\,\mathbf{q}_2^T\mathbf{q}_2 = (\mathbf{q}_2 \cdot \mathbf{v}) - (\mathbf{q}_2 \cdot \mathbf{v}) = 0.
\end{aligned}$$

This idea has been apply for the well known *Gramm-Schmidt orthogonalization process*. A set of linearly independent non-zero vectors $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\} \in \mathbb{R}^m$ $(m \geq n)$, can be transform into a set of orthogonal (or orthonormal) vectors $\{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n\} \in \mathbb{R}^m$. The first vector is set as

$$\mathbf{q}_1 = \mathbf{x}_1/\|\mathbf{x}_1\|_2.$$

Then set

$$\mathbf{r}_2 = \mathbf{x}_2 - \left(\mathbf{q}_1^T\mathbf{x}_2\right)\mathbf{q}_1, \quad \mathbf{q}_2 = \mathbf{r}_2/\|\mathbf{r}_2\|_2.$$

It is trivial to notice that $\mathbf{q}_2$ is orthonormal to $\mathbf{q}_1$. This process is repeated as

$$\begin{aligned}
\mathbf{r}_j &= \mathbf{x}_j - \left(\mathbf{q}_1^T\mathbf{x}_j\right)\mathbf{q}_1 - ... - \left(\mathbf{q}_{j-1}^T\mathbf{x}_j\right)\mathbf{q}_{j-1},\\
\mathbf{q}_j &= \mathbf{r}_j/\|\mathbf{r}_j\|_2, \quad 3 \leq j \leq n.
\end{aligned}$$

MATLAB programming note 1.2.

We will perform the Gramm-Schmidt Orthogonalization process over the column vectors **x**, **y** and **z**. For the first step

```
>> q1 = x/norm(x)
q1 =
    0.1348
    0.2697
    0.4045
    0.5394
    0.6742
```

for the next step

```
>> r = y - (q1'*y)*q1;
>> q2 = r/norm(r)
q2 =
   -0.9714
    0.0465
    0.2119
    0.0930
   -0.0258
```

Finally we obtain

```
>> r = z - (q1'*z)*q1 - (q2'*z)*q2;
>> q3 = r/norm(r)
q3 =
    0.0416
   -0.2437
   -0.1772
    0.8296
   -0.4682
```

We can check the orthonormality

```
>> q1'*q2
ans =
   2.5674e-16

>> q1'*q3
ans =
   1.1102e-16

>> q2'*q3
ans =
  -2.4807e-16
```

Notice that the inner product is not exactly zero, but is a very small number that approximates 0 (is usually called the "machine epsilon").

Corollary 1.1. *An orthonormal set of vectors* $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m\} \in \mathbb{R}^{m\times 1}$ *it is basis for* $\mathbb{R}^m$.

Homework 1.4. *Determine if the following set of vectors are linear independent*

$$
\begin{aligned}
&a) \qquad \mathbf{x_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{x_2} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \\
&b) \qquad \mathbf{x_1} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \mathbf{x_2} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \\
&c) \qquad \mathbf{x_1} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{x_2} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{x_3} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}
\end{aligned}
$$

Homework 1.5. *Determine if the following set of vectors are orthogonal*

$$
\mathbf{x_1} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{x_2} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad \mathbf{x_3} = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}
$$

Homework 1.6. *Show that the resultant set of vectors* $\{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n\} \in \mathbb{R}^{m\times 1}$ *from the Gramm-Schmidt Orthogonalization Process is orthogonal.*

Homework 1.7. *The Pythagorean theorem asserts that for a set of* n *orthogonal vectors* $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$,

$$
\left\| \sum_{j=1}^{n} \boldsymbol{x}_j \right\|_2^2 = \sum_{j=1}^{n} \|\boldsymbol{x}_j\|_2^2 .
$$

Project 1.1. *Prove Corollary 1.1.*

1.3 Basic Matrix Theory

We denote as $[\mathbf{A}] \in \mathbb{R}^{m\times n}$, the matrix

$$
[\mathbf{A}] = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & ... & a_{mn} \end{pmatrix},
$$

with m rows and n columns. A linear system can be written as

$$[\mathbf{A}]\,\mathbf{x} = \mathbf{b}$$

where $[\mathbf{A}] \in \mathbb{R}^{m\times n}$, $\mathbf{x} \in \mathbb{R}^{n\times 1}$ and $\mathbf{b} \in \mathbb{R}^{m\times 1}$. The mapping $\mathbf{x} \mapsto [\mathbf{A}]\,\mathbf{x}$ is linear, which means that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n\times 1}$ and any scalar α,

$$\begin{aligned} [\mathbf{A}]\,(\mathbf{x}+\mathbf{y}) &= [\mathbf{A}]\,\mathbf{x} + [\mathbf{A}]\,\mathbf{y}, \\ [\mathbf{A}]\,(\alpha\mathbf{x}) &= \alpha\,[\mathbf{A}]\,\mathbf{x}. \end{aligned}$$

For $[\mathbf{A}] \in \mathbb{R}^{m\times p}$, $[\mathbf{B}] \in \mathbb{R}^{p\times n}$ we define the multiplication

$$[\mathbf{C}] = [\mathbf{A}]\,[\mathbf{B}]\,, \quad c_{ij} = (\mathbf{a}_i \cdot \mathbf{b}_j)$$

using the vector notation

$$[\mathbf{A}] = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{pmatrix}, \quad [\mathbf{B}] = (\mathbf{b}_1, ..., \mathbf{b}_n)$$

where $\mathbf{a}_i \in \mathbb{R}^{1\times p}$ are row vectors and $\mathbf{b}_j \in \mathbb{R}^{p\times 1}$ are column vectors. Although in general matrix multiplication is not commutative, we have this property in the special case when $[\mathbf{A}]$ and $[\mathbf{B}]$ are diagonal matrices

$$[\mathbf{A}] = \begin{pmatrix} a_{11} & 0 & ... & 0 \\ 0 & a_{22} & 0 & \vdots \\ \vdots & & & \vdots \\ 0 & ... & 0 & a_{mm} \end{pmatrix}, \quad [\mathbf{B}] = \begin{pmatrix} b_{11} & 0 & ... & 0 \\ 0 & b_{22} & 0 & \vdots \\ \vdots & & & \vdots \\ 0 & ... & 0 & b_{mm} \end{pmatrix}.$$

The *range* of a matrix $[\mathbf{A}]$, written as $\mathrm{range}\,([\mathbf{A}])$, is the set of vectors that can be express as $[\mathbf{A}]\,\mathbf{x}$ for some $\mathbf{x}$.

Definition 1.1. *The* $\mathrm{range}\,([\mathbf{A}])$ *is the space spanned by the columns of* $[\mathbf{A}]$*.*

The $\mathrm{range}\,([\mathbf{A}])$ is also called the *column space* of $[\mathbf{A}]$. This means that for any $\mathbf{y}$ that belongs to $\mathrm{range}\,([\mathbf{A}])$ we have that

$$\mathbf{y} = c_1\mathbf{a}_1 + ... + c_n\mathbf{a}_n.$$

The *null-space* of a matrix $[\mathbf{A}] \in \mathbb{R}^{m\times n}$ or $\mathrm{null}\,([\mathbf{A}])$ is the set of vectors $\mathbf{x} \in \mathbb{R}^n$ that satisfy

$$[\mathbf{A}]\,\mathbf{x} = \mathbf{0}.$$

The *column rank* of a matrix $[\mathbf{A}]$ is the dimension of its column in space. Similarly the *row rank* of a matrix $[\mathbf{A}]$ is the dimension of the space spanned by its rows. The row rank always equals the column rank (we will see this later), so we refer to this number simply as the *rank* of a matrix $[\mathbf{A}]$ or as $\operatorname{rank}([\mathbf{A}])$.

An $m \times n$ matrix of *full rank* is one that has the maximal possible rank (the lesser of m and n). This means that a matrix of full rank with $m \geq n$ must have n linearly independent columns. Such matrix can also be characterized by the property that the map it defines is one-to-one

Example 1.10. *Find the null-space and rank of the matrix*

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

Solution: For the null-space and rank we solve the homogeneous system

$$\begin{pmatrix} 2 & 3 & 1 \\ 4 & 2 & 1 \\ -4 & 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and use the simplified form to apply Gauss elimination

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 4 & 2 & 1 & 0 \\ -4 & 3 & 4 & 0 \end{array}\right)$$

From example 1.4 the Gauss-Elimination process yields

$$\left(\begin{array}{ccc|c} 2 & 1 & 1 & 0 \\ 0 & 4 & 1 & 0 \\ 0 & 0 & -15 & 0 \end{array}\right)$$

The row echelon form shows immediately that the null space is $\{0\}$ and rank is 3. Since the rank is 3, then the matrix is of full rank.

Example 1.11. *Find the null-space and rank of the matrix*

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

Solution: For the null-space and rank we use Gauss-Elimination to obtain

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

Here the rank is 2 and the null-space is given by

$$\begin{pmatrix} c \\ 0 \\ -2c \end{pmatrix}$$

for a given number c.

Definition 1.2. *An square matrix* $[\mathbf{A}] \in \mathbb{R}^{m\times m}$ *is invertible if there exists an inverse matrix* $[\mathbf{A}]^{-1} \in \mathbb{R}^{m\times m}$ *such that*

$$[\mathbf{A}]\,[\mathbf{A}]^{-1} = [\mathbf{A}]^{-1}\,[\mathbf{A}] = [\mathbf{I}]$$

where $[\mathbf{I}] \in \mathbb{R}^{m\times m}$ *is the identity matrix with* 1*'s along the main diagonal and* 0*'s elsewhere.*

We have the property that for $[\mathbf{A}]$, $[\mathbf{B}]$ (of compatible dimensions)

$$([\mathbf{A}]\,[\mathbf{B}])^{-1} = [\mathbf{B}]^{-1}\,[\mathbf{A}]^{-1}\,.$$

MATLAB programming note 1.3.

MATLAB is the best environment to easily perform matrix computations. Define the matrices [**A**], [**B**]

```
>> A = [2 0 1; 4 6 2; -1 1 4];
>> B = [2 3 1; 1 6 2; 2 12 4];
>> C = [3 4; 12 -1; 0 1];
```

Matrix multiplications can be computed by the following commands

```
>> A*C
ans =
     6     9
    84    12
     9    -1
>> C*B
Error using  *
Inner matrix dimensions must agree.
```

Notice that the matrix multiplication should satisfy the dimension requirements. Notice that here we can check that matrix multiplication is not commutative since

```
>> A*B
ans =
     6    18     6
    18    72    24
     7    51    17

>> B*A
ans =
    15    19    12
    24    38    21
    48    76    42
```

Finally the identity matrix is defined using the “eye” command, and this matrix commutes with any other matrix

```
>> I=eye(3)
I =
     1     0     0
     0     1     0
     0     0     1

>> A*I
ans =
     2     0     1
     4     6     2
    -1     1     4
```



```
>> I*A
ans =
     2     0     1
     4     6     2
    -1     1     4
```

Theorem 1.2. *For a square matrix $[\mathbf{A}] \in \mathbb{R}^{m\times m}$ the following are equivalent*

1. $[\mathbf{A}]$ *has an inverse.*
2. $\operatorname{rank}([\mathbf{A}]) = m$.
3. $\operatorname{range}([\mathbf{A}]) = \mathbb{R}^m$.
4. $\operatorname{null}([\mathbf{A}]) = \boldsymbol{0}$.

Example 1.12. *From example 1.4 compute the inverse matrix of*

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

Solution: We set the simplified notation

$$[\mathbf{A}] = \left(\begin{array}{ccc|ccc} 2 & 3 & 1 & 1 & 0 & 0 \\ 4 & 2 & 1 & 0 & 1 & 0 \\ -4 & 3 & 4 & 0 & 0 & 1 \end{array}\right)$$

and by reducing to an upper triangular matrix we get

$$\left(\begin{array}{ccc|ccc} 2 & 3 & 1 & 1 & 0 & 0 \\ 0 & 4 & 1 & 2 & -1 & 0 \\ 0 & 0 & -15 & 10 & -9 & -4 \end{array}\right)$$

then by reduce row echelon form

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1/6 & 3/10 & -1/30 \\ 0 & 1 & 0 & 2/3 & -2/5 & -2/30 \\ 0 & 0 & 1 & -2/3 & 3/5 & 4/15 \end{array}\right)$$

Then

$$[\mathbf{A}]^{-1} = \begin{pmatrix} -1/6 & 3/10 & -1/30 \\ 2/3 & -2/5 & -2/30 \\ -2/3 & 3/5 & 4/15 \end{pmatrix}$$

MATLAB programming note 1.4.

The inverse of a matrix of example 1.12 can be found using the function

```
>> A = [2 3 1; 4 2 1; -4 3 4];
>> inv(A)
ans =

  -0.1667    0.3000   -0.0333
   0.6667   -0.4000   -0.0667
  -0.6667    0.6000    0.2667
```

We denote as the transpose of the matrix $[\mathbf{A}]^T \in \mathbb{R}^{n\times m}$ as

$$[\mathbf{A}]^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}.$$

Given two matrices $[\mathbf{A}]$, $[\mathbf{B}]$ (of compatible dimensions) have the property

$$([\mathbf{A}]\,[\mathbf{B}])^T = [\mathbf{B}]^T\,[\mathbf{A}]^T\,,$$

which is analogous to the properties for the matrix inversion. The notation $[\mathbf{A}]^{-T}$ is shorthand for $\left([\mathbf{A}]^T\right)^{-1}$ or $\left([\mathbf{A}]^{-1}\right)^T$, where these two are equal. If $[\mathbf{A}] = [\mathbf{A}]^T$, $[\mathbf{A}]$ is *symmetric*. By definition, a symmetric matrix must be square. As we observe the transpose operator interchanges the rows with the columns of $[\mathbf{A}]$.

Using Fig.1.4 we can define the *rotation matrices*

$$\begin{aligned} [\mathbf{R}]_x(\alpha) &:= \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ [\mathbf{R}]_y(\beta) &:= \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \\ [\mathbf{R}]_z(\gamma) &:= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & \sin\gamma \\ 0 & -\sin\gamma & \cos\gamma \end{pmatrix} \end{aligned}$$

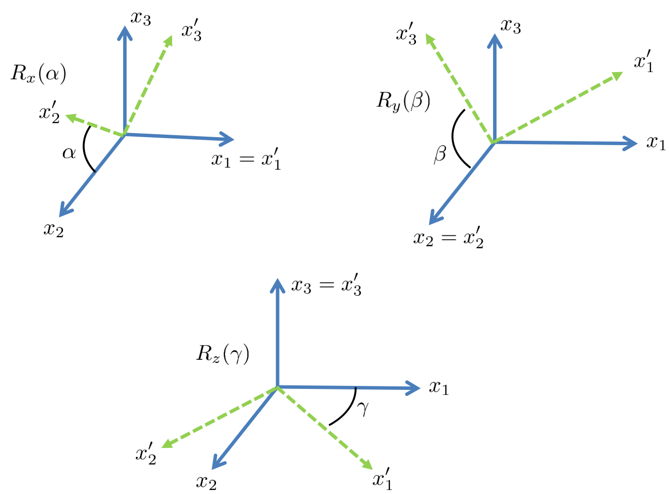

Figure 1.4: Rotation matrices and Euler angles.

The rotation angles (α, β, γ) are called the Euler angles. Notice that the rotation matrix preserves the length of any vector $\mathbf{x}$. Is easy to check this property since

$$\mathbf{b} = [\mathbf{R}]_x (\alpha)\mathbf{x},$$

and

$$\begin{aligned} \|\mathbf{b}\|_2^2 &= (\cos\alpha x_1 + \sin\alpha x_2)^2 + (-\sin\alpha x_1 + \cos\alpha x_2)^2 + x_3 \\ &= x_1^2 + x_2^2 + x_3^2 = \|\mathbf{x}\|_2^2. \end{aligned}$$

From this result we can argue that the only eigenvalues of the rotation matrices are 1. Notice another interesting property

$$\begin{aligned} & [\mathbf{R}]_x (\alpha)^H [\mathbf{R}]_x (\alpha) \\ = & \begin{pmatrix} \cos^2\alpha + \sin^2\alpha & \cos\alpha\sin\alpha - \cos\alpha\sin\alpha & 0 \\ \cos\alpha\sin\alpha - \cos\alpha\sin\alpha & \cos^2\alpha + \sin^2\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} = [\mathbf{I}]. \end{aligned}$$

It is also true that $[\mathbf{R}]_x (\alpha) [\mathbf{R}]_x (\alpha)^H = [\mathbf{I}]$. The same property holds for any rotation matrix. The rotation matrices belong to a more general type of matrices known as *unitary matrices.* A square matrix $[\mathbf{Q}] \in \mathbb{R}^{m\times m}$ is unitary (or orthogonal) if $[\mathbf{Q}]^T = [\mathbf{Q}]^{-1}$ or

$$[\mathbf{Q}]^T [\mathbf{Q}] = \mathbf{I}.$$

Homework 1.8. *Use theorem 1.2 to determine that the following matrices are invertible*

$$a)\begin{pmatrix} 2 & 3 \\ 1 & 0 \end{pmatrix}, \quad b)\begin{pmatrix} 2 & 0 & 1 \\ 4 & 6 & 0 \\ -1 & 1 & 4 \end{pmatrix}, \quad c)\begin{pmatrix} 2 & 3 & 1 \\ 1 & 0 & 1 \\ -1 & 2 & 0 \end{pmatrix}$$

Homework 1.9. *Use Gaussian-elimination to find the inverse of the following matrices*

$$a)\begin{pmatrix} 2 & 3 \\ 1 & -2 \end{pmatrix}, \quad b)\begin{pmatrix} 2 & 0 & 1 \\ 4 & 6 & 2 \\ -1 & 1 & 4 \end{pmatrix}, \quad c)\begin{pmatrix} 2 & 3 & 1 \\ 1 & -2 & 1 \\ -1 & 2 & 4 \end{pmatrix}$$

Homework 1.10. *Show that for a unitary matrix* $[\mathbf{Q}] \in \mathbb{R}^{m\times m}$*,*

$$\|[\mathbf{Q}]\,\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2\,,$$

for any $\boldsymbol{x} \in \mathbb{R}^m$*.*

Homework 1.11. *Prove that*

$$\left([\mathbf{A}]^T\right)^{-1} = \left([\mathbf{A}]^{-1}\right)^T.$$

Homework 1.12. *Complex numbers* $a + bi$*, with* a *and* b *real, may be represented by* 2×2 *matrices:*

$$a + bi \leftrightarrow \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

Show that his matrix representation is valid for (i) addition and (ii) multiplication. Can you find the matrix that corresponds to $(a + bi)^{-1}$*?*

Homework 1.13. *Let*

$$[\mathbf{i}] = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad [\mathbf{j}] = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad [\mathbf{k}] = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$$

Show that

1. $[\mathbf{i}]^2 = [\mathbf{j}]^2 = [\mathbf{k}]^2 = -[\mathbf{I}]$
2. $[\mathbf{i}][\mathbf{j}] = -[\mathbf{j}][\mathbf{i}] = [\mathbf{k}]$
3. $[\mathbf{j}][\mathbf{k}] = -[\mathbf{k}][\mathbf{j}] = [\mathbf{i}]$
4. $[\mathbf{k}][\mathbf{i}] = -[\mathbf{i}][\mathbf{k}] = [\mathbf{j}]$

These three matrices plus the identity $[\mathbf{I}]$ form a basis for quaternions[4].

Homework 1.14. *Define the matrix "Wronskian"*[5] *as*

$$\langle [\mathbf{A}], [\mathbf{B}] \rangle := [\mathbf{A}][\mathbf{B}] - [\mathbf{B}][\mathbf{A}]$$

Since the multiplication between matrices is not commutative ($[\mathbf{A}][\mathbf{B}] \neq [\mathbf{B}][\mathbf{A}]$) we find that in general $\langle [\mathbf{A}], [\mathbf{B}] \rangle \neq [\mathbf{0}]$ where $[\mathbf{0}]$ is the matrix with only 0 components. Also for the identity matrix $[\mathbf{I}] \in \mathbb{R}^{m\times m}$ we have that $\langle [\mathbf{A}], [\mathbf{I}] \rangle = [\mathbf{0}]$. Verify the Jacobi identity for square matrices

$$\langle [\mathbf{A}], \langle [\mathbf{B}], [\mathbf{C}] \rangle \rangle = \langle [\mathbf{B}], \langle [\mathbf{A}], [\mathbf{C}] \rangle \rangle - \langle [\mathbf{C}], \langle [\mathbf{A}], [\mathbf{B}] \rangle \rangle$$

This is useful in matrix descriptions of elementary particles.

Homework 1.15. *Show that for square matrices*

$$([\mathbf{A}] + [\mathbf{B}])([\mathbf{A}] - [\mathbf{B}]) = [\mathbf{A}]^2 - [\mathbf{B}]^2$$

if and only if $[\mathbf{A}]$ and $[\mathbf{B}]$ commute, i.e.,

$$\langle [\mathbf{A}], [\mathbf{B}] \rangle = [\mathbf{0}].$$

Homework 1.16. *Show that the matrices*

$$[\mathbf{A}] = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad [\mathbf{B}] = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad [\mathbf{C}] = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Satisfy the commutation relations

$$\langle [\mathbf{A}], [\mathbf{B}] \rangle = [\mathbf{C}], \quad \langle [\mathbf{A}], [\mathbf{C}] \rangle = [\mathbf{0}], \quad \langle [\mathbf{B}], [\mathbf{C}] \rangle = [\mathbf{0}].$$

These matrices have been used in a Lie algebra approach to Hermite polynomials.

Project 1.2. *Prove theorem 1.2.*

[4] In mathematics, the quaternions are a number system that extends the complex numbers. They were first described by Irish mathematician William Rowan Hamilton in 1843 and applied to mechanics in three-dimensional space.

[5] This definition is my own, but we encounter a similar terminology in the classical linear algebra books

1.4 Norms and Eigenvalues

A *norm* is a function $\|.\| : \mathbb{R}^m \mapsto \mathbb{R}$ that assigns a real-valued length to each vector. In order to conform a reasonable notion of length, a norm must satisfy the following conditions

$$\begin{aligned}
&\|\mathbf{x}\| \geq 0, \quad \text{and } \|\mathbf{x}\| = 0 \text{ only if } \mathbf{x} = 0,\\
&\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|,\\
&\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|.
\end{aligned} \tag{1.3}$$

for any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ and scalar α.

In the previous sections we only use the euclidian norm $\|.\|_2$, but in reality there exist other norms, which define different notions of length and sometimes its useful to have this flexibility. The most important class of vector norms, the p-norms, are defined for $\mathbf{x} \in \mathbb{R}^m$

$$\begin{aligned}
\|\mathbf{x}\|_1 &= \sum_{j=1}^{m} |x_j|,\\
\|\mathbf{x}\|_2 &= \left(\sum_{j=1}^{m} |x_j|^2\right)^{1/2} = \sqrt{\mathbf{x}^T\mathbf{x}},\\
\|\mathbf{x}\|_\infty &= \max_{1\leq j\leq m} |x_j|,\\
\|\mathbf{x}\|_p &= \left(\sum_{j=1}^{m} |x_j|^p\right)^{1/p}, \quad (1 \leq p < \infty).
\end{aligned} \tag{1.4}$$

In Fig.1.5 we show the geometrical interpretation in $\mathbb{R}^2$ for the different norms. The 2-norm is the Euclidean length function. The 1-norm is used by airlines to define the maximal allowable size of a suitcase. The Sergel plaza in Stockholm, Sweden has the shape of the unit ball in the 4-norm.

Aside from the p-norms, the most useful norms are the *weighted* p-norms, where each of the coordinates of a vector space is given its own weight. In general, given any norm $\|.\|$, a weighted norm can be written as

$$\|\mathbf{x}\|_W = \|\,[\mathbf{W}]\,\mathbf{x}\|.$$

Here $[\mathbf{W}]$ is the diagonal matrix in which the ith diagonal is the weight $w_i \neq 0$. One can generalize the idea of a weighted norm by allowing $[\mathbf{W}]$ to be an arbitrarily non-singular matrix, not necessarily diagonal.

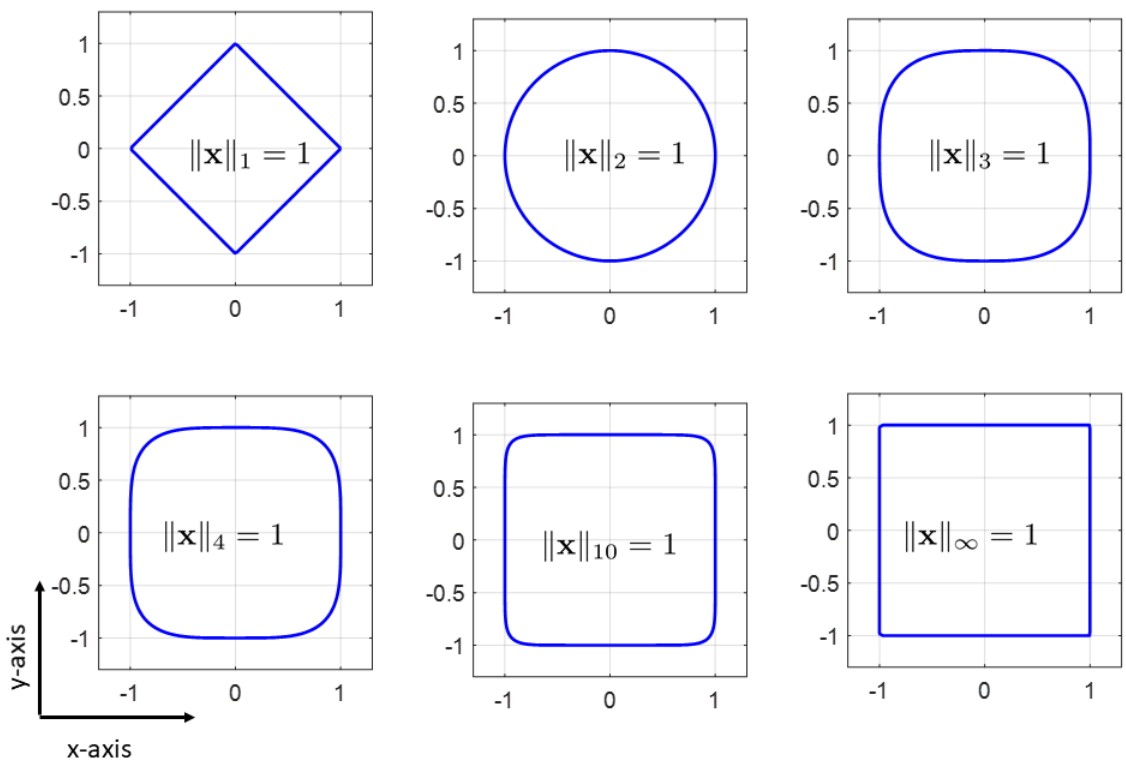

Figure 1.5: Geometrical representation in $\mathbb{R}^2$ of the unit vectors for different norms.

As in the case for vectors we can assign the concept of length using a norm for a $m \times n$ matrix. This application do not result in a explicit formula as in the case of vectors, but instead is defined through a maximization process. First notice that for a matrix $[\mathbf{A}] \in \mathbb{R}^{m\times n}$ we have the following inequality for any $\mathbf{x} \in \mathbb{R}^n$

$$\|[\mathbf{A}]\,\mathbf{x}\|_{(m)} \le C\,\|\mathbf{x}\|_{(n)}\,, \tag{1.5}$$

where $C > 0$ is a constant. Here we understand $\|.\|_{(m)}$ is any norm that satisfies Eq.(1.3) for a vector in $\mathbb{R}^m$, and a similar definition for $\|.\|_{(n)}$. We assume that the *induced matrix norm* $\|\,[\mathbf{A}]\,\|$ is the smaller number C for which the inequality Eq.(1.5) holds. Therefore, the norm can be defined equivalently as

$$\|[\mathbf{A}]\| = \sup_{\mathbf{x}\in\mathbb{R}^m} \frac{\|[\mathbf{A}]\,\mathbf{x}\|_{(m)}}{\|\mathbf{x}\|_{(n)}} = \sup_{\mathbf{x}\in\mathbb{R}^m, \|\mathbf{x}\|=1} \|[\mathbf{A}]\,\mathbf{x}\|_{(m)}\,. \tag{1.6}$$

Example 1.13. *The matrix*

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

maps $\mathbb{R}^2$ *into* $\mathbb{R}^2$*. We have that* $\|\,[\mathbf{A}]\,\|_1 = 4$*,* $\|\,[\mathbf{A}]\,\|_2 \approx 2.9208$*,* $\|\,[\mathbf{A}]\,\|_\infty = 3$*. See Fig.1.6.*

Example 1.14. *Let* $[\mathbf{D}] \in \mathbb{R}^{m\times m}$ *be a diagonal matrix*

$$[\mathbf{D}] = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m \end{pmatrix}$$

What is $\|\,[\mathbf{D}]\,\|_p$*?*

We notice that denoting as $d = \max |d_i|$, is trivial that

$$\|\,[\mathbf{D}]\,\mathbf{x}\|_p^p = \sum_i |d_i|^p = d^p \sum_i \left(\frac{|d_i x_i|}{d}\right)^p \le d^p \|\mathbf{x}\|_p^p$$

This means that $\|\,[\mathbf{D}]\,\|_p \le d$. But indeed, is not hard to check that this an equality.

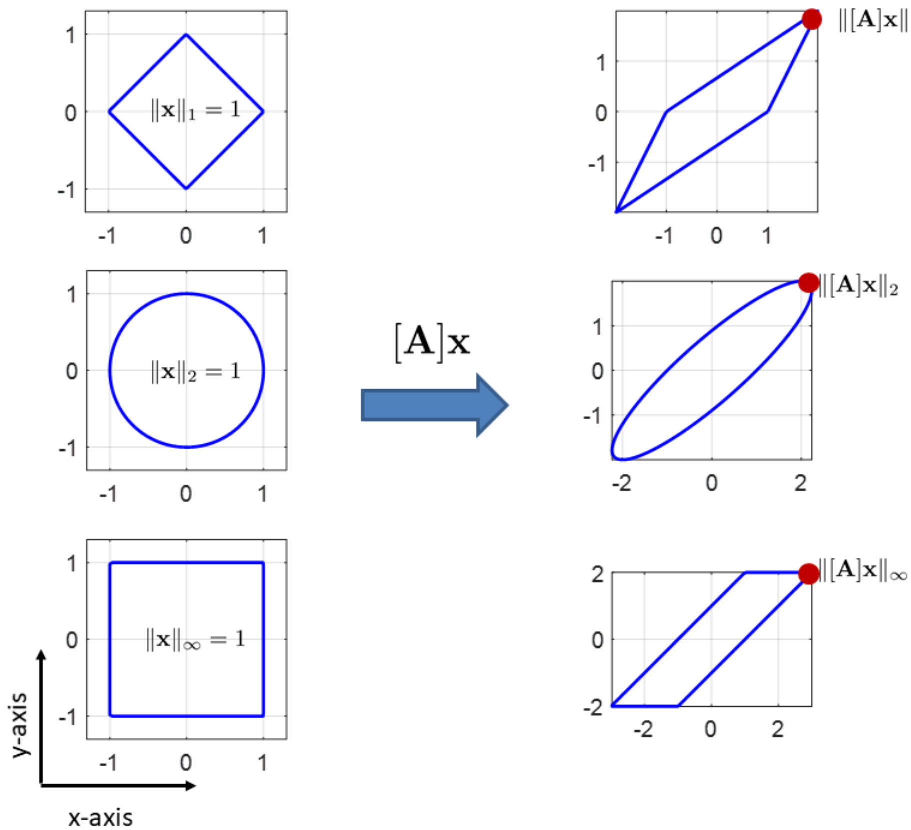

Figure 1.6: Geometrical representation in $\mathbb{R}^2$ of the projection of the unit vectors into the matrix $[\mathbf{A}]$ for different norms.

MATLAB programming note 1.5.

We can define the column vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$

```
>> x = [1;2;3;4;5];
>> y = [-2;3;5;6;7];
>> z = [2;0;1;8;1];
```

The corresponding norms $\|.\|_1$, $\|.\|_2$ and $\|.\|_\infty$ are computed as

```
>> norm(x,1)
ans =
    15
>> norm(x,2)
ans =
    7.4162
>> norm(x,inf)
ans =
     5
```

We can verify the matrix norms

```
>> A = [1 2;0 2];
>> norm(A)
ans =
    2.9208
>> norm(A,1)
ans =
     4
>> norm(A,inf)
ans =
     3
```

The *eigenvalue* λ and *eigenvector* $\mathbf{x}_\lambda$[6] of a square matrix $[\mathbf{A}] \in \mathbb{R}^{m\times m}$ are defined as

$$[\mathbf{A}]\,\mathbf{x}_\lambda = \lambda \mathbf{x}_\lambda \tag{1.7}$$

for a vector $\mathbf{x}_\lambda \neq 0$.

[6] In the 18th century Euler studied the rotational motion of a rigid body and discovered the importance of the principal axes. Lagrange realized that the principal axes are the eigenvectors of the inertia matrix. In the early 19th century, Cauchy saw how their work could be used to classify the quadric surfaces, and generalized it to arbitrary dimensions. Cauchy also coined the term racine caractéristique (characteristic root) for what is now called eigenvalue; his term survives in characteristic equation.

Example 1.15. *The matrix*

$$\begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$$

have the eigenvalue and eigenvector pairs $\lambda = 3$, $\mathbf{x}_\lambda = (1, 2)^T$ *and* $\lambda = -1$, $\mathbf{x}_\lambda = (1, -2)^T$.

This can be verified directly by

$$\begin{aligned} \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} &= \begin{pmatrix} 3 \\ 6 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \\ \text{and } \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} &= \begin{pmatrix} -1 \\ 2 \end{pmatrix} = -1 \begin{pmatrix} 1 \\ -2 \end{pmatrix} \end{aligned}$$

The idea is to write Eq.(1.7) in the form

$$[\mathbf{A}]\,\mathbf{x} = \lambda\,[\mathbf{I}]\,\mathbf{x}, \quad \text{or } \left(\lambda\,[\mathbf{I}] - [\mathbf{A}]\right)\mathbf{x} = \mathbf{0}, \quad \mathbf{x} \neq \mathbf{0}.$$

Hence the matrix $(\lambda\,[\mathbf{I}] - [\mathbf{A}])$ must be singular, and its determinant must therefore be 0, so we solve $p_A(\lambda) = 0$ where

$$p_A(\lambda) = \det\left(\lambda\,[\mathbf{I}] - [\mathbf{A}]\right)$$

is known as the *characteristic polynomial* of the matrix $[\mathbf{A}]$. p_A is a polynomial in the variable λ, whose roots are thus the eigenvalues. For a specific λ, the general eigen-vector $\mathbf{x}$ is found by solving the homogeneous equation

$$\left(\lambda\,[\mathbf{I}] - [\mathbf{A}]\right)\mathbf{x} = \mathbf{0}.$$

Since the eigenvalue problem can be reduced to finding the roots of $p_A(\lambda) = 0$ we encounter that if a matrix is real it can contain complex eigenvalues. Physically , this is related to the phenomenon that real dynamical systems can have motions that oscillate as well as growth or decay.

Example 1.16. *Find the eigendata from*

$$[\mathbf{A}] = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$$

Solution: We follow the procedure

$$\det(\lambda\,[\mathbf{I}] - [\mathbf{A}]) = \begin{vmatrix} \lambda-1 & -1 \\ -4 & \lambda-1 \end{vmatrix} = \lambda^2 - 2\lambda - 3 = (\lambda-3)(\lambda+1).$$

Solving $(\lambda-3)(\lambda+1) = 0$, clearly the eigenvalues are 3 and -1.

To obtain the eigenvectors we choose $\lambda = 3$, and $(\lambda\,[\mathbf{I}] - [\mathbf{A}])\mathbf{x} = 0$ becomes

$$\begin{pmatrix} 2 & -1 \\ -4 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which is equivalent to $2x_1 = x_2$. There are infinitely many solutions to $2x_1 = x_2$ but a simple one is $x_1 = 1$, $x_2 = 2$, or $\mathbf{x} = (x_1, x_2)^T = (1, 2)^T$. The construction for $\lambda = -1$ is similar.

In general we can always find that for a square matrix $[\mathbf{A}]$

$$[\mathbf{A}]\,[\mathbf{X}] = [\mathbf{X}]\,[\mathbf{\Lambda}] \tag{1.8}$$

where $[\mathbf{X}]$ is a matrix with columns given by the eigenvectors $\mathbf{x}_\lambda$ and $[\mathbf{\Lambda}]$ a diagonal matrix with eigenvalues in the diagonal.
From the example 1.15 notice that the eigendata provides the following relations

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

Notice that we can join the relations in the following form

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

We multiply by the matrix inverse to obtain

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

Theorem 1.3 (Spectral Theorem)**.** *A **non-degenerate matrix** (or non-defective)* $[\mathbf{A}] \in \mathbb{R}^{m\times m}$ *have a decomposition of the form*

$$[\mathbf{A}] = [\mathbf{X}]\,[\mathbf{\Lambda}]\,[\mathbf{X}]^{-1} \tag{1.9}$$

where $[\mathbf{X}]^{-1} \in \mathbb{R}^{m\times m}$ *exists.*

A *non-defective matrix* is a matrix where all its columns are linearly independent. From Theorem 1.2 we get that a *non-defective matrix* is invertible.

In general we can always find Eq.(1.8), but Eq.(1.9) does not hold.

Example 1.17. *Find the eigenvalue decomposition of*

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

Solution: We have the repeated eigenvalue $\lambda = 2$ and we can find the corresponding eigenvector $\mathbf{x}_\lambda = (0, 1)^T$. Similarly another eigenvector is given by $\mathbf{x}_\lambda = (0, -1)^T$. Notice that Eq.(1.9) is true

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

But what happens to relation Eq.(1.9)?

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

Here the problem is that

$$\begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}^{-1}$$

does not exist. When does the matrix $[\mathbf{X}]^{-1}$ exist? This happens when the eigenvectors are linearly independent.

MATLAB programming note 1.6.

Define the matrices that we use in example 1.14 and 1.15

```
>> A = [1 1; 4 1];
>> B = [2 0; 1 2];
```

We can compute the eigenvalues from the command

```
>> [X,D] = eig(A)
X =
    0.4472   -0.4472
    0.8944    0.8944
D =
    3.0000         0
         0   -1.0000
```

Notice that the eigenvalues coincide with our calculations in example 1.15, but the eigenvectors appear to be different. This reason for this difference is that MATLAB returns unitary eigenvectors (this is a common procedure of numerical eigenvalues calculations. Notice that our results can be normalized too

```
>> x1 = [1;2]; x2 = [1,-2];
>> x1 = [1;2]; x2 = [1;-2];
>> x1/norm(x1)
ans =
    0.4472
    0.8944

>> x2/norm(x2)
ans =
    0.4472
   -0.8944
```

The second eigenvector differs by a sign. Do this affect any of our results? Notice that for example 1.15

```
>> [X,D] = eig(B)
X =
         0    0.0000
    1.0000   -1.0000

D =
     2     0
     0     2
```

Homework 1.17. *Find the eigendata from*

$$a)\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix},\quad b)\begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix},\quad c)\begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix},$$

Homework 1.18. *Find the eigendata from*

$$a)\begin{pmatrix} 4 & 0 \\ 3 & 1 \end{pmatrix}\quad b)\begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}\quad c)\begin{pmatrix} 4 & 8 & 1 \\ 0 & 2 & 0 \\ 0 & 6 & 7 \end{pmatrix}$$

Homework 1.19. *Use MATLAB to compute the eigendata from the following matrices*

$$a)\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}\quad b)\begin{pmatrix} 1 & \sqrt{2} & 0 \\ \sqrt{2} & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\quad c)\begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}\quad d)\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

Homework 1.20. *Use MATLAB to compute the eigendata from the following matrices*

$$a)\begin{pmatrix} 1 & \sqrt{8} & 0 \\ \sqrt{8} & 1 & \sqrt{8} \\ 0 & \sqrt{8} & 1 \end{pmatrix}\quad b)\begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$c)\begin{pmatrix} \sqrt{2}/2 & 0 & \sqrt{2}/2 \\ 0 & 1 & 0 \\ -\sqrt{2}/2 & 0 & \sqrt{2}/2 \end{pmatrix}\quad d)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \sqrt{2} \\ 0 & \sqrt{2} & 0 \end{pmatrix}$$

Homework 1.21. *Assume that* $\boldsymbol{x} \in \mathbb{R}^m$ *and* $[\mathbf{A}] \in \mathbb{R}^{m\times m}$ *then show that*

$$\begin{aligned} a) \quad & \|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\|_2 \\ b) \quad & \|\boldsymbol{x}\|_2 \leq \sqrt{m}\|\boldsymbol{x}\|_\infty \end{aligned}$$

Homework 1.22. *For a symmetric matrix* $[\mathbf{A}] \in \mathbb{R}^{2\times 2}$ *prove*

1. *The eigenvalues are real*
2. *if* $\lambda_1 \neq \lambda_2$ *then the corresponding eigen-vectors are linear independent*

Homework 1.23. *Assume that an unitary matrix* $[\mathbf{U}]$ *satisfies the eigenvalue equation* $[\mathbf{U}]\,\boldsymbol{x} = \lambda\boldsymbol{x}$*,*

1. *Show that the eigenvalues of the unitary matrix have unit magnitude.*
2. *If* $\boldsymbol{x}_1$ *and* $\boldsymbol{x}_2$ *are the eigenvectors of a unitary matrix corresponding to two distinct eigenvalues, show that the eigenvectors are orthogonal, i.e.,* $\boldsymbol{x}_1^T\boldsymbol{x}_2 = 0$.

1.5 Eigenvalue Problems

As we define in the previous the eigenvalue problem for a square matrix $[\mathbf{A}] \in \mathbb{R}^{m\times m}$

$$[\mathbf{A}]\,\mathbf{x}_\lambda = \lambda \mathbf{x}_\lambda.$$

The set of all eigenvalues λ of the matrix $[\mathbf{A}]$ the *spectrum* of $[\mathbf{A}]$ and we denote it as $\Lambda([\mathbf{A}])$.

The trace of a square matrix $[\mathbf{A}]$ is the sum of all it's diagonal elements

$$\mathrm{tr}([\mathbf{A}]) = \sum_{j=1}^{m} a_{jj}.$$

We have the following important properties

$$\det([\mathbf{A}]) = \lambda_1 \cdot \ldots \cdot \lambda_m, \quad \mathrm{tr}([\mathbf{A}]) = \sum_{j=1}^{m} \lambda_j. \tag{1.10}$$

Given any invertible matrix $[\mathbf{B}]$ we can always define a similarity transformation for a square matrix $[\mathbf{A}]$

$$[\mathbf{C}] = [\mathbf{B}]\,[\mathbf{A}]\,[\mathbf{B}]^{-1}. \tag{1.11}$$

Notice that if the matrix $[\mathbf{A}]$ and $[\mathbf{C}]$ have a similarity transformation Eq.(1.11), then if $[\mathbf{A}]$ has the decomposition Eq.(1.9) then

$$\begin{aligned}
[\mathbf{C}] &= [\mathbf{B}]\,[\mathbf{A}]\,[\mathbf{B}]^{-1} \\
&= [\mathbf{B}]\,[\mathbf{X}]\,[\mathbf{\Lambda}]\,[\mathbf{X}]^{-1}\,[\mathbf{B}]^{-1} \\
&= ([\mathbf{B}]\,[\mathbf{X}])\,[\mathbf{\Lambda}]\,([\mathbf{B}]\,[\mathbf{X}])^{-1}.
\end{aligned}$$

This equation shows that the matrices $[\mathbf{A}]$ and $[\mathbf{C}]$ have the same eigenvalues.

It sometimes happens that not only a matrix contains linearly independent eigenvectors, but these can be chosen to be orthogonal.

Definition 1.3 (Unitary Decomposition)**.** *A matrix* $[\mathbf{A}] \in \mathbb{R}^{m\times m}$ *is unitary diagonalizable if*

$$[\mathbf{A}] = [\mathbf{Q}]\,[\mathbf{\Lambda}]\,[\mathbf{Q}]^T \tag{1.12}$$

where $[\mathbf{Q}] \in \mathbb{R}^{m\times m}$ *is unitary.*

Notice that this decomposition is stronger than Eq.(1.9), so we should expect that will only work with a smaller class of matrices from the non-defective matrices. It is well known that the symmetric matrices are unitary diagonalizable but are not the only ones (for example unitary matrices). In general, the class of matrices that are unitary diagonalizable have an elegant characterization. By definition, we say that a matrix $[\mathbf{A}]$ is *normal* if $[\mathbf{A}]^T[\mathbf{A}] = [\mathbf{A}][\mathbf{A}]^T$.

Theorem 1.4. *A matrix* $[\mathbf{A}] \in \mathbb{R}^{m\times m}$ *is unitary diagonalizable if and only if is normal.*

For a unitary diagonalizable matrix $[\mathbf{A}]$ we have that

$$\begin{aligned} \|[\mathbf{A}]\|_2 &= \max_{\|x\|_2=1} \|[\mathbf{A}]\,\mathbf{x}\|_2 \\ &= \max_{\|\mathbf{x}\|_2=1} \|[\mathbf{Q}]\,[\mathbf{\Lambda}]\,[\mathbf{Q}]^T\,\mathbf{x}\|_2 \\ &= \max_{\|\mathbf{x}\|_2=1} \|[\mathbf{\Lambda}]\,[\mathbf{Q}]^T\,\mathbf{x}\|_2 \\ &= \max_{\|\mathbf{y}\|_2=1} \|[\mathbf{\Lambda}]\,\mathbf{y}\|_2 \end{aligned}$$

which means that

$$\|[\mathbf{A}]\|_2 = \|[\mathbf{\Lambda}]\|_2 = \rho([\mathbf{A}]) \tag{1.13}$$

where $\rho([\mathbf{A}])$ is the *spectral radius* of $[\mathbf{A}]$, i.e., the largest absolute value $|\lambda|$ of an eigenvalue λ of $[\mathbf{A}]$.

There are some important square matrices $[\mathbf{A}] \in \mathbb{R}^{m\times m}$ classes that we will be interested. A *projection* matrix is a matrix such that $[\mathbf{A}]^2 = [\mathbf{A}]$. Finally, a symmetric is *positive definite* if for every non-zero $\mathbf{x} \in \mathbb{R}^{m\times 1}$,

$$\mathbf{x}^T\,[\mathbf{A}]\,\mathbf{x} \geq 0.$$

MATLAB programming note 1.7.

We define the random matrices [**A**] and [**B**]

```
>> A = rand(10,10);
>> B = randn(10,10);
```

The instruction "rand" yields a matrix 10×10 of uniformly distributed random number coefficients. Each random number coefficient is in [01]. Similarly "randn" produces a matrix of normally distributed random numbers. So how the eigenvalue look like?

```
>> [DA(1:5), DB(1:5)]
ans =
    5.2897 + 0.0000i   0.2558 + 3.2543i
   0.4240 + 0.9379i   0.2558 - 3.2543i
   0.4240 - 0.9379i   2.4225 + 0.0000i
   0.6325 + 0.0000i   1.2391 + 0.0000i
   0.2537 + 0.0000i   0.7026 + 1.6653i
```

As we can observe the eigenvalues are not necessarily real. We can visualize the eigenvalues by using

```
>> figure; plot(real(DA),imag(DA),'o')
>> figure; plot(real(DB),imag(DB),'o')
```

Here we can visualize the spectrum $\Lambda([\mathbf{A}])$ and $\Lambda([\mathbf{B}])$.

Homework 1.24. *Use MATLAB to compute the eigendata from the following matrices*

$$a)\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad b)\begin{pmatrix} 1 & \sqrt{2} & 0 \\ \sqrt{2} & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad c)\begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \quad d)\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

Determine if the given matrices are normal, symmetric, positive definite, unitary or else.

Homework 1.25. *Use MATLAB to compute the eigendata from the following matrices*

$$a)\begin{pmatrix} 1 & \sqrt{8} & 0 \\ \sqrt{8} & 1 & \sqrt{8} \\ 0 & \sqrt{8} & 1 \end{pmatrix} \quad b)\begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$c)\begin{pmatrix} \sqrt{2}/2 & 0 & \sqrt{2}/2 \\ 0 & 1 & 0 \\ -\sqrt{2}/2 & 0 & \sqrt{2}/2 \end{pmatrix} \quad d)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \sqrt{2} \\ 0 & \sqrt{2} & 0 \end{pmatrix}$$

Determine if the given matrices are normal, symmetric, positive definite, unitary or else.

Homework 1.26. *Determine if the following statements are true or false. Justify your answer (if false, provide a counter-example)*

1. *A normal matrix is symmetric*
2. *A symmetric matrix is invertible*
3. *A symmetric matrix is normal*

Homework 1.27. *Foe each of the following statements, prove that it is true or give an example to show it is false.*

1. *If λ is an eigenvalue of $[\mathbf{A}]$ and $\mu \in \mathbb{R}$, then $\lambda - \mu$ is an eigenvalue of $[\mathbf{A}] - \mu\,[\mathbf{I}]$.*
2. *if λ is an eigenvalue of $[\mathbf{A}]$ and $\mu \in \mathbb{R}$ and $[\mathbf{A}]$ is not singular, then λ^{-1} is an eigenvalue of $[\mathbf{A}]^{-1}$.*
3. *If all eigenvalues of $[\mathbf{A}]$ are zero, then $[\mathbf{A}] = \boldsymbol{0}$.*

Homework 1.28. *The same similarity transformation diagonalizes each of two matrices. Show that the original matrices must commute. This is particularly important in the matrix (Heisenberg) formulation of quantum mechanics.*

Homework 1.29. *Two Hermitian matrices $[\mathbf{A}]$ and $[\mathbf{B}]$ have the same eigenvalues. Show that $[\mathbf{A}]$ and $[\mathbf{B}]$ are related by a unitary similarity transformation.*

Homework 1.30. *A matrix $[\mathbf{P}]$ is a projector operator and λ its corresponding eigenvalue. Show that the only possibilities for the eigenvalue λ are 0 and 1.*

Homework 1.31. *Let $[\mathbf{A}]$ be a 10×10 random matrix with entries from the standard normal distribution, minus twice the identity. Write a program to plot $\|e^{t[\mathbf{A}]}\|_2$ against t for $0 \leq t \leq 20$ on a log scale, comparing the result with the straight line $e^{t\alpha([\mathbf{A}])}$, where $\alpha([\mathbf{A}]) = \max_j \Re(\lambda_j)$ is the spectral abcissa of $[\mathbf{A}]$. Run the program for 10 random matrices $[\mathbf{A}]$ and comment the results. What property of a matrix leads to a $e^{t[\mathbf{A}]}\|_2$ curve that remains oscillatory as $t \to \infty$?*

1.6 The Singular Value Decomposition

The singular value decomposition (SVD) is a matrix factorization whose computation is a setup in many algorithms. Equally important is the use of the SVD for conceptual purposes.

As we observe in Fig.1.6 in $\mathbb{R}^2$ the unit sphere under a 2×2 matrix is an ellipse. In a general case a unit sphere under any $m \times n$ matrix is an *hyperellipse*. The term hyperellipse may be unfamiliar, but this is just the m-dimensional generalization of an ellipse. We may define the hyperellipse in $\mathbb{R}^m$ as the surface obtained by stretching the unit sphere in $\mathbb{R}^m$ by some factors $\sigma_1, ..., \sigma_r$ (possibly zero) in some orthogonal directions $u_1, ..., u_m \in \mathbb{R}^m$. For convenience, let us take the u_i to be unit vectors, i.e., $\|u_i\| = 1$. The vectors $\{\sigma_i u_i\}$ are the principal semiaxes of the hyperellipse, with lengths $\sigma_1, ..., \sigma_m$. If $[\mathbf{A}]$ has rank r, exactly r of the lengths σ_i will turn out to be nonzero, and in particular, if $m \geq n$, at most n of them will be nonzero.

Let $\mathbb{S}^{n-1}$ be a unit sphere in $\mathbb{R}^n$, and take any $[\mathbf{A}] \in \mathbb{R}^{m \times m}$ with $m \geq n$. For simplicity, for the moment that $[\mathbf{A}]$ has a full rank n. The image of $[\mathbf{A}]\,\mathbb{S}^{n-1}$ is a hyperellipse in $\mathbb{R}^m$. We now define some properties of $[\mathbf{A}]$ in terms of the shape of $[\mathbf{A}]\,\mathbb{S}^{n-1}$.

First, we define the n singular values of $[\mathbf{A}]$. these are the lengths of the n principal axes of $[\mathbf{A}]\,\mathbb{S}^{n-1}$, written $\sigma_1, \sigma_2, ..., \sigma_n$. It is conventional to assume that the singular values are numbered in descending order, $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_n > 0$.

Next, we define the n left singular vectors of $[\mathbf{A}]$. These are the unit vectors $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\} \in \mathbb{S}^{n-1}$ oriented in the directions of the principal semiaxes of $[\mathbf{A}]\,\mathbb{S}^{n-1}$, numbered to correspond with the singular values. Thus the vector $\sigma_i \mathbf{u}_i$ is the ith largest principal semiaxes of $[\mathbf{A}]\,\mathbb{S}^{n-1}$.

Finally, we define the n right singular vectors of $[\mathbf{A}]$. These are the unit vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\} \in \mathbb{S}^{n-1}$ that are the preimages of the principal semiaxes of $[\mathbf{A}]\,\mathbb{S}^{n-1}$, numbered so that $[\mathbf{A}]\,\mathbf{v}_j = \sigma_j \mathbf{u}_j$.

Similar to the eigenvalue-eigenvector relation Eq.(1.7) we have the relation between the right and left singular vectors for a matrix $[\mathbf{A}] \in \mathbb{R}^{m \times n}$, $m \geq n$

$$[\mathbf{A}]\,\mathbf{v}_j = \sigma_j \mathbf{u}_j, \quad 1 \leq j \leq n, \tag{1.14}$$

and similar to the matrix expression Eq.(1.8) we obtain

$$[\mathbf{A}]\,[\mathbf{V}] = \left[\hat{\mathbf{U}}\right]\left[\hat{\mathbf{\Sigma}}\right].$$

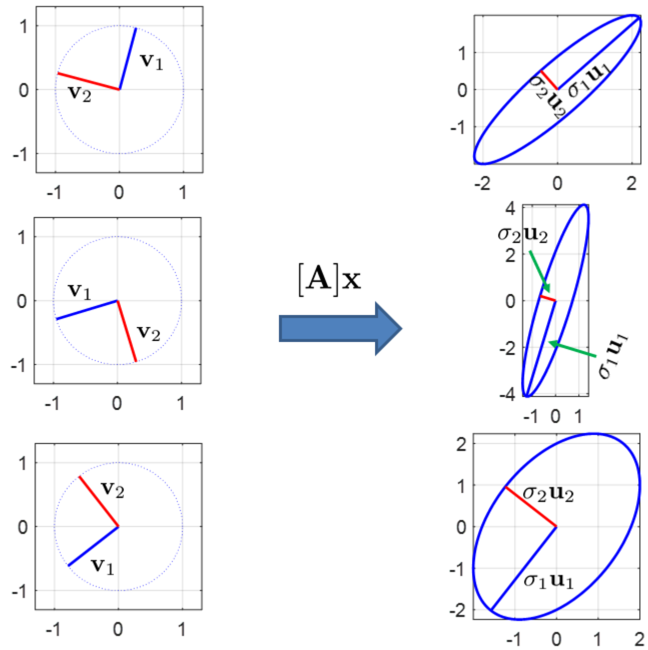

Figure 1.7: Geometrical representation in $\mathbb{R}^2$ of the SVD projection of the matrix $[\mathbf{A}]$ for different norms.

Here $[\mathbf{V}] \in \mathbb{R}^{n\times n}$ is a unitary matrix and $\left[\hat{\mathbf{\Sigma}}\right] \in \mathbb{R}^{n\times n}$ is a diagonal matrix that contains the singular values. $\left[\hat{\mathbf{U}}\right] \in \mathbb{R}^{m\times n}$ is a matrix with orthonormal columns. Since $[\mathbf{V}]$ is unitary we can multiply both sides of Eq.(1.15) to obtain the *reduced singular value decomposition*

$$[\mathbf{A}] = \left[\hat{\mathbf{U}}\right]\left[\hat{\mathbf{\Sigma}}\right][\mathbf{V}]^T. \tag{1.15}$$

Example 1.18. *We will study the SVD projections of the matrices*

$$[\mathbf{A_1}] = \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix}, \quad [\mathbf{A_2}] = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}, \quad [\mathbf{A_3}] = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$$

in Fig.1.7.

MATLAB programming note 1.8.

We define the matrices used in Example 1.18

```
>> A1 = [1 2; 0 2];
>> A2 = [1 1; 4 1];
>> A3 = [2 0; 1 2];
```

The SVD is calculated using the function "svd"

```
>> [U,S,V] = svd(A1);
U =
    0.7497   -0.6618
    0.6618    0.7497
S =
    2.9208         0
         0    0.6847
V =
    0.2567   -0.9665
    0.9665    0.2567
```

The resultant SVD decomposition is used in Fig.1.7.

Notice that by the procedure in Eq.(1.13) we get that

$$\| [\mathbf{A}] \|_2 = \sigma_1.$$

We can also observe that for a square matrix $[\mathbf{A}]$, we can add the following items in Theorem

5. 0 is not an eigenvalue of $[\mathbf{A}]$
6. 0 is not a singular value of $[\mathbf{A}]$
7. $\det([\mathbf{A}]) \neq 0$.

Homework 1.32. *Determine SVDs of the following matrices (by hand calculation)*

$$a)\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}, \quad b)\begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix}, \quad c)\begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix},$$

Homework 1.33. *Determine SVDs of the following matrices (by hand calculation)*

$$a)\begin{pmatrix} 4 & 0 \\ 3 & 1 \end{pmatrix} \quad b)\begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} \quad c)\begin{pmatrix} 4 & 8 & 1 \\ 0 & 2 & 0 \\ 0 & 6 & 7 \end{pmatrix}$$

Homework 1.34. *Use MATLAB to compute the SVD from the following matrices*

$$a)\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad b)\begin{pmatrix} 1 & \sqrt{2} & 0 \\ \sqrt{2} & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad c)\begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \quad d)\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

Homework 1.35. *For an arbitrary* $[\mathbf{A}] \in \mathbb{R}^{m\times m}$ *and norm* $\|.\|$ *prove that*

$$\lim_{n\to\infty} \| [\mathbf{A}]^n \| = 0,$$

if and only if $\rho([\mathbf{A}]) < 1$.

Homework 1.36. *Two matrices* $[\mathbf{A}], [\mathbf{B}] \in \mathbb{R}^{m\times m}$ *are unitary equivalent if* $[\mathbf{A}] = [\mathbf{Q}]\,[\mathbf{B}]\,[\mathbf{Q}]^T$ *for some unitary* $[\mathbf{Q}] \in \mathbb{R}^{m\times m}$. *Is it true or false that* $[\mathbf{A}]$ *and* $[\mathbf{B}]$ *are unitary equivalent if and ony if they have the same singular values?*