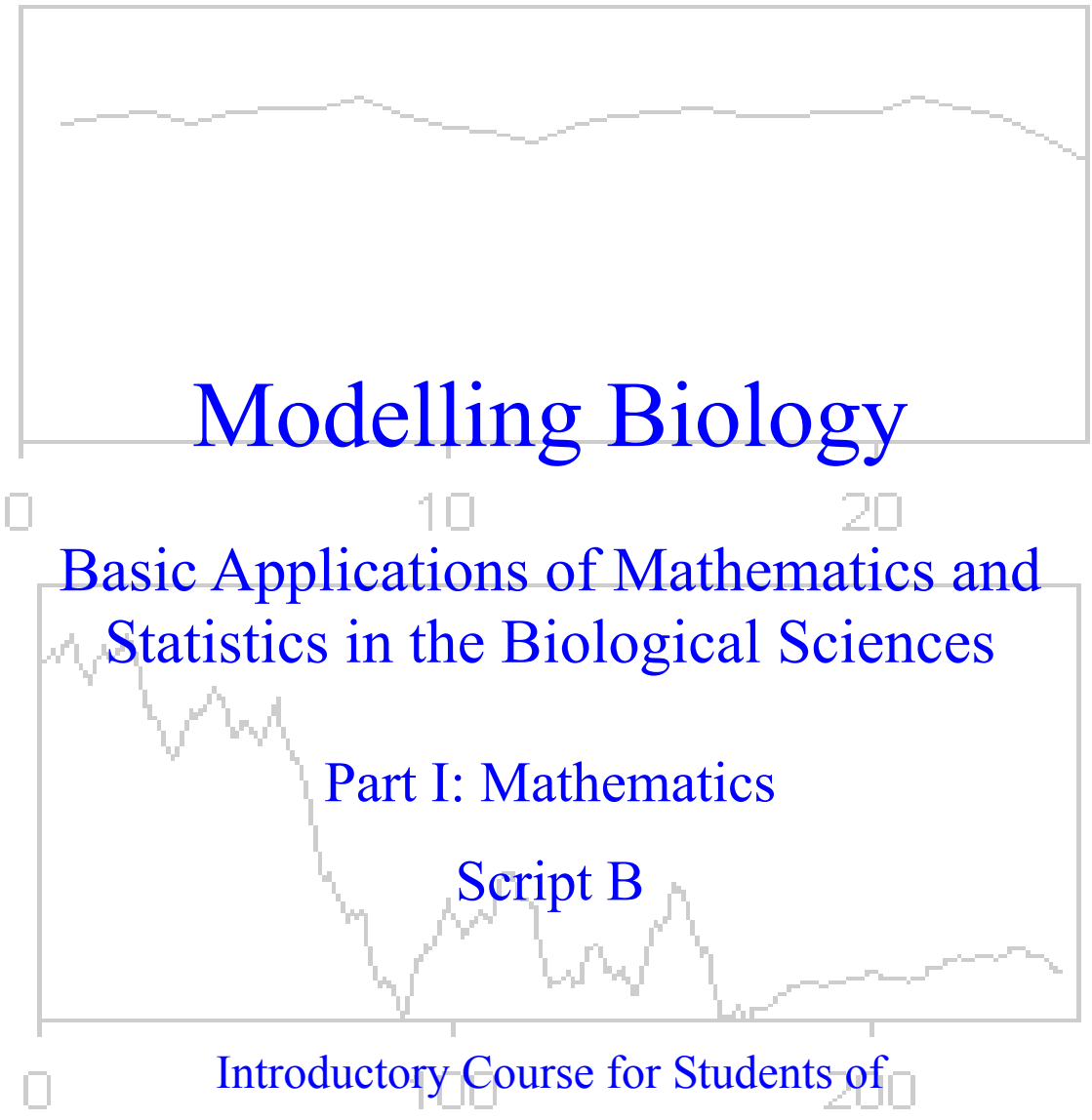


Modelling Biology

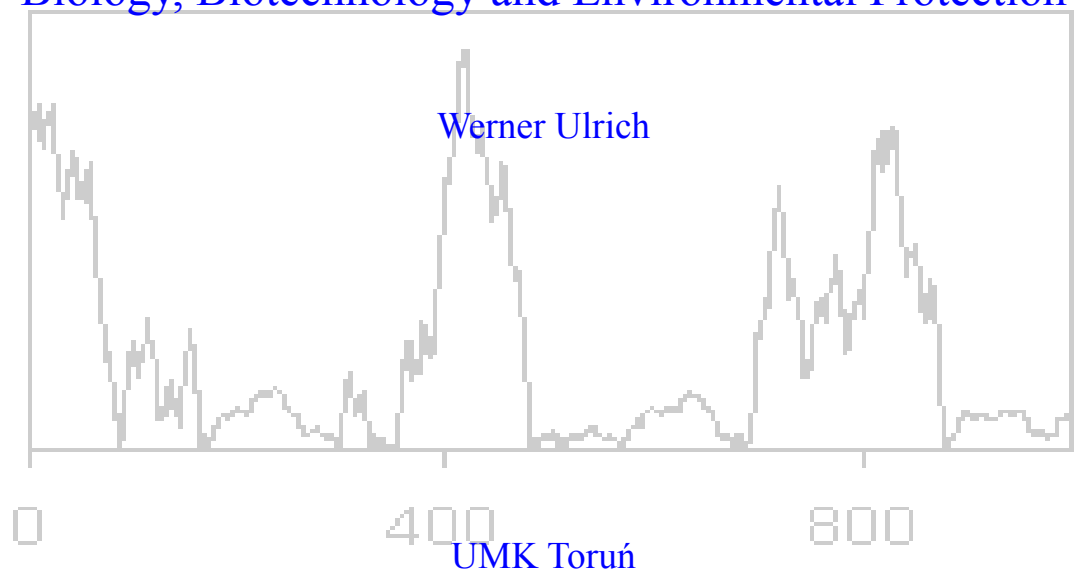
Basic Applications of Mathematics and Statistics in the Biological Sciences

Part I: Mathematics

Script B

Introductory Course for Students of

Biology, Biotechnology and Environmental Protection



Werner Ulrich

UMK Toruń
2007

Contents

Introduction 3
1. Combinations, variations and probabilities 4
2. Periodic processes 14
3. Vectors 23
4. Matrix algebra 29
5. Panta rhei 42
6.1. Partial derivatives and stationary points 52
6.2. Lagrange multipliers 57
Online archives and textbooks 59
Literature 60
Mathematical software 61

Introduction

The following text is the second part of a lecture in basic mathematics and statistics for biologists. It contains what might be considered an international standard of basic knowledge although many readers will surely miss important branches. It deals with basic probability theory, trigonometry, matrix algebra, and differential equations. However, a one year course that has to deal with mathematics modelling must be to a certain extent eclectic. Emphasis was especially paid to basic mathematical techniques and principles of biological modelling. Again, many examples are included that show how to program simple tasks with a spreadsheet program and how to use advanced mathematics software. The text does not repeat school mathematics.

The following text in not a textbook. It is intended as a script to present the contents of the lecture in a condensed form. There is no need to write a textbook again. Today, the internet took over many former tasks textbooks had. The end of this text contains therefore a small overview over important internet pages where students can find mathematics glossaries, textbooks, and program collections.

1. Combinations, variations and probability

Combinatory

With this lecture we start to deal with probability theory. Both fields are of major importance for biologists. At the beginning we will learn about the mathematics of counting, about variations, combinations and permutations. The computation of possible states of a biological systems and the comparison of this number with theoretical expectations becomes more and more a basis of many methods of hypothesis testing and modelling.

Let's start with a simple example. How many numbers with exactly 4 digits exist? These are of course

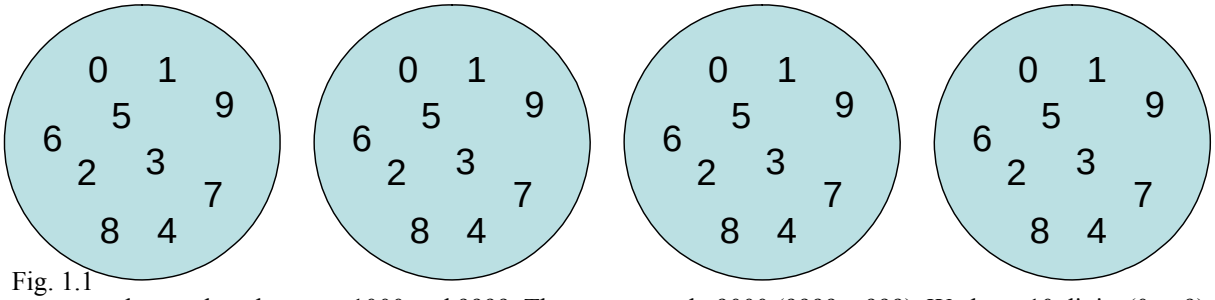


Fig. 1.1

the numbers between 1000 and 9999. There are exactly 9000 (9999 – 999). We have 10 digits (0 to 9) to combine. There are 10 one digit numbers, $10*10 = 10^2$ two digit numbers and of course $10*10*10*10 = 10^4$ four digit numbers. But we have to exclude all numbers that start with a zero. There are $10*10*10 = 10^3$ such numbers. Therefore, the total number of variations of four digit numbers is $10^4 – 10^3 = 9000$.

A convenient way to visualize the principle of counting is a so-called **polling box model** (Fig. 1.1). In our case we have 4 such boxes, each containing the digits 0 to 9. To make four digit numbers we take one digit from each box. In the first box we have 9 possibilities (1 to 9 because a zero as a starting point is not allowed). In the next box we have 10 possibilities. The total number of variations is therefore 9*10. The total number of variations from all four boxes is then 9*10*10*10 = 9000 variations. We obtain the same result if we consider only 1 box and take 4 times a digit while all digits remain in the box.

In general we can write that the number of **variations** of n elements into sets of k elements is

$$V_k^n = n^k \tag{1.1}$$

if repetition is allowed. In our case, it is allowed to form numbers like 1111, 2021, 3443 and so on.

What is if repetition is not allowed? In how many ways can we take numbers from the first box? There are 10 possibilities for the first digit. If we take one, 9 numbers remain and for the second digit these 9 possibilities remain. In total, we have 10*9 so-called permutations for the first two digits. For the third digit 8 numbers remain and we have of course 10*9*8 permutations for the first three digits and so on. We see that the total number of permutations of 10 digits is

$$P = 10*9*8*7*6*5*4*3*2*1 = 10! \tag{1.2}$$

In general, the number of **permutations** of n elements is

$$P^n = n! \tag{1.3}$$

What about our four digit numbers? How many permutations of four digit numbers without repetition are possible? If we take a number out of the first box (10 possibilities) in the second box 9 possibilities remain. Because repeating is not allowed we indeed took the first number out of all four boxes. Now we take a number from the second box. Again we have to take this number also out of the remaining two boxes. Therefore, the third box contains only 8 numbers. After taking one of them, the fourth box contains only seven numbers. In total, the number of four digit numbers without repetition is therefore P = 10*9*8*7 (if a zero as first number is allowed). We can immediately generalize this result and notify that the number of **permutations** of n elements into sets of r elements is exactly

$$P_r^n = n*(n-1)*(n-2)*...(n-r+1) = \prod_{i=n-r+1}^{n} i$$

We can write this result in a slightly different but more convenient form

$$P_r^n = n(n-1)(n-2)...(n-r+1)\frac{(n-r)(n-r-1)...1}{(n-r)(n-r-1)...1} = \frac{n!}{(n-r)!} \qquad (1.4)$$

This is the form permutations are most often given. Note that the last equation is a generalization of the above equation for simple permutations. If we consider only a set of n elements we have n! / (n-n)! = n!. By definition 0! = 1.

A next example. How many four letter words can be formed from the word permutation? If these words can contain the same letter two or several times we can consider a letter from the box but leave it. Then, we have of course 10^4 possibilities (permutation contains 10 letters, the t occurs twice). If all the letters of the four letter words have to be different there are 10*9*8*7 permutations or

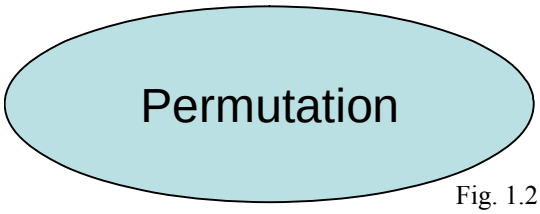


Fig. 1.2

$$P_4^{11} = \frac{10!}{(10-4)!} = 5040$$

In our above example the ordering of elements was decisive. The four letter word *perm* is of course different from the (non-)word *prem.*

What is if the ordering of elements is not decisive? In Duży Lotek every week 6 numbers are taken from a total of 49 numbers. How many six number combinations exist?

Our permutation equation predicts 49! / (49-6)! However, in this case the ordering of numbers is irrelevant. It does not matter at which step a number falls. Our permutation equation gives therefore a prediction that is too high. To obtain the correct result we have to consider the number of ways 6 numbers can be varied. This number is of course 6!. So, the correct answer of our Duży Lotek example is

$$C = \frac{\frac{49!}{(49-6)!}}{6!} = \frac{49!}{6!(49-6)!} = 13983816$$

Again, we can generalize this result. The number of **combinations** of k objects out of a total of n elements, when the ordering of objects is irrelevant, is

$$C_k^n = \frac{n!}{k!(n-k)!} = \binom{n}{k} \tag{1.5}$$

Again, we deal with so-called **binomial coefficients**. Binomial coefficients have some remarkable features. We know already about the Newtonian series

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}$$

The binomial coefficients give the values of each row of the **Pascal triangle**. We also get immediately some important relations

$$\begin{aligned}
&\binom{n}{k} = \binom{n}{n-k} \\
&\binom{n}{0} = 1 \\
&\binom{0}{0} = 1 \\
&\binom{k}{n} = 0;\ (k < n) \\
&\sum_{i=0}^{n} \binom{n}{i}^2 = \binom{2n}{n} \\
&\sum_{i=0}^{n-1} \binom{k+i}{k} = \binom{n+k}{k+1};\ (k \geq 0, n \geq 1)
\end{aligned} \tag{1.6}$$

Combinations are important in the planning of biological experiments. For instance, we have a strain of a bacterium that is able to digest a certain substrate. We know that 5 genes or combinations of these genes are responsible for this ability. By controlled inhibition of gene expression we have to find out the exact gene combination that is responsible for digestion. How many experimental treatments C do we need? We have a total of 5 elements. We have to consider them, all combinations of two genes, of three genes of four, and of five. Therefore

$$C = \binom{5}{1} + \binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5} = \sum_{i=1}^{d} \binom{5}{i}$$

For a = b = 1 we get

$$(a+b)^5 = \sum_{i=0}^{5} \binom{5}{i} a^i b^{5-1} = \sum_{i=0}^{5} \binom{5}{i} = \sum_{i=1}^{5} \binom{5}{i} + 1$$

Therefore

$C = 2^5 - 1 = 31$.

We need 31 different experimental treatments.

A last example. We consider a plant species in which blue, white and yellow flowers occur. We sample 12 plants, three with white, four with yellow and five with blue flowers. How many permutations of these 12

plants are possible? Of course, the number of permutations of 12 elements is 12!. But 3! times white will repeat at the same place, 4! times yellow, and 5! times blue. Therefore, the total number of permutations is

$$P\begin{pmatrix} 12 \\ 3,4,5 \end{pmatrix} = \frac{12!}{3!4!5!} = 27720$$

Again, we can generalize the result. The number of permutations of a set of n elements that is made up of k classes of n_i elements each ($n_{1+n2+\ldots+nk} = n$) is

$$P\begin{pmatrix} n \\ n_1, n_2, \ldots, n_k \end{pmatrix} = \frac{n!}{n_1! n_2! \ldots n_k!} \tag{1.7}$$

Probabilities

Everybody knows intuitively what probability is. Probability is when always the others have luck!

But you also know a more stringent every day's definition of probability. If you have a dice, the probability to get the event six is 1 out of six possibilities. This definition contains two expressions: **possibilities** and **events**. The event while throwing a dice is to have a six. For this outcome you have 6 possibilities.

From this we define the probability p(k) as the quotient of the realized or possible number of events we consider k and the total number of events or possibilities n

$$p(k) = \frac{k}{n} \tag{1.8}$$

The so defined probability has the same value as the **relative number** or the **frequency** of an event k.

Additionally, we define a complementary event ¬k (read not k) as

$$p(\neg k) = \frac{n-k}{n} = 1 - \frac{k}{n} = 1 - p(k) \tag{1.9}$$

Intuitively, it is also clear that

$$p(k) + p(\neg k) = 1 \tag{1.10}$$

and

$$0 \le p \le 1 \tag{1.11}$$

An impossible event has the probability p = 0, a sure event the probability p = 1. The same holds of course for the relative frequencies of two complementary events.

Hence, what is the probability to win in Duży Lotek? If you consider the event that you have all 6 numbers correctly this probability is

$$p = \frac{1}{\frac{49!}{6!(49-6)!}} = \frac{1}{13983816}$$

However if you consider any winning it is the sum of the probabilities to have 3, 4, 5, or 6 six numbers correctly. Intuitively, it seems that we have to look how many combinations of three, four, five, and six objects

out of 49 elements exist. This number is $\binom{n}{k}$. Therefore

$$p = \frac{1}{\binom{49}{6}} + \frac{1}{\binom{49}{5}} + \frac{1}{\binom{49}{4}} + \frac{1}{\binom{49}{3}} = 0.00006$$

But our result is wrong although it looks quite plausible. Why? Look at the event to have 3 numbers correctly. The total number of combinations is n over k. However, there are not only 3 numbers taken, there are always six numbers taken. Thus, you have 49 over 6 combinations and from 6 numbers taken you should have 3 correct. And of the remaining 43 numbers the other three have to be incorrect. For both cases you have 6 over 3 and 43 over 3 possibilities. The total number of possibilities for 3 numbers to be correct is therefore

$$C_{n,k}^{n,K} = \frac{\binom{N}{n}}{\binom{K}{k}\binom{N-K}{n-k}} \tag{1.12}$$

where N is the total number of entities or objects, n the sample size, K the number of objects having the desired property A, and k the number of elements with property A in the sample. This is the co-called **hypergeometric distribution** of which the equation for combinations is only a special case for K = k = n.

Now we can compute the probability to win in Duży Lotek

$$C_{n,k}^{n,K} = \frac{\binom{49}{6}}{\binom{6}{3}\binom{49-6}{6-3}} + \frac{\binom{49}{6}}{\binom{6}{4}\binom{49-6}{6-4}} + \frac{\binom{49}{6}}{\binom{6}{5}\binom{49-6}{6-5}} + \frac{\binom{49}{6}}{\binom{6}{6}\binom{49-6}{6-6}} \approx 54$$

Our probability is 1/54 = 0.0186.

Computing this probability by hand is time consuming, but with a spreadsheet program it is done very quickly. The Table below shows us an Excel solution.

	A	B	C	D	E	F	G	H	I
1	N	49		49		49		49	
2	K	6	=+KOMBINACJE(B1;B2)	6	=+KOMBINACJE(D1;D2)	6	=+KOMBINACJE(F1;F2)	6	=+KOMBINACJE(H1;H2)
3	n	6	=+KOMBINACJE(B2;B4)	6	=+KOMBINACJE(D2;D4)	6	=+KOMBINACJE(F2;F4)	6	=+KOMBINACJE(H2;H4)
4	k	3	=+KOMBINACJE(B1-B2;B3-B4)	4	=+KOMBINACJE(D1-D2;D3-D4)	5	=+KOMBINACJE(F1-F2;F3-F4)	6	=+KOMBINACJE(H1-H2;H3-H4)
5	Combinations		=+C2/(C3*C4)		=+E2/(E3*E4)		=+G2/(G3*G4)		=+I2/(I3*I4)
6	Probability		=1/C5		=1/E5		=1/G5		=1/I5
7	Sum		=+SUMA(C6:I6)						

This example leads us to two important relations between probabilities. First of all, we added probabilities. Our events were independent. If events are independent the combined probabilities are the sum of the single probabilities of each of the event. This is best shown by a simple Figure from set theory (Fig. 1.3).

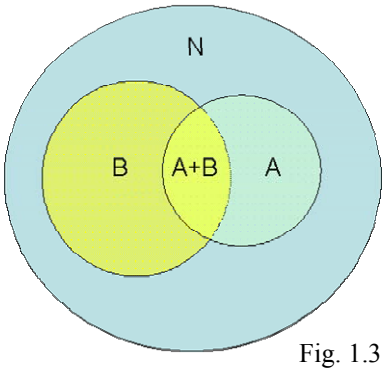


Fig. 1.3

We have a set of all possibilities N. From this we consider two subsets of events A and B. The combined probability of A and B is now the set

p(A + B) and we find

$$p\ (A \vee B)\ =\ p(A)\ +\ p(B)\ -\ p(AB) \tag{1.13}$$

We have to subtract p(AB) because otherwise we would count this area two times, one time in A and one time in B. If the subsets A and B are independent the set p(AB) is empty and we have

$$p\ (A \vee B)\ =\ p(A)\ +\ p(B) \tag{1.14}$$

This is the **summation law of probabilities**. It tells that the combined probability of two independent events is the sum of both events.

Therefore the probability to get either a six or a two throwing a dice is 1 / 6 + 1 / 6 = 1 / 3. But, what is the probability to throw with two dices exactly one time a six? Look at Fig. 1.3. In the set theoretical model the probability is represented through the given subsets. In our case it is the combined subset A and B. This subset is

$$p(A \vee B) = p(A) + p(B) - p(AB)$$

Our probability is therefore

P(A or B) = 1/6 + 1/6 – 1/36 = 0.31.

Why 1/36? This is the probability to throw two times a six with two dices. Intuitively we feel it should be 1 / 36 = (1/6) * (1/6). Indeed, this time we deal with **dependent events**. We ask what is the probability to get a second time a six if we have already one six. For such cases we need the law for dependent events. It is

$$p(B/A) = \frac{p(A \wedge B)}{p(A)} \tag{1.15}$$

Read: The probability of B under the condition A is the quotient of the probability of A and B and the probability of A.

Of course, it is also

$$p(A/B) = \frac{p(A \wedge B)}{p(B)}$$

What does this mean? P(A) is in our case 1/6, the probability to have a six. p(B/A) is the probability to have a six if already one six fall. This is of course again 1/6. The probability p(AB) we look for is therefore (1/6) * (1/6). This is the **multiplication law for probabilities**.

$$p(A \wedge B) = p(A)\,p(B/A) = p(B)\,p(A/B) \tag{1.16}$$

It tells that the combined probability of two independent events is the product of the probabilities of both events. If A and B are independent eq. 1.16 simplifies to

$$p(A \wedge B) = p(A)\,p(B) \tag{1.17}$$

Our above definition has certain but important limitations. It is confined to situations where we are able (theoretically or by experiments) to define and compute event numbers and possibilities. However, in most cases, for instance if we deal with complicated experiments, such values can’t be computed. For these cases we need a more stringent and general definition.

Look at the next Figure 1.4. It shows the relative abundance (the frequency) of newborn male infants during 30 years of study. What is the probability of a male to be born? Intuitively, we think this probability should be the number of newborn males in relation to all infants born, therefore the frequency of males. This frequency can be computed from the mean frequencies of all newborn infants. This mean is shown by the blue line in the Figure. It is about 0.486. This is the **statistical definition of probability**, it is **the frequency of a certain event**.

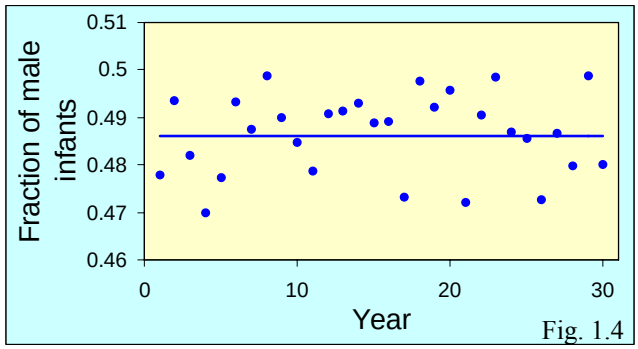


Fig. 1.4

Another way to derive probabilities is from logic (Fig. 1.5). We can now apply the summation and the multiplication law to reach in other laws about probabilities.

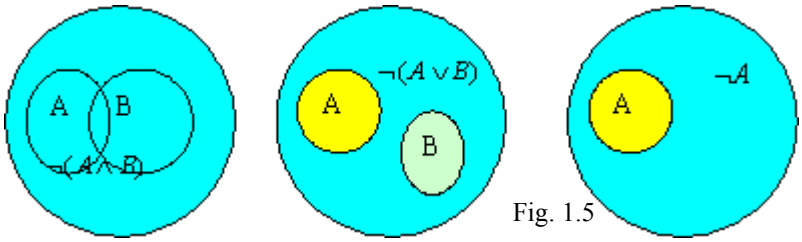


Fig. 1.5

For instance, logic tells us

$$\neg(A \wedge B) = \neg A \vee \neg B$$
$$\neg(A \vee B) = \neg A \wedge \neg B$$
$$\neg(\neg A) = A \qquad (1.18)$$

From this we get

$$p(A) = 1 - p(\neg A)$$
$$p(\neg(A \vee B)) = p(\neg A) * p(\neg B) = (1 - p(A)) * (1 - p(B)) = 1 - p(A) - p(B) + p(AB)$$
$$= 1 - (+p(A) + p(B) - p(AB)) = 1 - (p(A + B)) = \neg p(A \vee B)$$
$$p(\neg(A \wedge B)) = p(\neg A) + p(\neg B) - p(\neg A \wedge \neg B) = (1 - p(A) + (1 - p(B)) - p(\neg A) * p(\neg B)$$
$$= 2 - p(A) - p(B) - (1 - p(A))(1 - p(B) = 1 - p(A) * p(B) = 1 - p(AB) = \neg p(A \wedge B)$$

The latter equation is immediately clear. We consider not (A and B) that means all but not the combined area of A and B. The probability p(AB) is p(A)*p(B). The complementary probability is therefore 1-p(AB).

Randomness

At the end of this chapter we have to deal with randomness. Probabilities have much in common with the concept of randomness, because in probability theory we assume that samples are taken at random. If we model a biological process we also rely on random numbers to compute certain so-called stochastic events. A **stochastic process** is a process that mostly or in total depends on random events. To model this we need **random numbers**. What is a random number? A random number is an event that has inside a predefined range a certain so-called **probability function**. The probability, that this event has the outcome X is defined by this function p(X). The simplest form of a probability function is a linear one. All events X inside a range (A,B) have the same probability. For instance, the numbers of a dice all have the same probability p(X) to be thrown. This probability is 1 / 6 and the range (A,B) goes from 1 to 6. The function p(X) is therefore a simple straight

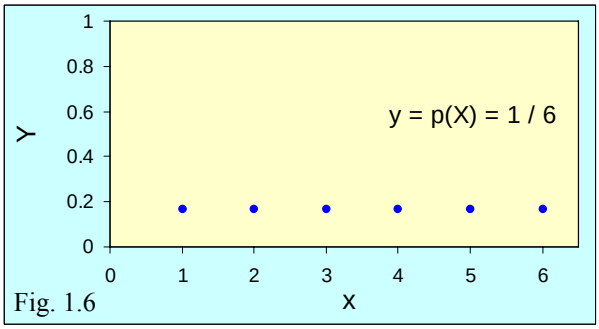


Fig. 1.6

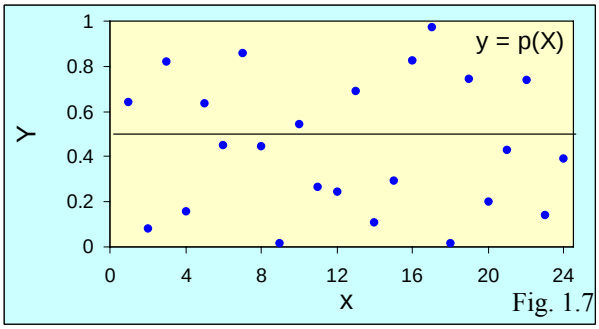


Fig. 1.7

line parallel to the x-axis (Fig. 1.6). Our probability is the same for every number, it is a linear random number. Our computer programs give such **linear random numbers** in the range between 0 and 1. For instance the command +los() or +ran() of the spreadsheet program Excel produce linear random numbers between 0 and 1 as shown above (Fig. 1.7). Plotted are 24 random numbers generated by Excel. The data points are quite 'randomly' distributed over the plotting area. The mean value is 0.5. If we would divide the y-axis into classes we would see that each class has the same probability to get a random number. Therefore, if we would divide the range into 10 intervals of width 0.1 each, the probability that a random number lies inside this interval would be exactly 0.1.

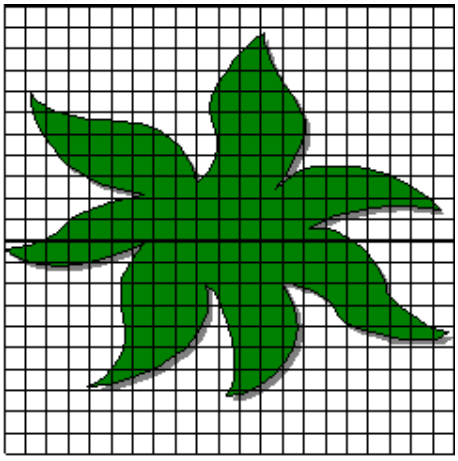


Fig. 1.8

Linear random numbers are the simplest but by no means the only and also not necessarily the most important form of random numbers. Consider a sample of leafs. To estimate the evapotranspiration of these leafs we have to measure the leaf area. We know already how to do this. By a double integral. Today, digitizers perform this measuring task for us, but own experience tells me that it is often faster to measure the area by hand after scanning the leafs and then using a raster as shown above (Fig. 1.8). The leaf area is then the area of all quadrates filled more than half by the leaf. If we sample and measure 1000 leafs we will get a picture as shown in Fig. 1.9. Most often leaf area takes values around a mean of about 12 to 14 (we use relative measures). Only sometimes very small or very large leafs occur. The next Figure (Fig. 1.10) shows the **relative frequencies** of leaf numbers per area class. We simply divided the number of leafs n_i per largeness class i through the total number of leafs N (in our case 1000). The relative frequency has therefore the form

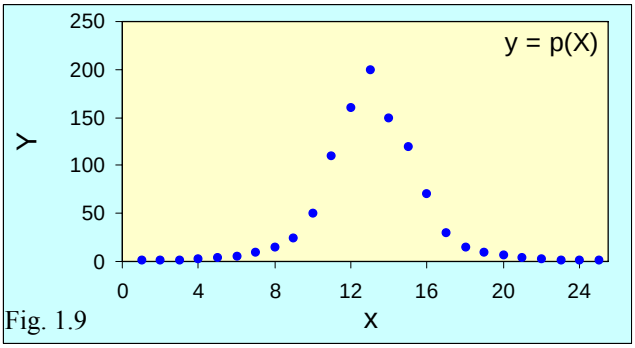


Fig. 1.9

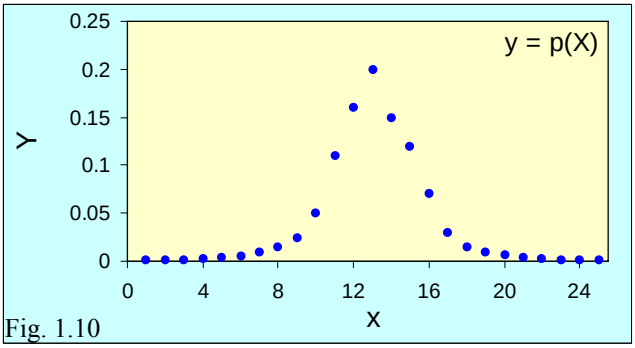


Fig. 1.10

$$f(X) = \frac{n_i}{N} \qquad (1.21)$$

The result is a so-called **frequency distribution**. In the statistics part we will deal with frequency distributions in detail. We may interpret this last Fig-

ure as to show the probability that a randomly taken leaf has a certain area. Therefore

$$p(X) = f(X) = n_i / N \tag{1.22}$$

Frequency or probability distributions have important properties. First they describe the distribution of probabilities of a certain event. Therefore, the sum of all these probabilities must add to unity, because one of the states must occur with necessity. In the case of discrete probabilities we have a **discrete distribution**.

$$\sum_{i=1}^{n} p(X_i) = 1 \tag{1.23}$$

In the continuous case we have a **continuous distribution**

$$\int_{min}^{max} p(x)dx = 1 \tag{1.24}$$

In turn any function having the property of eq. 1.24 can be viewed as being a probability function. For example eq. 1.12 introduced the discrete hypergeometric distribution. Fig. 1.11 shows the respective probability distribution of k desired objects in a sample of n = 50 for a total sample space (population) of N = 1000 having K = 500 desired objects. Fig. 1.12 shows the respective cumulative frequency distribution that goes to unity. In the statistics part we will deal with other important frequency (probability) distributions.

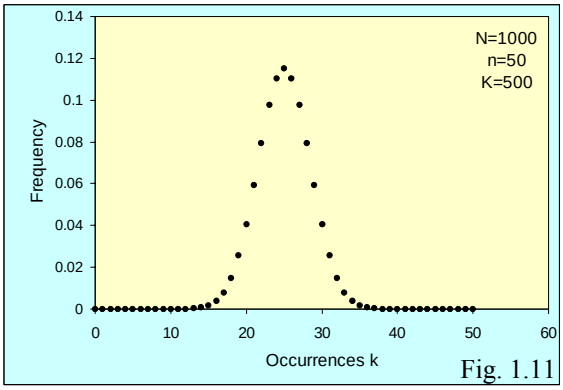


Fig. 1.11

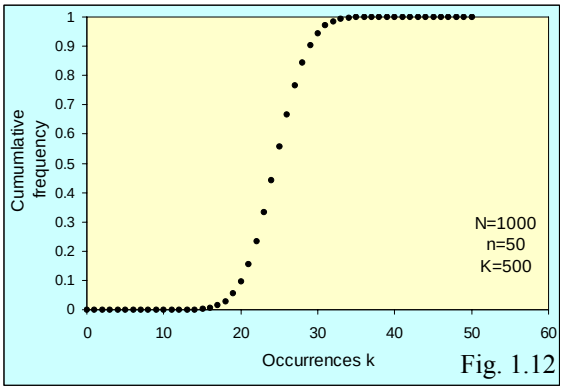


Fig. 1.12

Random numbers

From the math part we know already that computer programs give linear random numbers (exactly pseudo random numbers) in the range between 0 and 1 (denoted as ran(0,1)). To transform such random numbers into linear random numbers between various a and b we have to transform

$$ran(a,b) = (b-a)ran(0,1) + a$$

However for modelling we often need random numbers that have other frequency distributions. For instance we need a random number having a power function distribution. First of all we need a range for which this random number should be defined. This might be any range. Second, from eq. 1.22 it is clear that the total area under the curve of a continuous random variable or the sum of all frequencies of a discrete random variable has to be 1. In Fig. 1.13 the function $y = x^{-0.5}$ is given. We want to mimic such a process by a random variable. We choose the range between 0 and 7. Now we are looking for a **random variate** that gives a similar frequency distribution. We need a random variate that follows a power function with a slope of –0.5. We get our random numbers as follows. For $ran_{lin} = 0$ $ran_{pow} = 1$, for $ran_{lin} = 1$ $ran_{pow} = 7$. We find

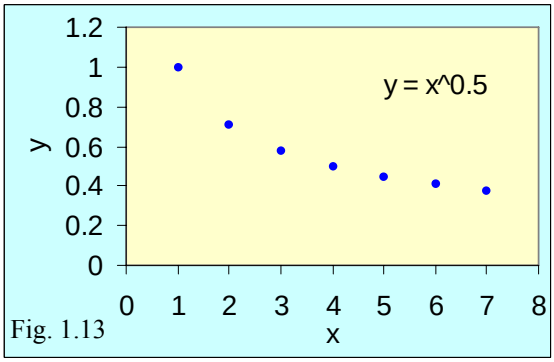


Fig. 1.13

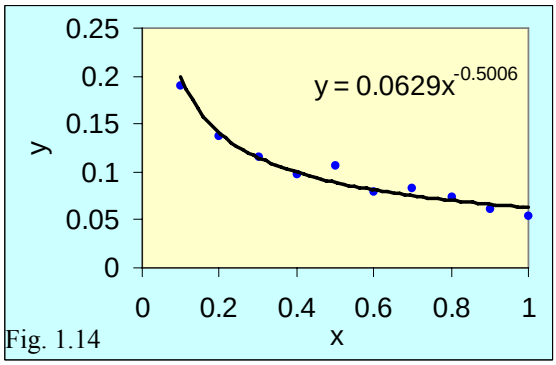


Fig. 1.14

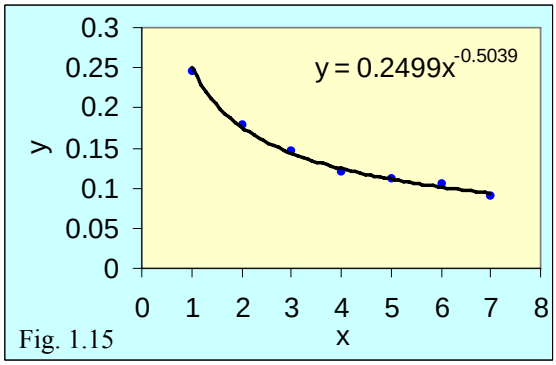


Fig. 1.15

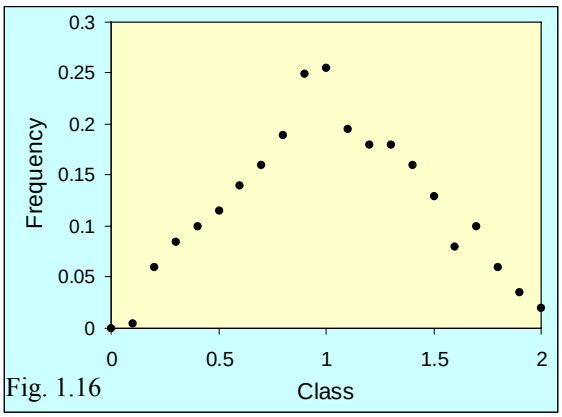


Fig. 1.16

$$ran_{pow}(0,1) = \frac{a^{ran(0,1)} - a}{1 - a} \tag{1.25}$$

and

$$ran_{pow}(0,7) = 7ran_{pow}(0,1)$$

Both functions are shown in Figs. 1.14 and 1.15.

The simple transformation

$$ran_{pow}(0,1) = ran(0,1)^a \tag{1.26}$$

might also be used to generate allometrically distributed random numbers but is not defined for ran = 0.

Important is a triangular distribution of random numbers. It can easily be obtained by the sum of two random numbers (Fig. 1.16).

$$\mathrm{ran}_{\mathrm{triangular}}(0,2) = \mathrm{ran}(0,1) + \mathrm{ran}(0,1) \tag{1.27}$$

2. Periodic processes

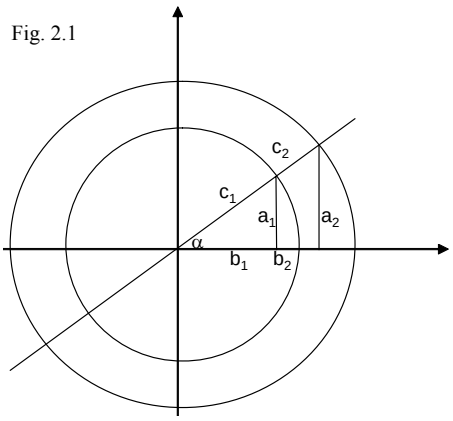
Trigonometric functions

Trigonometry is one of the oldest mathematical disciplines. It deals with the description of angels and geometric objects. Look at the next Figure. It shows us two circles of different radia c_1 and c_2. The law of Pythagoras tells us that $c_1^2 = a_1^2+b_1^2$ and $c_2^2 = a_2^2 + b_2^2$. If the smaller circle would have a radius of 1 it would follow that $a_1^2+b_1^2=1$. Now consider the angle α. Elementary geometry tells us that $a_1 / c_1 = a_2 / c_2$ and $b_1 / c_1 = b_2 / c_2$. α is therefore fully described by any set of quotients between c, a, and b. We can therefore define

$$\sin(\alpha) = \frac{a_1}{c_1};$$
$$\cos(\alpha) = \frac{b_1}{c_1};$$
$$\tan(\alpha) = \frac{a_1}{b_1};$$

(2.1)

Fig. 2.1



to name the three most important **trigonometric functions**, the **sine**, the **cosine** and the **tangent**. Three other, the **cotangent**, the **secans** and the **cosecans** are of minor importance. These above definitions imply also the possibility to measure angles in a new way. Most often, angles are measured from 0° to 360°. But we may also use the perimeter of a circle with radius 1. This perimeter is exactly 2π. This implies that 180° are equivalent to π, and 90° equivalent to π/2. This measure of angles are termed **radian** and is most often used in science. Nearly all computer and math programs use radians as the default. In general we get

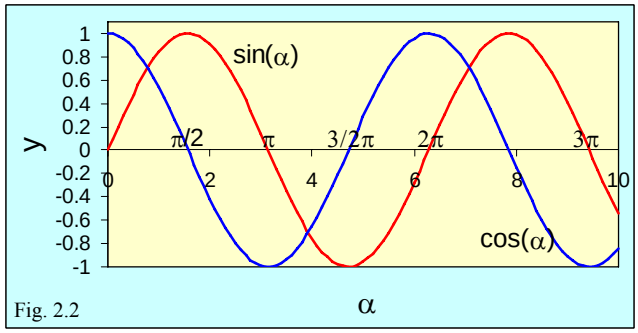


Fig. 2.2

$$\frac{x}{360^\circ} = \frac{y}{2\pi}$$

(2.2)

Hence

$$30^\circ \triangleq \frac{\pi}{6}; 45^\circ \triangleq \frac{\pi}{4}; 60^\circ \triangleq \frac{\pi}{3}; 90^\circ \triangleq \frac{\pi}{2}; 180^\circ \triangleq \pi$$

Now, we plot our three trigonometric functions against the angle measured in radians. We see that sine and cosine are periodic function that both have a period of 2π (Fig. 2.2). Their maximum is 1 and their minimum –1. The sine function is shifted π/2 to the right. The tangent functions is more complicated because for π/2, 3π /2, 5π /2 and so on the tangent goes to infinity (Fig. 2.3). The tangent is therefore not a continuous function.

Fig. 2.2 reveals two important relationships

$$\sin(x) = \cos(x - \frac{\pi}{2})$$

and

$$\cos(x) = \sin(x + \frac{\pi}{2})$$

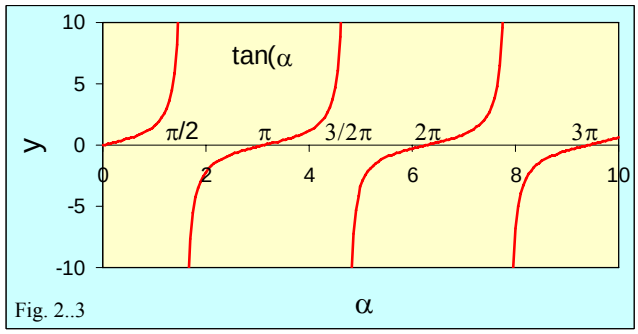


Fig. 2..3

There are many other relationships between the trigonometric functions. Most of them can be derived from simple geometry and are tabulated in mathematical compilations.

One of the more important is the so-called **cosine law**: For any triangle with sides a, b, and c, and angle γ between a and b holds:

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma) \tag{2.3}$$

A look back to Fig. 2.1 gives us immediately another important relationship. We deal with rectangular triangles and r is by definition set to 1. Therefore

$$r^2 = x^2 + y^2 \rightarrow 1 = \sin^2(\alpha) + \cos^2(\alpha)$$

Two other relationships might be still known from school and are very important

$$\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$$
$$\sin(\alpha + \beta) = \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta)$$

There is another class of functions that is closely related to the trigonometric ones. These are the **hyperbolic functions**. The hyperbolic cosine and sine are defined via exponential functions

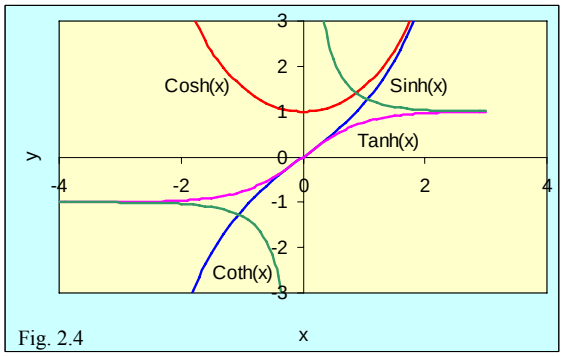


Fig. 2.4

$$\sinh x = \frac{1}{2}(e^x - e^{-x})$$
$$\cosh x = \frac{1}{2}(e^x + e^{-x})$$
$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} \tag{2.4}$$

Fig. 2.4 shows these functions.

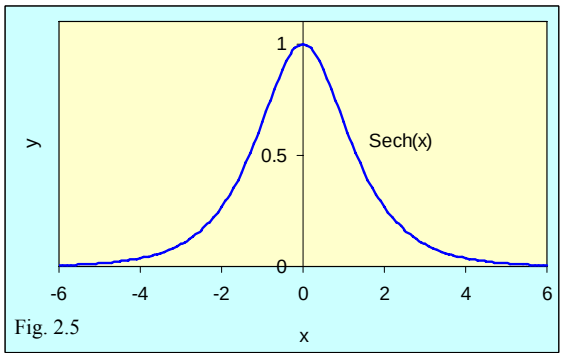


Fig. 2.5

An important function is the inverse of the cosh, the secans hyperbolicus function (Fig. 2.5). As in trigonometry it is defined as

Sech(x) = 1 / cosh(x). Hence

$$\sec h(x) = \frac{1}{\cosh(x)} = \frac{2}{e^x + e^{-x}} = \frac{2e^x}{1 + e^{2x}} \tag{2.5}$$

Sech(x) is therefore a function that describes a bell shaped or Gaussian function. Later we will deal with this type of function in detail. From the definition in eq. 2.4 we get two important relationships

$$\sinh(x)+\cosh(x)=e^{x}$$
$$\cosh(x)-\sinh(x)=e^{-x}$$

There are other relationships that are very similar to those of trigonometric functions

$$\cosh^2(x)+\sinh^2(x)=\frac{1}{4}(e^{2x}+2+e^{-2x})+\frac{1}{4}(e^{2x}-2+e^{-2x})$$
$$=\frac{1}{2}(e^{2x}+e^{-2x})=\cosh(2x)$$
$$\cosh^2(x)-\sinh^2(x)=\frac{1}{4}(e^{2x}+2+e^{-2x})-\frac{1}{4}(e^{2x}-2+e^{-2x})=1$$

The latter expression explains why the hyperbolic functions are similar to the trigonometric ones. Substituting u for $\cosh^2(x)$ and v for $\sinh^2(x)$ we get $u^2 - v^2 = 1$. This is the function of a hyperbola (Fig. 2.6) in contrast to a circle in the case of ordinary trigonometric functions.

The inverse hyperbolic functions result immediately from the definitions

$$y=arc\sinh(x)=\ln[x+\sqrt{x^2+1}) \quad ; x,y\in\mathbb{R}$$
$$y=arc\cosh(x)=\ln[x+\sqrt{x^2-1}) \quad ; x,y\in\mathbb{R}; x,y\geq 1$$
$$y=arc\tanh(x)=\frac{1}{2}\ln(\frac{1+x}{1-x}) \quad ; x,y\in\mathbb{R}; -1<x<1$$

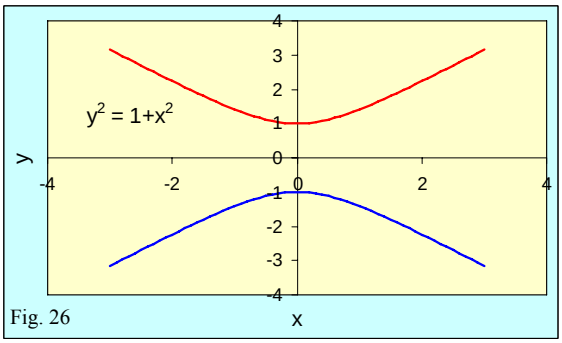


Fig. 26

Applications

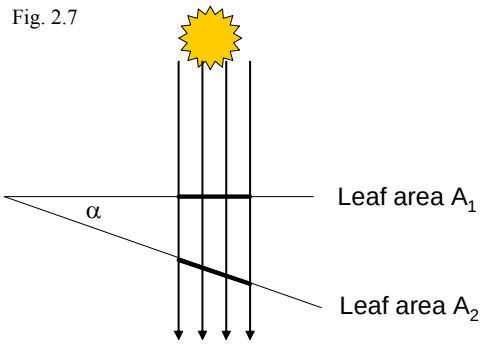
Why is it necessary to know about trigonometric functions? A simple example. Let the potential energy uptake of a leaf exposed to the sun at different angles be I_1 and I_2. Let A_1 and A_2 be the total leaf area (Fig. 2.7). Hence $A_1 / A_2 = \cos(\alpha)$. The total amount of energy falling on A_1 and A_2 remains constant. $A_1 / A_2 = I_1 / I_2$. Hence

$I_1 = I_2 \cos(\alpha)$.

If light reaches a leaf perpendicular the potential amount of energy is maximal. If we turn the leaf by an angle α, this energy decreases by $\cos(\alpha)$.

Trigonometric functions provide also a way to standardize measurements. If we have three series of measurements with values from, say, 0 to 100, -100 to 10, and 1 to 5, it is sometimes desired to standardize these measurements for further comparison. To do this we assume that these values are angles measured conventionally in degrees. Now we take the inverse of the tangent, called the arctan or atan (the inverse of the co-

Fig. 2.7



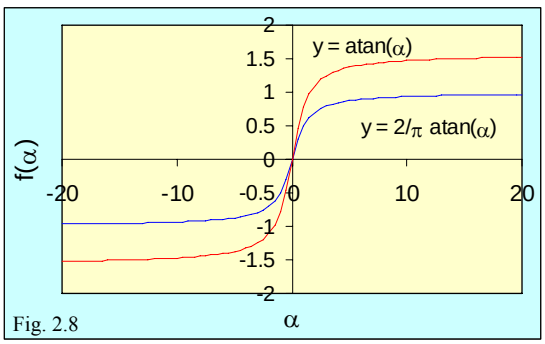


Fig. 2.8

sine is then the arccosine or acos, and the inverse of the sine the arcsine or asin). This transformation forces our initial values into a range from –1 to 1. We transform

$$y = \frac{2}{\pi}\arctan(x)$$

The result is shown in Figure 2.8. A variable x ranging from –20 to 20 is forced by the above transformation into a range between –1 and 1.

More important is the ability of combinations of trigonometric functions to mimic many other distributions. This is an indispensable tool for modelling cyclic or even irregular biological processes. Consider the next example. Figure 2.9 shows abundance data of the *Canada lynx* in Canada provided by the Hudson Bay Company over more than 100 years (1821-1934; they had excellent bookkeepers). We immediately notice an obvious regularity in the data. *Canada lynx* fluctuated regularly in density with a period of about 12 to 15 years.

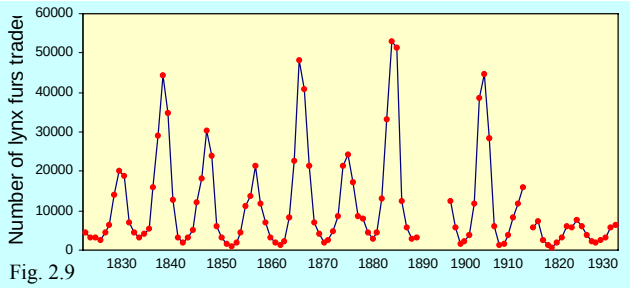


Fig. 2.9

We now mimic this process by a trigonometric function of the form

$$y = a * \sin(bt + c) + d \qquad (2.4)$$

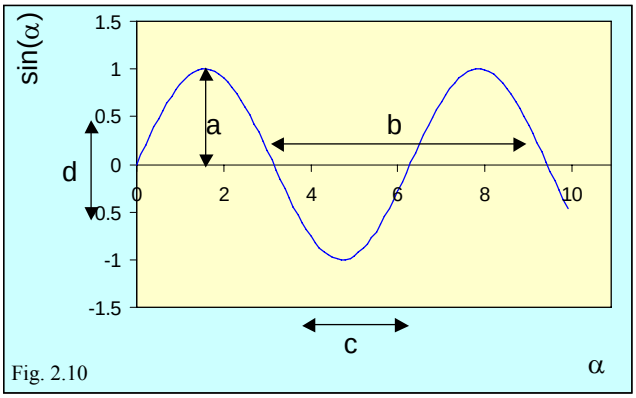


Fig. 2.10

This is the simplest form of a trigonometric model (Fig. 2.10). We have four parameters a, b, c, and d. a is responsible for the total **amplitude** of the model, b triggers the **period** (the **wavelength**), c shifts the function to the left our to the right, it triggers the **phase**, and d shifts the function up or down. The frequency of a wave is then defined as

f = 1 / wavelength. (2.5)

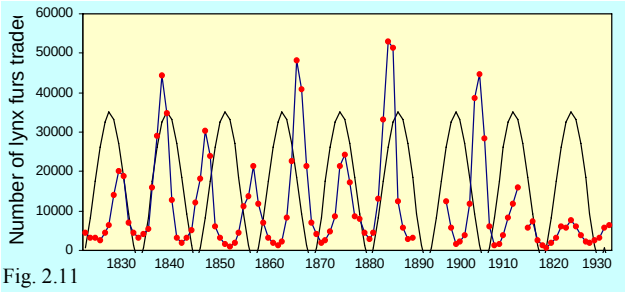


Fig. 2.11

Lets try to model the *Canada lynx* data (Fig. 2.11). Our aim is to foresee future densities of the *Canada lynx*. With a simple try and error method I found a model that fits more or less to the data. (Fig. 2.11) Indeed, it mimics the cycling behaviour and the period of 12 years detectable in the real data is followed by our model. However, the model is far from being perfect. And foreseeing the future seems to be impossible.

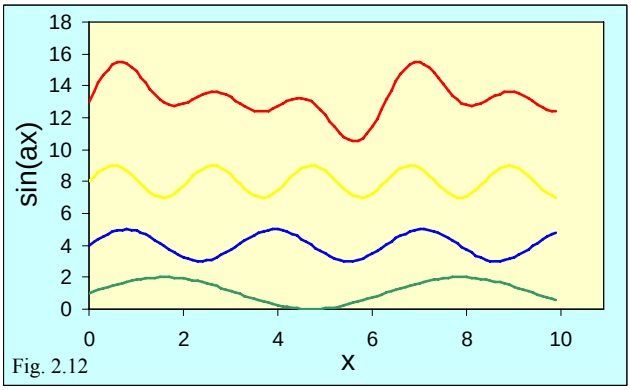


Fig. 2.12

But we can modify the model. Look at the next Figure (Fig. 2.12). There, three simple sine-

functions of the form y = sin(ax)+ b are shown. The b-values are only introduced to make the Figure readable. a triggers the **wavelength** of the function. The red line is the sum of these three waves. It is therefore computed from

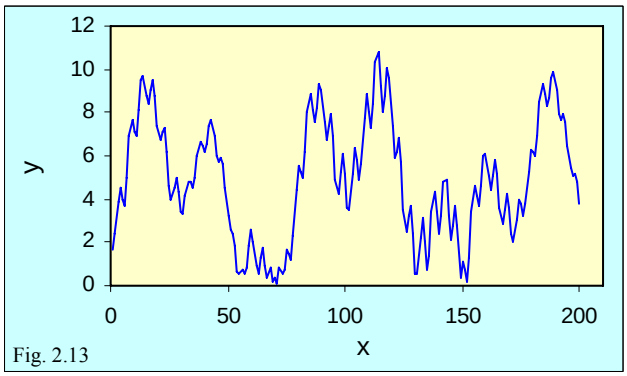


Fig. 2.13

$$y = \sum_{i=1}^{3} \sin(a_i x) + b_i$$

We see that an irregular pattern appears, a pattern that is not obvious from the three functions alone. In other words, it seems possible to study irregular series (for instance time series) by a sum of independent trigonometric functions. This is best exemplified by the next example. A simple linear combination of 8 sine and cosine functions gives Fig. 2.13.

$$y = \sum_{i=1}^{8} a_i \left(\cos(b_i x) + \sin(c_i x)\right) + d$$

To generate such an irregular pattern we had to define the a, b, and c constants as power functions depending on n. They had the general form

$$a_i = a_0^{-i}; b_i = b_0^{-i}; c_i = c_0^{-i}$$

where i runs from 1 to 8.

The above two examples showed us the possibility to mimic irregular patterns by a simple linear combination of trigonometric functions. It can be proven that it is always possible to approximate any continuous function in such a way. A general description of this approximation is the so-called **Fourier series** (Jean Baptiste Joseph de Fourier, French mathematician, 1768-1830). A Fourier series is defined by the following equation

$$y = f(x) = \frac{a_0}{2} + \sum_{i=1}^{\infty} \left(a_i \cos(cx) + b_i \sin(cx)\right) \tag{2.6}$$

The coefficients a_i, b_i and c can be solved exactly, but in praxis it is much more easy to approximate the series by a numerical solution. Most statistics packages and math programs contain Fourier analyses and very often, the first 5 to 10 a_i and b_i values are sufficient for a very good approximation. A Fourier series can mimic any pattern - even random ones. However, to predict the future, for instance future densities of species, pest outbreaks, physiological time series, we need something more. We have to detect regularities. Something that allows a simpler model to be constructed and predictions to be made. There are special computer programs available for pattern seeking and we will hear in the statistics part about some basic principles of modelling time series.

.

Polar Coordinates

After the second world war the German biologist Karl von Frisch (1886-1982) found that bees are able to tell other bees direction, distance and amount of food resources. They do this via their famous Schwaenzeldance (Fig. 2.14). The intensity of the dance tells about the amount of food. The number of moves

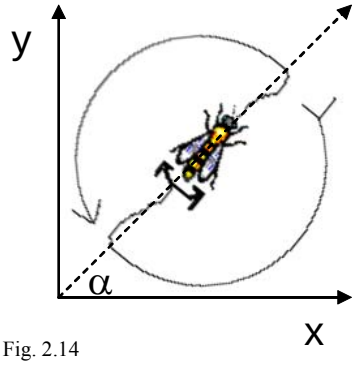


Fig. 2.14

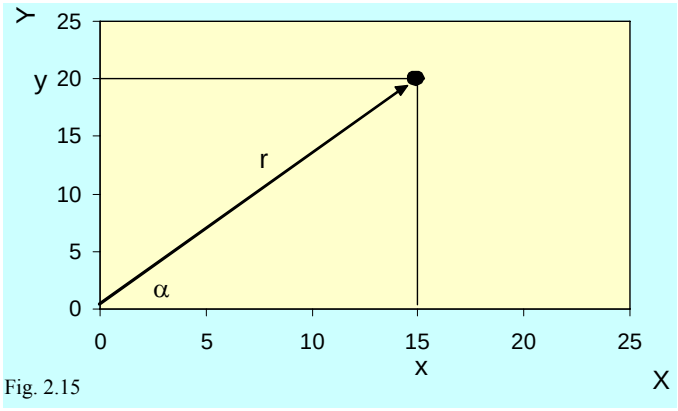


Fig. 2.15

(right—left) per unit time, (the frequency) gives the distance and the direction of the dance in relation to the sun the direction of the resource. Hence bees point to the direction of the resource and define this direction not via x- and y– axes but via an angle between the direction of the food and the direction of the sun.

This is very similar to the mathematician's concept of **polar coordinates**. What is this? The traditional way to denote points in a coordinate system is by giving the x- and y-coordinates. Hence the point in Fig. 2.15 has the coordinates x = 15 and y = 20, shortly (x;y) = (15;20). This is not the way organisms look at a point. They look in the direction of an object and try to estimate its distance. This is what polar coordinates do. They give a direction measured by an angle α in relation to the x- and y-axis. The distance is given by the length r. This can be written in the following way

$$\begin{aligned} x &= r\cos(\alpha) \\ y &= r\sin(\alpha) \end{aligned} \tag{2.7}$$

These are the polar coordinates of the point in Fig. 2.15. Instead of x and y we use r and α to describe a point. If we have r and α x and y are fixed. We can also compute r and α from the knowledge of x and y.

$$\left.\begin{aligned} x &= r\cos(\alpha) \\ y &= r\sin(\alpha) \end{aligned}\right\} x^2 + y^2 = r^2(\sin^2(\alpha) + \cos^2(\alpha)) = r^2 \tag{2.8}$$

α is given from cos(α) = x/r and sin(α) = y/r. We get

$$\alpha = \arg(x,y) = \begin{cases} \arccos(x/r) & \text{if } y \geq 0 \\ \arcsin(y/r) & \text{if } x \geq 0 \end{cases} = \begin{cases} -\arccos(x/r) & \text{if } y < 0 \\ -\arcsin(x/r) & \text{if } x < 0 \end{cases} \tag{2.9}$$

The latter is called the argument or the polar angle of the point (x,y).

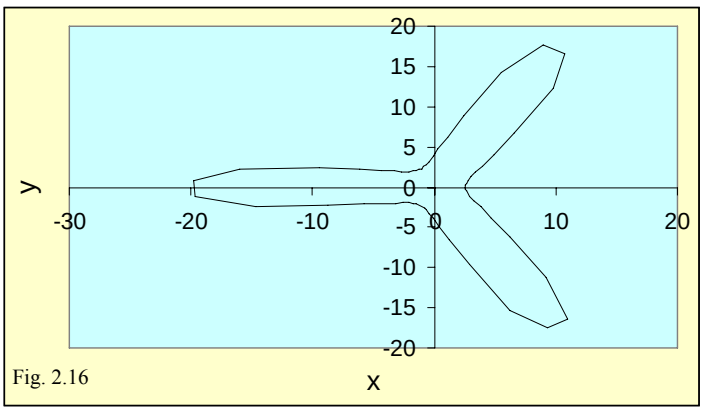


Fig. 2.16

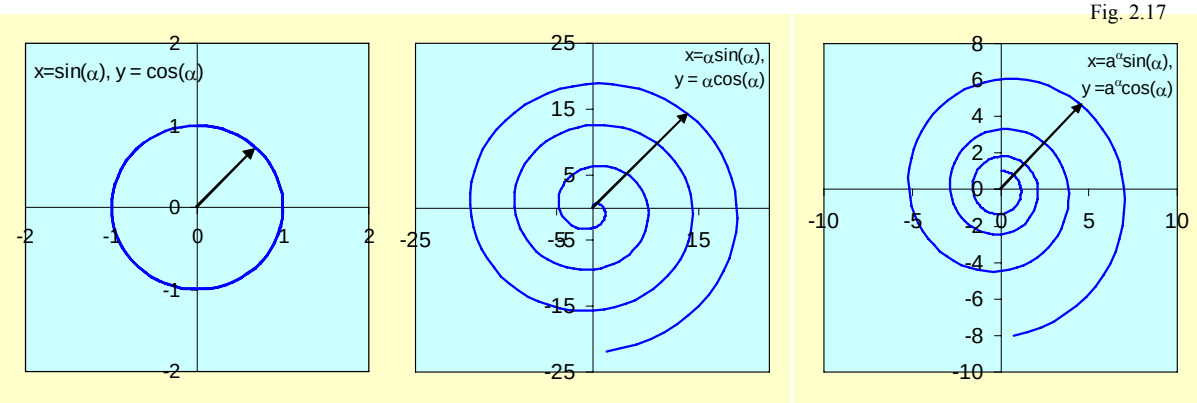
It is possible to express functions in polar coordinates. A polar function has the general form

$$\begin{aligned} r &= f(\alpha) \\ x &= r\cos(\alpha) \\ y &= r\sin(\alpha) \end{aligned} \tag{2.10}$$

The task is to find the function f(α) that defines a certain shape. In 2003 the Belgian

	A	B	C	D	E	F	G	H	I
1	Parameters	a	b	c	l	m	k		
2		2	6	6	4	5	3		
3	Theta	r(theta)	cos-Term	Sin-Term				x	y
4	0	1	0.5	0	0.0625	0	2.5198421	2.5198421	0
5	+A4+0.1	1	((1/$B$2*COS($A5/4*$D$2))^2)^0.5	((1/$C$2*SIN($A5/4*$D$2))^2)^0.5	+C5^$E$2	+D5^$F$2	1/((E5+F5)^(1/$G$2))	+$G5*COS($A5)	+$G5*SIN($A5)

Fig. 2.17



mathematician Johan Gielis derived a so-called **superformula** and showed that every plane geometric object (two dimensional object) can be modelled with this equation.

$$r = \frac{f(\theta)}{\sqrt[k]{\left(\left|\frac{1}{a}\cos\left(\frac{c}{4}\theta\right)\right|\right)^{l} + \left(\left|\frac{1}{b}\sin\left(\frac{c}{4}\theta\right)\right|\right)^{m}}} \tag{2.11}$$

where a, b, c, k, l, and m are real parameters and f(θ) is a deliberate function depending on the angle θ.

An example how to apply equation 2.11 shows Fig. 2.16. The table shows the Excel model of this figure.

Polar coordinates are a very simple tool to describe periodic process and curves. Fig. 2.17 shows simpler shapes. The function x=sin(α) and y = cos(α) describe a circle, x=αsin(α) and y = αcos(α) an Archimedic spiral,and x=a^{α}sin(α) and y = a^{α}cos(α) a logarithmic spiral. Archimedic spirals are found in nature for instance in Nummulites, logarithmic spirals were typical for Ammonites. Many example of famous curves can be found at http://www-history.mcs.st-andrews.ac.uk/Curves/Curves.html

Periodic processes in nature

Many biological process occur periodically, the oscillate. The simplest process is a harmonic oscillation, a simple wavelike process that can be described by a sine or cosine function. They have the general form

$$y = a\sin(\omega t + \varphi) \tag{2.12}$$

with a defining the amplitude, ω the wavelength and φ the phase. We know already that if we have two functions that superimpose (add to each other) the resulting wave might look irregular.

If two waves have different phases we speak of **phase lags**. Of course, a phase lag of 2π makes both functions again identical. However, the wavelength remains unchanged. This follows from

$$\cos(\omega)+\cos(\omega+\varphi)=2\cos(\omega+\varphi/2)$$

The wavelength is still defined by ω. If their phase differs by π, the sum of both waves becomes zero. Both annihilate.

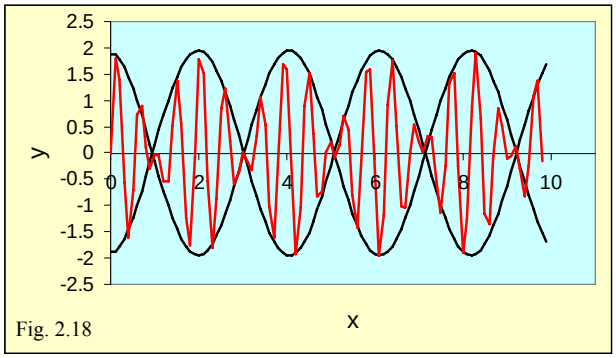


Fig. 2.18

Another important feature of superimposed waves are **compound oscillations** or **beats**, compound oscillations that generate new waves (Fig. 2.18). The superimposition of two simple sine waves generates a new wave with higher amplitude

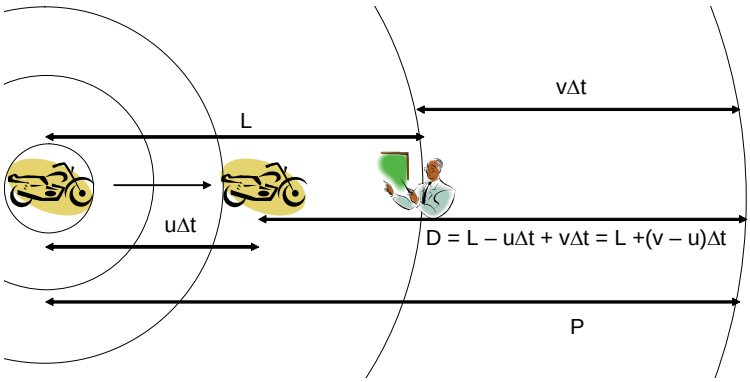
$$y=\sin(\omega)+\sin(\omega+\Delta\omega)=2\sin\left(\frac{2\omega+\Delta\omega}{2}\right)\cos\left(\frac{\Delta\omega}{2}\right)\approx 2\sin(\omega+\Delta\omega/2)$$

The beat frequency is given by

$$f=\frac{\Delta\omega}{2\pi}$$

with Δω being the initial difference in phase. Beats only arise when both waves are nearly in phase, hence when Δω is small.

Fig. 2.19



Lastly, we consider the so called **Doppler effect** (after the Austrian mathematician Christian Doppler, 1803-1853). Suppose we have a sound wave with a frequency of f waves per second. In time t_0 it travels from the emitter to a observer, which is a distance L away. Hence L * f * t_0 waves are emitted. Because wavelength λ is a distance we also notice that wavelength per time must be a measure of speed. If the speed of the wave is v and T the period (T = 1 / f) we get

$$\frac{\lambda}{T}=v=\lambda f \qquad (2.13)$$

At a time difference Δt the emitter moves in the direction of the observer. The way is given by uΔt, with u being its speed. In the same time the wave travelled further by a way vΔt, with v > u. How does the observer perceive the wave front?

The distance D = [L + (v—u)Δt] contains D / λ waves lengths. In Δt fΔt new waves have been generated. Hence

$$\frac{L+(v-u)\Delta t}{\lambda}-\frac{L}{\lambda}=f\Delta t\rightarrow f=\frac{v-u}{\lambda}$$

Now assume that the observer records p waves per second. Therefore p = v / λ (eq. 2.13). Hence

$$p = \frac{v}{\lambda} = f\,\frac{v}{v-u} = f\,\frac{1}{1-\frac{u}{v}}$$

The perceived frequency p differs from the emitted frequency f by a constant factor. If u > 0 (travelling in the direction of the observer) the perceived frequency is larger and the wavelength smaller than emitted. If u < 0 the perceived frequency is smaller and the wavelength larger than emitted. The Doppler effect also explains the so called red shift in astronomy. The wavelengths of very distant and fast travelling galaxies appear to have longer wavelengths (a red spectrum) than more nearby galaxies.

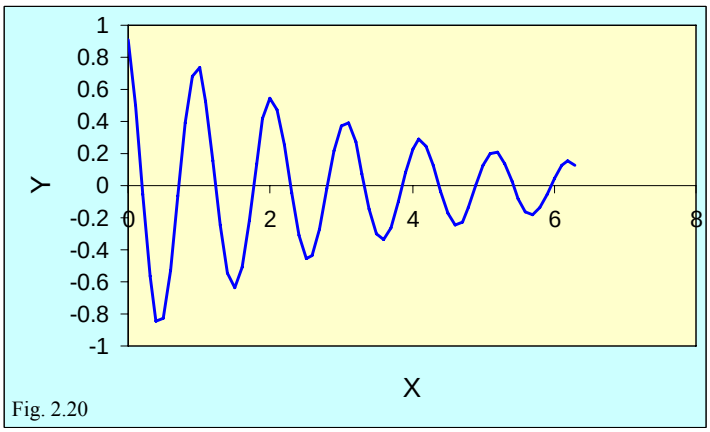


Fig. 2.20

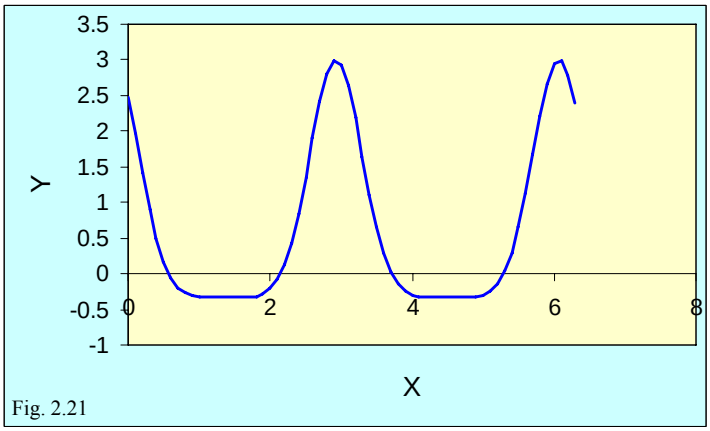


Fig. 2.21

Fig. 2.20 shows an important example of periodic processes in nature A **damped oscillation** is given by the superposition of an exponential function with a wave function

$$y \propto e^{-ax};\, y \propto sin(x)$$
$$y = be^{-ax} sin(x) \qquad (2.14)$$

Damped oscillators occur in physiological or bacterial growth processes.

The combination of exponential or power functions with trigonometric functions can mimic various oscillating process in nature.

An oscillation with time lag shows for instance Fig. 2.21. It is generated by the function

$$y = A^{\sin(\omega x+\varphi)} \sin(\omega x+\varphi)$$

3. Vectors

Given a point in space we can shift this point to another place. The arrow that goes from the original point to its new place is called a **vector**. In Fig. 3.1 we have three vectors A, and B, and a vector C that points back to itself. The latter is called a **null vector**. In a Cartesian system vectors are given by the coordinates of the endpoint minus the coordinates of the starting point. Hence in Fig. 3.1

$$A = \begin{pmatrix} 20-5 \\ 20-10 \end{pmatrix} = \begin{pmatrix} 15 \\ 10 \end{pmatrix}$$

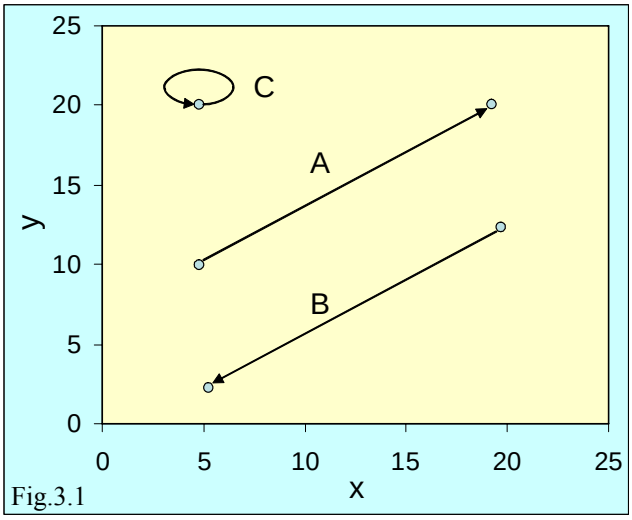


Fig.3.1

Hence, a vector V is defined by two points A = $\{x_1,y_1\}$ and B = $\{x_2,y_2\}$.

$$V = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix} \quad (3.1)$$

All vectors with identical x and y values are identical. Further

$$B = \begin{pmatrix} 5-20 \\ 2-12 \end{pmatrix} = \begin{pmatrix} -15 \\ -10 \end{pmatrix} = -1\begin{pmatrix} 15 \\ 10 \end{pmatrix} = -A$$

The vector B is parallel to A but points in the opposite direction. That means B = -A. In Fig. 3.2 the vectors i and j are given by

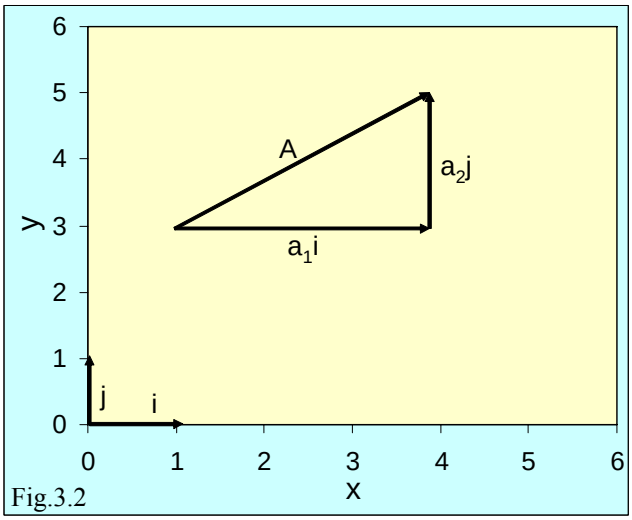


Fig.3.2

$$i = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; j = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The vector A can be seen as the multiplication of a number a_1 with i and a_2 with j. In vector algebra numbers are called **scalars**. Hence

$$A = \begin{pmatrix} a_1 i \\ a_2 j \end{pmatrix}$$

We call the scalars a_1 and a_2 the **coordinates of the vectors A**. The null vector is therefore defined by **o = {0,0}** and the unity (or normality) vectors are **I = {1,0} and J = {0,1}**.

We can define vectors in more than two dimensions. The general form of a vector in an n-dimensional space is

$$V = \begin{pmatrix} a_1 \\ \dots \\ \dots \\ a_n \end{pmatrix}$$

The examples above provide a natural introduction to basic vector operations. The addition and subtraction of vectors are defined as

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix} \tag{3.2}$$

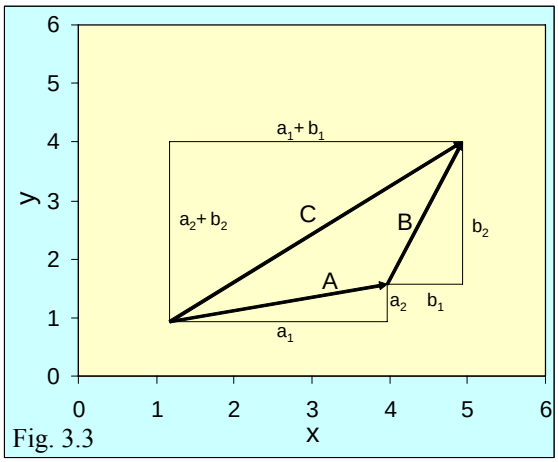


Fig. 3.3

In two-dimensional space this can be interpreted as generating a parallelogram from the vectors A and B that has the longer diagonal of C (Fig. 3.3). Consequently a subtraction A - B is defined as an addition of the antivector of -B and A (Fig. 3.4).

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} -b_1 \\ -b_2 \end{pmatrix} = \begin{pmatrix} a_1 - b_1 \\ a_2 - b_2 \end{pmatrix} \tag{3.3}$$

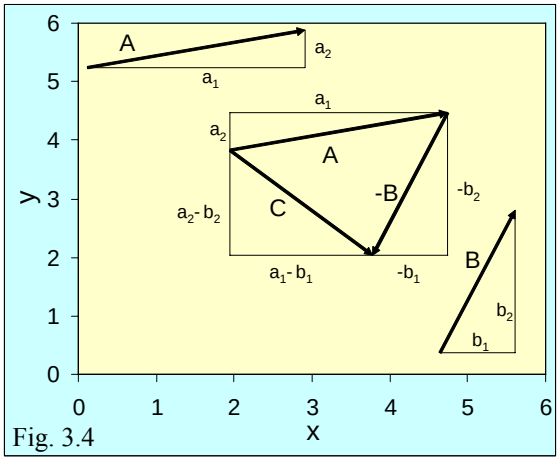


Fig. 3.4

Both definitions of course hold for additional dimensions too.

Basic theorems for addition and subtractions hold for vectors too. Both operations are commutative, and associative. Hence

$$\begin{aligned} &A + B = B + A \\ &A + (B + C) = (A + B) + C \\ &A + o = A \\ &A - A = o \end{aligned} \tag{3.4}$$

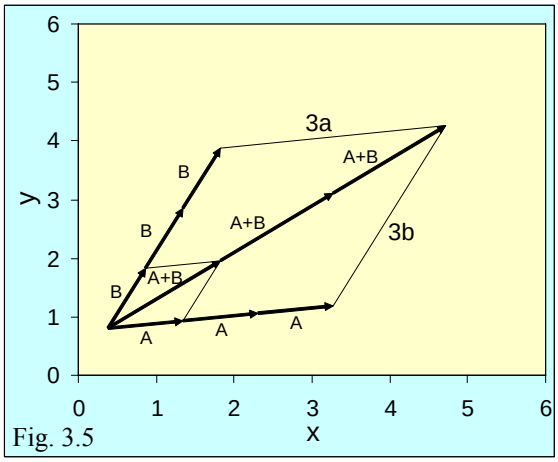


Fig. 3.5

An addition of A with the null vectors gives A and A - A gives the null vector *o*.

Next we define the multiplication. There are three types of vector multiplication. The first is the multiplication with a scalar, the **scalar multiplication (or S-product)**. Fig. 3.5 shows a natural definition of the S-product. We get

$$\lambda A = \lambda \begin{pmatrix} a_1 \\ \dots \\ a_n \end{pmatrix} = \begin{pmatrix} a_1 \\ \dots \\ a_n \end{pmatrix} + \dots + \begin{pmatrix} a_1 \\ \dots \\ a_n \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \dots \\ \lambda a_n \end{pmatrix} \tag{3.5}$$

The S-multiplication if commutative and distributive (Fig. 3.5).

$$\begin{aligned}
&\gamma(\lambda A) = (\gamma\lambda)A \\
&\lambda(A + B) = \lambda B + \lambda A \\
&(\lambda + \gamma)A = \lambda A + \gamma A \\
&oA = o
\end{aligned} \tag{3.6}$$

The length of a vector is of course given by the law of Pythagoras

$$|A| = \sqrt{a_1^2 + \ldots + a_n^2} \tag{3.7}$$

This definition implies that the length of $|A + B| \le |A| + |B|$.

We can multiply two vectors to get a scalar. This is called the **scalar product** and is defined as

$$\begin{pmatrix} a_1 \\ \ldots \\ a_n \end{pmatrix} \bullet \begin{pmatrix} b_1 \\ \ldots \\ b_n \end{pmatrix} = a_1 b_1 + \ldots + a_2 b_2 = \sum_{i=1}^{n} a_i b_i \tag{3.8}$$

The commutative, distributive and the mixed associative laws hold

$$\begin{aligned}
&A \bullet B = B \bullet A \\
&A \bullet (B+C) = (A \bullet B) + A \bullet C \\
&\lambda(A \bullet B) = (\lambda A) \bullet B = A \bullet (\lambda B) \\
&A \bullet o = o
\end{aligned} \tag{3.9}$$

The simple associative law does not hold $A \bullet (B \bullet C) \neq (A \bullet B) \bullet C$ because one time we get a vector in direction A and the other a vector in direction C.

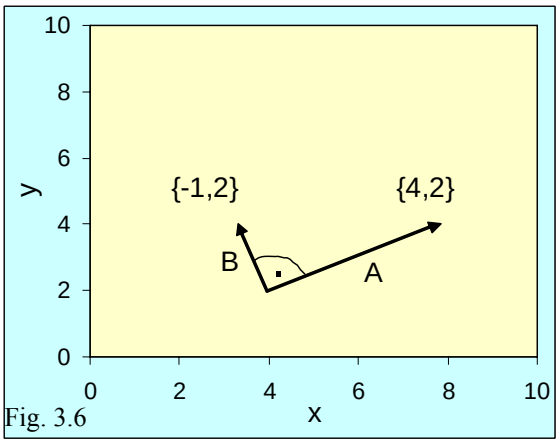


Fig. 3.6

Important is the case when we multiply two vectors that are perpendicular (orthogonal). Two-dimensional perpendicular vectors have the folowing structure (Fig. 3.6)

$$A = \begin{pmatrix} x \\ y \end{pmatrix}; B = \begin{pmatrix} \lambda y \\ -\lambda x \end{pmatrix} \rightarrow A \bullet B = \lambda xy - \lambda xy = 0 \tag{3.10}$$

Hence if the scalar product of two non-zero vectors is zero they are orthogonal.

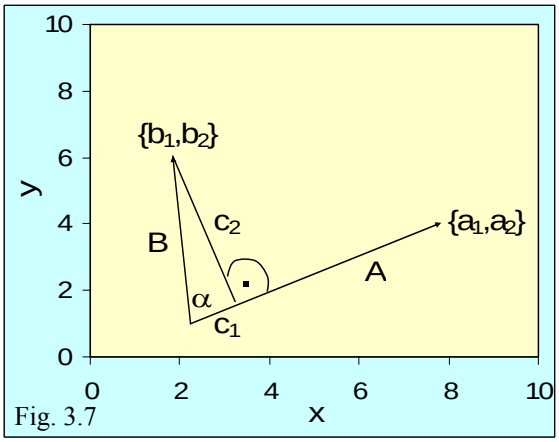


Fig. 3.7

Important is also that the equation $A \bullet X = b$ has an indefinite number of solutions (why?). **Further the division through a vector is not defined.**

The scalar product has a simple geometrical interpretation. It is the product of the length of A with the perpendicular projection c_1 of B on A (Fig. 3.7)(why?). Because of $c1/|B| = \cos(\alpha)$ we get

$$A \bullet B = |A||B|\cos(\alpha) = c_1|A| \tag{3.11}$$

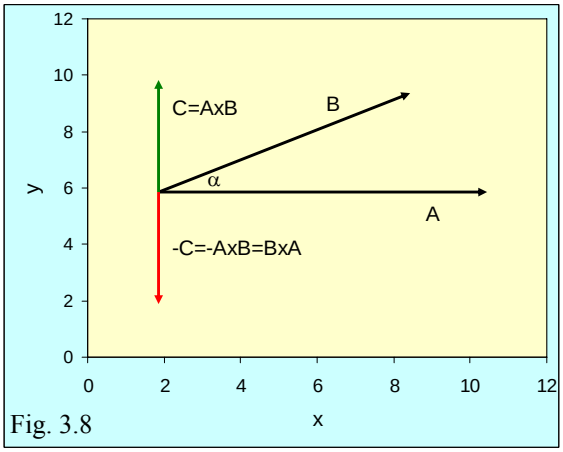


Fig. 3.8

If $\alpha = \pi/2$ A•B = 0. Further we get A•A=$|A|^2$.

There is another product of two vectors, the so-called **cross, or outer or vector product**. The vector product AxB of two vectors A and B gives a new vector C that stands perpendicular on the plane made of A and B (Fig. 3.8). We have two possibilities for a vector to be perpendicular. We get the correct orientation from the right hand rule (Fig. 3.9). Hence we have

$$AxB = -BxA$$

The vector product is therefore not commutative. But it is still distributive and associative.

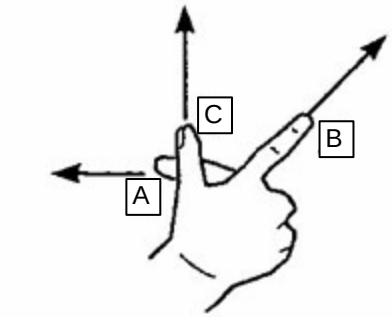


Fig. 3.9

In vector notation the cross product is defined as

$$C = AxB = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} x \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_1b_3 - a_3b_1 \\ a_1b_2 - a_2b_1 \end{pmatrix} \tag{3.12}$$

The length of C is given by

$$|C| = |A||B|\sin(\alpha) \tag{3.13}$$

The use of vectors allow for some easy proofs of geometrical theorems.

For instance the cosine theorem can be derived using a triangle made of three vectors C=A-B. and A•B = |B| |B| cos(γ). Hence

$$C^2 = c^2 = (A-B)^2 = A^2 - 2AB + B^2 = c^2 = a^2 + b^2 - 2ab\cos(\lambda)$$

For $\gamma = \pi/2$ we get the theorem of Pythagoras.

Important fields to use vectors are trigonometry and analytical geometry. Using vectors of unit length (Fig. 3.10) we can define an angle α from

$$E_1 \bullet E_2 = |E_1||E_2|\cos(\alpha) = \cos(\alpha)$$

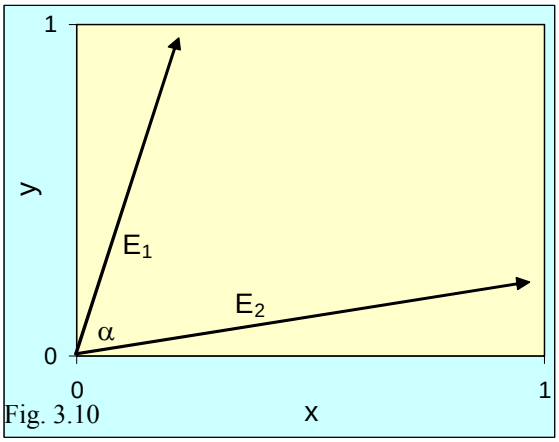


Fig. 3.10

Further we can define projections of geometrical objects. For instance a parallel shift of a length A by a vector V (Fig. 3.11) is given by

$$A = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \rightarrow A' = a_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} a_1 + v_1 \\ a_2 + v_2 \end{pmatrix}$$

Using vectors we can define straight lines in space (Fig. 3.12). The line A to point P is given by

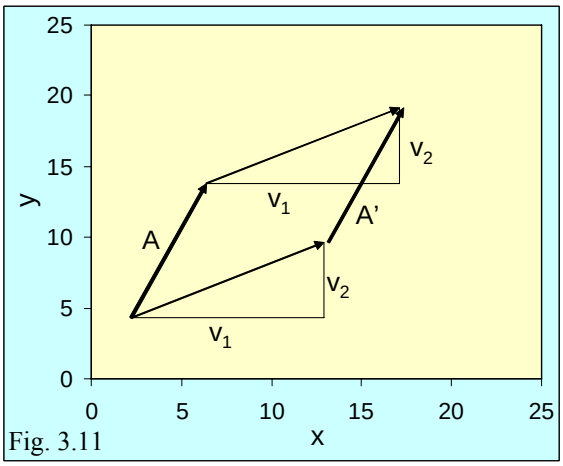


Fig. 3.11

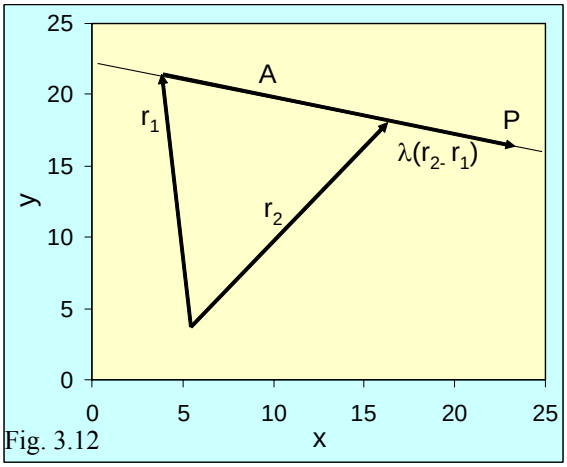


Fig. 3.12

$$A = r_1 + \lambda(r_2 - r_1)$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_{r1} \\ y_{r1} \end{pmatrix} + \lambda \begin{pmatrix} x_{r2} - x_{r1} \\ y_{r2} - y_{r1} \end{pmatrix} \quad (3.14)$$

This equation makes is easy to calculate geometrical relationships. For instance do the straight lines defined by the points $A_1 = \{1,2,3\}$ and $A_2 = (4,5,6\}$ and $B_1 = \{2,3,4\}$ and $B_2 = \{5,4,3\}$ intersect? We compute

$$A = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \lambda \begin{pmatrix} 4-1 \\ 5-2 \\ 6-3 \end{pmatrix} = B = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} + \gamma \begin{pmatrix} 5-2 \\ 4-3 \\ 3-4 \end{pmatrix}$$

$$\lambda \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix} - \gamma \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

and see that $\gamma = 0$ and $\lambda = 1/3$ fulfil this equation. Both straight lines cross in the point $P = \{2,3,4\}$.

Assume we have two points A and B on a straight line and the direction vector u that defines the straight line. U is given by the vector P from A to B by

$$P = \lambda u$$

where λ is a scalar that defines the length of the vector u.

The **normal vector n** of a straight line is a vector that stands perpendicular on this vector u. The norm vectors of any vector $Y=\{a,b\}$ are of course $n_1 = \{-b,a\}$ and $n_1 = \{b,-a\}$. The normal vector to two vectors A and B in space is given by the cross product $n = AxB$.

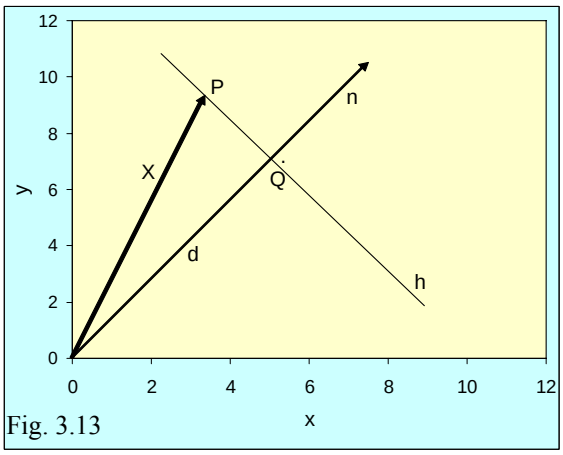


Fig. 3.13

From Fig. 3.13 we see that the dot product nX equals the projection d of X on n multiplied with the length of n. This gives a short notation for a straight line (the Hesse form, after the German mathematician Otto Hesse, 1811-1874))

$$n \bullet X = d|n| \quad (3.15)$$

$|n|$ is the length of the normal vector and d is the orthogonal distance of the line from the origin. How to translate the traditional equation of the straight line into a vector notation? From

$$y = mx + b \rightarrow y - mx - b = 0$$

we get a general form of

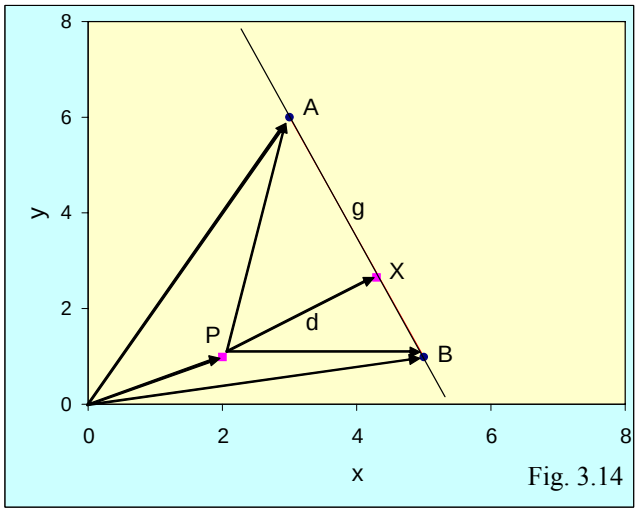


Fig. 3.14

$$a_1 x + a_2 y - b = 0$$
$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \bullet \begin{pmatrix} x \\ y \end{pmatrix} - b = 0 \qquad (3.16)$$

This equals eq. 3.15 and we get the norm vector n from a_1 and a_2.

This can be extended to any point P as the origin of n. (Fig. 3.14) For instance what is the distance of P = {2,1} from a straight line g defined by A = {3,6} and B = {5,1}? We obtain the two point form of the straight line and get the coordinate form by eliminating the parameter λ. This gives the normal vector n = {5,2}. The distance is then given by

$$d = \frac{\left|\overline{OP} \bullet n - b\right|}{|n|}$$

Hence

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix} + \lambda \begin{pmatrix} 3-5 \\ 6-1 \end{pmatrix}$$
$$x = 3 - 2\lambda; y = 6 + 5\lambda$$
$$\frac{x-3}{-2} = \frac{y-6}{5} \rightarrow 5x - 15 = -2y + 12$$
$$2y + 5x - 27 = 0$$
$$\left| \begin{pmatrix} 2 \\ 1 \end{pmatrix} \bullet \begin{pmatrix} 5 \\ 2 \end{pmatrix} - 27 \right| = d|n| \rightarrow d = \left| \frac{12 - 27}{\sqrt{29}} = \right| = 2.79$$

4. Matrix algebra

Biological data bases are most often structured in form of a matrix. Typical examples are our spreadsheet matrices, for instance *Excel*, *Access*, or *Matlab* (short for Matrix laboratory). The *Excel* examples below show typical biological data sets. Species are in rows and these are described by a set of nominally, ordinally or metrically scaled variables (descriptors). In the first matrix below we have species at four sites and the values are total catches. In ecology we often have only data about the absence or presence of a certain species. In this case we deal with presence absence matrices and presences are coded with a 1and absence with a 0.

In general we write matrices in form of rows and columns. An important special case is a matrix that has the same number of rows and columns. This is a square matrix. Matrices with only one row or one column (row or column matrices) are vectors. Hence matrices can be seen as being composed of several vectors.

Species	Taxon	Guild	Mean length (mm)	Site 1	Site 2	Site 3	Site 4
Nanoptilium kunzei (Heer, 1841)	Ptiliidae	Necrophagous	0.60	0	0	0	0
Acrotrichis dispar (Matthews, 1865)	Ptiliidae	Necrophagous	0.65	13	0	4	7
Acrotrichis silvatica Rosskothen, 1935	Ptiliidae	Necrophagous	0.80	16	0	2	0
Acrotrichis rugulosa Rosskothen, 1935	Ptiliidae	Necrophagous	0.90	0	0	1	0
Acrotrichis grandicollis (Mannerheim, 1844)	Ptiliidae	Necrophagous	0.95	1	0	0	1
Acrotrichis fratercula (Matthews, 1878)	Ptiliidae	Necrophagous	1.00	0	1	0	0
Carcinops pumilio (Erichson, 1834)	Histeridae	Predator	2.15	1	0	0	0
Saprinus aeneus (Fabricius, 1775)	Histeridae	Predator	3.00	13	23	4	9
Gnathoncus nannetensis (Marseul, 1862)	Histeridae	Predator	3.10	0	0	0	2
Margarinotus carbonarius (Hoffmann, 1803)	Histeridae	Predator	3.60	0	5	0	0
Rugilus erichsonii (Fauvel, 1867)	Staphylinidae	Predator	3.75	8	0	5	0
Margarinotus ventralis (Marseul, 1854)	Histeridae	Predator	4.00	3	2	6	1
Saprinus planiusculus Motschulsky, 1849	Histeridae	Predator	4.45	0	5	0	0
Margarinotus merdarius (Hoffmann, 1803)	Histeridae	Predator	4.50	5	0	6	0

Species	Taxon	Guild	Mean length (mm)	Site 1	Site 2	Site 3	Site 4
Nanoptilium kunzei (Heer, 1841)	Ptiliidae	Necrophagous	0.60	0	0	0	0
Acrotrichis dispar (Matthews, 1865)	Ptiliidae	Necrophagous	0.65	1	0	1	1
Acrotrichis silvatica Rosskothen, 1935	Ptiliidae	Necrophagous	0.80	1	0	1	0
Acrotrichis rugulosa Rosskothen, 1935	Ptiliidae	Necrophagous	0.90	0	0	1	0
Acrotrichis grandicollis (Mannerheim, 1844)	Ptiliidae	Necrophagous	0.95	1	0	0	1
Acrotrichis fratercula (Matthews, 1878)	Ptiliidae	Necrophagous	1.00	0	1	0	0
Carcinops pumilio (Erichson, 1834)	Histeridae	Predator	2.15	1	0	0	0
Saprinus aeneus (Fabricius, 1775)	Histeridae	Predator	3.00	1	1	1	1
Gnathoncus nannetensis (Marseul, 1862)	Histeridae	Predator	3.10	0	0	0	1
Margarinotus carbonarius (Hoffmann, 1803)	Histeridae	Predator	3.60	0	1	0	0
Rugilus erichsonii (Fauvel, 1867)	Staphylinidae	Predator	3.75	1	0	1	0
Margarinotus ventralis (Marseul, 1854)	Histeridae	Predator	4.00	1	1	1	1
Saprinus planiusculus Motschulsky, 1849	Histeridae	Predator	4.45	0	1	0	0
Margarinotus merdarius (Hoffmann, 1803)	Histeridae	Predator	4.50	1	0	1	0

	Descriptors				
Objects	V1	V2	V3	V4	V5
A	$v_{1,1}$	$v_{1,2}$	$v_{1,3}$	$v_{1,4}$	$v_{1,5}$
B	$v_{2,1}$	$v_{2,2}$	$v_{2,3}$	$v_{2,4}$	$v_{2,5}$
C	$v_{3,1}$	$v_{3,2}$	$v_{3,3}$	$v_{3,4}$	$v_{3,5}$
D	$v_{4,1}$	$v_{4,2}$	$v_{4,3}$	$v_{4,4}$	$v_{4,5}$
E	$v_{5,1}$	$v_{5,2}$	$v_{5,3}$	$v_{5,4}$	$v_{5,5}$
F	$v_{6,1}$	$v_{6,2}$	$v_{6,3}$	$v_{6,4}$	$v_{6,5}$
G	$v_{7,1}$	$v_{7,2}$	$v_{7,3}$	$v_{7,4}$	$v_{7,5}$

There are several types of matrices that have

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad V = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad V = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} \quad V = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \end{pmatrix}$$

special properties. Let's first look again to our species x sites matrix. We want to infer whether the site abundances and species occurrences per site are related. In order to do so we use different measures of association (distance).

Distance measures can operate on presence absence matrices and count site overlap or operate on metrically scaled variables. The simplest count measure is the well known **Soerensen index** of species overlap (= Czekanowski index ≈ Jaccard measure):

$$D = \frac{2S_{jk}}{S_j + S_k} \tag{4.1}$$

where S_j and S_k are the species number of site j and k and S_{jk} is the number of shared species.

The general metrically based distance measure that includes the **Manhattan** (= taxi driver) distance (z = 1) and the **Euclidean distance** (z = 2) is the **Minkowski metric**:

$$D = \sqrt[z]{\sum_{i=1}^{n} (x_{i,j} - x_{i,k})^z} \tag{4.2}$$

Other important distance measures are the **Bray - Curtis** (Czekanowski) metric

$$D = \frac{\sum_{i=1}^{n} \left| x_{i,j} - x_{i,k} \right|}{\sum_{i=1}^{n} x_{i,j} + \sum_{i=1}^{n} x_{i,k}} \tag{4.3}$$

This metric can be used on raw and ranked data. In ecology the **index of proportional similarity of Colwell and Futuyma** is often used

$$I_{j,k} = 1 - 0.5 \sum_{1=1}^{n} \left| p_{i,j} - p_{i,k} \right| \tag{4.4}$$

where $p_{i,j}$ and $p_{i,k}$ denote the frequencies $p_i = N_i / N_{total}$ of species i at sites j and k.

Finally, the Pearson and Spearman correlation coefficients provide measures of distance. An example is shown in Fig. 4.1. Data from four sites are compared by the Soerensen measure to give a symmetric 4x4 distance matrix with diagonal elements of one. Such a matrix is the basis for many multivariate statistical techniques.

Association matrices are always square. Square matrices are those with equal numbers of rows and columns. In statistics square matrices are of great importance. A special type of square matrices are **diagonal matrices** where all elements apart from the diagonal are zero

$$V = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$$

If all values of a in a diagonal matrix are 1 we speak of a **unit or identity matrix**

Site 1	Site 2	Site 3	Site 4
0	0	0	0
13	0	4	7
16	0	2	0
0	0	1	0
1	0	0	1
0	1	0	0
1	0	0	0
13	23	4	9
0	0	0	2
0	5	0	0
8	0	5	0
3	2	6	1
0	5	0	0
5	0	6	0

$$D = \frac{2S_{jk}}{S_j + S_k}$$

Soerensen matrix

	Site 1	Site 2	Site 3	Site 4
Site 1	1	0.14	0.43	0.29
Site 2	0.14	1	0.14	0.14
Site 3	0.43	0.14	1	0.21
Site 4	0.29	0.14	0.21	1

$$V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.5}$$

Identity matrices are equivalent to the 1 of ordinary numbers.

The **transpose matrix** A' is a matrix (m*n) obtained from an original matrix A (n*m) where rows and columns are changed

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix} \rightarrow A' = \begin{pmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{pmatrix}$$

A **transpose matrix that is identical to the original is square and symmetric**

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 6 & 10 \\ 3 & 10 & 11 \end{pmatrix} \rightarrow A' = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 6 & 10 \\ 3 & 10 & 11 \end{pmatrix}$$

The Table below shows again important matrix types.

We now start to define matrix operations. First look at a vector, that means a matrix with only one column (or row). Note that a scalar (a simple number) can be viewed as a vector with only one row or a matrix with only one row and one column. Further we can view a matrix as a grouping of single vectors. A vector denotes a point in space. It length is given by the law of Pythagoras.

Skalar	3	Diagonal matrix	$\begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$
Column vector	$\begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$	Identity matrix (unit matrix)	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
Row vector	$(3 \quad 1 \quad 2)$	Symmetric matrix	$\begin{pmatrix} 2 & 4 & 6 \\ 4 & 3 & 7 \\ 6 & 7 & 1 \end{pmatrix}$
Null vector	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	Orthogonal matrix	$\begin{pmatrix} 3 & 3 \\ 3 & -3 \end{pmatrix}$
Square matrix	$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$	Upper triangular matrix	$\begin{pmatrix} 2 & 4 & 6 \\ 0 & 3 & 7 \\ 0 & 0 & 1 \end{pmatrix}$

$$V = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \rightarrow L = \sqrt{1^2 + 2^2 + 3^2} = 14^{1/2} \tag{4.6}$$

We see immediately that any vector can be normalized by dividing its elements through the length. The new vector will have a length of 1.

The first operation to introduce is matrix addition. Assume you have insect counts of 4 species (rows) at 3 sites (columns) during 3 months. This can be formulated in matrix notation

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 5 & 7 \\ 3 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 4 & 0 \\ 1 & 2 & 0 \\ 6 & 9 & 1 \\ 1 & 1 & 4 \end{pmatrix} + \begin{pmatrix} 2 & 8 & 1 \\ 7 & 5 & 5 \\ 0 & 0 & 1 \\ 5 & 6 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 14 & 4 \\ 10 & 9 & 9 \\ 9 & 14 & 9 \\ 9 & 8 & 5 \end{pmatrix}$$

The total catch per site and species is the sum of the respective matrix elements. We see that matrix addition is only defined for matrices with identical numbers of rows and columns.

Matrix addition immediately leads to the first type of multiplication, the **S-product**. We have

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 5 & 7 \\ 3 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 5 & 7 \\ 3 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 5 & 7 \\ 3 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 6 & 9 \\ 6 & 6 & 12 \\ 9 & 15 & 21 \\ 9 & 3 & 0 \end{pmatrix} = 3 \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 5 & 7 \\ 3 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 \cdot 1 & 3 \cdot 2 & 3 \cdot 3 \\ 3 \cdot 2 & 3 \cdot 2 & 3 \cdot 4 \\ 3 \cdot 3 & 3 \cdot 5 & 3 \cdot 7 \\ 3 \cdot 3 & 3 \cdot 1 & 3 \cdot 0 \end{pmatrix}$$

In other words multiplying a matrix with a scalar means multiplying each matrix element with that scalar. For matrix addition and the S-product hold the commutative, the distributive and the associative laws

$$A - B = -B + A = -1B + A$$
$$A + B = B + A$$
$$A + (B + C) = (A + B) + C$$
$$\lambda A = A\lambda$$
$$\lambda(A + B) = \lambda A + \lambda B$$
$$A(\lambda + \kappa) = A\lambda + A\kappa$$

The next example introduces the **multiplication of matrices**. Assume you have production data (in tons) of winter wheat (15 t) , summer wheat (20 t), and barley (30 t). In the next year weather condition reduced the winter wheat production by 20%, the summer wheat production by 10% and the barley production by 30%. How many tons do you get the next year? Of course (15*0.8 + 20* 0.9 + 30 * 0.7) t = 51 t. In matrix notion

$$P = \begin{pmatrix} 15 & 20 & 30 \end{pmatrix} \bullet \begin{pmatrix} 0.8 \\ 0.9 \\ 0.7 \end{pmatrix} = 15 * 0.8 + 20 * 0.9 + 30 * 0.7 = 51$$

this type of multiplication is called a **scalar (or dot) product** because it results in a number (a scalar in matrix terminology). In general

$$A \bullet B = \begin{pmatrix} a_1 & \dots & a_n \end{pmatrix} \bullet \begin{pmatrix} b_1 \\ \dots \\ b_n \end{pmatrix} = \sum_{i=1}^{n} a_i b_i = \text{scalar} \tag{4.7}$$

We can easily extend this example to deal with matrices. We add another year and ask how many cereals we get if the second year is good and gives 10 % more of winter wheat, 20 % more of summer wheat and 25 % more of barley. For both yes we start counting with the original data and get a vector with one row that is the result of a two step process. First we compute the first year value and then the second year value and combine both scalars in a new row vector with two columns denoting both years.

$$P = \begin{pmatrix} 15 & 20 & 30 \end{pmatrix} \bullet \begin{pmatrix} 0.8 & 1.1 \\ 0.9 & 1.2 \\ 0.7 & 1.25 \end{pmatrix} = \begin{pmatrix} 15*0.8+20*0.9+30*0.7 & 15*1.1+20*1.2+30*1.25 \end{pmatrix} = \begin{pmatrix} 51 & 78 \end{pmatrix}$$

Now we consider three sites with different harvest. Recall that species are in columns, sites in rows. We get an intuitional definition of the **scalar or dot or inner product** of matrices. The final values give total production at three sites and two years. The result is not a scalar but a matrix.

$$P=\begin{pmatrix}15 & 20 & 30\\ 10 & 15 & 20\\ 5 & 10 & 15\end{pmatrix}\bullet\begin{pmatrix}0.8 & 1.1\\ 0.9 & 1.2\\ 0.7 & 1.25\end{pmatrix}=\begin{pmatrix}15*0.8+20*0.9+30*0.7 & 15*1.1+20*1.2+30*1.25\\ 10*0.8+15*0.9+20*0.7 & 10*1.1+15*1.2+20*1.25\\ 5*0.8+10*0.9+15*0.7 & 5*1.1+10*1.20+15*1.25\end{pmatrix}=\begin{pmatrix}51 & 78\\ 35.5 & 54\\ 23.5 & 36.25\end{pmatrix}$$

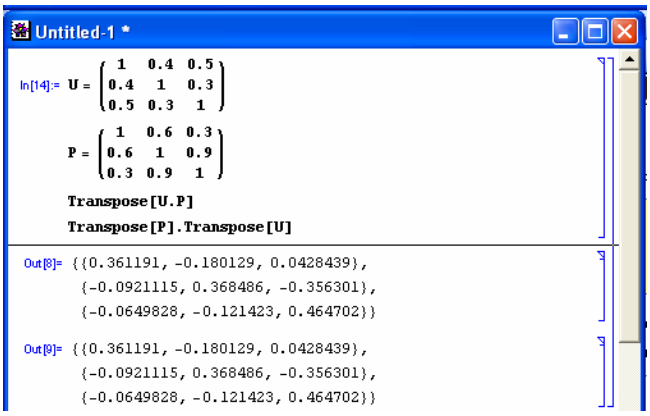
In general we get

$$A\bullet B=\begin{pmatrix}a_{11} & \dots & a_{1m}\\ \dots & \dots & \dots\\ a_{n1} & \dots & a_{nm}\end{pmatrix}\bullet\begin{pmatrix}b_{11} & \dots & b_{1k}\\ \dots & \dots & \dots\\ a_{m1} & \dots & a_{mk}\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^{m}a_{1i}b_{i1} & \dots & \sum_{i=1}^{m}a_{1i}b_{ik}\\ \dots & \dots & \dots\\ \sum_{i=1}^{m}a_{ni}b_{i1} & \dots & \sum_{i=1}^{m}a_{ni}b_{ik}\end{pmatrix}=\begin{pmatrix}A_1B_1 & \dots & A_1B_k\\ \dots & \dots & \dots\\ A_mB_1 & \dots & A_mB_k\end{pmatrix} \tag{4.8}$$

Hence the dot product can be viewed as a step by step procedure where each row and each column are subjects to single dot products of two vectors. Further, a dot product is only defined if the number of columns of A is equal to the number of rows in B. The new matrix has the same number of columns than in B and the same number of rows than in A. Further, in most cases **A• B ≠ B• A**. Next, B•B only exists if B is a square matrix.

The above definition implies that if the matrix B is a simple scalar the dot product simplifies to

$$A\bullet B=\begin{pmatrix}a_{11} & \dots & a_{1m}\\ \dots & \dots & \dots\\ a_{n1} & \dots & a_{nm}\end{pmatrix}\bullet c=\begin{pmatrix}a_{11}c & \dots & a_{1m}c\\ \dots & \dots & \dots\\ a_{n1}c & \dots & a_{nm}c\end{pmatrix} \tag{4.9}$$



As for scalars we have to look whether the dot product is distributive, associative and commutative. The dot product is generally **not commutative but associative and distributive**.

$$A\bullet B\neq B\bullet A$$

$$(A\bullet B)\bullet C=A\bullet(B\bullet C)=A\bullet B\bullet C$$

$$(A+B)\bullet C=A\bullet C+B\bullet C$$

In the case of a symmetric matrix an important relation exists

$$(A\bullet B)'=B'\bullet A' \tag{4.10}$$

It should be noted that the **outer or cross product** is not used in matrix algebra.

The **trace of a symmetric matrix** is defined as the sum of all diagonal elements of this matrix.

$$Tr(A_{n,n})=\sum_{i=1}^{n}a_{ii} \tag{4.11}$$

An important transformation of square matrices A_{nn} is the **determinant det A or |A|**. The determinant is

a scalar that enables to transform matrices A into new ones B. The associated function B = f(A) has to confirm to three basic rules:

1. B is a linear transformation of A that means any change in A results in a linear change in B.
2. Any change in the ordering of rows or columns in A should cause a change of sign in f(A).
3. f(A) is determined by a scalar, called the **norm or value of A** in such a way that the norm of the identity matrix is 1. Hence f(I) = 1

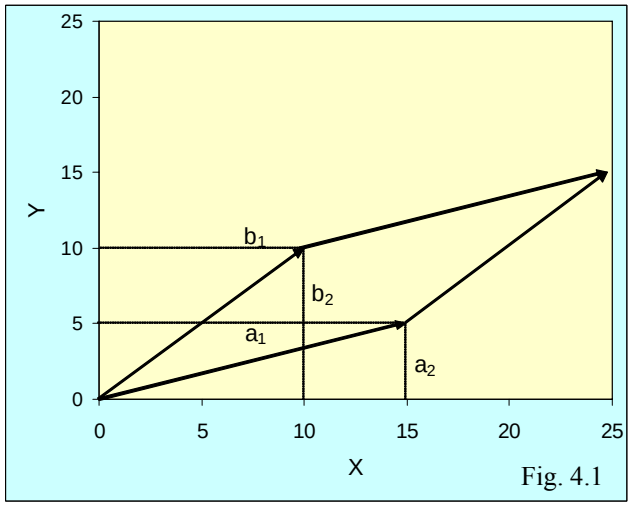


Fig. 4.1

The value of a determinant is calculated as the sum of all possible products containing one, and only one, element from each row and each column. These products receive a sign (+ or -) according to a predefined rule. The simplest determinant is

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \qquad (4.12)$$

Determinants have a simple graphical interpretation (Fig. 4.1). The area under the parallelogram spanned by the vectors $\{a_1,a_2\}$ and $\{b_1,b_2\}$ is identical to the determinant of the matrix A.

Determinants of matrices of higher order become increasingly time-consuming to be calculated because we get n! permutations of single products. However with today's Math programs this is an easy task. Above is the respective norm of an example matrix calculated by *Mathematica*.

Why are determinants important? Determinants are used to solve systems of linear equations. They allow to infer some properties of a given matrix. Particularly holds

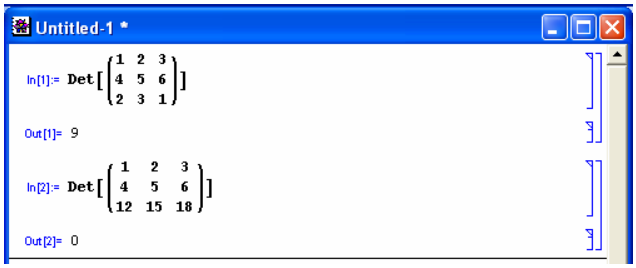
1. If a row or a column of a matrix A is zero det (A) = 0.

2. If a row or a column of a matrix A is linearly dependent on another row or column (is proportional to another row or column) then det(A) = 0.

3. If a row or a column of A is multiplied by a scalar k to result in another matrix B then det(B) = k det(A).

$$\det\begin{pmatrix} 1 & 5 & 0 \\ 2 & 6 & 0 \\ 3 & 7 & 0 \end{pmatrix} = 0$$

$$\det\begin{pmatrix} 1 & 3\cdot 5 & 2 \\ 2 & 3\cdot 6 & 3 \\ 3 & 3\cdot 7 & 4 \end{pmatrix} = 3\det\begin{pmatrix} 1 & 5 & 2 \\ 2 & 6 & 3 \\ 3 & 7 & 4 \end{pmatrix}$$



If the determinant of a matrix is zero the matrix is called **singular**. Look for instance at the next matrices

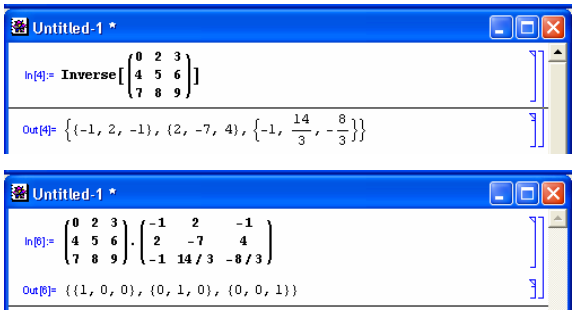
$$P=\begin{pmatrix}1&2&3\\4&5&6\\2&3&1\end{pmatrix};Q=\begin{pmatrix}1&2&3\\4&5&6\\12&15&18\end{pmatrix}$$

In P row 2 can be obtained from row one by the transformation $r_2 = r_1 + 3$. However, this is not a linear transformation. In Q, in turn, row 3 is three times row 2 (row 3 is proportional to row 2). This is a **linear transformation** and the respective determinant is zero.

Determinants of upper or lower triangular matrices are easy to compute. The determinant of

$$\det\begin{pmatrix}1&2&3\\0&-2&-4\\0&0&2\end{pmatrix}=\prod_{i=1}^{3}a_{ii}=1\cdot-2\cdot2=-4$$

Triangular distance matrices are important in multivariate statistics so is the method for computing the determinant.



For ordinary numbers (scalars) the product of a number with its inverse gives always 1. Hence $a \bullet a^{-1} = 1$. The extension of this principle to matrices looks at follows

$$A \bullet A^{-1} = I \qquad (4.13)$$

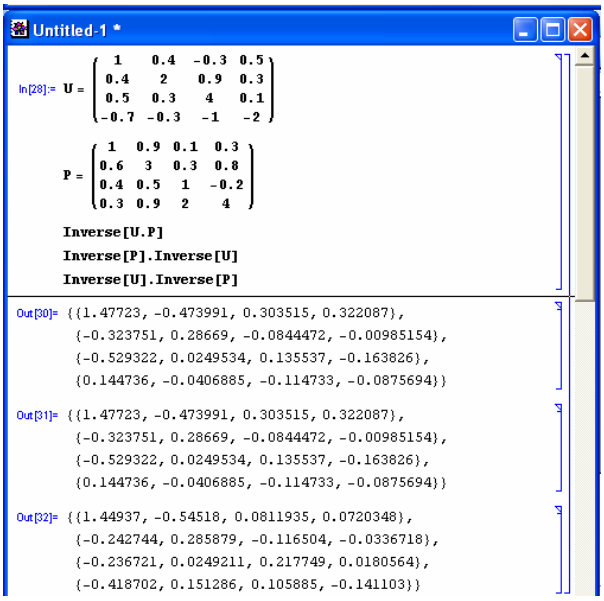
To solve this equation we need the inverse of a matrix A^{-1}. We see immediately that **this operation is only defined for square matrices** (why?).

The inverse of a matrix is closely related to its determinant.

$$A \bullet A^{-1} = \frac{1}{|A|}\begin{pmatrix}|A|&0&\dots&0&0\\0&|A|&&&0\\\dots&&|A|&&\\0&&&|A|&0\\0&0&\dots&0&|A|\end{pmatrix}=\begin{pmatrix}1&0&\dots&0&0\\0&1&&&0\\\dots&&1&&\dots\\0&&&1&0\\0&0&\dots&0&1\end{pmatrix}=I \qquad (4.14)$$

The equation tells that an inverse only exists if a matrix is not singular, that is if it has a non-zero determinant.

Computing the inverse of a matrix is quite tricky. A formal method provides the **Gauß algorithm** that is implemented in standard matrix software. *Mathematica* calculates inverse matrices with the inverse command. We can then check whether the calculation conforms to our definition of the inverse. Indeed the dot product of both matrices equals the identity matrix.



Inverse matrices have several important properties.

1. $(B^{-1})^{-1} = B$
2. $B^{-1} \bullet B = B \bullet B^{-1} = I$

3. $|B^{-1}| = 1 / |B|$

4. $(A \bullet B)^{-1} = B^{-1} \bullet A^{-1} \neq A^{-1} \bullet B^{-1}$

How to apply matrices? A first well known examples deals with systems of linear equations. Take

$$\left.\begin{matrix} a_{11}x + a_{12}y = b_1 \\ a_{21}x + a_{22}y = b_2 \end{matrix}\right\} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \bullet \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \tag{4.14}$$

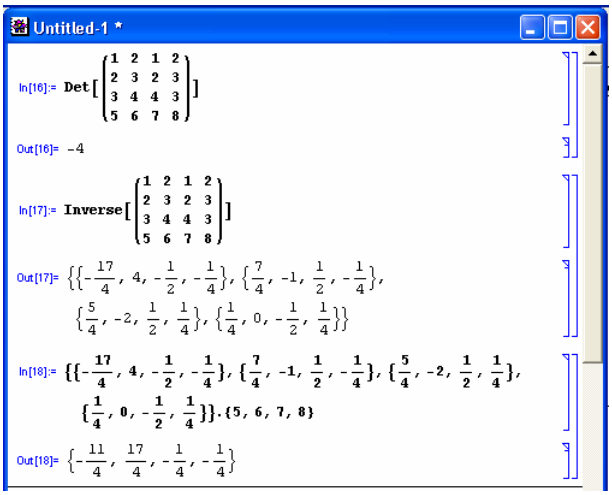
Solving the system conventionally gives

$$X = \frac{b_1 a_{22} - b_2 a_{12}}{a_{11}a_{22} - a_{12}a_{21}} = \frac{\det A_1}{\det A}; Y = \frac{b_2 a_{11} - b_1 a_{21}}{a_{11}a_{22} - a_{12}a_{21}} = \frac{\det A_2}{\det A}$$

$$A_1 = \begin{pmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{pmatrix}; A_2 = \begin{pmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{pmatrix} \tag{4.15}$$

We see why the determinant was defined in such a curious way. It has to match the requirements for solving linear algebraic equations. The general solutions for systems with n equations and n unknown variables X_i is:

$$X_i = \frac{|A_i|}{|A|} \tag{4.16}$$



However, we can make things easier and solve linear systems without using determinants. Assume you have a system of four linear equations

$$\begin{aligned} a_1 + 2a_2 + a_3 + 2a_4 &= 5 \\ 2a_1 + 3a_2 + 2a_3 + 3a_4 &= 6 \\ 3a_1 + 4a_2 + 4a_3 + 3a_4 &= 7 \\ 5a_1 + 6a_2 + 7a_3 + 8a_4 &= 8 \end{aligned}$$

This system can be written in matrix notation

$$\begin{pmatrix} 1a_1 & 2a_2 & 1a_3 & 1a_4 \\ 2a_1 & 3a_2 & 2a_3 & 3a_4 \\ 3a_1 & 4a_2 & 4a_3 & 3a_4 \\ 5a_1 & 6a_2 & 7a_3 & 8a_4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 3 & 2 & 3 \\ 3 & 4 & 4 & 3 \\ 5 & 6 & 7 & 8 \end{pmatrix} \bullet \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \\ 7 \\ 8 \end{pmatrix}$$

This equation has the formal structure of A•X=B. Because $A^{-1} \bullet A = A \bullet A^{-1} = I$ we can multiply both sides with the inverse of A. We get $X = A^{-1} \bullet B$ and the solution for the coefficients a_i.

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 3 & 2 & 3 \\ 3 & 4 & 4 & 3 \\ 5 & 6 & 7 & 8 \end{pmatrix}^{-1} \bullet \begin{pmatrix} 5 \\ 6 \\ 7 \\ 8 \end{pmatrix}$$

The respective *Mathematica* solution looks as follows. First we check for singularity. The determinant of A is -4. Therefore our system should have a solution. Then we compute the inverse and in a last step we multiply with the vector B. We get $a_1 = -11/4$, $a_2 = 17/4$, $a_3 = -1/4$, and $a_4 = -1/4$. The matrix notation makes solving a linear system an easy task.

However things are not as easy. First, the inverse matrix has to exist as in our example. If it is singular no inverse exists and the respective determinant is zero. For non-square matrices we need another feature of matrices, the **rank**. Take the next examples of matrices

$$A=\begin{pmatrix}1&4&8&3\\2&5&10&5\\3&6&12&7\end{pmatrix};B=\begin{pmatrix}1&5&8&3\\2&3&10&5\\3&1&12&7\end{pmatrix};C=\begin{pmatrix}1&4&8&3\\2&5&10&5\\0&0&0&0\end{pmatrix}$$

In A column C is two times column B and both square submatrices with three rows/columns (the **order of the matrix**) have zero determinants. Except of two the 2x2 submatrices have determinants > 0. The rank is therefore 2. In B both 3x3 submatrices have determinants > 0 and the rank is 3. In C the 2x2 submatrices have determinants > 0 and the rank is 2. Hence the rank of a matrix equals the order of the largest submatrix whose determinant > 0.

We further need to know what an **augmented matrix** is. An augmented matrix is a simple combination of two matrices

$$A=\begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{pmatrix};B=\begin{pmatrix}b_{11}&b_{12}\\b_{21}&b_{22}\\b_{31}&b_{32}\end{pmatrix}\rightarrow A:B=\begin{pmatrix}a_{11}&a_{12}&a_{13}&b_{11}&b_{12}\\a_{21}&a_{22}&a_{23}&b_{21}&b_{22}\\a_{31}&a_{32}&a_{33}&b_{31}&b_{32}\end{pmatrix}$$

The question whether a linear system has solutions is best explained from simple examples.

$$\left.\begin{array}{l}2x_1+6x_2+5x_3+9x_4=10\\2x_1+5x_2+6x_3+7x_4=12\\4x_1+4x_2+7x_3+6x_4=14\\5x_1+3x_2+8x_3+5x_4=16\end{array}\right\}\begin{pmatrix}2&6&5&9\\2&5&6&7\\4&4&7&6\\5&3&8&5\end{pmatrix}\bullet\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}10\\12\\14\\16\end{pmatrix}$$

$$\left.\begin{array}{l}2x_1+3x_2+4x_3+5x_4=10\\4x_1+6x_2+8x_3+10x_4=20\\4x_1+5x_2+6x_3+7x_4=14\\5x_1+6x_2+7x_3+8x_4=16\end{array}\right\}\begin{pmatrix}2&3&4&5\\4&6&8&10\\4&5&6&7\\5&6&7&8\end{pmatrix}\bullet\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}10\\20\\14\\16\end{pmatrix}$$

$$\left.\begin{array}{l}2x_1+3x_2+4x_3+5x_4=10\\4x_1+6x_2+8x_3+10x_4=12\\4x_1+5x_2+6x_3+7x_4=14\\5x_1+6x_2+7x_3+8x_4=16\end{array}\right\}\begin{pmatrix}2&3&4&5\\4&6&8&10\\4&5&6&7\\5&6&7&8\end{pmatrix}\bullet\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}10\\12\\14\\16\end{pmatrix}$$

$$\left.\begin{array}{l}2x_1+3x_2+6x_3+9x_4=10\\2x_1+4x_2+5x_3+6x_4=12\\4x_1+5x_2+4x_3+7x_4=14\end{array}\right\}\begin{pmatrix}2&3&6&9\\2&4&5&6\\4&5&4&7\end{pmatrix}\bullet\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}10\\12\\14\end{pmatrix}$$

$$\left.\begin{array}{l}2x_1+3x_2+4x_3+5x_4=10\\4x_1+6x_2+8x_3+10x_4=12\\4x_1+5x_2+6x_3+7x_4=14\\5x_1+6x_2+7x_3+8x_4=16\\10x_1+12x_2+14x_3+16x_4=16\end{array}\right\}\begin{pmatrix}2&3&4&5\\4&6&8&10\\4&5&6&7\\5&6&7&8\\10&12&14&16\end{pmatrix}\bullet\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}10\\12\\14\\16\\16\end{pmatrix}$$

$$\left.\begin{array}{l}2x_1+3x_2+4x_3+5x_4=10\\4x_1+6x_2+8x_3+10x_4=12\\4x_1+5x_2+6x_3+7x_4=14\\5x_1+6x_2+7x_3+8x_4=16\\10x_1+12x_2+14x_3+16x_4=32\end{array}\right\}\begin{pmatrix}2&3&4&5\\4&6&8&10\\4&5&6&7\\5&6&7&8\\10&12&14&16\end{pmatrix}\bullet\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}10\\12\\14\\16\\32\end{pmatrix}$$

In the first example we have four unknown variables and four equations. The ranks of A and the aug-

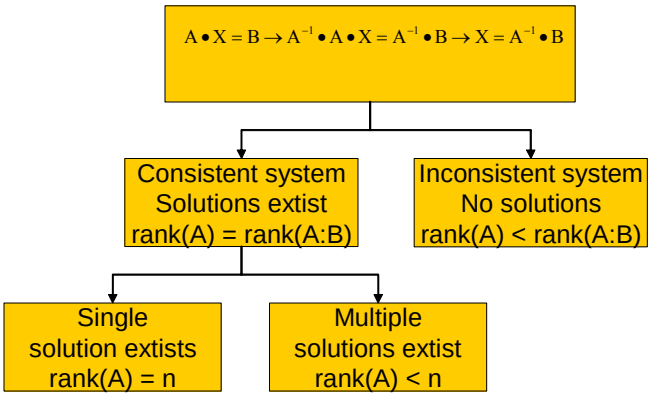


Fig. 4.2: Modified from A. K. Kaw (2002): Introduction to Matrix Algebra.

mented A:B are 4. The system has a single solution. In the second case the second equation is simply two times the first. Therefore we have only three independent equation but four variables. The ranks of A and A:B are three and therefore less then n, the number of variables. We have multiple solutions. In the third example the rank of A is three (rows one and two are proportional) but the rank of the augmented matrix is four. In such a case no solution exists. In the fourth example we have only three equations but four variables. The ranks of A and A:B are three and less than n. An infinite number of solutions exist. In the last two examples we have more equations than variables. In the first of these examples the last equations is inconsistent to the previous. The rank of the augmented matrix is higher than the rank of A. In turn, in the last example the last equation is simple two times the previous. It does not contain additional information. The rank of A equals the rank of A:B.

In general we can assess whether a linear system has solutions from the ranks of A and A:B and the number of variables n (Fig. 4.2). A system has no solution (is inconsistent) if rank(A) < rank (A:B). If Rank (A) = n a single solution exists, if rank(A) < n infinite solutions exists. It is easy to show that a system cannot have a finite number of solutions. Assume we have two solutions. We multiply both equations with with the scalars r and 1-r.

$$A \bullet X = B \rightarrow A \bullet rX = rB$$
$$A \bullet Y = B \rightarrow A \bullet (1-r)Y = (1-r)B$$
$$\downarrow$$
$$A \bullet rX + A \bullet (1-r)Y = rB + (1-r)B$$
$$\downarrow$$
$$A \bullet [rX + (1-r)Y] = B$$

We got a new equation of the structure A•Z = B. Hence [rX+(1-r)Y] = Z must be a solution of our initial system. Because r can be any scalar their must be infinite solutions of A Important is that a solution as in Fig. 4.3 is only possible if we can compute the inverse of A•Z = B. The inverse is only defined for square matrices. Hence, the number of equations must be identical to the number of variables.

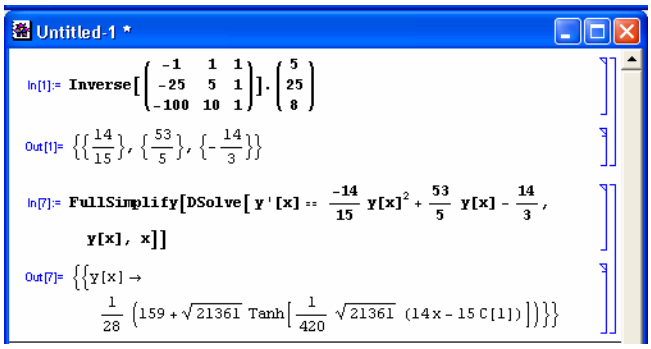
Matrix algebra can also be used to solve algebraic equations. For instance logistic population growth follows a second order algebraic (quadratic) model

$$\frac{dN}{dt} = -aN^2 + bN + c$$

At three populations sizes N (1, 5, 10) we got the following rates of increase dN/dt: 5, 25, 8 individuals generation^{-1}. Now our model looks as follows

$$\left.\begin{aligned} -a+b+c&=5\\ -a5^2+b5+c&=25\\ -a10^2+b10+c&=8 \end{aligned}\right\}\begin{pmatrix} -1 & 1 & 1\\ -25 & 5 & 1\\ -100 & 10 & 1 \end{pmatrix}\begin{pmatrix} a\\ b\\ c \end{pmatrix}=\begin{pmatrix} 5\\ 25\\ 8 \end{pmatrix}$$

rank(A) = rank(A:B) = n. Hence a single solution exists. This is shown beside. We se again how simple it is to solve such a system using math programs. The logistic growth model is



$$\frac{dN}{dt}=-\frac{14}{15}N^2+\frac{53}{5}N-\frac{14}{3}$$

and the respective solution of N(t) is obtained from the solution of the respective differential equation. The model is a tangens hyperbolicus function (Fig. 4.3) equivalent to the logistic growth equation. We get the maximum abundance K from the Fig. K = 10.89.

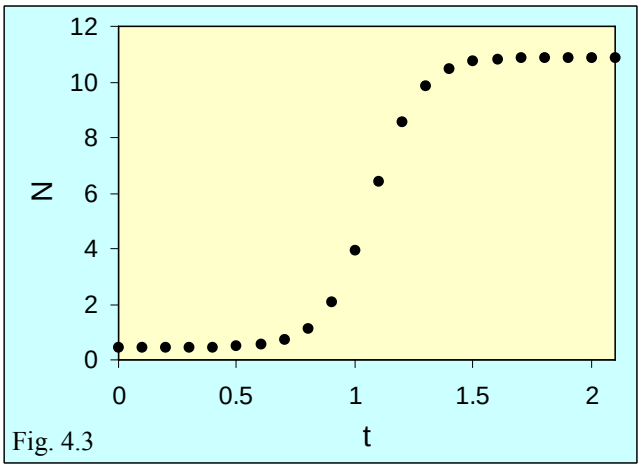


Fig. 4.3

Next we have to deal with **eigenvalues** and **eigenvectors**. Assume we have a matrix A in two dimensional space defining data points in an x,y system (Fig. 4.4). We can now define new orthogonal axes u_1 and u_2 that minimize the distances δ_{i1} and δ_{i2} to our data points. These new axes (vectors) are called **principal axes** of the matrix A. The length of the vectors u_1 and u_2 are called the eigenvalues λ_1 and λ_2. Fig. 4.5 and 4.6 also show how to interpret eigenvalues and eigenvectors. Assume our data points to be enclosed by an ellipse. The longer diagonal marks the first eigenvector. The longer this vector is the closer scatter the data points along this vector. Hence λ is a measure of how well the principal axis describes our data matrix.

The X,Y matrix in Figure 4.4 is not symmetrical. But using the respective distance matrix of X and Y instead we get a symmetrical 2x2 matrix. Eigenvalues λ_i and eigenvectors u_i are now connected to the distance matrix A in such a way that the dot product of $A \bullet u_i$ gives u_i multiplied with λ. Hence

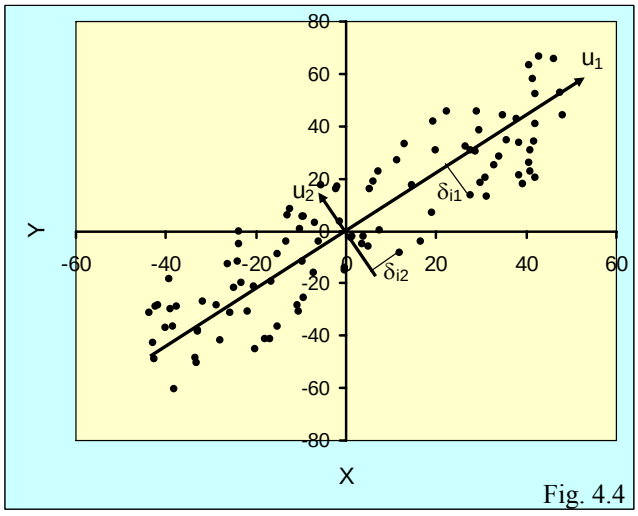


Fig. 4.4

$$A \bullet u_i = \lambda_i u_i \tag{4.17}$$

Eq, 5.26 can be transformed to

$$A \bullet u_i = \lambda_i u_i \rightarrow A \bullet u_i - \lambda_i u_i = 0 \rightarrow [A - \lambda_i I] \bullet u_i = 0 \tag{4.18}$$

The eigenvector u is the principle axis vector. Eq

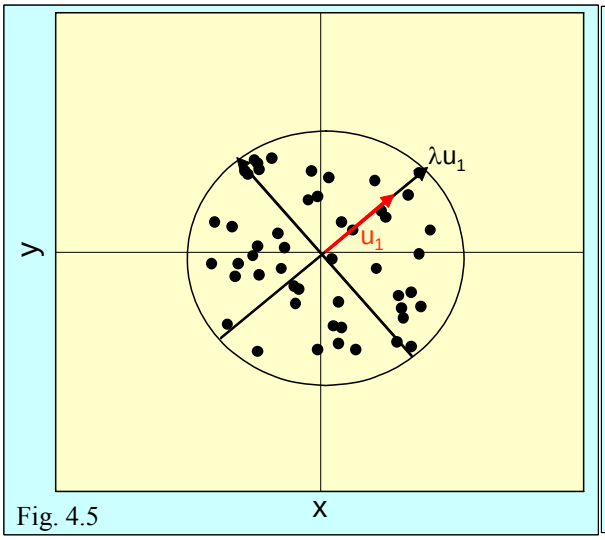


Fig. 4.5

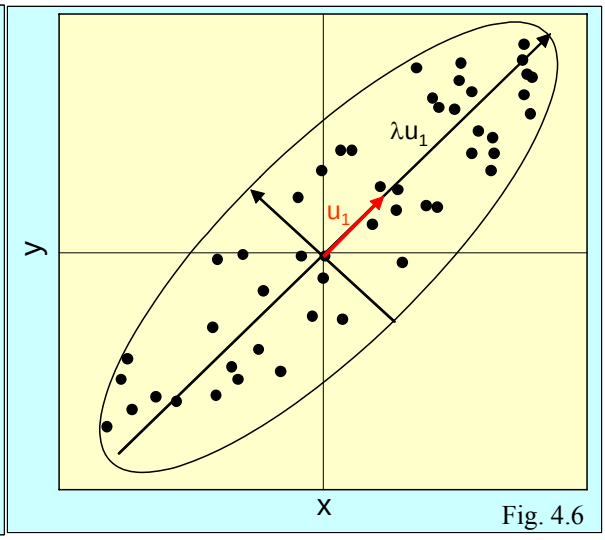


Fig. 4.6

4.18 might be interpreted as follows. How long do I have to stretch the vector u to get a new matrix a•u from A. Eq. 4.17 is of course solved by a null vector u. The equation is also zero if A-λ_iI is a null vector. This is equivalent to det (A-λ_iI) = 0. This property gives a solution for λ_i. In other words we are looking for such a λ that makes (A-λI) to a vector that is **orthogonal** to u. An example. For the matrix A = {{2,1},{3,4} we need

$$\left|\begin{pmatrix}2 & 1\\ 3 & 4\end{pmatrix}-\lambda\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right|=0\rightarrow\begin{vmatrix}2-\lambda & 1\\ 3 & 4-\lambda\end{vmatrix}=0\rightarrow$$

$$(2-\lambda)(4-\lambda)=3\rightarrow\lambda_1=1;\lambda_1=5$$

W get the associated eigenvectors from both λ values by solving

$$\begin{pmatrix}2 & 1\\ 3 & 4\end{pmatrix}-\lambda\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\bullet\begin{pmatrix}u_1\\ u_2\end{pmatrix}=0\rightarrow\begin{pmatrix}2-\lambda & 1\\ 3 & 4-\lambda\end{pmatrix}\bullet\begin{pmatrix}u_1\\ u_2\end{pmatrix}=0\rightarrow\left.\begin{matrix}(2-\lambda)u_1+u_2=0\\ 3u_1+(4-\lambda)u_2=0\end{matrix}\right\}\rightarrow u=\begin{pmatrix}-1 & 1\\ 1 & 3\end{pmatrix}$$

We check

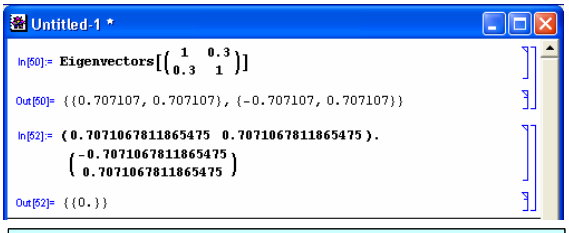
$$\begin{pmatrix}2 & 1\\ 3 & 4\end{pmatrix}\bullet\begin{pmatrix}-1\\ 1\end{pmatrix}=\begin{pmatrix}-1\\ 1\end{pmatrix}=1\begin{pmatrix}-1\\ 1\end{pmatrix}$$

$$\begin{pmatrix}2 & 1\\ 3 & 4\end{pmatrix}\bullet\begin{pmatrix}1\\ 3\end{pmatrix}=\begin{pmatrix}5\\ 15\end{pmatrix}=5\begin{pmatrix}1\\ 3\end{pmatrix}$$

These vectors are not orthogonal. $u_1 \bullet u_2 \neq 0$. However, one special type of matrices gives **orthogonal eigenvectors**. These are symmetrical matrices. Symmetrical matrices are often used in biology. These are for instance most association and all dispersion matrices where $D_{ij} = D_{ji}$. The diagonal elements are 1. Look at the next *Mathematica* solution. A symmetrical association matrix has orthogonal eigenvectors. We check for this by multiplying both. The result is zero. The Fig. 4.7 shows both eigenvectors. They are indeed orthogonal. Hence symmetrical matrices have orthogonal eigenvectors. Further, it holds

$$A \bullet U = U \bullet \Lambda$$

that means the matrix of eigenvectors U transoses a matrix



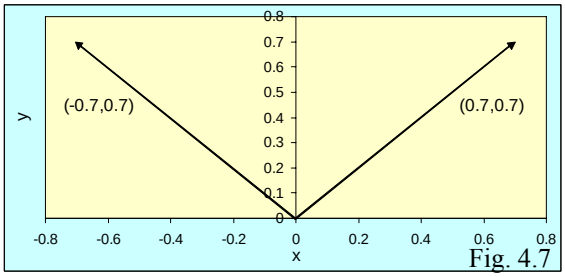


Fig. 4.7

A into the matrix of eigenvalues Λ.

Another application of eigenvectors and eigenvalues. For Markov chains it is often necessary to calculate the power of a matrix A^n. We got the diagonal matrix L from A•U=Λ•U=U•Λ. Look at the example below. We see that A•U = U•Λ with U being the matrix of eigenvectors and Λ the diagonal matrix of eigenvalues. We get a simple way to compute the power of a matrix

$$\begin{aligned}
&A \bullet U = U \bullet \Lambda \rightarrow A = U \bullet \Lambda \bullet U^{-1} \rightarrow \\
&A^n = (U \bullet \Lambda \bullet U^{-1})(U \bullet \Lambda \bullet U^{-1})(U \bullet \Lambda \bullet U^{-1}) \ldots \rightarrow \\
&A^n = U \bullet (\Lambda \bullet U^{-1} \bullet U \ldots) \bullet U^{-1} = U \bullet \Lambda^n \bullet U^{-1} = U \bullet \lambda_i^n \bullet I^n \bullet U^{-1} = U \bullet \left(\lambda_i^n\right) \bullet U^{-1}
\end{aligned} \tag{4.20}$$

We need the matrix of eigenvectors, its inverse and the vector of eigenvalues where each element λ_i is raised to power n.

A next important property is that the determinant of a square matrix is identical to the product of its eigenvalues.

$$\left|P_{nn}\right| = \prod_{i=1}^{n} \lambda_i$$

Hence a determinant is zero if at least one of its eigenvalues is zero.

5. Panta rhei

In the last lectures we got the necessary knowledge abort calculus, series and functions to deal with more complicated but also more realistic problems. Consider a taxonomic group of animals or plants. Denote the probability that during evolution a species becomes extinct with e. Let the probability of a speciation event in that group be c. Empirically, we found that the general fitness of that group decreased with time. The number of speciation events slightly decreased, the number of extinctions increased with time. Assume that both relations can be described by a linear dependence, it est that speciation and extinction rates are reciprocal and direct proportional to time. We can now try to describe the change of the number of species in time. This change in time is the difference between total speciation and extinction events. Both are proportional to the absolute number of species and to time. Such a process might be described by a difference equation, or if we consider very small time intervals by a so-called **differential equation**.

$$\Delta N = N_{t+1} - N_t = (ctN_t - etN_t)\Delta t$$

$$\frac{dN}{dt} = ctN - etN$$

Is it possible to compute the number of species at any time t? Obviously, we have to solve the above equation for N, we have to solve the differential equation. By a simple rearrangement we get

$$\frac{dN}{N} = (c-e)tdt$$

We integrate both sides as already shown in the last chapter and get

$$\ln(N) + c = \frac{c-e}{2}t^2 + c \;\rightarrow\; N = Ce^{\frac{c-e}{2}t^2} = N_0 e^{\frac{c-e}{2}t^2}$$

As in chapter 9 we obtained N_0 by setting t = 0. Two examples of this function show Figs. 5.1 and 5.2. We notice an astonishing effect. Even a small difference in extinction and speciation rate would drive a taxon either to very large species numbers or to rapid extinction. From this we conclude that in nature both rates must be quite tightly adjusted by additional mechanisms not included in our simple model.

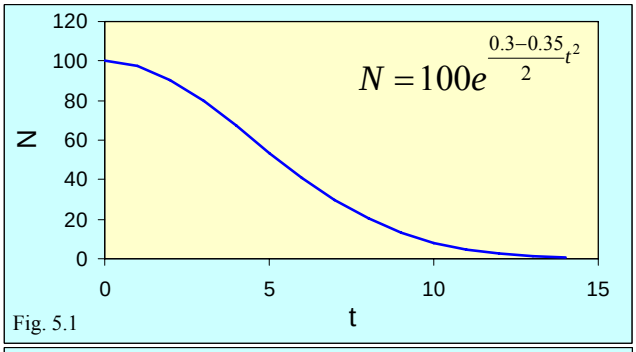


Fig. 5.1

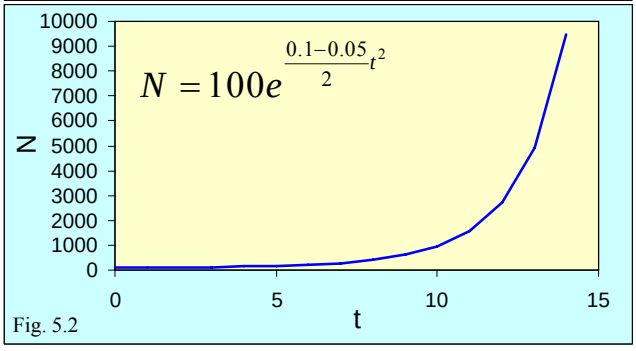


Fig. 5.2

We solved a differential equation. It is a simple case of a **first order linear differential equation**. First order because the derivative was of the first order and linear because no higher exponents than 1 occurred for N.

First order linear differential equations are very important in biology and we have to deal with them in more detail. The most general expression for them is

$$\frac{dy}{dx} + f(x)y = g(x) \tag{5.1}$$

where f and g are functions depending on x. The equation is called **homogeneous** if g(x) is 0 for all x. Let's first consider only this homogeneous case. We get

$$\frac{dy}{y} = -f(x)dx \tag{5.2}$$

We integrate as before and have the general solution

$$y = Ce^{\int -f(x)dx} \tag{5.3}$$

The constant C can often be determined by setting x to an appropriate value (often 0) and solving for C.

If our equation is inhomogeneous the situation is more complicated. We start with eq. 5.3 and use a small trick, known as **variation of constants**. Instead of C we introduce a function c(x) and get

$$y = c(x)e^{\int -f(x)dx} \rightarrow$$
$$y' = c(x)'e^{\int -f(x)dx} + c(x)f'(x)e^{\int -f(x)dx}$$

We introduce this into our eq. 5.1 and get

$$c'(x)e^{\int -f(x)dx} + c(x)\left(e^{\int -f(x)dx}\right)' + f(x)c(x)e^{\int -f(x)dx} = g(x)$$

This reduces to

$$c'(x)e^{\int -f(x)dx} - c(x)f(x)e^{\int -f(x)dx} + f(x)c(x)e^{\int -f(x)dx} = g(x)$$
$$c'(x)e^{\int -f(x)dx} = g(x)$$
$$c'(x) = g(x)e^{\int f(x)dx}$$
$$c(x) = \int g(x)e^{\int f(x)dx}dx + C$$

We introduce this into equation 5.3 and get

$$y = e^{-\int f(x)dx}\int e^{\int f(x)dx}g(x)dx + Ce^{-\int f(x)dx} \tag{5.4}$$

This is the most general solution of a first order linear differential equation. It looks quite complicated but is in most cases easy to apply if the functions f(x) and g(x) are not too complicated.

Let's consider an important example. The concentration of certain hormones and of many medicaments in the bloodstream can be modelled as the sum of production or induction rate and of degradation rate. Let's assume that a medicament is provided at a more or less constant rate g. At the same time, it is metabolized proportional to its concentration. The process can be described by the following equation

$$\frac{dF}{dt} = g - fF \tag{5.5}$$

The change of medicament concentration F in time is a simple difference between provision rate g and

usage rate fF. This is an inhomogeneous first order differential equation. To compute the medicament concentration at time t we use the general solution. We need

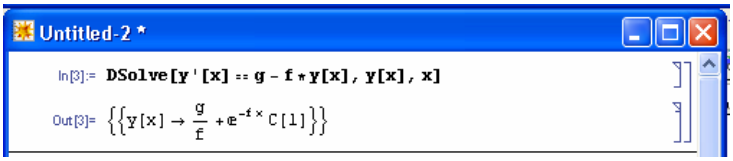
$$-\int f(t)dt = -\int f dt = -(ft + C_1)$$

$$\int e^{\int f(t)dt} g(t)dt = \int (e^{ft+C_1}) g\, dt = g\,\frac{1}{f} e^{ft+C_1} + C_2$$

We introduce these results into our general solution and get

$$F = Ce^{-ft+C_1} + e^{-(ft+C_1)}\left(\frac{g}{f}e^{ft+C_1}\right) = Ke^{-ft} + \frac{g}{f}$$

The constants C contain all the resulting integration constants. The last remaining constant K can be determined by setting t to zero. K is therefore K = F(0)-c/g. We can therefore write our ultimate result in the form



$$F(t) = \frac{g}{f} + \left(F(0) - \frac{g}{f}\right)e^{-ft} \qquad (5.6)$$

This result is important because it is a general solution of the first order linear differential equation where the functions f(t) and g(t) are both constants.

We can also apply a math program to solve our problem. The *Mathematica* solution is shown beside. It contains C(0) the starting condition (K in our notation). Again we have to obtain C(1) by setting t = 0.

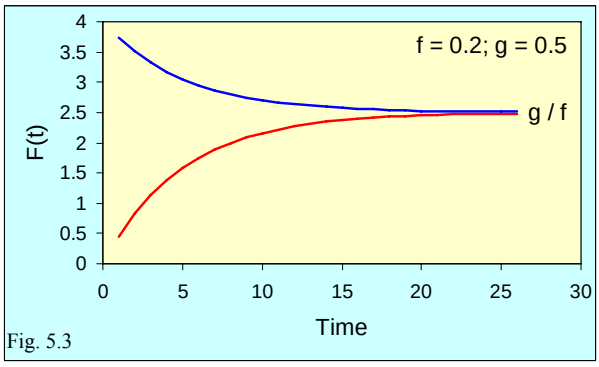


Fig. 5.3

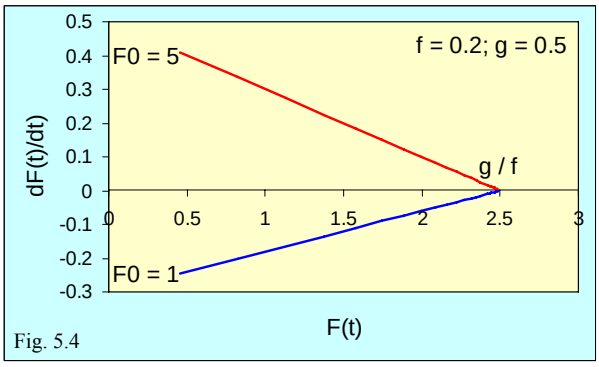


Fig. 5.4

Let's study our solution in detail. Fig. 5.3 shows a plot of F(t) against time. We see that irrespective of the initial concentration the process goes towards a stable medicament concentration. This concentration is called the **equilibrium point**. We compute this point by setting dF / dt = zero. This gives us immediately $F_{equilibrium}$ = g / f as shown in the Figure. Fig. 5.4 shows the so called **phase diagram** of the process. dF / dt is a linear function and irrespective of the initial concentration the root of this function is given by F(t) = g / f.

Another example. Above we dealt with the exponential growth equation. This equation is realistic only at low population sizes. In reality there will be upper limits, so-called **carrying capacities**, and these upper limits cause the real population growth to becoming lower at higher population sizes.

Look at Fig. 5.5. A population of *Saccharomyces cerevisiae* incubated in a chemostat (I took the data from G. F. Gause, 1934, The struggle for existence) grew initially fast according to an exponential growth model. But the higher the population size (the volume of the colony) was the slower grew brewers yeast.

How to model such a process? We take our initial exponential growth function and add a second term

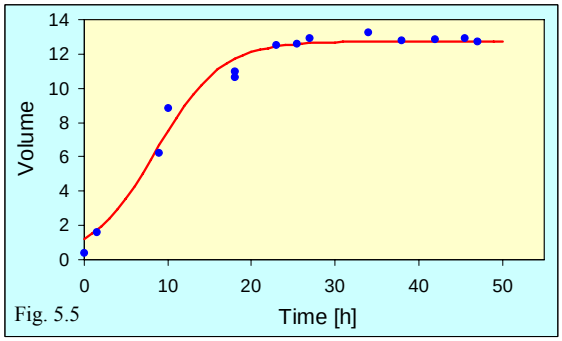


Fig. 5.5

that reduces population growth the nearer N comes to the limit. Let us denote this upper limit of population size with K. We might model as follows

$$\frac{dN}{dt} = rN\left(\frac{K-N}{K}\right) = rN - rN\frac{N}{K} \qquad (5.7)$$

If N becomes larger and larger K-N goes to zero and the second multiplicand also goes to zero, the population growth decreases. For small N the term (K - N) / K is nearly 1 and the whole process resembles our initial exponential growth. The above equation is the simplest form to model such a process and is called the **logistic growth process** or **Verhulst-Pearl equation** (after the Belgian mathematician Pierre F. Verhulst, 1804-1849 and the American mathematician Raymond Pearl, 1879-1940). The Figure beside (Fig. 5.6) shows us that the growth rate is very small for small t and becomes zero at N = K. The maximum growth rate occurs at N = K/2. Our logistic growth equation gives us no way to compute N. For this task we have to solve the differential equation

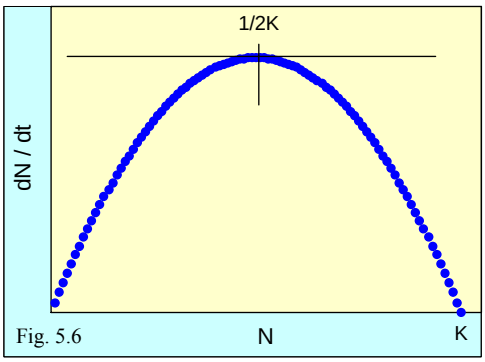


Fig. 5.6

$$\frac{dN}{dt} = rN\left(\frac{K-N}{K}\right)$$

We approximate the function from a Taylor series expansion. We assume that the function N(t) can be developed into a Taylor series

Therefore dN/dt must be an algebraic function of the form

$dN(t)/dt = a_1 + a_2N(t) + a_3N(t)^2 + a_4N(t)^3 \ldots\ldots$

Our population growth process has two roots, at N = 0 and at N = K. The simplest equation with two roots is a quadratic function and we cut the Taylor series therefore after the third term to get

$$\frac{dN}{dt} \approx a_1 + a_2N + a_3N^2$$

For N = 0 we get $a_1 = 0$. We get

$$\frac{dN}{dt} \approx a_2N + a_3N^2$$

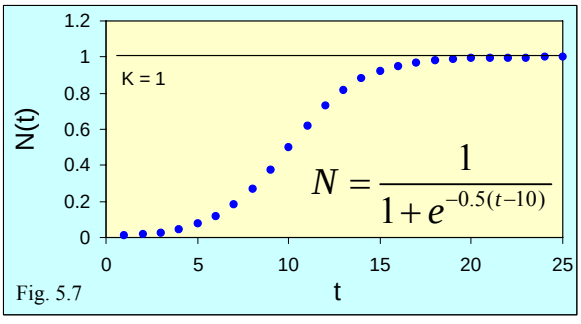


Fig. 5.7

At this stage we use a trick. We divide the whole equation through $N(a_2 + a_3N)$. We then get after some rearrangement

$$\frac{dN}{N} - \frac{a_3dN}{a_2 + a_3N} = a_2dt$$

This equation is more easy to differentiate

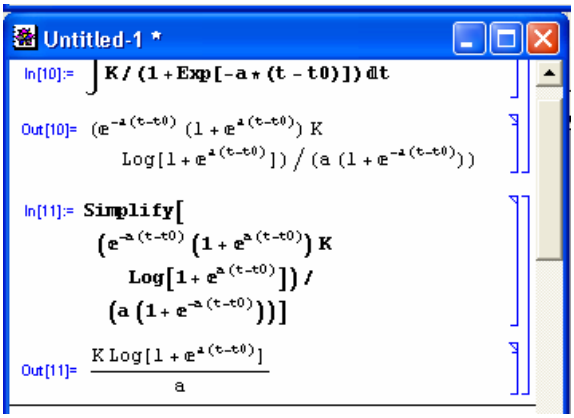
$$\int\frac{dN}{N} - \int\frac{a_3dN}{a_2 + a_3N} = \ln(N) - \frac{a_3\ln(a_2 + a_3N)}{a_3} = \ln\left(\frac{N}{a_2 + a_3N}\right) = \int a_2dt = a_2t + c$$

Therefore

$$\frac{N}{a_2 + a_3 N} = Ce^{a_2 t} \tag{5.8}$$

Now we can rearrange this equation and combine the constants a_2, a_3, and C into one constant K. We get

$$N(t) = \frac{K}{1 + e^{-a_2 (t - t_0)}} = \frac{K}{1 - Ce^{-a_2 t}} = \frac{K}{1 - (K / N_0 - 1)e^{-a_2 t}} \tag{5.9}$$



In the latter equation t_0 gives the population size N_0 at time $t = t_0$. K, the carrying capacity, is the upper limit of N and is easily obtained by setting dN / dt = 0. An example of this function shows Fig. 5.7. In the yeast example of Figure 5.5 I fitted the data with the function

$$N(t) = \frac{12.74}{1 - 9.32e^{-0.26t}}$$

In genetics or epidemiology mutations or infection rates are sometimes mass effects. That means the total number of mutation or infection inducing agents determine the outcome. Hence we have to know the sum of all these agents. Assume the process under study can be described by a logistic growth model. Then the total number of agents would be the integral under the logistic growth curve. We ask a math program and get the following answer

$$\int \frac{K}{1 + e^{-a(t - t_0)}} dt = \frac{K \ln(1 + e^{a(t - t_0)})}{a} + C$$

To obtain this equation we first computed the integral and then simplified the result.

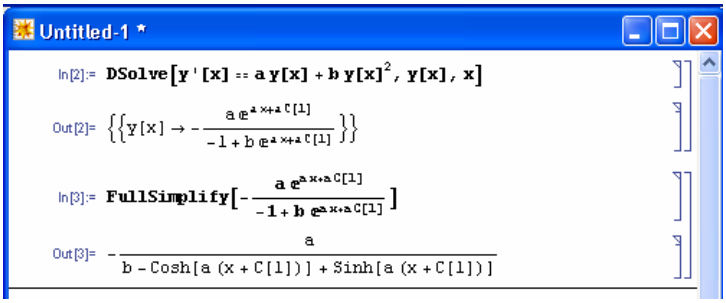
The last example calls for a general solution of a quadratic first order differential equation. Let's define such an equation in the following way

$$\frac{dy}{dx} = ay + by^2 \tag{5.10}$$

We first determine the equilibrium point and get for y' = 0

$$y(x) = -\frac{a}{b}$$

The solution of eq. 5.10 is similar to the solution above. But now we apply again the math program that gives us a general solution. C[1] is our starting value c



$$y = \frac{ae^{ac}e^{ax}}{1 - be^{ac}e^{ax}} \tag{5.11}$$

This is a general logistic growth function and the solution of the so-called **autonomous differential equation** (eq. 5.11). We see that for large x this function asymptoti-

cally goes to the equilibrium value of –a / b.

Mathematica shows us also that we can reformulate this solution in terms of hyberbolic sine and cosine functions.

The next example leads us to a general process that is important in genetics and ecology. Assume we have a population of an animal or plant species that colonizes a series of habitat patches (Fig. 5.8). Now let the initial frequency of colonized habitat patches be p, the initial frequency of empty patches q = (1-p). Note that frequencies are always chosen in such a way that the sum of all frequencies is 1 (or 100% if you want). Members of this population migrate. They are able to colonize empty patches (light green in Fig. 5.8) or populations at colonized patches die out. Such a population that is divided among a series of patches connected by migration events is called a **metapopulation.** Our problem is now to model the change of p (the number of colonized patches) in time. The change of p through time is dp/dt. Let μ denote the colonization (immigration) and ν the local extinction rate (probability). This change in p can now be modelled by a sum of two independent processes, the change due to colonization and due to local extinction. The increase in the number of colonized patches is assumed to be proportional to the actual number of colonized patches p and to the number of empty patches q = 1-p. Hence dp / dt = μp (1-p). Decrease in p is assumed to be proportional to the local extinction rate. Hence dp / dt = -νp We need the negative sign because p decreases in this process. We can now model the whole process by a simple equation

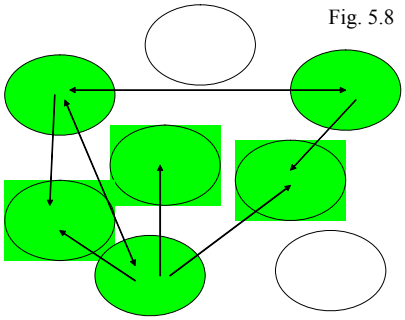


Fig. 5.8

$$\frac{dp}{dt} = \mu p(1-p) - \nu p = -\mu p^2 + p(\mu - \nu) \tag{5.12}$$

This is the well known **metapopulation model** of Richard Levins proposed in 1969. Extensions of it are currently intensively studied and debated.

To determine the equilibrium number of colonized patches we set dp/dt to 0 and get immediately

$$p = 1 - \frac{\nu}{\mu}$$

The model gives us a key prediction. Our metapopulation only persists if ν < μ. But implicitly it gives other key predictions. Assume that extinction probability ν is inversely related to patch area A. That means at larger patch size extinction probabilities decrease due to higher population sizes. Hence: ν ∝ 1/A. Assume further that colonisation rates μ decrease when patch distances D increase. Hence: m ∝ 1/D. We get

$$\frac{dp}{dt} = \frac{1}{D} p(1-p) - \frac{1}{A} p \tag{5.13}$$

Two new predictions emerge. First the number of colonized habitats will increase with increasing mean patch area (due to decreasing extinction rates). Second the number of colonized habitats with decrease will increasing distances between patches (due to reduced colonization events) (try to model this).

According to our classification eq. 5.12 is a **quadratic first order differential equation**. We are inclined to apply the general solution of eq. 5.13. We get

$$p(t) = \frac{(\mu-\nu)e^{\nu C}e^{\mu t}}{\mu e^{\nu C}e^{\mu t} - e^{\mu C}e^{\nu t}} = \frac{1-\frac{\nu}{\mu}}{1-e^{(\mu-\nu)C}e^{(\nu-\mu)t}} \tag{5.14}$$

We still need our starting value C. At t = 0 p equals p_0, the initial frequency. We get for C

$$p_0 = \frac{1-\frac{\nu}{\mu}}{1-e^{(\mu-\nu)C}} \rightarrow C = \frac{\ln\left(\frac{1-p_0-\nu/\mu}{-p_0}\right)}{\mu-\nu}$$

Note that the logarithm must be larger than 0. Hence for a rationale solution must hold $p_0 > 1 - \nu/\mu$.

Now we can plot the fraction of patches occupied against time. This shows Figure 5.9 and the respective Excel routine gives the following Table. After only ten generations the process approaches its equilibrium at 1—0.3 / 0.5 = 0.4. Hence under this starting conditions we expect that 40% of the habitat patches will be colonized.

	A	B	C	D
1	Parameter:	My	Ny	p0
2			0.5	0.3	0.45
3	Time			C
				=LN((1-$D$3-$C$3/$B$3)/-$D$3)/($B$3-$C$3)
4	1	3.840029564		
		EXP(($B$3-$C$3)*$D$3)*EXP(($C$3-$B$3)*A6))		
5	+a4+1			

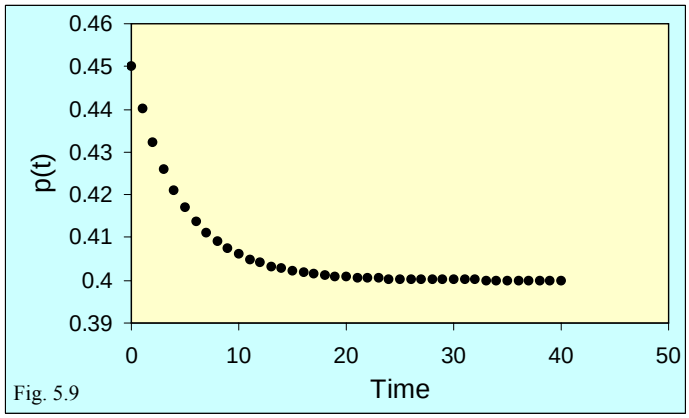
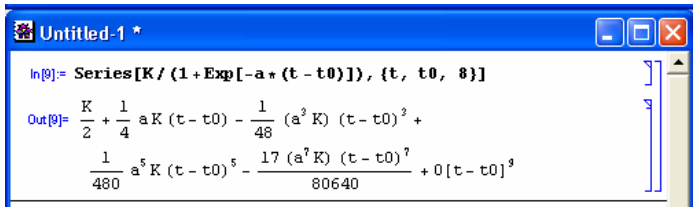


Fig. 5.9

However, often it is not so simple and integrals are either very complicated or it is even not possible to express an integral in a so called **closed form**. Then we can try another possibility. We approximate the integral. To do this we evaluate the function into a Taylor series and cut the resulting series beyond an appropriate element. Computing the integral is then easily done by hand because you need only the integral of

$$\int a(x-b)^n dx = \frac{a}{n+1}(x-b)^{n+1} + C$$



In the case of our logistic growth equation *Mathematica* computes the Taylor series expansion for us. We cut after the third element and get as an approximation of the integral

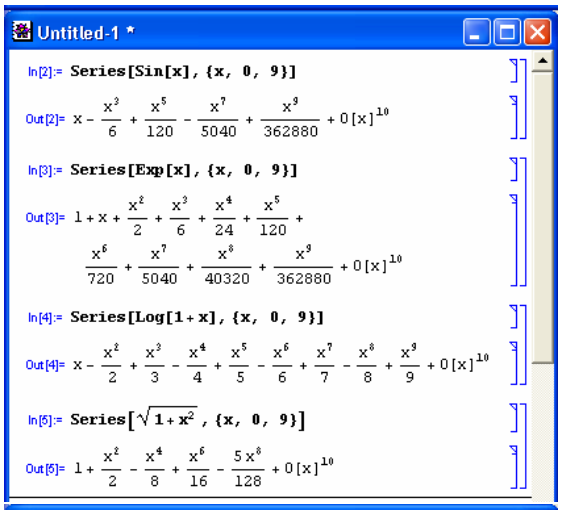
$$\int \frac{K}{1+e^{-a(t-t_0)}}dt \approx \int \frac{K}{4} + \frac{aK}{4}(t-t_0) - \frac{a^3K}{48}(t-t_0)^3 dt$$

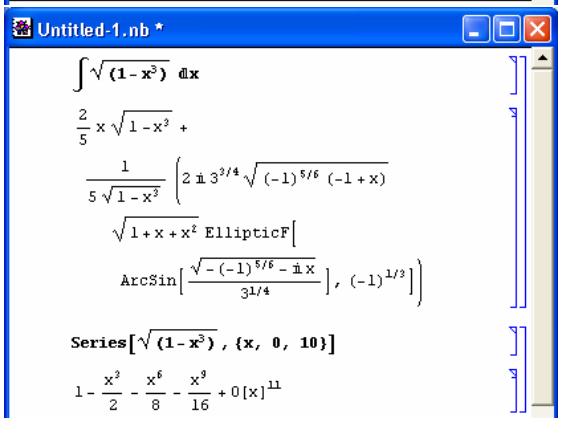
$$\approx \frac{K}{4}(t-t_0) + \frac{aK}{8}(t-t_0)^2 - \frac{a^3K}{192}(t-t_0)^4 + C$$

We can compare the results. The exact solution is for a = 1, K = 1, and $t_0 = 0$ between t = 0 and t = 3 a value of N = 2.355. The approximation gives 1.45. Our error is therefore (2.355 - 1.45) / 2.355 = 0.38 ≡ 38%. Might be better.

Our whole approach heavily depends on how fast our Taylor series **converges** that means approaches

values near to the exact solution. Hence applying this method has to be accompanied by an analysis how fast the series converges to its limes, the true function value f(t). We have to plot the sum of a series of elements against element number.

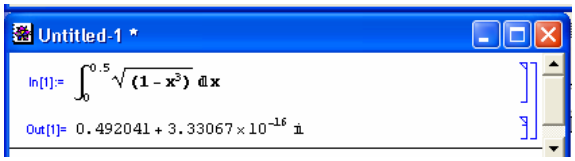




Now we return to the problem of approximating an integral using a Taylor series. Consider the equation $y = (1 - x^3)^{0.5}$ with $0 < x < 1$. *Mathematica* computes a very complicated integral. However the Taylor series is very simple and converges very quickly. The integral from 0 to 0.5 becomes

$$\int_0^{0.5} \sqrt{1-x^3}\,dx \approx \int_0^{0.5} 1-\frac{x^3}{2}-\frac{x^6}{8}\,dx = x-\frac{x^4}{8}-\frac{x^7}{56}\Big|_0^{0.5} = 0.492$$

The numerical *Mathematica* solution is nearly identical.



Next we deal with second order differential equations. These have the homogeneous form

$$\frac{d^2y}{dx^2} + c\frac{dy}{dx} + ax = 0 \qquad (5.15)$$

First order homogeneous linear equations have a solution of the type $y = e^{bx}$. We test

$$\frac{d(e^{bx})}{dx} = be^{bx}; \frac{d(be^{bx})}{dx} = b^2e^{bx}$$

Hence

$$\frac{d^2y}{dx^2} + c\frac{dy}{dx} + ax = b^2e^{bx} + cbe^{bx} + ae^{bx} = e^{bx}(b^2 + cb + a)$$

The term on the right side has to be zero. Therefore

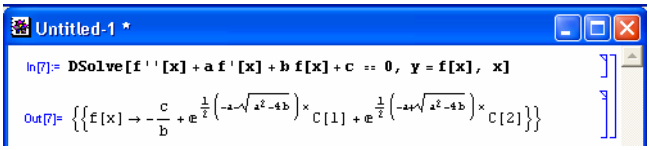
$$b^2 + cb + a = 0 \rightarrow b_{1,2} = \frac{1}{2}\left(-c \pm \sqrt{c^2 - 4a}\right) \qquad (5.16)$$

Eq. 5.16 is called the characteristic equation and gives us solutions of second order homogeneous linear differential equations. In the case where $c^2 > 4a$ we get two simple and a class of compound solutions of eq.

5.15

$$y = e^{b_1 x}$$
$$y = e^{b_2 x}$$
$$y = re^{b_1 x} + se^{b_2 x} \tag{5.17}$$

with r and s being arbitrary constants. In the case where $c^2 < 4a$ the solutions are more complicated and we don't deal with this case because the solutions involve complex numbers. The general second order equation is even more complicated to solve. However, today math programs do the job for us



$$\frac{d^2y}{dx^2} + a\frac{dy}{dx} + by + c = 0 \tag{5.18}$$

An important special case of differential equations are those of the separable type. This means that we can separate the variables x and y on both sides of the equations. Hence

$$\frac{dy}{dx} = f(x)g(y) \rightarrow \frac{dy}{g(y)}\frac{dx}{f(x)} \rightarrow \int \frac{1}{g(y)}dy = \int \frac{1}{f(x)}dx + C \tag{5.19}$$

For instance what is the solution of dy/dx = xy? We need

$$\frac{dy}{y} = xdx \rightarrow \ln(y) = \frac{1}{2}x^2 + C \rightarrow y = y(0)e^{x^2/2}$$

Another type of easy to solve equations are those without explicit y-term. In this case we set dy/dx = u and get an equation of lower order.

$$\frac{d^2y}{dx^2} + \frac{dy}{dx} - a = 0 \rightarrow \frac{du}{dx} + u - a = 0 \rightarrow u = a + Ce^{-x}$$
$$\frac{dy}{dx} = a + C_1e^{-x} \rightarrow y = ax - C_1e^{-x} + C_2$$

In biology second order differential equations appear for instance when we want to infer how the points where a process changes fastest depend on the parameters involved.

Next we consider the second order linear equation

$$\frac{d^2f(x)}{dt^2} + \omega^2 f(x) = 0 \tag{5.20}$$

This eq. has the solution

$$f(x) = C\cos(\omega t + \varphi) = C\cos[\omega(t + \frac{\varphi}{\omega})] \tag{5.21}$$

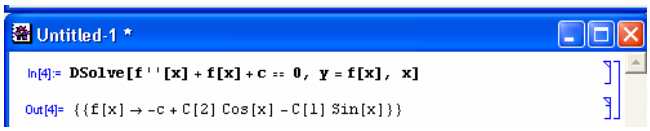
The solution is therefore a class of cosine functions with t shifted to the left by ϕ/ω (ϕ and C are constants). Eq, 5.20 describes a harmonic oscillator. They have the frequency of $\omega/2\pi$.

A special problem of all differential equations is the **initial value problem**. In other words we need val-

ues for the constants c that appear in our solutions. In the case of a first order equation we obtain the value of the single constant by setting x = 0.

$$\frac{df(x)}{dx} + af(x) + b = 0 \rightarrow f(x) = \frac{-b}{a} + Ce^{-ax} \rightarrow C = f(0) + \frac{b}{a} \tag{5.22}$$

In the case of a second order equation we get two constants C_1 and C_2 (cf eq . 5.17). The solutions are therefore an infinite number of functions. To pick out particular functions we need additional constraints that determine the constants C_1 and C_2. The first condition might be f(x = 0)= y_0. The second constraint is then d(f(x = 0)/dx = =y_1.



An example. What is the solution of $d^2y/dx^2 + dy/dx + c = 0$ for which y = 1 and y' = 1 at x = 0? We get

$$f(0) = c + C_2\cos(0) - C_1\sin(0) \rightarrow C_2 = f(0) - c = 1 - c$$
$$f'(0) = -C_2\sin(0) - C_1\cos(0) \rightarrow C_1 = -f(0) = -1$$

6.1 Partial derivatives and stationary points

In biology derivatives are used to

- Assess rates of change
- To infer points of maximal change (speed)
- To infer minima and maxima of processes
- To infer points of equilibrium

The latter points are called **stationary points**. We find these by solving f'(x) = 0.

Up to now we dealt with models containing one dependent and one independent variable. The derivative is

$$y = f(x) \rightarrow \frac{dy}{dx} = \frac{d(f(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \qquad (6.1.1.$$

However many biological models contain more than one independent variable. In this case we have functions of the type

$$y = f(x_1, x_2 ... x_n) \qquad (6.1.2)$$

For instance, a plane in space has the function z = ax + by +c. A hemisphere is given by $z = (x^2 + y^2)^{1/2}$ (Fig. 6.1.1). The saddle has $z = x^2 - y^2$ (Fig. 6.1.2).

Fig.6.1.1

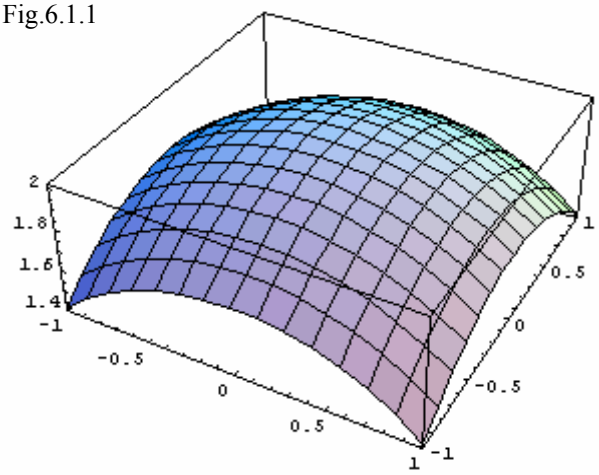
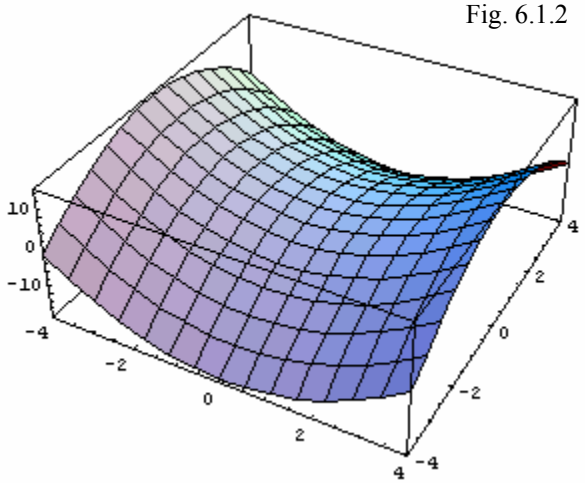


Fig. 6.1.2



In general a function z = f(x,y) defines a **surface** in space. This surface can change in three dimensions. In this case we can defines **partial derivatives** for the variables x and y

$$\frac{\delta z}{\delta x} = \frac{\delta f(x,y)}{\delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - (f(x,y)}{\Delta x}$$
$$\frac{\delta z}{\delta y} = \frac{\delta f(x,y)}{\delta y} = \lim_{\Delta y \rightarrow 0} \frac{f(x, y + \Delta y) - (f(x,y)}{\Delta y} \qquad (6.1.3)$$

Handling partial derivatives is not more difficult than using ordinary ones. We have to choose the variable that changes and hold the other (or others if we have more than one) constant. A simple example

$$y = xe^{-t} \quad \rightarrow \quad \frac{\partial y}{\partial t} = -xe^{-t}; \quad \frac{\partial y}{\partial x} = e^{-t}$$

The partial derivatives of z = sin(x)cos(y) are

$$\frac{\delta z}{\delta x} = \cos(x)\cos(y)$$
$$\frac{\delta z}{\delta y} = -\sin(x)\sin(y)$$

Note that for dz/dx the variable y is treated as a constant. Hence

$$\frac{\delta(x^2 + y)}{\delta x} = 2x$$

Of course we can also calculate higher order derivatives

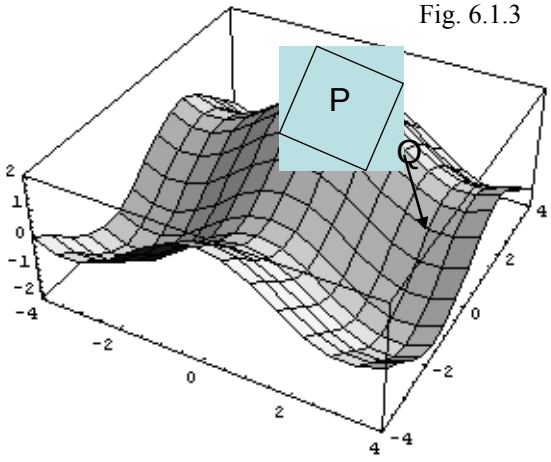
$$\frac{\delta z}{\delta x}\left(\frac{\delta z}{\delta x}\right) = \frac{\delta^2 z}{\delta x^2}$$
$$\frac{\delta z}{\delta x}\left(\frac{\delta z}{\delta y}\right) = \frac{\delta^2 z}{\delta x \delta y} \qquad (6.1.4)$$

The latter case is called a **mixed derivative**. An important special case is the function z = f(ax-cy). The inner term is therefore a linear function. We get

$$\frac{\delta^2 z}{\delta x^2} = a^2 f''(ax - cy)$$
$$\frac{\delta^2 z}{\delta y^2} = (-c)^2 f''(ax - cy) \rightarrow \frac{\delta^2 z}{\delta x^2} = \frac{a^2}{c^2}\frac{\delta^2 z}{\delta y^2}$$

and this function is called the **wave equation**. Because this is a partial differential equation z = f(ax-cy) is always a solution independent of the function f.

Fig. 6.1.3



A function y = f(x) in a two dimensional coordinate system has a tangent line that gives the slope at any point P. Any surface in space has then a tangent plane. Any plane touching the surface z = f(x,y) at the point P ={p,q, f(p,q)} is given by the function z = Ax+By+c (Fig. 6.1.3). In direction x the slope of the plane is given by

$$A = \frac{\delta z}{\delta x}; B = \frac{\delta z}{\delta y}$$

c is then

$$c = f(p,q) - Ap - Bq$$

Therefore the tangent plane through the point {p, q, f(p,q)} has the general function

$$z = \frac{\delta z}{\delta x}(x - p) + \frac{\delta z}{\delta y}(y - q) + c \qquad (6.5)$$

In Fig. 6.1.3 we detect two local minima and one maximum. How to detect them? In two dimensions the

maximum and minimum were characterized by dy/dx = 0. The same holds for the tangent plane. Hence, we get maxima and minima from

$$\frac{\delta z}{\delta x} = 0$$
$$\frac{\delta z}{\delta y} = 0 \qquad (6.1.6)$$

For instance find one maximum of $z = x^2 + y^2$ and the function of the plane at the maximum. We need

$$\frac{\delta z}{\delta x} = 2x; \frac{\delta z}{\delta y} = 2y$$
$$2x = 0$$
$$2y = 0$$

Because we deal with a hemisphere, the points x = 0 and y = 0 denote the maximum. The plane at this point has of course a zero slope. However, in three dimensions we have a third possibility. A point might be a minimum for the x-axis and a maximum for the y-axis or vice versa. Then we have a **saddle** (Fig. 6.1.2). To decide whether we have a minimum, a maximum or a saddle (the so-called **stationary points**) we use similar criteria as for simple functions and involve second order derivatives

$$\frac{\delta^2 z}{\delta x^2}\frac{\delta^2 z}{\delta y^2} - \left(\frac{\delta^2 z}{\delta x \delta y}\right)^2 < 0 \rightarrow \text{saddle}$$
$$\frac{\delta^2 z}{\delta x^2}\frac{\delta^2 z}{\delta y^2} - \left(\frac{\delta^2 z}{\delta x \delta y}\right)^2 > 0; \frac{\delta^2 z}{\delta x^2} < 0 \vee \frac{\delta^2 z}{\delta y^2} < 0 \rightarrow \text{maximum}$$
$$\frac{\delta^2 z}{\delta x^2}\frac{\delta^2 z}{\delta y^2} - \left(\frac{\delta^2 z}{\delta x \delta y}\right)^2 > 0; \frac{\delta^2 z}{\delta x^2} > 0 \vee \frac{\delta^2 z}{\delta y^2} > 0 \rightarrow \text{minimum} \qquad (6.1.7)$$

An example. Find the stationary points of $z = x^3 + 3y^3 - 2x^2 - y^2$. We need

$$\frac{\delta z}{\delta x} = 3x^2 - 4x = 0$$
$$\frac{\delta z}{\delta y} = 9y^2 - 2y = 0$$
$$\frac{\delta^2 z}{\delta x^2} = 6x - 4$$
$$\frac{\delta^2 z}{\delta y^2} = 18y - 2$$
$$\frac{\delta^2 z}{\delta x \delta y} = 0$$

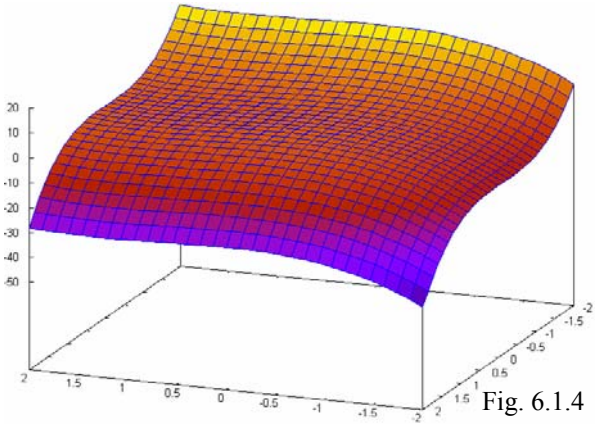


Fig. 6.1.4

From the first two equations we get the stationary points (0,0), (0,2/9), (4/3,0), and (4/3,2/9). Hence

$$\frac{\delta^2 z}{\delta x^2} = -4 \vee -4 \vee 4 \vee 4$$

$$\frac{\delta^2 z}{\delta y^2} = -2 \vee 2 \vee -2 \vee 2 \quad \text{and} \quad \frac{\delta^2 z}{\delta x^2}\frac{\delta^2 z}{\delta y^2} = 0$$

The first point (0,0) is therefore a local maximum, (0,2/9) and (4/3,0) are saddles, and (4/3,2/9) is a minimum.

Because a plane can be curved in any direction we can obtain derivatives and stationary points in any direction. To obtain a function that gives us the **directional derivation** we first look at **small changes**. Suppose a function y = f(x). The derivative of f(x) at a given point a is defined as

$$\frac{\delta y}{\delta x} = \frac{f(a+\delta x) - f(a)}{\delta x} = f'(a)$$

Hence

$$\delta y = f'(a)\delta x = \frac{dy}{dx}\delta x \qquad (6.1.8)$$

This equation gives the approximation how y changes at small changes of x. For instance how does a ball of 10 cm radius increase in volume at 11 cm radius? We need

$$V = \frac{4}{3}\pi r^3 \rightarrow V' = 4\pi r^2$$

Hence $\delta V = 4\pi 100 = 1256.6$. The new Volume is Vnew = 4188.8 + 1256,6 = 5445,4. This can be compared to the exact value of Vnew = 5576.3. Even at a difference in radius of 10% eq. 6.8 gives a quite precise estimate. We now apply eq. 6.8 to the problem of directional change. The change δz of the function z = (f(x,y) giving changes δx and δy is

$$\delta z = \frac{dz}{dx}\delta x + \frac{dz}{dy}\delta y \qquad (6.1.9)$$

If we go from point Q to point P (Fig. 6.1.5) we have a change of δx in x direction and δy in y direction. The change in the joint direction is given by δc. Dividing the above equation by δc gives

$$\frac{\delta z}{\delta c} = \frac{\delta z}{\delta x}\frac{\delta x}{\delta c} + \frac{\delta z}{\delta y}\frac{\delta y}{\delta c} = \frac{\delta z}{\delta x}\cos(\alpha) + \frac{\delta z}{\delta y}\sin(\alpha) \qquad (6.1.10)$$

This is the equation of directional change. For instance what is the slope of the tangent plane of $z = xy - x^2 - y^2$ at P = (1,1) in direction $\alpha = \pi/6$? We need

$$\frac{\delta z}{\delta x} = y - 2x = -1; \frac{\delta z}{\delta y} = x - 2y = -1$$

$$\cos(\pi/6) = \sqrt{3}/2$$

$$\sin(\pi/6) = 0.5$$

$$\downarrow$$

$$\frac{\delta z(\alpha = \pi/6)}{\delta c} = -\sqrt{3}/2 - 1/2 = -\frac{\sqrt{3}+1}{2}$$

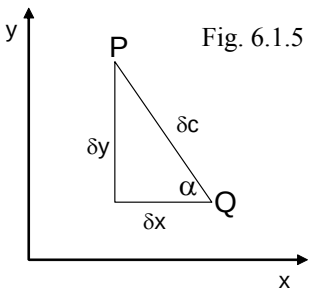


Fig. 6.1.5

Planes have directions where the slope is steepest. Again we need the second derivative to find these points. We first compute δz/δc and then we need the derivative with respect to the angle α to obtain the direction. For instance find the directions of the steepest slopes of the function z = sin(x) * cos(y) (Fig. 6.1.6). We get

$$\frac{\delta z}{\delta c} = \cos(x)\cos(y)\cos(\alpha) - \sin(x)\sin(y)\sin(\alpha)$$
$$\frac{\delta(\delta z/\delta c)}{\delta\alpha} = -\cos(x)\cos(y)\sin(\alpha) - \sin(x)\sin(y)\cos(\alpha) = 0$$
$$\frac{\sin(\alpha)}{\cos(\alpha)} = \tan(\alpha) = \frac{-\cos(x)\cos(y)}{\sin(x)\sin(y)}$$

and obtain a general equation that gives the directions of the steepest slopes for any point (x,y).

The last point we have to discuss is implicit differentiation. Consider a function f(x,y) = c. This function describes a curves within a x,y plane. Often the function f(x,) cannot be solved for y = g(x). However it is always possible to get the slope dy/dx. We have

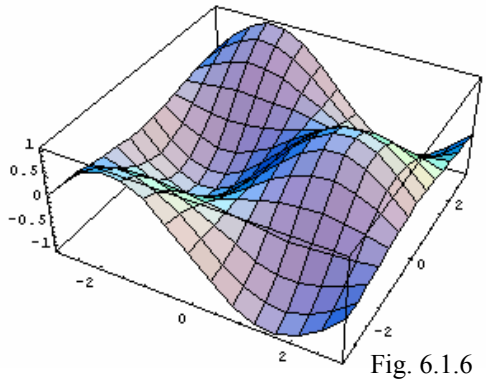


Fig. 6.1.6

$$\frac{dy}{dx} = \lim_{\delta x \to 0} \frac{\delta y}{\delta x}$$

Because we are moving on the curve described by f(x,y) we have no change in f, hence δf = δz = 0. Now we use the function of incremental increase (eq. 6.1.9) and get

$$\frac{dy}{dx} = -\frac{\delta f(x,y)}{\delta x} \Big/ \frac{\delta f(x,y)}{\delta y} \qquad (6.1.11)$$

For instance find the derivative dy/dx of the ellipse given by $ax^2 + by^2 = c$. We get

$$\frac{\delta f}{\delta x} = 2ax; \frac{\delta f}{\delta y} = 2bx \rightarrow \frac{dy}{dx} = -\frac{ax}{by}$$

This is are remarkable result that tells how to obtain slopes of an ellipse (a ≠ b) and a circle (a = b) for any point (x,y). Further this is a differential equation and we learn that the general solution of a differential equation of the type $dy/dx = -cxy^{-1}$ is given by closed geometric objects like ellipses and circles of the form $ax^2 + by^2 = c$. We can solve by rearranging

$$\frac{dy}{dx} = -\frac{ax}{by} \rightarrow \int bydy = -\int axdx \rightarrow \frac{b}{2}y^2 + c_1 = -\frac{a}{2}x^2 - c_2$$
$$ax^2 + by^2 = C$$

6.2 Lagrange multipliers

Sometimes it is quite difficult to find stationary points of functions with two or more variables. In cases where we have an additional condition that constraints our function z = f(x,y) we can apply a so-called **Lagrange multiplier** named after the Italian mathematician Joseph-Luis Lagrange (1736-1813). Assume you need the maximum or minimum of z = f(x,y) and you have a second function g(x,y) = 0 that constraints the values of f. Remember that stationary points fulfil the condition

$$\frac{\delta f(x,y)}{\delta x}dx + \frac{\delta f(x,y)}{\delta y}dy = 0 \rightarrow \frac{\delta f(x,y)}{\delta x} = 0; \frac{\delta f(x,y)}{\delta y} = 0$$

The differential g(x,y) = 0 is

$$\frac{\delta g(x,y)}{\delta x}dx + \frac{\delta g(x,y)}{\delta y}dy = 0$$

The Langrange method now adds λ times the derivative of g to f. The value λ is called the **Langrange multiplier**. Hence

$$\begin{aligned}\frac{\delta f(x,y)}{\delta x} + \lambda\frac{\delta g(x,y)}{\delta x} &= 0\\ \frac{\delta f(x,y)}{\delta y} + \lambda\frac{\delta g(x,y)}{\delta y} &= 0\end{aligned} \qquad (6.2.1)$$

Now we have three unknown variables x, y and λ. Solving both equations for λ gives an equation y = h (x) that contains only x and y. But we have a second function g, the constraint. Substituting y = h(x) into g(x,y) gives an equation that contains only y and can be solved. Having y we also get x.

A simple example to explain the method. Find the coordinates x and y for the shortest distance of a circle $f(x,y) = x^2 + y^2 - r^2$ from the point P = (1,2). A graphical evaluation of the problem (Fig. 6.2.1) leads to two circles that touch at the point Q. We are looking for the coordinates of Q. Hence the function $g(x,y) = (x-1)^2 + (y-2)^2 - (\sqrt{5}-r)^2 = 0$ constraints the function f(x,y) in an appropriate way. Thus we need

$$\frac{\delta f(x,y)}{\delta x} + \lambda\frac{\delta g(x,y)}{\delta x} = 2x + \lambda 2(x-1) = 0$$

$$\frac{\delta f(x,y)}{\delta y} + \lambda\frac{\delta g(x,y)}{\delta y} = 2y + \lambda 2(y-2) = 0$$

We get

$$2x + \lambda 2x - 2\lambda = 0 \rightarrow \lambda = \frac{-x}{x-1}$$

$$2y + \lambda 2y - 4\lambda = 0 \rightarrow \lambda = \frac{-y}{y-2}$$

$$\frac{-x}{x-1} = \frac{-y}{y-2} \rightarrow x = \frac{y}{2}$$

Now we use the original constraint g(x,y) and substitute x = y/2.

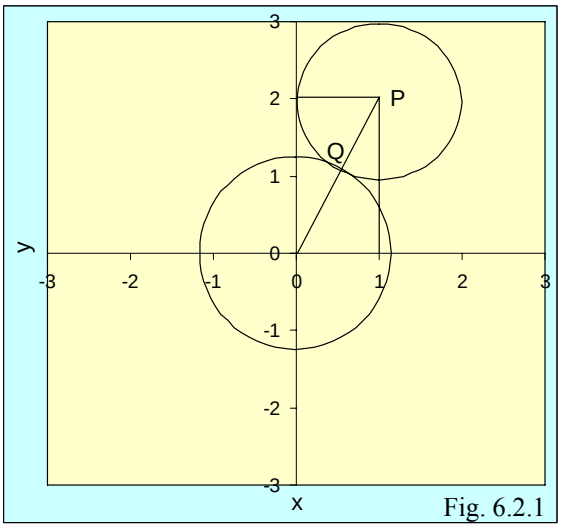


Fig. 6.2.1

$$(x-1)^2+(y-2)^2-\left(\sqrt{5}-r\right)^2=0\rightarrow\left(\frac{y}{2}-1\right)^2+(y-2)^2-\left(\sqrt{5}-r\right)^2=0$$

$$y=\frac{2}{3}\left(3+\sqrt{3}\sqrt{-5+2r\sqrt{5}-r^2}\right)$$

$$x=\frac{2}{6}\left(3+\sqrt{3}\sqrt{-5+2r\sqrt{5}-r^2}\right)$$

Another example. What is the rectangle of maximum area within the ellipse $g(x,y) = ax^2 + by^2 = 1$ (Fig. 6.2.2)? The function f(x,y) that defines the area of the rectangle within the ellipse is f(x,y) = 4xy. The constraint is given by g(x,y). We have therefore three equations

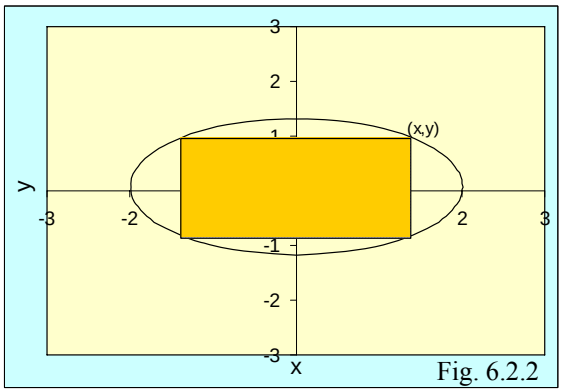


Fig. 6.2.2

$$g(x,y)=ax^2+by^2-1=0$$

$$\frac{\delta f}{\delta x}+\lambda\frac{\delta g}{\delta x}=4y+\lambda 2ax=0$$

$$\frac{\delta f}{\delta y}=4x+\lambda 2by=0$$

From the eq. 2 and 3 we get

$$\lambda=\frac{-2y}{ax}=\frac{-2x}{by}\rightarrow y=\pm\sqrt{\frac{ax^2}{b}}$$

In our case we need only the positive solution. We introduce this into the first eq. and get

$$ax^2+ax^2=1\rightarrow x=\sqrt{\frac{1}{2a}}\rightarrow y=\sqrt{\frac{2}{b}}$$

The area is then $A = 4xy=4(1/ab)^{1/2}$.

Lagrange multipliers are therefore useful when dealing with maximum problems. A last simple example. What is the maximum of a rectangle restricted by a fixed perimeter? The perimeter is given by c = 2x + 2y, the area has A = xy. Hence

$$\frac{\delta f}{\delta x}+\lambda\frac{\delta g}{\delta x}=y+2\lambda;\frac{\delta f}{\delta y}+\lambda\frac{\delta g}{\delta y}=x+2\lambda$$

This is only possible if x = y. Thus x = y = c/4.

Online archives and textbooks

Online Mathematical textbooks (A large collection of textbooks) http://www.math.gatech.edu/~cain/textbooks/onlinebooks.html

General mathematics (a collection of online lecture scripts and basic text on mathematics) http://www.geocities.com/alex_stef/mylist.html

Mathematics online (a source of educational online texts) http://www.glencoe.com/sec/math/

Mathematics Virtual Library (Many links to interesting web pages and programs) http://www.math.fsu.edu/Science/math.html

Math on the web (Search engine for all sorts of mathematics) http://www.ams.org/mathweb/mi-mathinfo07.html

The Math Archive (Many links to interesting web pages and programs) http://archives.math.utk.edu/

Eric Weisstein's Mathematics (a large online mathematics dictionary, with many examples) http://mathworld.wolfram.com/

The Internet Mathematics library (a large collections of topics for pupils and students, math-beginners) http://mathforum.org/library/

Mathematic resources (a large compilation of math internet pages) http://www.clifton.k12.nj.us/cliftonhs/chsmedia/chsmath.html

Kolegium nauczyczielski. Materiały z wykładów. (Online scripts on various topics) http://info.fuw.edu.pl/~ajduk/lect.html

Johannes Müller. 2003. Mathematical models in biology. Lecture term at TU Munich. http://www-m12.ma.tum.de/lehre/model_2003/skript/skript.pdf

Population growth models (a nice collection of growth models) http://www.math.duke.edu/education/postcalc/growth/contents.html.

Population growth models (A collection of growth model an animations) http://members.optusnet.com.au/exponentialist/Growth_Models.htm

Competition models (for persons who are interested In a discussion of the Lotka Volterra models) http://www.ub.rug.nl/eldoc/dis/fil/r.c.looijen/c11.pdf

The MacTutor history of mathematics (a very nice page on historical topics) http://www-history.mcs.st-andrews.ac.uk/.

Excel Turorials (Many macros) http://www.herber.de/index.html?http://www.herber.de/forum/archiv/104to108.htm.

Computational molecular Biology. (a very good side with examples how to use mathematics in molecular biology). http://www.cs.bc.edu/~clote/ComputationalMolecularBiology/

Literature

(Coloured titles are available in the Institute or the library, red titles are of major importance)

Cornish Bowden A. 1999—Basic Mathematics for Biochemists—Oxford Univ. Press 2nd. Ed.

Ennos R. 1999—Statistical and Data Handling Skills in Biology—Longman.

Foster P. C. 1998—Easy Mathematics for Biologists— Taylor and Francis.

Jordan D. W., Smith P. 2002. Mathematical techniques. 3rd. Ed. Oxford Univ. Press

Grossman S., Tuner J. E. 1974—Mathematics for the Biological Sciences—Macmillan.

Martin J. 1972—Podstawy Matematyki i Statystyki—Warszawa.

Portenier C., Gromes W. 2003. Mathematik für Biologen und Humanbiologen. Script. Marburg

Scheiner S. M., Gurevitch J. (eds.) 2001—Design and Analysis of Ecological Experiments— Oxford Univ. Press (2nd. Ed.).

Wilson E. O., Bossert W. H. 1971, A Primer of Population Biology. Sinauer (Stamford).

Murray J. D. 2003. Mathematical Biology. 3rd ed. Parts I and II. Springer New York.

Science Magazine special feature. 2004. Mathematics in biology. Science 303.

Napiórkowski K. 2001. Matematyka. http://info.fuw.edu.pl/~ajduk/FUW/matnkf/matematyka01_nkf.pdf

Mathematical software

The Windows software collection (public domain and freeware)
http://archives.math.utk.edu/software/.msdos.directory.html (contains many very nice programs)

The mathematics virtual library (a collection of software pages) http://www.math.fsu.edu/Virtual/index.php?f=21.

Guide to mathematical software (a search engine for math programs) http://gams.nist.gov//

Step by step derivatives (a very good program for computing derivatives) http://www.calc101.com/webMathematica/derivatives.jsp#topdoit

Derivative calculator (a nice small but quite effective program for computing derivatives) http://cs.jsu.edu/mcis/faculty/leathrum/Mathlets/derivcalc.html

JAVA Mathlets for Math Explorations (a nice collection of small math programs for everybody) http://cs.jsu.edu/mcis/faculty/leathrum/Mathlets/

The integrator (a small but effective integration program)
http://www.integrals.com/index.en.cgi

The MathServ Calculus toolkit (a collection of Math applets for calculus computation)
http://www.math.vanderbilt.edu/~pscrooke/toolkit.html

Modelowanie reczwistości (a nice Polish page with a program collection and many further links) http://www.wiw.pl/modelowanie/

Maple homepage. http://www.maplesoft.com/

Mathematica homepage (Wofram research) http://www.wri.com/

Mathworks homepage (Matlab) http://www.mathworks.com/

Mathtype (Office build in tool for mathematics writing) http://www.mathtype.com/en/products/mathtype/