

MATHEMATICS FOR SCIENCE STUDENTS

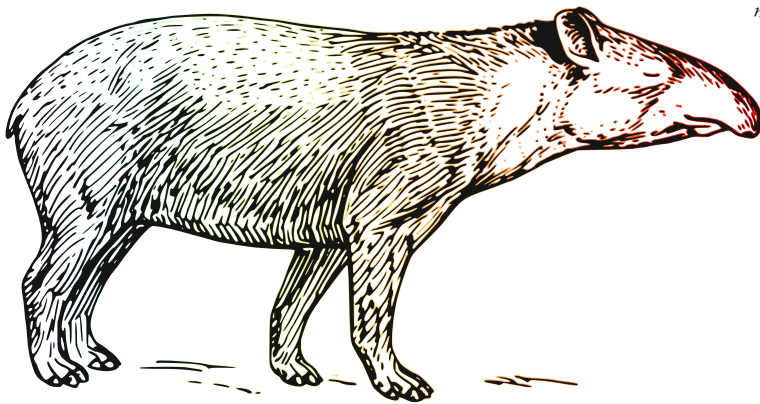
An open-source book

Written, illustrated and typeset (mostly) by

PELEG BAR SAPIR

with contributions from others

$$\begin{array}{l} a^b = e^{b \log(a)} \\ (a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k \\ \binom{n}{k} = \frac{n!}{k!(n-k)!} \\ T(\alpha \vec{u} + \beta \vec{v}) = \alpha T(\vec{u}) + \beta T(\vec{v}) \\ R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \\ A = Q \Lambda Q^{-1} \\ \Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \\ \frac{df}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \\ \langle \hat{e}_i, \hat{e}_j \rangle = \delta_{ij} \\ \vec{v} = \sum_{i=1}^n \alpha_i \hat{e}_i \\ e^{\pi i} + 1 = 0 \\ \int_a^b f(x) dx = F(b) - F(a) \\ \cos(x) = \sum_{n=0}^\infty \frac{(-1)^n}{(2n)!} x^{2n} \end{array}$$



PUBLISHED IN THE WILD

! To be written: Rights, lefts, etc. will be written here in the future **!**

HERE BE TABLE

CHAPTER

0



LINEAR ALGEBRA

(INTUITIVE APPROACH)

Linear algebra is one of the most important and often used fields, both in theoretical and applied mathematics. It brings together the analysis of systems of linear equations and the analysis of linear functions (in this context usually called linear transformations), and is employed extensively in almost any modern mathematical field, e.g. approximation theory, vector analysis, signal analysis, error correction, 3-dimensional computer graphics and many, many more.

In this book, we divide our discussion of linear algebra into two chapters: the first (this chapter) deals with a wider, birds-eye view of the topic: it aims to give an intuitive understanding of the major ideas of the topic. For this reason, in this chapter we limit ourselves almost exclusively to discussing linear algebra using 2- and 3-dimensional analysis (and higher dimensions when relevant) using real numbers only. This allows us to first create an intuitive picture of what is linear algebra all about, and how to use correctly the tools it provides us with.

The next chapter takes the opposite approach: it builds all concepts from the ground-up, defining precisely (almost) all basic concepts and proving them rigorously, and only

then using them to build the next steps. This approach has two major advantages: it guarantees that what we build has firm foundations and does not fall apart at any future point, and it also allows us to generalize the ideas constructed during the process to such extent that they can be used as foundation to build ever newer tools we can apply in a wide range of cases.

0.1 FUNCTIONS AS VECTORS

Note 0.1 Calculus ahead!

This section uses ideas discussed in the chapter about 1-dimensional real calculus (??). While strict knowledge of calculus is not necessary, some concepts would be rather difficult to understand without it.

In addition, this section is merely a brief introduction to a broader idea, and is not meant in any way to be an exhaustive analysis of the topic.



0.1.1 Properties of vectors

Up until now we used a rather informal definition for vectors (??). While the formal definition will be discussed in the next chapter, it is worth while to review the basic properties of vectors we saw so far. We start with the scaling of vectors:

- **It is ‘close’:** the result of scaling a vector is itself a vector.
- **The scalar 1 is neutral to scaling:** i.e. $1 \cdot \vec{v} = \vec{v}$.
- **It is associative:** for any $\alpha, \beta \in \mathbb{R}$ and $\vec{v} \in \mathbb{R}^n$,

$$\alpha \cdot (\beta \vec{v}) = (\alpha \cdot \beta) \vec{v}.$$

And for vector addition:

- **It is close:** the sum of any two vectors is also a vector.
- **It is commutative:** the order of addition does not change its result.
- **It is associative:** while not discussed directly, it is obvious that when adding together any three vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$, we can change the order of addition and that too will not change its result: calculating $\vec{u} + \vec{v}$ first and then adding \vec{w} to the result is the same as calculating $\vec{v} + \vec{w}$ first and then adding \vec{u} to the result, i.e.

$$(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w}).$$

- **$\vec{0}$ is neutral to addition:** i.e. the addition of $\vec{0}$ to any vector \vec{v} results in \vec{v} .
- **Any vector has an additive inverse:** given a vector $\vec{u} \in \mathbb{R}^n$, there is always a vector $\vec{v} \in \mathbb{R}^n$ such that

$$\vec{u} + \vec{v} = \vec{0},$$

namely the vector $\vec{u} = -\vec{v}$.

The two operations also have two important properties together:

- **Vector addition is distributive:** for any $\alpha \in \mathbb{R}$ and $\vec{u}, \vec{v} \in \mathbb{R}^n$,

$$\alpha(\vec{u} + \vec{v}) = \alpha\vec{u} + \alpha\vec{v}.$$

- **Scalar multiplication is distributive:** for any $\alpha, \beta \in \mathbb{R}$ and $\vec{v} \in \mathbb{R}^n$,

$$(\alpha + \beta)\vec{v} = \alpha\vec{v} + \beta\vec{v}.$$

All these properties seem rather obvious, but not all mathematical structures actually have them: for example, in a later chapter in the book we learn about **groups**, which do not have a scaling operator and don't necessarily have some of these properties, e.g. not all of them are commutative under their own operation.

Real functions, on the other hand, do have all these properties. We can therefore apply to real functions all the stuff we learned about vectors in this chapter. While we're not promised that everything will *look* the same, their general behavior is identical to that of vectors: the two obvious examples are scaling and addition:

- **Scaling:** given a real function f we can always scale it by multiplying its output by any real number α : if $f(x) = y$, then

$$\alpha f(x) = \alpha y.$$

For example, we can scale the polynomial $P(x) = x^3 - 2x + 5$ by a factor of 7, yielding

$$7P(x) = 7 \cdot x^3 - 7 \cdot 2x + 7 \cdot 5 = 7x^3 - 14x + 35.$$

And indeed we see that the result is itself a real function.

- **Addition:** any two real functions f, g can be added together. For example, given the functions $f(x) = e^x$ and $g(x) = 5 \sin(x)$, their sum is then

$$[f + g](x) = e^x + 5 \sin(x),$$

which is indeed a real function by itself.

You should go over all the above properties of vectors and verify for yourself that real functions do indeed possess them.

0.1.2 Smooth functions

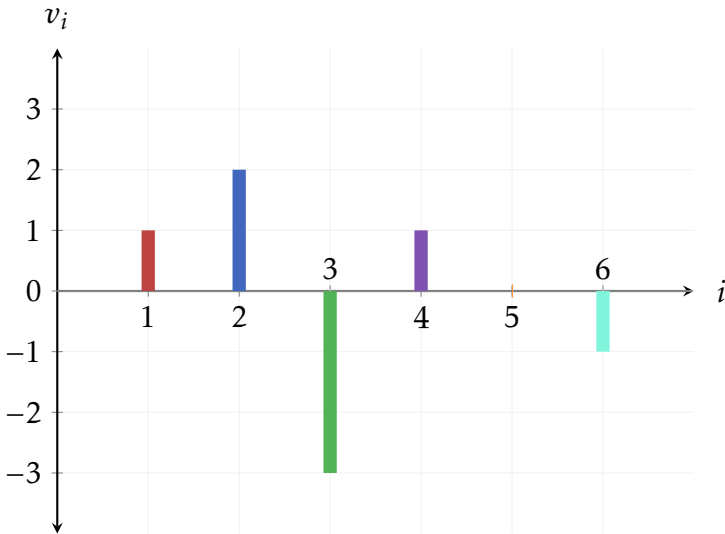
Before we move on, we should limit our further discussion to a specific set of real functions in order to avoid some annoying details which might rise later, and are not critical for the understanding of the idea of functions as vectors. Thus, from now on in the section we will only discuss functions of the set C^∞ . **Reminder:** a function f which is continuous on an interval $I \subseteq \mathbb{R}$ is said to be in C^k ($k \in \mathbb{N}$) if $f', f'', f''', \dots, f^{(k)}$ are all continuous on I as well. The set C^∞ is thus the set of all functions which are infinitely many times differentiable in some interval $I \subseteq \mathbb{R}$.

0.1.3 Function values as vector components

The first important thing to do when using vectors is to choose a basis set to represent them. For any \mathbb{R}^n it is rather easy to understand how this representation looks like: we simply write the vector with n components. How can we do that with functions? Well, consider the following vector in \mathbb{R}^6 :

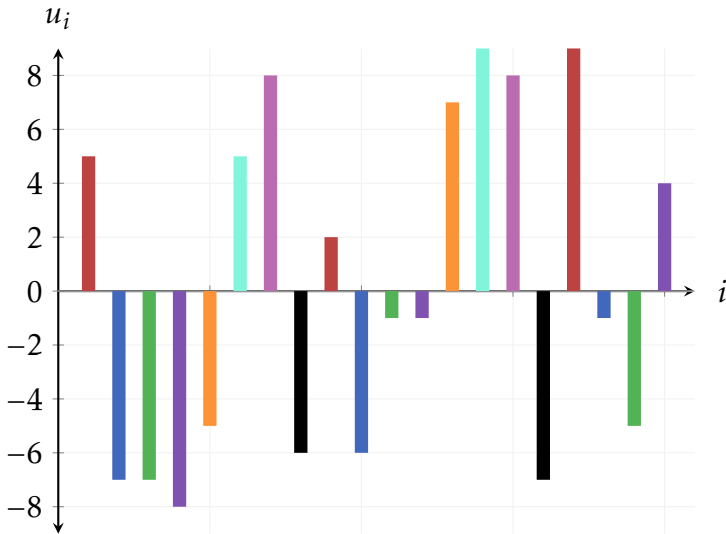
$$\vec{v} = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 1 \\ 0 \\ -1 \end{bmatrix}. \quad (0.1.1)$$

While drawing 6-dimensional spaces is rather difficult, we can draw as a bar chart each of the components v_i as a function of the component index i :

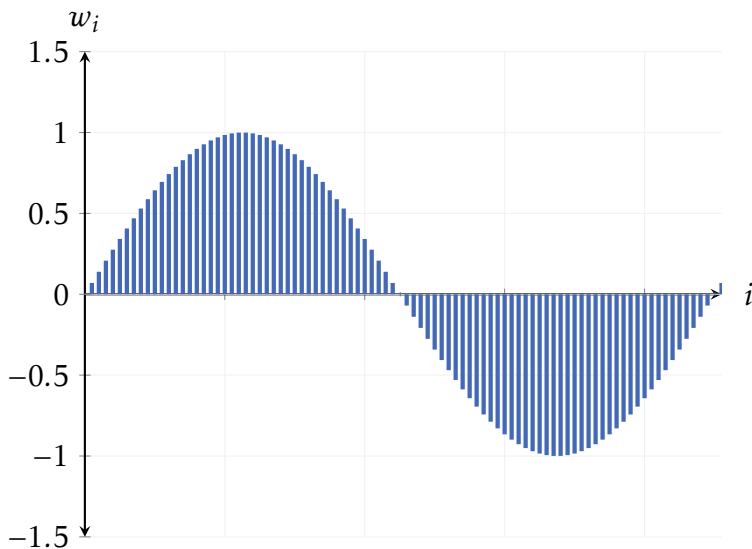


Now let's do the same with a vector in \mathbb{R}^{20} , e.g.

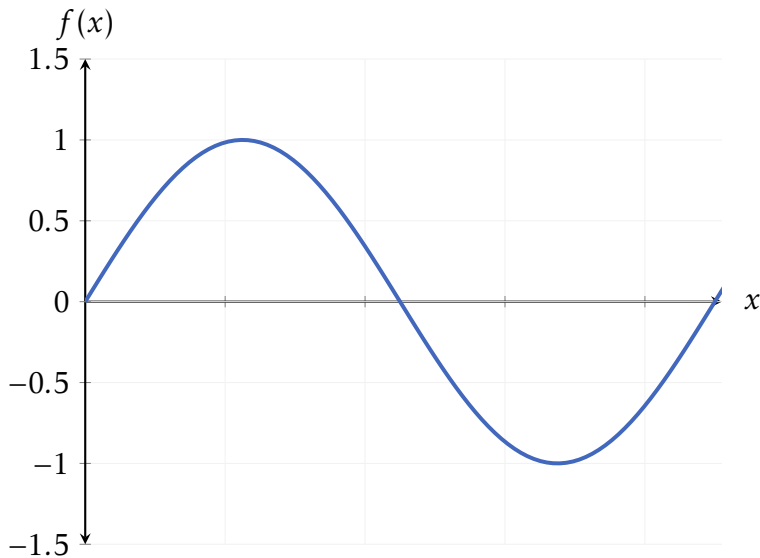
$$\vec{u} = [5, -7, -7, -8, -5, 5, 8, -6, 2, -6, -1, -1, 7, 9, 8, -7, 9, -1, -5, 4]^\top : \quad (0.1.2)$$



...and for some vector $\vec{w} \in \mathbb{R}^{90}$:



What we see here is that we can use this method for any natural number. We can also easily generalize it for any integer by shifting our indexing to include negative integers. In fact, we can expand it to any real number by taking this idea to the continuous limit: given any $x \in \mathbb{R}$, we assign it some $f(x) \in \mathbb{R}$ and say that $f(x)$ is the component corresponding to the index x . . . and we just defined a real function! For example, the function $f(x) = \sin(x)$:



We saw in the comparison between vectors in functions how scaling and addition behave in functions in a similar way to their behavior in vectors. Now we will describe how to view these operations in a similar way to their component-wise depiction in vectors (?? and ??):

- **Scaling:** given a scalar $\alpha \in \mathbb{R}$ and a function f which is smooth on an interval I , for each component $x \in I$, if x is mapped by f to $y \in \mathbb{R}$, then the function αf maps x to αy , i.e.

$$x \xrightarrow{f} y \Rightarrow x \xrightarrow{\alpha f} \alpha y.$$

Example: for $f(x) = x^2$, $x \xrightarrow{f} x^2$. Therefore for αf , $x \xrightarrow{\alpha f} \alpha x^2$.

- **Addition:** given two functions f, g which are smooth on an interval I , for any $x \in I$ if $x \xrightarrow{f} y_f$ and $x \xrightarrow{g} y_g$, then

$$x \xrightarrow{f+g} y_f + y_g.$$

Example: for $f(x) = 3x^2$ and $g(x) = -7x$, $x \xrightarrow{f+g} 3x^2 - 7x$.

0.1.4 Scalar product, norm

Given two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, the scalar product $\vec{u} \cdot \vec{v}$ is defined using the respective components of the vectors as

$$\vec{u} \cdot \vec{v} = \sum_{i=1}^n u_i v_i.$$

The natural extension of this sum of products to the continuous realm is via an integral (see ??), i.e. given two functions f, g which are smooth on the interval (a, b) (where $a, b \in \mathbb{R}$), then

$$f \cdot g = \int_a^b f(x)g(x)dx. \quad (0.1.3)$$

Example 0.1 Scalar product of two functions

Given the functions $f(x) = x^3 + 3x^2$ and $g(x) = \frac{1}{x}$, the scalar product $f \cdot g$ in the interval $I = [-1, 1]$ is

$$\begin{aligned} f \cdot g &= \int_{-1}^1 f(x)g(x)dx \\ &= \int_{-1}^1 \frac{x^3 + 3x^2}{x} dx \\ &= \int_{-1}^1 (x^2 + 3x) dx \\ &= \left(\frac{1}{3}x^3 + \frac{3}{2}x^2 \right) \Big|_{-1}^1 \\ &= \frac{1}{3} + \frac{3}{2} - \frac{-1}{3} - \frac{3}{2} \\ &= \frac{2}{3}. \end{aligned}$$



This brings up an interesting question: what does it mean when the scalar product of two functions is 0? For example, consider the two functions $f(x) = \sin(x)$ and $g(x) = \cos(x)$. Their scalar product $f \cdot g$ on the interval $I = (-\pi, \pi)$ is

$$\begin{aligned} f \cdot g &= \int_{-\pi}^{\pi} f(x)g(x)dx = \int_{-\pi}^{\pi} \sin(x)\cos(x)dx = \frac{\sin^2(x)}{2} \Big|_{-\pi}^{\pi} \\ &= \frac{1}{2} (\sin^2(\pi) - \sin^2(-\pi)) = \frac{1}{2} (0 - 0) = 0. \end{aligned}$$

Much like with vectors in \mathbb{R}^n we say that these two functions are orthogonal. In turn, this means that we can use them to form an orthonormal basis set to express all smooth functions on the same interval. In fact - a very similar idea, i.e. using infinite series of trigonometric functions to construct other functions, is used to form the **Fourier series**, which we explore in greater depth in ??).

The **Legendre polynomial** are an infinite set of orthogonal polynomials defined on the interval $[-1, 1]$, which arise in many different situations throughout science such as the

wave functions of the electrons in a Hydrogen-like atom. The first few Legendre polynomials are

$$\begin{aligned}
 P_0(x) &= 1, \\
 P_1(x) &= x, \\
 P_2(x) &= \frac{1}{2}(3x^2 - 1), \\
 P_3(x) &= \frac{1}{2}(5x^3 - 3x), \\
 P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3), \\
 P_5(x) &= \frac{1}{8}(63x^5 - 70x^3 + 15x), \\
 P_6(x) &= \frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5), \\
 &\vdots
 \end{aligned} \tag{0.1.4}$$

with the general formula for the n -th Legendre polynomial being

$$P_n(x) = \frac{1}{2^n} \sum_{k=0}^n \binom{n}{k} (x-1)^{n-k} (x+1)^k. \tag{0.1.5}$$

Let's check two combinations from the first 7 Legendre polynomials to see that they functions are indeed orthogonal:

- P_1 and P_3 :

$$P_1 \cdot P_3 = \frac{1}{2} \int_{-1}^1 (5x^4 - 3x^2) dx = \frac{1}{2} (x^5 - x^3) \Big|_{-1}^1 = 1 - 1 - 1 + 1 = 0.$$

- P_2 and P_5 :

$$\begin{aligned}
 P_2 \cdot P_5 &= \frac{1}{16} \int_{-1}^1 (3x^2 - 1)(63x^5 - 70x^3 + 15x) dx \\
 &= \frac{1}{16} \int_{-1}^1 (189x^7 - 210x^5 + 45x^3 - 63x^5 + 70x^3 - 15x) dx \\
 &= \frac{1}{16} (27x^8 - 42x^6 + 15x^4 - 12x^6 + 23x^4 - 7.5x^2) \Big|_{-1}^1 \\
 &= \frac{1}{16} [(27 - 42 + 15 - 12 + 23 - 7.5) - (27 - 42 + 15 - 12 + 23 - 7.5)] \\
 &= 0.
 \end{aligned}$$

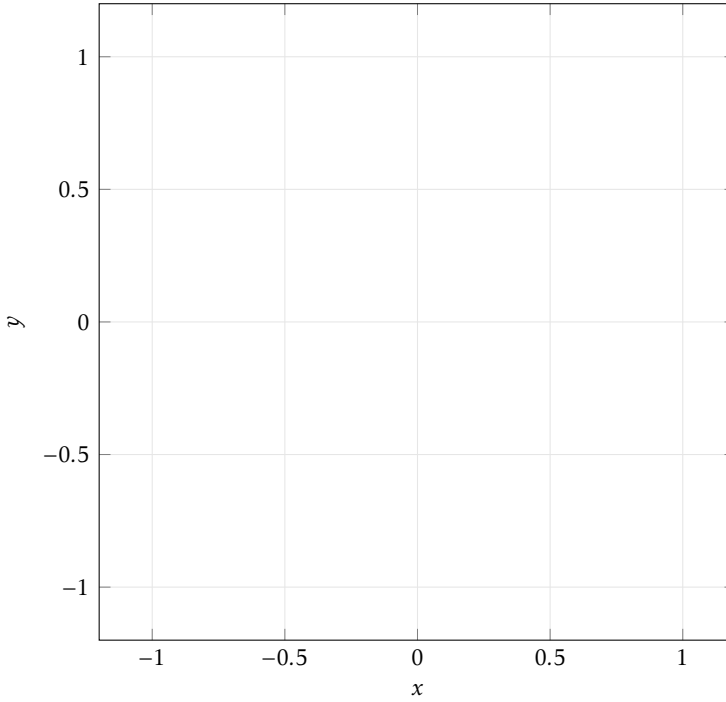


Figure 0.1 The first 6 Legendre polynomials.

0.1.5 Basis sets

0.1.6 Operators

In the context of vectors, linear transformations are functions that obey the two properties of scalability and additivity (??). With smooth real functions we can introduce a similar idea: **linear operators**. In our context, an operator is itself a function, however instead of being applied to vectors it is applied to smooth functions in some interval $[a, b]$, i.e.

$$O : C^\infty \rightarrow C^\infty. \quad (0.1.6)$$

Example 0.2 Operator

Let O be an operator $O : C^\infty \rightarrow C^\infty$ defined as

$$O(f(x)) = 2[f(x)]^2.$$

The following table shows some results of applying O to different smooth functions:

$f(x)$	$O(f(x))$
x	$2x^2$
x^2	$2x^4$
$x^2 + 1$	$2(x^2 + 1)^2$
$\sqrt{\frac{x}{2}}$	x
e^x	$2e^{2x}$
$\sin(x)$	$2\sin^2(x)$



A **linear operator** is therefore an operator which behaves like a linear transformation, i.e. given the operator $O : C^\infty \rightarrow C^\infty$ it is linear if it obeys the following two criteria:

- **Scalability:** for any $\alpha \in \mathbb{R}$ and $f \in C^\infty$,

$$O(\alpha f(x)) = \alpha O(f(x)).$$

- **Additivity:** for any $f, g \in C^\infty$,

$$O(f(x) + g(x)) = O(f(x)) + O(g(x)).$$

One very well known operator which obeys these two criteria is the derivative operator: recall that for any real number α and smooth functions f, g ,

$$\begin{aligned} f'(\alpha x) &= \alpha f'(x), \\ (f(x) + g(x))' &= f'(x) + g'(x). \end{aligned} \tag{0.1.7}$$

We can therefore apply to derivatives all the general ideas we learned about linear transformations! This is very useful, and will be used in later chapters.

In fact, in some cases we can even express the derivative operator as a matrix: consider the following set of real smooth functions:

$$B = \{1, x, x^2, x^3, \dots\}. \tag{0.1.8}$$

We can write any smooth function as an infinite sum of the functions in B , i.e. a Maclaurin expansion (??). Therefore, the set spans C^∞ . For example,

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \dots$$

Since all the terms of the series are needed in order to express all the possible functions in C^∞ , it is also has a minimal set which spans C^∞ - and thus B is a basis set of C^∞ .

Note 0.2 Orthogonality of B

B is not an orthogonal set on \mathbb{R} , since e.g.

$$x \cdot x^2 = \int_{-\infty}^{\infty} x^3 dx = \frac{1}{4} x^4 \Big|_{-\infty}^{\infty} = \infty + \infty = \infty.$$



We can represent any function spanned by B as an infinite column vector, where each component is the respective coefficient of $1, x, x^2, \dots$ in the function's Maclaurin expansion. For example,

$$e^x = \begin{bmatrix} 1 \\ 1 \\ \frac{1}{2} \\ \frac{1}{3!} \\ \frac{1}{4!} \\ \vdots \end{bmatrix}$$

The basis functions in B can then be written as

$$1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \quad x = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \quad x^2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix}, \quad x^3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix}, \quad \dots$$

Recall that the first derivative of a general n -degree polynomial

$$P(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots + a_n x^n$$

is

$$P'(x) = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \dots + na_n x^{n-1}.$$

(??)

We can describe the derivative as follows: for any n we set the coefficient of x^{n-1} in the derivative to na_n , and set the coefficient of x^n to be zero:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \end{bmatrix} \mapsto \begin{bmatrix} a_1 \\ 2a_2 \\ 3a_3 \\ 4a_4 \\ \vdots \end{bmatrix},$$

and the effect on the basis vector is then

$$\begin{array}{cccc}
 P(x) & \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} & \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix} \\
 & \downarrow & \downarrow & \downarrow & \downarrow \\
 P'(x) & \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} & \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} & \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \\ \vdots \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 3 \\ 0 \\ \vdots \end{bmatrix}
 \end{array}$$

Now we can easily write the matrix representation of the derivative in this basis:

$$D = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 2 & 0 & \dots \\ 0 & 0 & 0 & 3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (0.1.9)$$

Since the derivative transforms the first basis vector into the zero vector (i.e. the derivative of all constants is always zero), D has a column of zeros. Having a zero column means that the determinant of the derivative is zero. This makes sense: the derivative “looses” a dimension, i.e. it isn’t an injection: as mentioned, all constants are transformed into zero - and thus the derivative isn’t invertible. The closest we get to recover an original function given its derivative is by using the anti-derivative (i.e. indefinite integrals) - but this gives us only a family of functions and not the any specific one we used.

Note 0.3 Infinite dimensions

It is important to note that we didn’t define what infinite column vectors and matrices are, and a-priori there is no guarantee that we can even perform any of the operations we use for finite-dimensional vectors and matrices. However, what we showed here is correct - but a more advanced and rigorous linear algebra is needed in order to prove it.



We can use the Gram-Schmidt process to orthogonalize the basis B . In order to avoid infinite scalar products we use the interval $I = [-1, 1]$. First, let’s quickly calculate the first three square norms of B :

$$\begin{aligned}
 \|1\|^2 &= \int_{-1}^1 1 \cdot 1 \, dx = x \Big|_{-1}^1 = 2, \\
 \|x\|^2 &= \int_{-1}^1 x \cdot x \, dx = \int_{-1}^1 x^2 \, dx = \frac{1}{3} x^3 \Big|_{-1}^1 = \frac{2}{3},
 \end{aligned}$$

$$\begin{aligned}\|x^2\|^2 &= \int_{-1}^1 x^4 \cdot x \, dx = \frac{1}{5} x^5 \Big|_{-1}^1 = \frac{2}{5}, \\ &\vdots \\ \|x^n\|^2 &= \int_{-1}^1 x^{2n} \cdot x \, dx = \frac{1}{2n+1} x^{2n+1} \Big|_{-1}^1 = \frac{2}{2n+1}.\end{aligned}$$

The following table summarizes the dot product of any two of the basis functions $1, x, x^2$ and x^3 on $[-1, 1]$:

	1	x	x^2	x^3
1	2	0	$\frac{2}{3}$	0
x	0	$\frac{2}{3}$	0	$\frac{2}{5}$
x^2	$\frac{2}{3}$	0	$\frac{2}{5}$	0
x^3	0	$\frac{2}{5}$	0	$\frac{2}{7}$

Using these values we can now apply the GSP (we use the bracket notation as to not confuse the scalar product with the regular product):

$$\begin{aligned}1 &\longrightarrow 1, \\ x &\longrightarrow x - \frac{\langle 1, x \rangle}{2} = x - 0 = x, \\ x^2 &\longrightarrow x^2 - \frac{\langle 1, x^2 \rangle}{2} - \frac{\langle x, x^2 \rangle}{\frac{2}{3}} x = x^2 - \frac{1}{3}, \\ x^3 &\longrightarrow x^3 - \frac{\langle 1, x^3 \rangle}{2} - \frac{\langle x, x^3 \rangle}{\frac{2}{3}} x - \frac{\langle x^2, x^3 \rangle}{\frac{2}{5}} x^2 = x^3 - \frac{2}{3} x - \frac{3}{5} x^2.\end{aligned}$$

The resulting orthogonal basis set is exactly the Legendre polynomials, up to scaling! This is not really suprising, as the Legendre polynomials are orthogonal, after all.

0.1.7 Eigenfunctions

Some operators have **eigenfunctions**: functions that are scaled when the operator is applied to them. In the case of the derivative operator, for example, one such function is the exponent function, since

$$\left[e^{\alpha x} \right]' = \alpha e^{\alpha x}. \quad (0.1.10)$$

The corresponding eigenvalue of $e^{\alpha x}$ is therefore α .