Essential Mathematics for Neuroscience

Fabian Sinz, Jakob Macke & Philipp Lies

October 16, 2009

Contents

Chapter 1

Basics

This chapter serves as an introduction to a few basic elements that will be needed troughout the course. We begin by reviewing basic families of functions like *linear functions*, *polynomials* and *trigonometric functions*, as well as some of their properties. Afterwards we will look at some elementary calculus on those types of functions.

1.1 Essential Functions

A function is a rule that relates to sets of quantities, the *inputs* and the *outputs*. Each input x is deterministially related to an output $f(x)$. For example, $f(x)$ might temperature on day x, or the firing rate of a neuron in response to a stimulus x. Thus, functions can be used as mathematical models of processes in which one quantity is transformed into another in a deterministic way. Even when the process of transformation is not deterministic, usually an underlying deterministic process corrupted by random noise can be used. In the example above, the firing rate of the neuron could be $f(x) + \varrho$, where ϱ is a noise-term. In contrast to a deterministic function, $f(x) + \varrho$ denotes a whole set of values for a given x since the random term ϱ can take different values for each trial. Therefore, $f(x) + \varrho$ is not a function in the strict sense. The reason is that, functions—by definition—are rules how to assign elements x of one set to *unique* elements $f(x)$ of another set. Only if the target elements are unique, the assignment rule is called *function*. When defining a function, we have to specify the two *sets* between the function is mapping and the *rule* that transforms an element of the target set to an element of the input set. For example, if we want to define a function f that is transforming elements of a set A into elements of a set B according to the rule r, we would write this as

$$\begin{aligned} f: \quad A &\to B \\ a &\mapsto r(a). \end{aligned}$$

Here, a is an element of A (written $a \in A$) and $r(b)$ is an element of B (i.e. $r(b) \in B$). The set A is usually called *domain of* f while B is called *the co-domain of* f. The arrow "$\to$" is used to denote the mapping between the two sets, while "$\mapsto$" denotes the mapping from an element of the domain to an specific element of the co-domain. This means that "$\to$" tells us what kind of objects are mapped into another and "$\mapsto$" specifies the assignment rule.

The rule r can be anything that can be done with elements of A. For example, if the function f simply doubles any real number, we would write

$$\begin{aligned} f: \quad \mathbb{R} &\to \mathbb{R} \\ x &\mapsto 2 \cdot x \quad x \in \mathbb{R}. \end{aligned}$$

In most cases, the inputs and outputs of function will be numbers, but this does not necessarily have to be the case (i.e. the elements of the domain A and the co-domain B do not need to be numbers).

Although, in principle, there are infinitely many functions on the real numbers, knowing only a few of them is usually enough to get along well in most natural sciences. The reason is that most complex functions are built by adding, multiplying, or composing simpler ones. It is important that you get comfortable with those simper functions since they are your toolbox to understand and build more complex functions. Once you have and intuition how those simple functions behave, it is often not too difficult to get a feeling for a more complicated one. In this section we will review the most important simple functions and present their most important properties.

1.1.1 Polynomials and Powers

Polynomials is a very common class of functions. The two most widely known kinds of polynomials are the parabola $f(x) = x^2$ and the more general quadratic function $f(x) = ax^2+bx+c$. In general, polynomials consist of a sum of positive integer powers k of x with coefficients a_k:

$$f(x) = a_n x^n + ... + a_1 x + a_0.$$

The single terms in the sum are called *monomials.* The *degree* of the polynomial is the largest exponent of its monomials. The polynomial above has a degree of n. Polynomials have nice properties like e.g. the *derivatives* and *anti-derivatives* of polynomials are easy to calculate and yield polynomials again. One frequent use of polynomials is to approximate any function at a certain location. This approximation is called *Taylor-Expansion.* We will discuss the Taylor-Expansion and many properties of polynomials in later chapters.

This is a good point to introduce the notation for sums over several elements: Instead of indicating the entire sum by three dots "..." we use the greek uppercase letter sigma Σ (like sum) to indicate a sum over all terms directly after the sigma. These terms are usually indexed and the range of the index is written below and above the Σ. Since $x^0 = 1$ for all $x \in \mathbb{R}$ we write the polynomial from above as

$$\begin{aligned} f(x) &= a_n x^n + ... + a_1 x + a_0 \\ &= \sum_{k=0}^{n} a_k x^k. \end{aligned}$$

While polynomials have exponents $k \in \mathbb{N}_0$ (where $\mathbb{N}_0$ denotes the set of natural number including 0), exponents can in principle be in $\mathbb{R}$ as well. There are two most important cases: when the exponent is negative and when it is a rational number (i.e. a number that can be written as a fraction). A negative exponent of a number is merely a shortcut for $x^{-a} = \frac{1}{x^a}$. In many cases, for example when calculating derivatives, the notation with negative exponent is useful. A fraction in the exponent is another way of writing roots. For example the square root $\sqrt{x}$ is equivalently written as $x^{\frac{1}{2}}$. In general, the nth root of x can be written as $\sqrt[n]{x} = x^{\frac{1}{n}}$.

We conclude this section by stating a few calculation rules for powers for $x, a \in \mathbb{R}$. You should know all of them by heart and be able to use them effortlessly.

Calculation Rules for Powers
The following rules apply to any $x, a \in \mathbb{R}$:

1. Anything to the power of zero is one: $x^0 = 1$
2. Multiplying two terms with the same basis is equivalent to adding their exponents: $x^a \cdot x^b = x^{a+b}$
3. Dividing two terms with the same basis is equivalent to subtracting their exponents: $\frac{x^a}{x^b} = x^a \cdot x^{-b} = x^{a-b}$
4. Exponentiating a term is equivalent to multiplying its exponents: $(x^a)^b = x^{a \cdot b}$
5. A special case of rule 3. is given by $\frac{1}{x^a} = x^{-a}$
6. The a^{th} root of x is given by $\sqrt[a]{x} = x^{\frac{1}{a}}$ for $x \geq 0$.

1.1.2 Linear Functions

Linear functions are among the simplest functions one can imagine. You can imagine a linear function as a line (plane, or hyperplane) through the origin. Algebraically, their key property is that the function value of a sum $x + y$ of elements x, y equals the sum of their function values $f(x) + f(y)$.The same is true for multiples of input elements, i.e. the function value of some multiple $a \cdot x$ of an element x from the domain is the multiple of the function value $a \cdot f(x)$. If any function fulfills these two properties, it is linear by definition.

Definition (Linear Function) A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is said to be *linear* if it fulfills the following two properties:

$$f(x + y) = f(x) + f(y) \qquad \text{for all } x, y \in \mathbb{R} \tag{1.1}$$
$$f(a \cdot x) = a \cdot f(x) \qquad \text{for all } a, x \in \mathbb{R}. \tag{1.2}$$

$\diamondsuit$

These properties have remarkable consequences. While for general functions, a single input-output pair of values $(x, f(x))$ does not tell anything about the value of the function at other locations $y \neq x$, a single such pair is enough to know the value of a linear function at any location: Assume we are given the input-output pair $(x, f(x))$ and we know that f is linear. In order to calucate the value of f at another location y, we search for a scalar a that scales x into y, i.e. $y = a \cdot x$. Clearly, this scalar is easy is given by $a = \frac{y}{x}$. Once we know a, we can compute $f(y)$ via

$$\begin{aligned} f(y) &= f(a \cdot x) \\ &= a \cdot f(x). \end{aligned}$$

Example (Mathematical Modelling of Receptive Fields) For some neurons, it is often assumed that their responses, i.e. the spike rate $r(x)$, depends linearly on the stimulus x.

Assume our cell responds to a visual image I_1 with a spike rate of $r_1 = 20$ spikes per second and to another image I_2 with $r_2 = 60$ spikes per second. What spike rate would we expect to the mean of I_1 and I_2, if our neuron was truely linear in the stimuli? The answer is easy to calculate. Let $r : \mathcal{I} \to \mathbb{R}$ denote the function from images (denoted by $\mathcal{I}$) to spike rate. We already know $r(I_1) = r_1 = 20\frac{sp}{s}$ and $r(I_2) = \nu_2 = 60\frac{sp}{s}$. Then the response to the mean of the two images is

$$\begin{aligned} r\left(\frac{1}{2}I_1 + \frac{1}{2}I_2\right) &= r\left(\frac{1}{2}I_1\right) + r\left(\frac{1}{2}I_2\right) \\ &= \frac{1}{2}r(I_1) + \frac{1}{2}r(I_2) \\ &= \frac{1}{2}r_1 + \frac{1}{2}r_2 \\ &= 10\frac{sp}{s} + 30\frac{sp}{s} \\ &= 40\frac{sp}{s} \end{aligned}$$

This property does not only hold for two input stimuli. It holds for an arbitrary number of stimuli. If the rate function r realized of our neuron is linear, then response to the mean of n images is just the mean response to the single images.

$$\begin{aligned} r\left(\frac{1}{n}\sum_{k=1}^{n} I_k\right) &= \frac{1}{n}\sum_{k=1}^{n} r\left(I_k\right) \\ &= \frac{1}{n}\sum_{k=1}^{n} \nu_k \end{aligned}$$

Question: $\boxed{?}$

Of course, real neurons are not truly linear. If a neuron was indeed linear, for inputs, this would lead to some very unrealistic conclusions. Name two of them!

Answer:

For some stimuli, the spike rate would be negative. Also, for stimuli with very high input, the spike-rate would be arbitrarily large, i.e. the neuron's rate would not saturate.

◁

1.1.3 Trigonometric Functions

Trigonometric functions are functions of an angle ϑ. The most common trigonometric functions are $\sin(\vartheta)$, $\cos(\vartheta)$, $\tan(\vartheta)$ and $\text{cotan}(\vartheta)$.

In general, there are two natural ways to think about trigonometric functions: the geometrical view quantities and the periodic signal view.

1.1.3.1 The geometric view

In the geometric view, $\cos(\vartheta)$ and $\sin(\vartheta)$ represent the x-coordinate and the y-coordinate of a point on the intersection between a circle with radius one centered at the origin, and a line through the origin that encloses an angle of ϑ with the x-axis. In this view, $\tan(\vartheta)$ and $\text{cotan}(\vartheta)$ have a natural interpretation as well, namely the length of the line, touching the circle, between the upper leg of the angle and the x-axis or the y-axis, respectively. Alternatively, $\tan(\theta)$ is the ratio between the $x-$and the $y-$coordinate (see Figure 1.1).

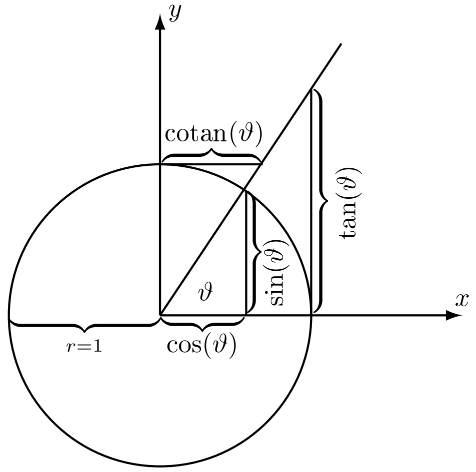
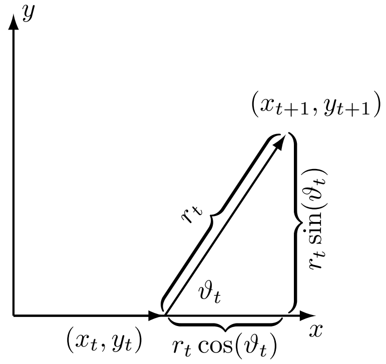


Figure 1.1: Geometrical view of $\sin(\vartheta)$, $\cos(\vartheta)$, $\tan(\vartheta)$ and $\text{cotan}(\vartheta)$. For a given angle ϑ, $(\cos(\vartheta), \sin(\vartheta))$ are the coordinates of the point on the intersection between a circle of radius $r = 1$ and the upper leg of the angle. $\tan(\vartheta)$ is the length of the line between the x-axis and the upper leg of the angle, that "touches" the unit circle (therefore the name *tangens* from *lat. tangere = to touch*). $\text{cotan}(\vartheta)$ is defined analogoulsy for the line touching the circle from above.

Example (Path Integration) The term *path integration* denotes the ability of moving organisms (such as ants) to remember the direction and length of the vector to its home while moving in the environment. We will now just look at a special case of updating the home vector in world coordinates, i.e. a global fixed coordinate system. This means that the home base is assigned the coordinates $(0, 0)$ and the moving organism stores its position according to a global coordinate frame.

Imagine you are an ant living in a completely flat world. For the sake of simplicity we further imagine that you have a compass and you know the length of your steps, so that you can measure the angles you turn and the distance you walked. Now imagine that you already explored your environment for a while and that you are now standing at position (x_t, y_t) looking along the x-axis. If you know turn by an angle of ϑ_t and move into the new direction by a distance of r_t, what is your new position (x_{t+1}, y_{t+1})?



Looking again at figure 1.1 and the figure above shows that the new direction in which you are walking is $(\cos(\vartheta_t), \sin(\vartheta_t))$. Since you are walking along that direction for a distance of r_t, the total displacement is $r \cdot (\cos(\vartheta_t), \sin(\vartheta_t))$. If we add this displacement to the old position, we get the new position

$$(x_{t+1}, y_{t+1}) \quad = \quad (x_t, y_t) + r_t \cdot (\cos(\vartheta_t), \sin(\vartheta_t))$$

$\triangleleft$

Due to the geometrical property of $\sin(\vartheta)$ and $\cos(\vartheta)$ we can derive a very useful equality by employing Pythagoras theorem.

Theorem (Pythagoras) For a right angle triangle with side lengths a, b and c, where c is the length of the longest leg, while the legs with length a and b enclose an angle of $90°$ (or $\frac{\pi}{2}$ in radians), the following equality holds:

$$a^2 + b^2 \quad = \quad c^2$$

$\diamondsuit$

In the case of $\sin(\vartheta)$ and $\cos(\vartheta)$, we know the value of c for any given ϑ. It is simply $c = 1$, since the point $p = (\cos(\vartheta), \sin(\vartheta))$ lies on the circle of radius $r = 1$. Therefore, we know that

$$\cos(\vartheta)^2 + \sin(\vartheta)^2 \quad = \quad 1$$

for all angles ϑ.

There is another useful equality that we can read of figure 1.1. Assume we want to know the scaling factor s that transforms $\sin(\vartheta)$ into $\tan(\vartheta)$, i.e. $\sin(\vartheta) \cdot s = \tan(\vartheta)$. From looking at figure 1.1 we know that it is the same scaling factor that scales $\cos(\vartheta)$ into 1, i.e. $\cos(\vartheta) \cdot s = 1$. In this case, s is easy to calculate: It is simply $s = \frac{1}{\cos(\vartheta)}$. Therefore we have found a formula how to calculate $\tan(\vartheta)$ from $\sin(\vartheta)$ and $\cos(\vartheta)$:

$$\tan(\vartheta) = \frac{\sin(\vartheta)}{\cos(\vartheta)}.$$

In an analogous manner we can also derive the equality

$$\text{cotan}(\vartheta) = \frac{\cos(\vartheta)}{\sin(\vartheta)}.$$

Since $\tan(\vartheta)$ and $\text{cotan}(\vartheta)$ can be written in terms of a quotient, the radius of the circle does not even have to be one. Assume we want to know the tangens between the x-axis and the leg $(0, p)$ for a point $p = (x, y)$ that lies on a circle with radius r. Since we can write p equivalently as $p = (x, y) = r \cdot (\cos(\vartheta), \sin(\vartheta))$ for an appropriate value of r, we have that

$$\frac{y}{x} = \frac{r \cdot \sin(\vartheta)}{r \cdot \cos(\vartheta)} = \tan(\vartheta).$$

Therefore, $\tan(\vartheta)$ is simply the quotient of the opposite leg and the adjacent leg of a right angle triangle. Similarily we can get

$$\frac{x}{y} = \frac{r \cdot \cos(\vartheta)}{r \cdot \sin(\vartheta)} = \text{cotan}(\vartheta).$$

Further, sine and cosine can also be defined in terms of quotients for circles with an arbitrary radius. According to the intercept theorem the ratio of $\cos(\vartheta)$ to 1 is equal to the ratio of x to r for every point $p = (x, y)$ on a circle wirh radius r. Equally, the ratio of $\sin(\vartheta)$ to 1 is equal to the ratio of y to r. Therefore, we get

$$\begin{aligned} \sin(\vartheta) &= \frac{y}{r} \\ \cos(\vartheta) &= \frac{x}{r} \end{aligned}$$

Other useful properties that can also be read off from the geometric view. Among them are the symmetry properties of $\sin(\vartheta)$, $\cos(\vartheta)$, $\tan(\vartheta)$ and $\text{cotan}(\vartheta)$:

$$\begin{aligned} \cos(-\vartheta) &= \cos(\vartheta) \\ \sin(-\vartheta) &= -\sin(\vartheta) \\ \tan(-\vartheta) &= -\tan(\vartheta) \\ \text{cotan}(-\vartheta) &= -\text{cotan}(\vartheta) \end{aligned}$$

Sometimes we do not want to compute the sine or cosine of an angle but the angle itself. For example, we take point $p = (x, y)$ and want to know what the angle between the x-axis and the line from the origin to p is. We know that $\tan(\vartheta) = \frac{y}{x}$, but what is ϑ? To solve for ϑ, we need the inverse trigonometric functions. For every trigonometric function there is an inverse trigonometric function: $\arccos(x)$, $\arcsin(x)$, $\arctan(x)$, $\text{arccotan}(x)$. In some texts they are confusingly written as $\cos^{-1}(\vartheta)$ whereas $\cos(\vartheta)^{-1}$ means $1/\cos(\vartheta)$, so you should always use the arc-notation. Now, by using the arctan function we can compute the angle between p and the x-axis as $\vartheta = \arctan(\frac{y}{x})$.

Remark There are two units for measuring angles which are commonly used and get quite often mixed up: *radians* and *degrees*. A full circle of 360° corresponds to 2π radians. In computer programs, the default unit is radians. Therefore to compute the sine of 45° one has to compute $\sin(\pi/4)$. Likewise if you computed an angle by using $\arccos(x)$ your output is in radians. To convert from radians to degrees, you have to multiply your result by $\frac{180}{\pi}$ and similarly converting from degrees to radians by multiplying with $\frac{\pi}{180}$.

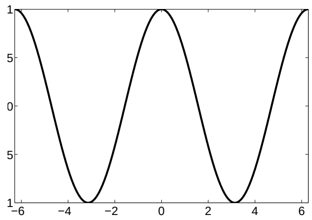
$\triangleleft$

1.1.3.2 The Periodic Signal View

The periodic signal view of thinking about $\sin(\vartheta)$ and $\cos(\vartheta)$ is especially useful when dealing with signals such as e.g. membrane potentials of neurons or stripes of natural images. It is also closely related to techniques such as Fourier- or Spectral Analysis, which are indispensable tools for the analysis of neurophysiological signals. In order to get from the geometric to the periodic signal view, just imagine a point on the unit circle that is moving with constant speed counterclockwise along the circle. If we plot the time t against the point's x-coordinate $x(t) = \cos(\vartheta(t))$, we get a strongly periodic function. Same applies to the y-coordinate $y(t) = \sin(\vartheta(t))$. Figure 1.2 shows the graphs of the two functions. For the moment we wrote ϑ as a function of time in order to be able to say that the point is moving with constant speed. The faster the point is moving along the circle the more cycles we can get in one fixed time interval. This is expressed by the frequency ω of the sine or cosine, respectively. It tells us how many cycles our point does in one unit time interval. We can therefore equivalently write $x(t) = \cos(\omega \cdot t)$ and $y(t) = \sin(\omega \cdot t)$. From now on we will drop the dependence on time. The frequency then tells us how many cycles of our point fit in the interval $[0, 2\pi]$.

Example (Sine and Cosine Gratings) Assume that you want measure the orientation selectivity a V1-cell you are recording from with an electrode. A simple experiment would be to present gratings with different orientations and varying amount of bars in one unit interval, i.e. different spatial frequency. The most common way of producing such patterns is to use *sine* and *cosine gratings*. Figure 1.3 shows such a grating.

Since an image can be seen as a function,that assigns a graylevel value to each pixel ($f(x)$is the gray value of pixel x) those gratings are simply sine or



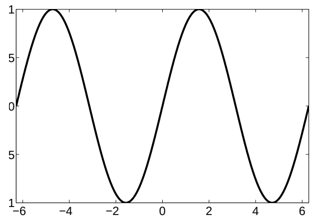


Figure 1.2: $\cos(\omega t)$ (left) and $\sin(\omega t)$ (right) in the interval of $t \in [-2\pi, 2\pi]$ with a frequency of $\omega = 1$.

cosine functions over $\mathbb{R}^2$. In order to get to the graylevel values at the different pixels, the function is discretized, i.e. is only evaluated at certain locations that correspond to the pixels. At the moment we shall only look at how to produce vertical or horizontal gratings. We will see how to produce gratings of arbitrary orientation later.

For a vertical grating, we know that the graylevel value along the vertical axis must be constant. If $I(x, y)$ denotes the function that represents the image, i.e. the functions that assigns a graylevel value to a position (x, y), we can produce a vertical grating by $I(x, y) = \sin(\omega x)$. Instead of using the sine we could as well have used the cosine function. A horizontal grating can be generated in exactly the same way by replacing x by y inside the sine function. Here is the matlab code to produce a horizontal grating of frequency $\omega = 2$:

```
>> [X,Y] = meshgrid([-2*pi:0.01:2*pi]); % get the sample points
>> omega = 2; % set the frequency
>> imagesc(sin(omega*X)) % display grating
>> colormap(gray)  % set colormap to gray values
>> axis off % switch off the axes
```

At the moment all our gratings have a fixed contrast, since $-1 \leq \sin(x), \cos(x) \leq 1$ for all $x \in \mathbb{R}$. However, we can vary the contrast by varying the amplitude of the sine. This is done by premultiplying an appropriate scaling factor $I(x, y) = A \cdot \sin(\omega x)$. In order to build that into the matlab code above, you must specify the maximum and the minimum gray value when calling the function `imagesc`, since it automatically scales the gray values otherwise

```
>> [X,Y] = meshgrid([-2*pi:0.01:2*pi]); % get the sample points
>> omega = 2; % set the frequency
>> A = 3; % set the contrast
>> maxA = 10; % set maximal contrast
>> imagesc(A*sin(omega*X),[-maxA,maxA]) % display grating
>> colormap(gray)  % set colormap to gray values
>> axis off % switch off the axes
```

$\triangleleft$

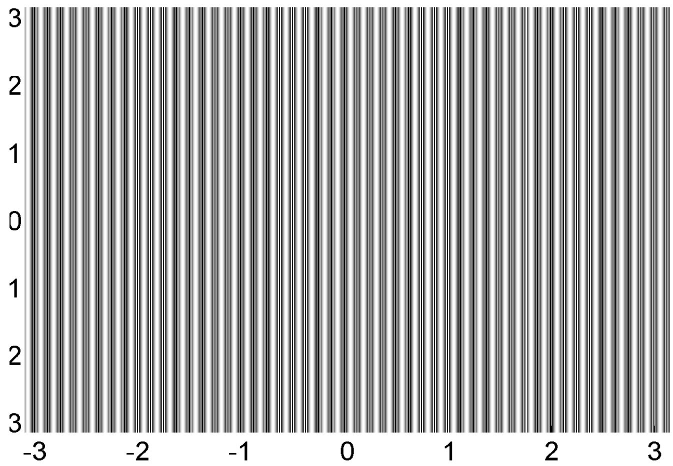


Figure 1.3: Example of a vertical sine grating with a spatial frequency of approx. 8Hz along the x-axis. The graylevel value at each position (x, y) is given by $I(x, y) = \sin(50 \cdot x)$.

Up to now we saw how to control the frequency and the amplitude of sine and cosine functions. In order to complete this subsection we will also see how to shift the sine and the cosine functions along the x-axis. Let us look at a sine function with a certain frequency $\sin(\omega t)$. At $t = 0$ also $\sin(\omega t) = 0$. If we want to shift the sine function along the x-axis by an offset of ϕ, we must ensure that the shifted version is zero at $t = \phi$ and not at $t = 0$. However, this will be the case, if we subtract ϕ from the argument of the sine. Therefore, a sine that is shifted by an offset of ϕ along the x-axis is given by $\sin(\lambda t - \phi)$. Cosine functions are shifted in exactly the same way. This offset ϕ is called *phase* of the signal. Now we are able to write down the general form of a sine or cosine signal with a given amplitude A and phase ϕ: It is

$$A \cdot \sin(\omega t - \phi) \quad \text{and} \quad A \cdot \cos(\omega t - \phi).$$

Before finishing this section about trigonometric functions we just want to mention two equalities that are useful when calculating with sine and cosine. These equalities are called the *Addition Theorems*.

Theorem (Addition Theorems) The following equalities hold for all $x, y \in \mathbb{R}$:

$$\begin{aligned} \cos(x + y) &= \cos(x)\cos(y) - \sin(x)\sin(y) \\ \sin(x + y) &= \sin(x)\cos(y) + \cos(x)\sin(y) \end{aligned}$$

$\Diamond$

Using the symmetry properties of $\sin(x)$ and $\cos(x)$ one can also derive similar expressions for $\cos(x-y)$ and $\sin(x-y)$.

Exercise

E

Write down the the corresponding expressions for $\cos(x-y)$ and $\sin(x-y)$.

$\lhd$

Important Rules for Trigonometric Functions
The following rules apply to any $x, y, \vartheta \in \mathbb{R}$:

- Pythagoras's Theorem: $\cos(\vartheta)^2 + \sin(\vartheta)^2 = 1$
- Symmetry Properties:

$$\begin{aligned}
\cos(-\vartheta) &= \cos(\vartheta) \\
\sin(-\vartheta) &= -\sin(\vartheta) \\
\tan(-\vartheta) &= -\tan(\vartheta) \\
\mathrm{cotan}(-\vartheta) &= -\mathrm{cotan}(\vartheta)
\end{aligned}$$

- Addition Theorems

$$\begin{aligned}
\cos(x+y) &= \cos(x)\cos(y) - \sin(x)\sin(y) \\
\sin(x+y) &= \sin(x)\cos(y) + \cos(x)\sin(y)
\end{aligned}$$

1.1.4 The e-function and the Logarithm

1.1.4.1 The Exponential Function

The exponential or e-function $f(x) = e^x = \exp(x)$ is one of the most frequently occurring and important function in the everyday life of a natural scientist. The number denoted by "e" is an irrational number called *Euler's number.* Its first digits are $e \approx 2.7183$. You should remember the approximate value of its inverse $\frac{1}{e} \approx 0.37$ because it is used to define time constants of neural signal transduction.

The exponential function appears in many probability distribution in statistics, in solution of differential equations and will appear in many mathematical model of neural processes. The exponential function has some very nice properties. One of them is that it is its own derivative $f'(x) = (e^x)' = e^x$.

The following examples show three cases, in which the exponential function naturally occurs.

Examples

1. One central probability density in statistics is the *Normal* or *Gaussian Distribution* $\mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. In many experiments that involve

noisy measurements, the noise is assumed to be Gaussian with mean $\mu = 0$. One possible justification for this assumption is that sums of random quantities with other probability distributions tend to be Gaussian if the total number of this quantities increases. Since we think of the noise in an experiment as the superposition of many other processes that we are not interested in, the assumption that there are a large number of them which are linearly superimposed motivates the Gaussian noise assumption.
Apart from that, the Gaussian distribution is frequently used because it has a lot of properties that make it possible to calculate analytical solutions of the respective statistical problems.

2. The potential change $\Delta V_m(t)$ over time at a passive neuron membrane after applying a rectangular current pulse can be described by the following equation

$$\Delta V_m(t) \quad = \quad I_m R \left(1 - e^{-\frac{t}{\tau}}\right).$$

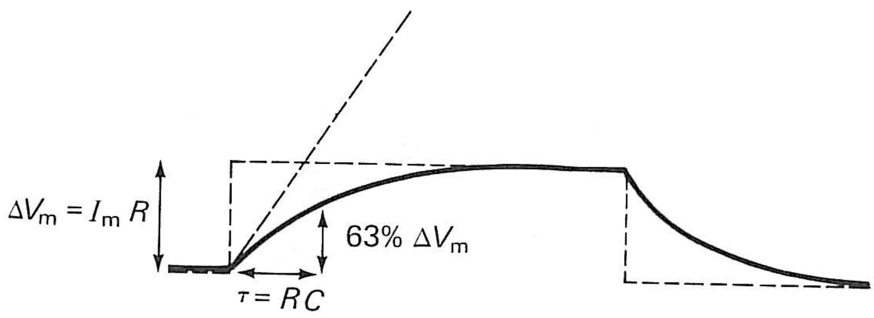


Figure 1.4: Potential change $\Delta V_m(t)$ (solid line) over time at a passive neuron membrane after applying a rectangular current pulse (dashed line) [Figure from [1]].

Figure 1.4 shows the time course of the potential. Here I_m is the current, that has been injected, R is the membrane resistance and τ is the *time constant* of the membrane. It tells us how fast the membrane potential follows the rectangular pulse. The greater τ is, the longer it takes until $\Delta V_m(t) = I_m R$, where $I_m R$ is the potential change induced by injecting the current I_m. In some sense τ defines a time scale for that membrane. τ is the time needed until the potential change reaches $0.63 \cdot I_m R = (1 - 0.37) \cdot I_m R = (1 - \frac{1}{e}) I_m R$, i.e. 63% of the potential change induced by the rectangular pulse. One can show that $\tau = RC$, where C is the membrane capacitance, i.e. its ability to buffer charge.

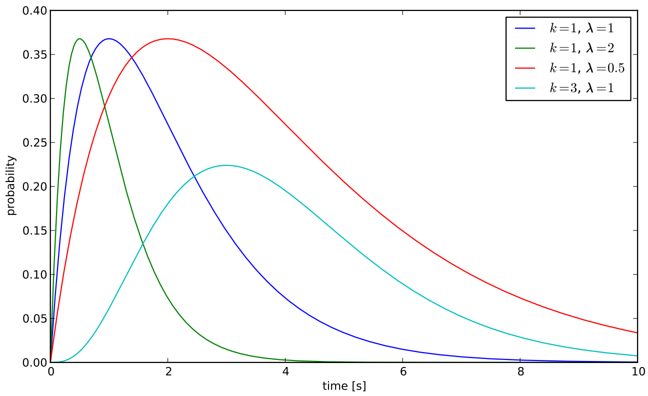


Figure 1.5: Poisson Process with 4 different sets of parameters.

3. A simple stochastical model for spike generation by a neuron is the *Poisson Process*. In this model, time is divided into a large number of bins. In each bin a spike occurs with probability p independent of whether a spike occurred in the bin before or not. If p is sufficiently small and the average spiking rate is λ, the probability of observing exactly k spikes in a time window of length Δt is given by the distribution of the Poisson Process with rate λ:

$$P(k, \Delta t) = \frac{e^{-\lambda \Delta t}(\lambda \Delta t)^k}{k!}.$$

Figure 1.5 gives an example how a Poisson Process looks like for different sets of parameters. The symbol expression "$k!$" denotes the *factorial function* $k! = k \cdot (k-1) \cdot ... \cdot 2 \cdot 1$. We can use a notation similar to the Σ for sums to denote a product of several components: The uppercase Greek letter Π (for product) denotes a product of all elements following it. The indices are written in the same way as for sums (see also Appendix 2.1). Therefore, we can write the factorial function as $k! = \prod_{n=1}^{k} n$.

4. If a spike train is generated by a *Poisson Process*, then the distribution of the *inter-spike-intervals (ISIs)* is an *exponential distribution*. That is, if we observe a spike at time 0, then the probability (density) of observing the next spike at time s is given by $p(s) = \mu e^{-\mu s}$, for $\mu = 1/\lambda$ in the example above.

$\triangleleft$

1.1.4.2 Logarithms

Logarithms and their derivatives often occur in statistics. Estimating the parameters of a statistical model is often done via maximum likelihood estimation.

This involves taking the derivative of the log of the likelihood function.

Here we introduce logarithms. More advanced examples will be discussed in later chapters. For now, we start with a small example:

Example In this example, we look again at the time course of the membrane potential after the application of a rectangular current pulse $\Delta V_m(t) = I_m R\left(1 - e^{-\frac{t}{\tau}}\right)$. Assume that we want to measure the time constant of a certain membrane. For this purpose we excited the membrane with a rectangular current pulse and measured $\Delta V_m(t)$ at several points t_k, $k = 1, ..., n$ in time. Now we want to solve $\Delta V_m(t) = I_m R\left(1 - e^{-\frac{t}{\tau}}\right)$ for τ at each time step t_k and get n values τ_k that we average to get your final estimation $\hat{\tau} = \frac{1}{n}\sum_{k=1}^{n}\tau_k$ of the membrane's time constant. How do we solve for τ_k? The first step is easy: You rearrange the the terms to get

$$\begin{aligned} \Delta V_m(t) = I_m R\left(1 - e^{-\frac{t}{\tau_k}}\right) \quad &\Leftrightarrow \quad I_m R - \Delta V_m(t_k) == I_m R \cdot e^{-\frac{t_k}{\tau_k}} \\ &\Leftrightarrow \quad \frac{I_m R - \Delta V_m(t_k)}{I_m R} = e^{-\frac{t_k}{\tau_k}}. \end{aligned}$$

Now we somehow have to extract $-\frac{t_k}{\tau_k}$ from the exponent. This means that you are searching for a number a, such that $e^a = \frac{I_m R - \Delta V_m(t_k)}{I_m R}$. This is exactly the definition of the *natural logarithm* $\ln(x)$: It is the number that you have to put in the exponent of e in order to obtain x. Now you can solve for τ:

$$\begin{aligned} \frac{I_m R - \Delta V_m(t_k)}{I_m R} = e^{-\frac{t_k}{\tau_k}} \quad &\Leftrightarrow \quad \ln\left(\frac{I_m R - \Delta V_m(t_k)}{I_m R}\right) = -\frac{t_k}{\tau_k} \\ &\Leftrightarrow \quad -\frac{\ln\left(\frac{I_m R - \Delta V_m(t_k)}{I_m R}\right)}{t_k} = \frac{1}{\tau_k} \\ &\Leftrightarrow \quad -\frac{t_k}{\ln\left(\frac{I_m R - \Delta V_m(t_k)}{I_m R}\right)} = \tau_k \end{aligned}$$

$\lhd$

We just saw that the natural logarithm is the function that cancels the exponential function, i.e. $\ln(e^x) = e^{\ln(x)} = x$. In general, a function g that cancels another function f is called *inverse function* of f and is denoted by $g = f^{-1}$. Not all functions have inverses. Some of them only have an inverse on a restricted range. We will discuss inverse function in more detail in the chapter about analysis.

So far we have seen how to solve e^x for x by using the natural logarithm. What if we want to solve an equation like 2^x for x? Here we cannot use the natural logarithm since $\ln(x)$ is only the inverse function of e, not 2. Here we must use a logarithm that fits to 2. This logarithm is denoted by $\log(x)$ and has the property that $2^{\log(x)} = \log(2^x) = x$. In general there is a logarithm for any number b that is the inverse of the function $f(x) = b^x$. The number b is called *base* of the logarithm. If the base is not $b = 2$ or $b = e$, we indicate the base in the subscript of $\log_b(x)$. However, the only frequently used logarithms are $\log(x) = \log_2(x)$ and $\ln(x) = \log_e(x)$.

Remark It happens quite often that the base of of the logarithm is not specified, either because it does not matter or it is clear from the context. In this case, the notation $\log(x)$ is usually used. Usually this means that the base is e (e.g. Physics), sometimes it means that the base is 2 (e.g. in Information Theory) and sometimes it is used for base 10 (e.g. Economics). In the remaining part of the script we simply use $\log(x)$ to denote any logarithm. The base should always be clear from the context or it does not matter. If we want to emphasize a certain base we will write it in the subscript or use the explicit notation $\ln(x) = \log_e(x)$ or $\lg(x) = \log_{10}(x)$. $\boxed{!}$

$\triangleleft$

There is a neat trick how to calculate logarithms of arbitrary bases by using logarithms of another base:

$$\log_b(x) = \frac{\ln(x)}{\ln(b)} = \frac{\log(x)}{\log(b)}.$$

Until now we skipped an important detail of logarithms. When only dealing with real numbers, the logarithm is only defined on the strictly positive part of $\mathbb{R}$. We denote this set by $\mathbb{R}^+$. The reason for this restriction is easy to see. If we remember that $\log_b x$ is the number that has to be put in the exponent of b in order to obtain x. If x is negative, there can generally be no such real number since b^x is positive. (The concept of logarithm can be extended to negative and imaginary numbers, but does lead to some complications, so we will not cover it here.)

We conclude this section with a few calculation rules. Most of them follow directly from the calculation rules of powers or the definition of the logarithm.

Calculation Rules for Logarithms
The following rules apply to any logarithm:

- $\log_b(b^x) = x$
- $\log_b(x \cdot y) = \log_b(x) + \log_b(y)$ for $x, y \in \mathbb{R}^+$
- $\log_b\left(\frac{x}{y}\right) = \log_b(x) - \log_b(y)$ for $x, y \in \mathbb{R}^+$
- $\log_b(x) = \frac{\ln(x)}{\ln(b)} = \frac{\log(x)}{\log(b)}$ for $x, b \in \mathbb{R}^+$
- $\log_b(x^y) = y \cdot \log_b(x)$ for $x \in \mathbb{R}^+$, $y \in \mathbb{R}$

1.1.5 Lines (Affine Functions)

Most people think about lines when they think about linear functions. However, lines are not generally linear functions. Only the lines that include the origin are strictly speaking linear functions. Lines versions of linear functions that are shifted along the y-axis. The general equation for a line is

$$f(x) = mx + t,$$

where m is called the *slope* of f. It is the first derivative or, equivalently, the amount about f changes if we increase x by one, i.e. $m = f(x+1) - f(x)$. The value of t determines the y coordinate of the point where f cuts through the y-axis. This can easily been seen by evaluating f at $x = 0$. Obviously, the function value $f(0) = t$ must be the location on the y-axis where f hits it.

From the general form of lines we can also see why they are not strictly linear if $t \neq 0$. If they were linear, $f(x + y) = f(x) + f(y)$ would have to hold for all $x, y \in \mathbb{R}$. However, it is easy to check that this is not the case:

$$\begin{aligned} f(x) + f(y) &= mx + t + my + t \\ &= m(x + y) + 2t \\ &\neq m(x + y) + t \\ &= f(x + y). \end{aligned}$$

Therefore, lines with $t \neq 0$ are not linear. But we can always make them linear by subtracting t. This yields a line with the same slope m, which is shifted along the y-axis such that it cuts through $(0, 0)$, which make it a truly linear function.

Functions of the form $f(x) = mx + t$ are also called *affine functions*.

1.1.6 Piecewise Defined Functions

Sometimes, it is convenient to define a function by using two or more other functions. This is useful if we want to change the behaviour of the function on certain parts of the domain. Achieving a certain behaviour with a single expression might be difficult. It is then usually easier to use an several expressions, one for each part of the domain. These functions are called *piecewise defined functions*. We just mention a few important examples here.

Examples

1. The Heaviside function is defined as:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}.$$

2. The absolute value is given by

$$f(x) = \begin{cases} -x & x \leq 0 \\ x & x > 0 \end{cases}$$

3. The maximum-function is defined as

$$f(x,y) = \begin{cases} y & x \leq y \\ x & x > y \end{cases}$$

$\triangleleft$

Piecewise defined functions can for example be used to define a more realistic model of neurons.

Example: Neurons that are linear over some range.

Earlier, we saw that neurons are linear for some stimuli x, but clearly not for all: Firstly, the firing rate of a neuron can not be negative, and secondly, there is a maximal firing rate that can not be exceeded. Let us suppose that a neuron responds to a one-dimensional stimulus x with firing rate $f(x)$.

$$f(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq f_{max} \\ f_{max} & x > f_{max} \end{cases}$$

$\triangleleft$

Piecewise functions are just like any other function. Within each interval, the functions can be differentiated and integrate like other functions. Nevertheless, there is some care needed at the points at which the intervals meet. In particular, it is often (but not always) the case that piecewise functions are not continuous of differentiable at these points.

1.1.7 Sketching Functions

Nowadays, computers are around almost everywhere allowing us to plot functions whenever needed. Nevertheless, being able to imagine how functions look like and sketch them is a useful ability because it gives us a better intuition for what those functions do. There are some simple tricks for imagining and drawing functions which we briefly present in this section. They can basically be classified into two categories. The first is for functions that are transformations of certain basic functions which one usually knows by, like the exponential function, sine and cosine function or easy polynomials like the parabola. The second is for functions that are compositions of known basis functions.

1.1.7.1 Adapting Functions

Everyone knows how to sketch the parabola $f(x) = x^2$: It is opened upwards, symmetric, equals one for $x = \pm 1$ and diverges to infinity for $x \to \pm\infty$. But what about the function $f(x) = -\frac{1}{2}(x-2)^2 + 5$? We will see that it is easy to adapt $f(x) = x^2$ to make it look like $f(x) = -\frac{1}{2}(x-2)^2 + 5$. The first question, we have to answer, is in which order we want to introduce the changes to x^2 to

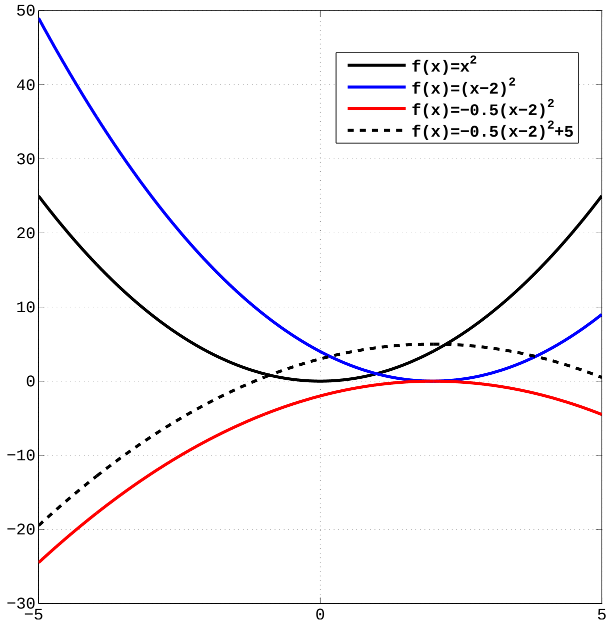


Figure 1.6: Different steps of transforming $f(x) = x^2$ into $f(x) = -\frac{1}{2}(x-2)^2+5$.

transform it into $-\frac{1}{2}(x-2)^2+5$. The answer is: We do that in exactly that order in which we would compute the result of $-\frac{1}{2}(x-2)^2+5$. This means, we first subtract 2, then square the result, then premultiply $-\frac{1}{2}$ and finally add 5. If we would not do that we would end up with a different function since we would violate calculation rules at some point in the process.

So let us start to transform x^2. You can follow the different steps graphically in Figure 1.6.As we just mentioned, the first step is to transform x^2 into $(x-2)^2$. Here, we get our first rule: The graph of $f(x-a)$ is simply the graph of $f(x)$ shifted by a to the right. Of course, if a is negative, shifting by a becomes a shift to the left. Why is that the case? If you think about it, $x-a$ can also be seen as shifting the whole x-axis by $-a$, i.e. a to the left. This, however, is equivalent to shifting f to the right by a. Applied to your example, this means that we have to shift x^2 to the right by 2 in order to obtain the graph of $(x-2)^2$.

The next step is to include the factor $-\frac{1}{2}$. We already know the graph of $(x-2)^2$, how does the graph of $-\frac{1}{2}(x-2)^2$ look like? Well, premultiplying -1 surely reflects the graph along the x-axis. The factor $\frac{1}{2}$ squeezes the result. For example, y-values that used to be -1 are now $-\frac{1}{2}$, y-values that used to be -2 are now -1, and so on.

Now, we are almost there. The last step is to include the additive constant $+5$. This is easy: It simply shifts the graph of $-\frac{1}{2}(x-2)^2$ upwards by 5. This is it! We arrived at the graph of $-\frac{1}{2}(x-2)^2+5$ by a few simple adaptation rules for the graph of x^2. With this few simple rules, you can already sketch a decent amount of functions.

Rules for adapting functions

- The graph of $f(x-a)$ is simply the graph of $f(x)$ shifted by a.
- The graph of $-f(x)$ is the graph of $f(x)$ flipped along the x-axis.
- The graph of $a \cdot f(x)$ is the graph of $f(x)$ stretched ($a > 1$) or squeezed ($0 \leq a < 1$) by a.
- The graph of $f(x) + b$ is the graph of $f(x)$ shifted by b along the y-axis.

1.1.7.2 Compositions of Functions

Sketching compositions of functions is a little bit more art, but for a lot of examples it is not so difficult. As an example, we use the function $f(x) = \exp\left(-(x-2)^2\right)$ which is just a nicer way of writing $f(x) = e^{-(x-2)^2}$. The rules from above are not sufficient to sketch this function. There is, however, one rule that we can use: If we know the graph of $f(x) = \exp\left(-x^2\right)$, we know that we arrive at $f(x) = \exp\left(-(x-2)^2\right)$ by shifting the graph to the right by 2. Therefore, we look at how to sketch the graph of $f(x) = \exp\left(-x^2\right)$ in the following.

The first step, you can always to is to check whether there are function values which are easy to compute and help drawing the graph. Usually, it is a good idea to look at the behavior of $f(x)$ at $x = 0$, $x = \pm 1$ and what $f(x)$ does if x goes to $\pm\infty$. In our case, the interesting cases are $x = 0$ and $x \to \pm\infty$. The position $x = 0$ is interesting since anything to the power of 0 equals 1. Therefore $f(0) = 1$. When we ask how $f(x)$ behaves for $x \to \pm\infty$ we can first observe that the behavior will be the same at both sides since the squaring operation cancels out the sign. So let us look at what happens when $x \to \infty$. Let us advance step by step, just as before. First, when $x \to \infty$, surely we will get $x^2 \to \infty$ as well. By that, we can immediately see that $-x^2 \to -\infty$. Therefore, we only need to know what happens to $\exp(z)$ if its argument z assumes a very large negative value. We can rewrite the problem a bit by using one of the calculation rules for exponentials: $e^{-x} = \frac{1}{e^x}$. Now, the answer should be easy. If $-x^2 \to -\infty$, there will be a large value in the denominator and, therefore, $f(x) = \exp\left(-x^2\right)$ will approach zero. We can even say a bit more, namely that it will approach zero from above since $\exp(z)$ can never become negative if $z \in \mathbb{R}$.

In a similar manner, you can sketch functions of the form $f(x) = g(x) + h(x)$ or $f(x) = \frac{g(x)}{h(x)}$. First, find a few points where the function value is easy to compute. Then check what happens if x approaches points, where one of the functions goes to zero or infinity. Then the question is usually, which of the functions "wins", i.e. which approaches zero or infinity faster. For example, $f(x) = x^2 - x$ will definitely diverge to infinity for $x \to \infty$ since x^2 grows much faster than $-x$ is able to drag it into the negative side. Similarly $f(x) = \frac{x}{\exp(x)}$ will approach zero for $x \to \infty$ since $\exp(x)$ grows faster than x does.

Drawing compositions of functions takes a bit of practice, but is a useful tool for understanding how functions look like and what they do.

Chapter 2

Appendix

2.1 Notation and Symbols

- The capital Greek letter sigma "$\sum$" (sigma like sum) denotes a sum over several elements. Usually the components of the sum are indexed with lower case roman letters starting from "i". The starting index is indicated below the "$\sum$" and final index is indicated on top. For example, the sum over n real numbers $x_1, ..., x_n \in \mathbb{R}$ is denoted by

$$x_1 + ... + x_n = \sum_{i=1}^{n} x_i.$$

  Sometimes, when summing over all elements of a set, the set is indicated below the sigma. For example, summing all elements of the set $A = \{1, 2, 3, 4, ..., 15\}$, could be written as $\sum_{x \in A} x$ as well as $\sum_{n=1}^{15} n$.

- The capital Greek letter pi "Π" is used in an analogous manner for products, i.e.

$$x_1 \cdot ... \cdot x_n = \prod_{k=1}^{n} x_k.$$

Bibliography

[1] Eric C. Kandel and James H. Schwartz. *Principles of Neural Science*. Elsevier Science Publishing Co., Inc., 2 edition, 1985.