Mathematical modelling

Lecture 4, March 12, 2021

Faculty of Computer and Information Science
University of Ljubljana

2020/2021

3. Nonlinear models

General formulation

Given is a sample of points $\{(x_1, y_1), \ldots, (x_m, y_m)\}$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$.

The mathematical model is nonlinear if the function

$$y = F(x, a_1, \ldots, a_p) \tag{1}$$

is a nonlinear function of the parameters a_i. This means it cannot be written in the form

$$y = a_1 f_1(x) + a_2 f_2(x) + \ldots + a_p f_p(x),$$

where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is some function.

Plugging each data points into (1) we obtain a system of nonlinear equations

$$\begin{aligned} y_1 &= F(x_1, a_1, \ldots, a_p), \\ &\vdots \\ y_m &= F(x_m, a_1, \ldots, a_p), \end{aligned} \tag{2}$$

in the parameters $a_1, \ldots, a_p \in \mathbb{R}$.

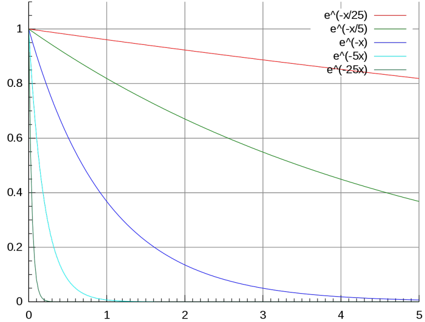
Examples

1. Exponential decay or growth: $F(x, a, k) = ae^{kx}$, a and k are parameters.

   A quantity y changes at a rate proportional to its current value, which can be described by the differential equation

$$\frac{dy}{dx} = ky.$$

   The solution to this equation (obtained by the use of separation of variables) is $y = F(x, a, k)$.

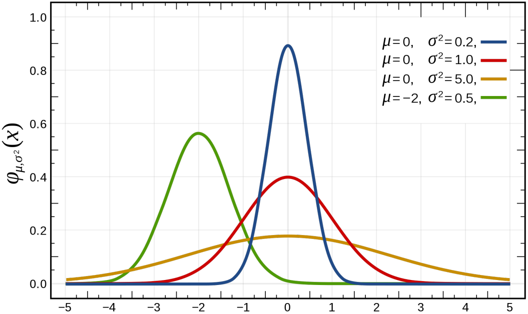


2. Gaussian model: $F(x, a, b, c) = ae^{-\left(\frac{x-b}{c}\right)^2}$, $a, b, c \in \mathbb{R}$ parameters.

   a is the value of the maximum obtained at $x = b$ and c determines the width of the curve.

   It is used in statistics to describe the normal distribution, but also in signal and image processing.

   In statistics $a = \frac{1}{\sigma\sqrt{2\pi}}$, $b = \mu$, $c = \sqrt{2\sigma}$, where μ, σ are the expected value and the standard deviation of a normally distributed random variable.

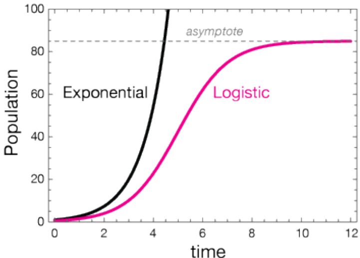


3. Logistic model: $F(x, a, b, k) = \frac{a}{(1+be^{-kx})}, k > 0$

The logistic function was devised as a model of population size by adjusting the exponential model which also considers the saturation of the environment, hence the growth first changes to linear and then stops.

The logistic function $F(x, a, b, k)$ is a solution of the first order non-linear differential equation

$$\frac{dy(x)}{dx} = ky(x)\Big(1 - \frac{y(x)}{a}\Big).$$



4. In the area around a radiotelescope the use of microwave owens is forbidden, since the radiation interferes with the telescope. We are looking for the location (a, b) of a microwave owen that is causing problems.

   The radiation intensity decreases with the distance from the source r according to $u(r) = \dfrac{\alpha}{1+r}$.

   Measured values of the signal at three locations are $z(0,0) = 0.27$, $z(1,1) = 0.36$ in $z(0,2) = 0.3$.

   This gives the following system of equations for the parameters α, a, b:

$$\begin{aligned}
\frac{\alpha}{1+\sqrt{a^2+b^2}} &= 0.27, \\
\frac{\alpha}{1+\sqrt{(1-a)^2+(1-b)^2}} &= 0.36, \\
\frac{\alpha}{1+\sqrt{a^2+(2-b)^2}} &= 0.3.
\end{aligned}$$

- Our goal is to fit the data points

$$\{(x_1, y_1), \ldots, (x_m, y_m)\}, \quad x_i \in \mathbb{R}^n, \; y_i \in \mathbb{R}.$$

- We choose a fitting function

$$F(x, a_1, \ldots, a_p)$$

which depends on the unknown parameters $a_1, \ldots, a_p$.

- Equivalent formulation of the system (2) (which will be more suitable for solving with numerical algorithms) is:
  1. For $i = 1, \ldots, m$ define the functions

     $$g_i : \mathbb{R}^p \to \mathbb{R} \qquad \text{by the rule} \qquad g_i(a_1, \ldots, a_p) = y_i - F(x_i, a_1, \ldots, a_p).$$

  2. Solve or approximate the following system by the least squares method

$$\begin{aligned} g_1(a_1, \ldots, a_p) &= 0, \\ &\vdots \\ g_m(a_1, \ldots, a_p) &= 0. \end{aligned} \tag{3}$$

In a compact way (3) can be expressed by introducing a vector function

$$G : \mathbb{R}^p \to \mathbb{R}^m, \quad G(a_1, \ldots, a_p) = (g_1(a_1, \ldots, a_p), \ldots, g_m(a_1 \ldots, a_p)), \quad (4)$$

and search for the tuples $(a_1, \ldots, a_p)$ that solve the system (or minimize the norm of the left-hand side)

$$G(a_1, \ldots, a_p) = (0, \ldots, 0). \quad (5)$$

Remark

Solving (5) *is a difficult problem. Even if the exact solution exists, it is not easy (or even impossible) to compute. For example, there does not even exist an analytic formula to determine roots of a general polynomial of degree 5 or more.*

Our plan: Learn some numerical algortihms to *approximate* the solutions of (5).

3.1 Vector functions of a vector variable

Neccessary terminology to achieve our plan

G from (4) is an example of

- a vector function: since it maps into $\mathbb{R}^m$, where m might be bigger than 1.
- a vector variable: since it maps from $\mathbb{R}^p$, where p might be bigger than 1.

Remark

- *If $m = 1$ and $p > 1$, then G is a usual multivariate function.*
- *If $m = 1$ and $p = 1$, then G is a usual (univariate) function.*

For easier reference in the continuation we call $g_1, \ldots, g_m$ from (4) the component (or coordinate) functions of G.

Examples

1. A linear vector function $G : \mathbb{R}^n \to \mathbb{R}^m$ is such that all the component functions g_i are linear:

$$g_i(x_1, \ldots, x_n) = a_{i1} \cdot x_1 + a_{i2} \cdot x_2 + \ldots + a_{in} \cdot x_n, \quad \text{where } a_{ij} \in \mathbb{R}. \quad (6)$$

In this case

$$G(x) = Ax,$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}.$$

2. Adding constants $b_i \in \mathbb{R}$ to the left side of (6) we get the definition of an affine linear vector function,

$$g_i(x_1, \ldots, x_n) = a_{i1}x_1 + a_{i2}x_2 + \ldots a_{in}x_n + b_i,$$

and then

$$G(x) = Ax + b, \qquad \text{where} \quad b = \begin{bmatrix} b_1 & b_2 & \ldots & b_n \end{bmatrix}^T.$$

3. Most of the (vector) functions are nonlinear, e.g.,

$$
\begin{aligned}
f : \mathbb{R}^3 \to \mathbb{R}^2, &\quad f(x, y, z) = (x^2 + y^2 + z^2 - 1, x + y + z), \\
g : \mathbb{R}^2 \to \mathbb{R}^3, &\quad g(z, w) = (zw, \cos z + w^2 - 2, e^{2z}), \\
h : \mathbb{R} \to \mathbb{R}^2, &\quad h(t) = (t + 3, e^{-3t}).
\end{aligned}
$$

Derivative of a vector function - is needed in the algorithms we will use

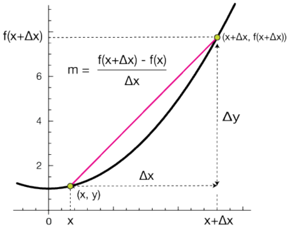
The derivative of a vector function $F : \mathbb{R}^n \to \mathbb{R}^m$ in the point

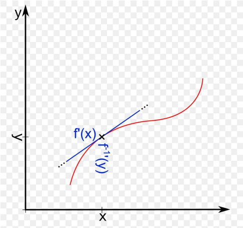
$$a := (a_1, \ldots, a_n) \in \mathbb{R}^n$$

is called the Jacobian matrix of F in a:

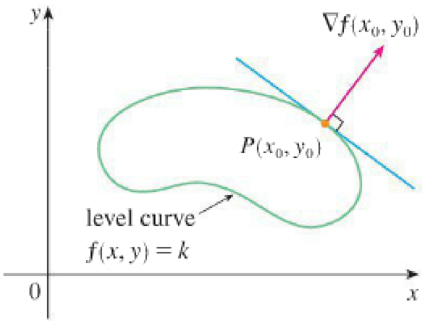
$$J_F(a) = DF(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{bmatrix}.$$

- If $n = m = 1$, the $Df(x) = f'(x)$ is the usual derivative.





- For general n and $m = 1$, f is a function of n variables and
$$Df(x) = \text{grad } f(x)$$
is its gradient.



- For general m and n, $Df(x) = \begin{bmatrix} \text{grad } f_1 \\ \vdots \\ \text{grad } f_m \end{bmatrix}$ is a vector of gradients of component functions.

1. For an affine linear function $f : \mathbb{R}^n \to \mathbb{R}^m$, given by $f(x) = Ax + b$, it is easy to check that
$$Df(x) = A.$$
2. For a vector function $f : \mathbb{R}^3 \to \mathbb{R}^2$, given by
$$f(x, y, z) = (x^2 + y^2 + z^2 - 1, x + y + z),$$
then
$$Df(x) = \begin{bmatrix} 2x & 2y & 2z \\ 1 & 1 & 1 \end{bmatrix}.$$

A linear approximation of the vector function $f : \mathbb{R}^n \to \mathbb{R}^m$ at the point $a \in \mathbb{R}^n$ is the affine linear function

$$L_a : \mathbb{R}^n \to \mathbb{R}^m, \quad L_a(x) = Ax + b$$

that satisfies the following conditions:

1. It has the **same value** as f in a: $L_a(a) = f(a)$.
2. It has the **same derivative** as f at a: $DL_a(a) = Df(a)$.

It is easy to check that

$$L_a(x) = f(a) + Df(a)(x - a).$$

- $n = m = 1$:
$$L_a(x) = f(a) + f'(a)(x - a)$$
The graph $y = L_a(x)$ is the tangent to the graph $y = f(x)$ at the point a.

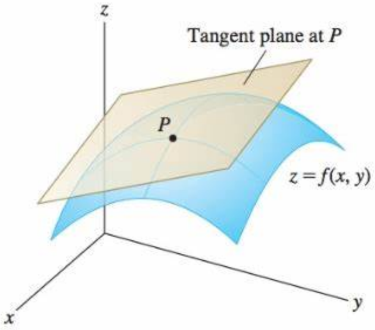
- If $n = 2$ and $m = 1$, then

$$L_{(a,b)}(x, y) = f(a, b) + \operatorname{grad} f(a, b) \begin{bmatrix} x - a \\ y - b \end{bmatrix}.$$

The graph

$$z = L_{(a,b)}(x, y)$$

is the tangent plane to the surface $z = f(x, y)$ at the point (a, b).



Example

The linear approximation of the function

$$f : \mathbb{R}^3 \to \mathbb{R}^2, \qquad f(x, y, z) = (x^2 + y^2 + z^2 - 1, x + y + z)$$

at $a = (1, -1, 1)$ is the affine linear function

$$\begin{aligned}
L_a(x, y, z) &= f(1, -1, 1) + Df(1, -1, 1) \begin{bmatrix} x-1 \\ y+1 \\ z-1 \end{bmatrix} \\
&= \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & -2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x-1 \\ y+1 \\ z-1 \end{bmatrix} \\
&= \begin{bmatrix} 2 + 2(x-1) - 2(y+1) + 2(z-1) \\ 1 + (x-1) + (y+1) + (z-2) \end{bmatrix} \\
&= \begin{bmatrix} 2 & -2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} -4 \\ 0 \end{bmatrix}.
\end{aligned}$$

3.2 Solving systems of nonlinear equations

Let $f : D \to \mathbb{R}^m$ be a vector function, defined on some set $D \subset \mathbb{R}^n$.

We will study the Gauss-Newton method to solve the system $f(x) = 0$, which is one of the numerical methods for searching approximate of this system. It is based on linear approximations of f.

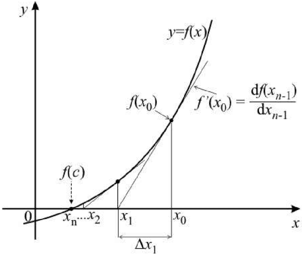
We are searching zeroes of the function $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}$, i.e., we are solving $f(x) = 0$.

Newton's or tangent method:

We construct a recursive sequence with:

- x_0 is an initial term,
- x_{k+1} is a solution of

$$L_{x_k}(x) = f(x_k) + f'(x_k)(x - x_k) = 0, \text{ so } x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$



Theorem

The sequence x_i converges to a solution α, $f(\alpha) = 0$, if:

(1) $0 \neq |f'(x)|$ *for all* $x \in I$, *where* I *is some interval containing* α,

(2) $x_0 \in I$.

Under these assumptions the convergence is quadratic, meaning that:

$$\textit{If we denote by } \varepsilon_j = |x_j - \alpha|, \textit{ then } \varepsilon_{i+1} \leq M\varepsilon_i^2,$$

where M *is some constant. If* f *is twice differentiable, then*

$$M \leq \max_{x \in I} |f''(x)| / \min_{x \in I} |f'(x)|.$$

Proof.

Condition (1) implies in particular that α is a simple zero of f. Plugging α in the Taylor expansion of f around x_i we get

$$\begin{aligned} 0 = f(\alpha) &= f(x_i) + f'(x_i)(\alpha - x_i) + \frac{f''(\eta)}{2}(\alpha - x_i)^2 \\ &= f(x_i) + f'(x_i)(\alpha - x_i) + \frac{f''(\eta)}{2}(\alpha - x_i)^2 \end{aligned} \tag{7}$$

where η is between α and x_i. Dividing (7) with $f'(x_i)$ we get

$$0 = \frac{f(x_i)}{f'(x_i)} - (\alpha - x_i) + \frac{f''(\eta)}{2f'(x_i)} e_i^2$$

and hence

$$\left(x_i - \frac{f(x_i)}{f'(x_i)} \right) - \alpha = x_{i+1} - \alpha = \frac{f''(\eta)}{2f'(x_i)} e_i^2.$$

Thus,

$$e_{i+1} = \left| \frac{f''(\eta)}{2f'(x_i)} \right| e_i^2$$

Now

$$\left|\frac{f''(\eta)}{2f'(x_i)}\right| \leq \frac{\max_{x\in I}|f''(x)|}{\min_{x\in I}|f'(x)|}.$$

To prove that the sequence converges note that there exists $\delta_0 > 0$ such that

$$M\delta_0 < \frac{1}{2}.$$

Hence, if $e_i \leq \delta_0$, then

$$e_{i+1} = \left|\frac{f''(\eta)}{2f'(x_i)}\right| e_i^2 = \frac{1}{2} e_i.$$

Therefore

$$\lim_{n\to\infty} e_n = \lim_{n\to\infty} \frac{1}{2^n} \cdot e_0 = 0.$$

□

Example

Use Newton's method in Matlab to find the square root of 3 with the initial approximation $x_0 = 3$ for different functions:

- $f_1(x) = x^2 - 3$.
- $f_2(x) = 1 - \dfrac{3}{x^2}$.
- $f_3(x) = x - \dfrac{3}{x}$.

What do you notice about convergence of the method?

Newton's method generalizes to systems of n nonlinear equations in n unknowns:

- x_0 – initial approximation,
- x_{k+1} – solution of

$$L_{x_k}(x) = f(x_k) + Df(x_k)(x - x_k) = 0,$$

so

$$x_{k+1} = x_k - Df(x_k)^{-1} f(x_k).$$

In practice inverses are difficult to calculate (require to many operations) and the linear system for $\Delta x_k = x_{k+1} - x_k$

$$Df(x_k)\Delta x_k = -f(x_k)$$

is solved at each step (using LU decomposition of $Df(x_k)$) and hence

$$x_{k+1} = x_k + \Delta x_k.$$

Example

Derive Newton's method for solving the system of quadratic equations:

$$x^2 + y^2 - 10x + y = 1,$$
$$x^2 - y^2 - x + 10y = 25.$$

We are searching for the zero of the vector function

$$F : \mathbb{R}^2 \to \mathbb{R}^2, \quad F(x, y) = (x^2 + y^2 - 10x + y - 1, x^2 - y^2 - x + 10y - 25).$$

The Jacobian of F in (x, y) is

$$DF(x, y) = \begin{bmatrix} 2x - 10 & 2x - 1 \\ 2y + 1 & -2y + 10 \end{bmatrix}.$$

Using Newton's metod we:

- Choose an initial term (x_0, y_0).
- Calculate $x_{r+1} = x_r + \Delta x_r$, where $DF(x_r, y_r)\Delta x_r = -F(x_r, y_r)^T$.