Mathematics Pre-requisites for “Theory and Applications in Bioinformatics”

Here we list some of the mathematical concepts and definitions required in order to follow the Theory and Applications in Bioinformatics module. The descriptions are quite concise and for a more detailed explanation you might find the book *Basic Mathematics for Biochemists* by Cornish-Bowden useful. There are also many other introductory textbooks covering these topics in greater detail.

1 Sums and products

We write a sum of N terms as,

$$\sum_{i=1}^{N} f(i) = f(1) + f(2) + \ldots + f(N) \; .$$

Examples of sums we can work out exactly are,

$$\sum_{i=1}^{N} i = \frac{N(N+1)}{2} \qquad \sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}$$

eg.

$$\begin{aligned} \sum_{i=1}^{4} i &= 1+2+3+4 \;=\; 10 \;=\; \frac{4\times 5}{2} \; , \\ \sum_{i=1}^{4} i^2 &= 1+2^2+3^2+4^2 \;=\; 30 \;=\; \frac{4\times 5\times 9}{6} \; . \end{aligned}$$

Sometimes we will write sums over a set of values which aren’t in a natural sequence from 1 to N. For example, we might wish to sum up the number of each nucleotide in a DNA sequence. Call this number $n(x)$ where x is one of A,C,T or G (in set notation we write $x \in \{A, C, T, G\}$). We could then write the total number of nucleotides as,

$$\sum_{x\in\{A,C,T,G\}} n(x) \;=\; n(A) + n(C) + n(T) + n(G) \; .$$

In some cases it will be cumbersome to write the terms which the sum is over and we may just write,

$$\sum_x n(x) \;=\; n(A) + n(C) + n(T) + n(G)$$

where it is understood that the sum is over all possible values that x may take. This allows us to write general expressions where the range of the sum will depend on the context and is particularly useful when writing mathematical expressions which hold for many different cases (eg. where x could be nucleotides, base-pairs, amino acids, codons etc.).

In a similar way to sums of many terms we can write a product of many terms as,

$$\begin{aligned}\prod_{i=1}^{N} f(i) &= f(1) \times f(2) \times \ldots f(N) \\ &= f(1)f(2)\ldots f(N)\;.\end{aligned}$$

Notice that we often drop the explicit $\times$ sign for compactness when terms are multiplied together, as in the second line above. As an example where a product symbol can be used consider the factorial function which is defined,

$$N\,! = N \times (N-1) \times (N-2) \ldots 2 \times 1 \qquad \text{eg.} \quad 4\,! = 4 \times 3 \times 2 \times 1 = 24\;.$$

This can be written using the product notation as,

$$N\,! = \prod_{i=1}^{N} i$$

2 Powers, logarithms and exponentials

Powers and logarithms are ubiquitous in the biological sciences and their importance cannot be over-emphasised. We write the nth power of a number a as a^n. For example,

$$2^4 = 2 \times 2 \times 2 \times 2 = 16\;, \qquad 10^6 = 1\,000\,000\;.$$

The raised number is called the power or exponent. The power may be negative in which case we define,

$$a^{-n} = \frac{1}{a^n} \qquad \text{eg. } 2^{-4} = \frac{1}{16}\;.$$

The power may also be fractional in which case it represents a *root*, eg. $a^{\frac{1}{2}} = \sqrt{a}$ is the squared root.

An important related function is the logarithm. In general we define this function by the equation,

$$\log_a a^n = n$$

where $\log_a x$ is called the logarithm of x to base a. The most usual logarithms are base 2 ($\log_2 x$), base 10 ($\log_{10} x$) and base e ($\log_{\mathrm{e}} x$, also written as $\ln x$ – we define the number e below).

Base 2 logarithms are particularly useful when describing a process involving doubling. For example, if every member of a cell population divides at each time-step, then the number of cells N after t time-steps will be 2^t. If we know the number of cells then we can determine the number of time-steps using logarithms, ie. $\log_2 N = t$. Often the logarithm is used when a quantity grows very rapidly, eg. in the cell division example the population soon becomes extremely large but the logarithm remains manageable (it grows in proportion with time, or linearly with time).

Base 10 logarithms are very useful when describing large decimal numbers, eg.

$$\log_{10}(10000) = \log_{10} 10^4 = 4 \; .$$

In general if x is a decimal with N digits on the left of the decimal place, then $\log_{10} x$ will be a number between $N-1$ and N. We see that the logarithm of a large number is much smaller than the number itself, and logarithms are therefore useful for describing huge numbers. We can also use logarithms to describe very small numbers,

$$\log_{10}(0.0001) = \log_{10} 10^{-4} = -4 \; .$$

If x is a decimal less than one with N consecutive zeros to the right of the decimal place, then $\log_{10} x$ is a negative number which lies between $-(N+1)$ and $-N$. This is particularly useful in the context of statistics of biological sequences, since the probabilities involved are often very tiny numbers.

Base e logarithms are called *natural logarithms* where $\mathrm{e} = 2.17828\ldots$ is an irrational number, a number which is not exactly representable as a decimal number. The power of e is so important it is given its own name, the *exponential function*,

$$\mathrm{e}^x = \exp(x) \; .$$

The reason the exponential is such an important function is that it is equal to its own derivative (see section 4), eg. it describes a quantity growing at a rate proportional to its own amount. This is a property common to many biological and chemical processes.

All logarithms are proportional to one another, ie. they only differ by a constant factor. In many formulas we will therefore not need to specify the base of logarithm used, since this will not usually affect the results. If one logged number is larger than another, this will remain the case when we change base.

Important formulas for manipulating powers and logarithms (to any base) are,

$$\begin{aligned}
(a^x)(a^y) &= a^{x+y} \; , \quad \frac{1}{a^x} = a^{-x} \; , \quad \frac{a^x}{a^y} = a^{x-y} \; , \\
\log(xy) &= \log x + \log y \; , \quad \log(x^n) = n \log x \; , \\
\log\left(\frac{x}{y}\right) &= \log x - \log y \; , \quad \log \prod_{i=1}^{N} f(i) = \sum_{i=1}^{N} \log f(i) \; .
\end{aligned}$$

Notice the last equation shows that a product of numbers can be converted into a sum of logarithms. Before the advent of calculators, log tables could be used to help with calculations involving the multiplication and division of large numbers. Logarithms are still useful now to help computers deal with arithmetic involving very large or very small numbers.

3 Vectors and matrices

Bioinformatics often involves data organised into columns and rows of a table. In this case it is natural to describe values as being collected into vectors and matrices. Matrices are also the basic data-type used in the Matlab programming language (standing for **mat**rix **lab**oratory).

We will mostly consider column vectors. For example, we could write a 3-dimensional vector $\boldsymbol{x}$ as,

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

where x_i is the value in the ith row of the vector. Similarly we can describe values in a table using a matrix. For example, we would write a 3×2 matrix $\boldsymbol{A}$ as,

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

where a_{ij} is the value in the ith row and jth column of the matrix. Notice that a column vector is just a matrix with a single column. This is important, because it means we can multiply vectors and matrices together and we can use matrix operations like transposition on vectors.

3.1 Addition and multiplication

We can add vectors and matrices if they are the same size. We just add all the corresponding elements,

$$\begin{pmatrix} 1 & 3 \\ 4 & 5 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} 5 & 4 \\ 10 & 2 \\ 1 & 6 \end{pmatrix} = \begin{pmatrix} 6 & 7 \\ 14 & 7 \\ 3 & 12 \end{pmatrix}.$$

Two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can be multiplied together if the number of rows in B equals the number of columns in A. In this case we write,

$$\boldsymbol{AB} = \boldsymbol{C} \quad \text{where the elements of } \boldsymbol{C} \text{ are } c_{ik} = \sum_{j=1}^{N} a_{ij} b_{jk}$$

where N is the number of columns in $\boldsymbol{A}$ and rows in $\boldsymbol{B}$. Eg.

$$\begin{aligned} \begin{pmatrix} 1 & 3 & 2 \\ 4 & 5 & 1 \end{pmatrix} \begin{pmatrix} 5 & 4 \\ 10 & 2 \\ 1 & 6 \end{pmatrix} &= \begin{pmatrix} (1 \times 5 + 3 \times 10 + 2 \times 1) & (1 \times 4 + 3 \times 2 + 2 \times 6) \\ (4 \times 5 + 5 \times 10 + 1 \times 1) & (4 \times 4 + 5 \times 2 + 1 \times 6) \end{pmatrix} \\ &= \begin{pmatrix} 37 & 22 \\ 71 & 32 \end{pmatrix}. \end{aligned}$$

3.2 Transpose and dot-product

The transpose of a matrix is a matrix where the rows and columns have been switched. We put a capital T at the top right of a matrix to show that we are taking the transpose, eg.

$$\begin{pmatrix} 1 & 3 & 2 \\ 4 & 5 & 1 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 1 & 4 \\ 3 & 5 \\ 2 & 1 \end{pmatrix} .$$

The transpose of a column vector is a row vector and from the previous section we see that it is possible to multiply together column and row vectors of the same length to get a single number (a scalar), eg.

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \qquad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = x_1 y_1 + x_2 y_2 + x_3 y_3 .$$

This particular form of multiplication is quite common and is given the name *dot-product* or *scalar product*. In general the dot-product of two length N column vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ with elements x_i and y_i is written,

$$\boldsymbol{x} \cdot \boldsymbol{y} \; = \; \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} \; = \; \sum_{i=1}^{N} x_i y_i .$$

The dot-product plays an important role in the definition of linear discriminant functions and neural networks.

3.3 Geometrical interpretation of vectors

It is sometimes helpful to think of vectors in geometrical terms. We can think of a vector as representing a point in an N-dimensional space which is connected to the origin of some axes. In figure 1 we show an example of two 2-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ with angle θ between them. One can show that the dot-product $\boldsymbol{x} \cdot \boldsymbol{y} = |\boldsymbol{x}||\boldsymbol{y}| \cos\theta$ where $|\boldsymbol{x}|$ is the the length (or *magnitude*) of vector $\boldsymbol{x}$, which is defined,

$$|\boldsymbol{x}| \; = \; \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}} \; = \; \sqrt{\sum_{i=1}^{N} x_i^2} .$$

4 Differential calculus

4.1 Differentiation

Differential calculus is the part of mathematics which deals with smooth functions. In figure 2 we show plots of examples which have already been introduced. Looking at the plots we can see some differences between the functions. All the functions shown except for $y = x^2$ increase with x for their entire range; we call them monotonically

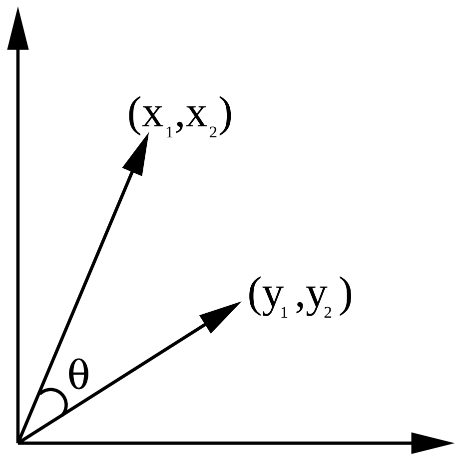

Figure 1: Vectors can be represented geometrically as arrows from the origin to a point with coordinates given by the vector elements. Here we show two 2-dimensional vectors $\boldsymbol{x} = (x_1, x_2)$ and $\boldsymbol{y} = (y_1, y_2)$. The *dot-product* of the two vectors is $\boldsymbol{x} \cdot \boldsymbol{y} = |\boldsymbol{x}||\boldsymbol{y}| \cos\theta$ where the length of vector $\boldsymbol{x}$ is defined $|\boldsymbol{x}| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}$.

increasing functions. Notice that $\ln(x)$ grows slower than $y = x$ while $\exp(x)$ grows much faster. An interesting feature of the quadratic function $y = x^2$ is that it has a minimum point at $x = 0$. Differential calculus can be used to describe these kinds of features for all sorts of smooth functions.

One of the most fundamental quantities in differential calculus is the *derivative* of a function. The derivative of a function at a point is the slope or *gradient* of the function at that point, ie. how much y changes for a small constant increase in x. In fact it is the ratio of the change in y over the increase in x. Points which are in regions of the plot which slope up to the right are points with positive derivative (gradient). Points which are in regions of the plot which slope down to the right are points with negative derivative (gradient). Points which are in flat regions or lie between sloping regions have zero derivative (gradient).

In figure 3 we have plotted $y = f(x)$ with $f(x) = 2x - x^3$, a cubic polynomial. The gradient of the curve at $x = 1/\sqrt{3}$ is one, ie. there is a slope of 45^o at this point. There are some equivalent notations for derivative,

$$f'(x) \quad \text{or} \quad \frac{d}{dx}f(x) \quad \text{or} \quad \frac{dy}{dx} \quad (\text{when } y = f(x)) \,.$$

The notation on the right reminds us that the gradient is the ratio of a small change in y, dy to a small change in x, dx. The notation on the left reminds us that the derivative is different at different parts of the function, so the derivative itself is a function of x. Below we will show how to calculate the derivative of a specific function.

One important use for derivatives is in determining the *stationary points* of a function. Examples of stationary points are the maxima and minima of the function. At these points the derivative vanishes. In figure 4 we show a maximum and minimum of $f(x) = 2x - x^3$ which are at $x = \pm\sqrt{2/3}$. These are actually *local* or *relative* minima

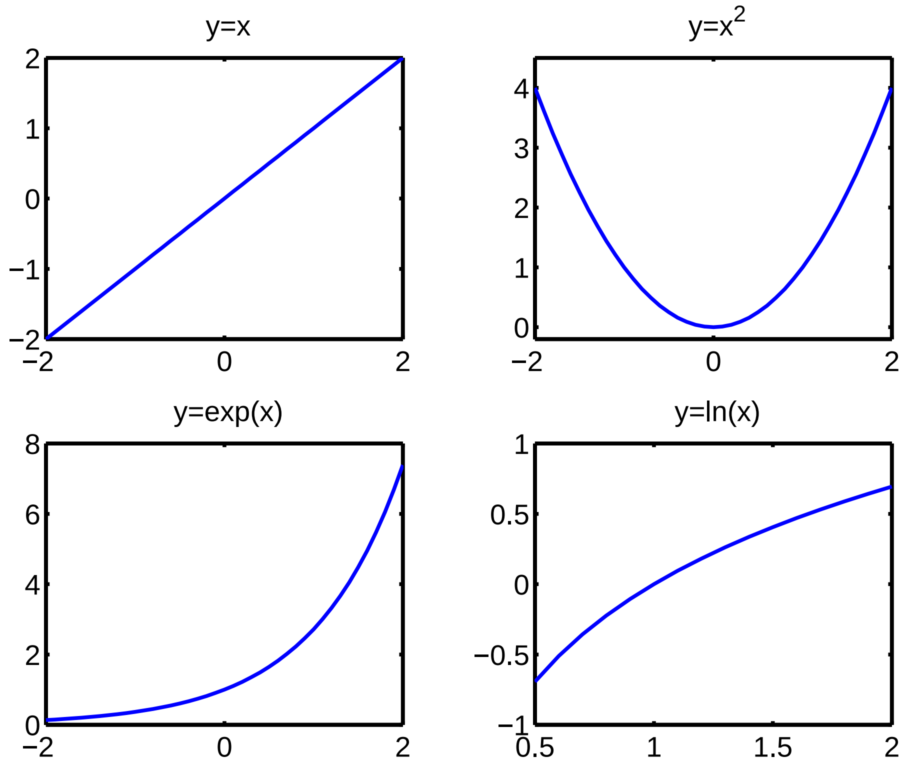

Figure 2: We plot some of the most fundamental functions described in the text. Notice that $\ln(x)$ grows slower than $y = x$ while $\exp(x)$ grows much faster. The quadratic function $y = x^2$ grows faster than linear for positive x but not as fast as $\exp(x)$, and has a minimum at $x = 0$. The other functions are monotonically increasing with x, ie. they increase continuously and do not have any maxima or minima.

and maxima because the function does take higher and lower values for other values of x. This should be contrasted with the minimum of $y = x^2$ shown in figure 2 which is a *global* minimum, ie. at $x = 0$ the function reaches its lowest point over the entire range of x. We can find stationary points by setting the derivative to zero and solving the corresponding equation for x.

We can calculate the derivative of any function using a small number of rules. For example,

$$\begin{aligned}
\frac{d}{dx}\left(a f(x) + b\, g(x)\right) &= a\,\frac{d}{dx} f(x) + b\,\frac{d}{dx} g(x)\ , \\
\frac{d}{dx}\left(x^n\right) &= n x^{n-1}\ , \\
\frac{d}{dx}\exp(ax) &= a\exp(ax)\ ,
\end{aligned}$$

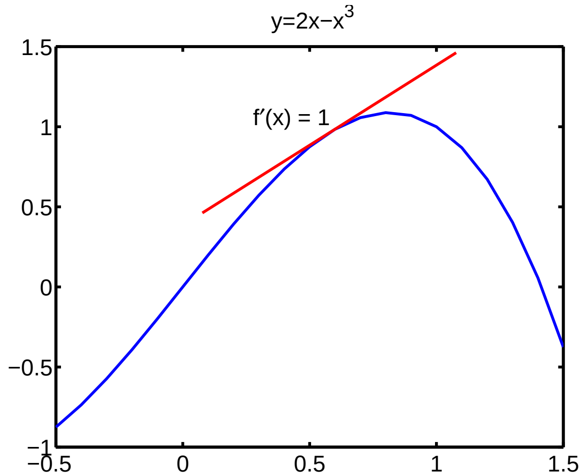

Figure 3: We plot a cubic polynomial. The gradient at the point $x = 1/\sqrt{3}$ is one ($f'(x) = 1$), corresponding to a 45^o slope to the right.

$$\frac{d}{dx}\ln(x) \quad = \quad \frac{1}{x} \, ,$$

where a and b are numbers or possibly other functions which are independent of x. We can use these rules to calculate the derivative of many functions. For example, consider the polynomial plotted in figures 3 and 4, $f(x) = 2x - x^3$,

$$\frac{d}{dx}\left(2x - x^3\right) \; = \; 2\frac{d}{dx}x - \frac{d}{dx}x^3 \; = \; 2 - 3x^2$$

where we have used the top two rules above. We can now work out why the points in the figures have the derivatives shown. Setting the derivative to one we get,

$$\begin{aligned} f'(x) &= 2 - 3x^2 \; = \; 1 \\ \rightarrow \; 3x^2 &= 1 \\ \rightarrow \quad x &= \pm\frac{1}{\sqrt{3}} \, . \end{aligned}$$

The positive solution $x = 1/\sqrt{3}$ is the point highlighted in figure 3. A similar equation can be used to determine positions for the maximum and minimum in figure 4. We just set the derivative to zero in this case.

Often we will want to take the derivative of a function containing another function. In this case it is useful to use the primed notation $f'(x)$ for the derivative,

$$\frac{d}{dx}g(f(x)) = f'(x)g'(f(x)) \, .$$

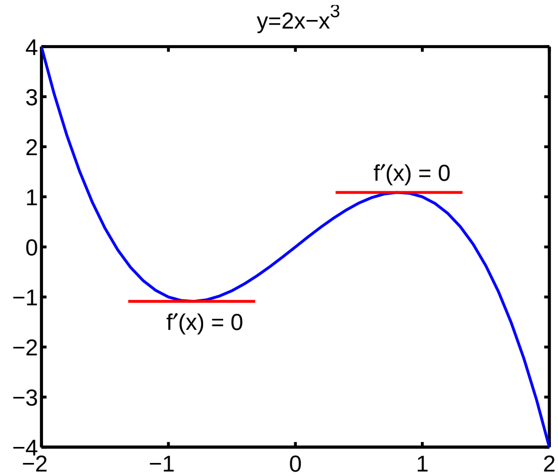

Figure 4: We plot a cubic polynomial which has a local maximum at $x = \sqrt{2/3}$ and a local minimum at $x = -\sqrt{2/3}$. The derivative is zero at these points.

For example,

$$\frac{d}{dx}\exp(f(x)) = f'(x)\exp(f(x)) \quad \text{and} \quad \frac{d}{dx}\ln(f(x)) = \frac{f'(x)}{f(x)} \; .$$

4.2 Integration

Integration is the inverse of differentiation. If $f(x)$ is a function of x with derivative $f'(x)$ then,

$$f(x) = \int f'(x)\; dx \; .$$

A definite integral is one where the limits of integration are specified. If we have a graph of a function $f(x)$ then we can integrate in the range a to b in order to calculate the area under the function in this range,

$$\text{Area under } f(x) \text{ between } a \text{ and } b \; = \; \int_a^b f(x) \; .$$

If the function becomes negative then the integral is the area above zero and below $f(x)$ minus the area below zero and above $f(x)$. For example, the integral of the polynomial $2x - x^3$ between $x = -1$ and $x = 1$ is given by the blue area minus the red area in figure 5.

$$\text{Blue area} - \text{red area} = \int_{-1}^{1} 2x - x^3 \; dx \; .$$

Since the function is anti-symmetric about $x = 0$ (ie. $f(x) = -f(-x)$ which we call an *odd function*) the red and blue areas cancel and the integral equals zero. In general

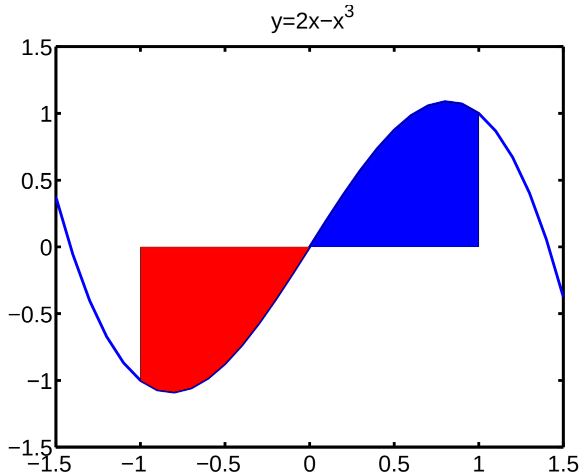

Figure 5: The integral of $f(x) = 2x - x^3$ between $x = -1$ and $x = 1$ is equal to the area shaded blue minus the area shaded red.

integrals are harder to compute than derivatives and we will not need to know how to compute integrals in this module.

4.3 Functions of many variables

More often than not the function we are interested in will involve many variables. In this case we can think of a function as a surface in a space of dimension $N+1$ where N is the number of variables. For example, we can picture a function of two variables $f(x_1, x_2)$ as a surface in three dimensions, ie. consider a 3-dimensional set of axes where the first two axes represent x_1 and x_2 while the third axis represents $f(x_1, x_2)$. For higher dimensions we can no longer picture the function, but the same principles apply. The gradient is no longer a number in this case, but a vector pointing in the steepest upwards direction along the surface. The length of the vector gives us some indication of the steepness.

In general we have a function $f(\boldsymbol{x})$ whose argument is a vector $\boldsymbol{x}$ with components $x_1, x_2 \ldots, x_N$. The gradient vector of $f(\boldsymbol{x})$ is written $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$ and is defined as,

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_N} \end{pmatrix} .$$

This is a generalisation of the gradient in one dimension which is a number (or scalar). The $\partial/\partial x_i$ is a *partial derivative* which means the derivative is with respect to x_i while treating all other variables as constants. We can determine stationary points (including maxima and minima) of a function with N variables by setting each element of the gradient vector to zero.