Mathematical modelling

Chapter 2

Nonlinear models and geometric models

Faculty of Computer and Information Science
University of Ljubljana

2017/2018

3. Nonlinear models

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* is the function

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

is a nonlinear function of the parameters a_i.

Each data point gives a nonlinear equation

$$y_i = F(x_i, a_1, \ldots, a_p), i = 1, \ldots, m.$$

Examples of nonlinear models:

1. *Exponential decay* $F(x, a, k) = ae^{-kx}$ ali *rast* $F(x, a, k) = a(1 - e^{-kx}),\ k > 0$
2. *Gaussian model*: $F(x, a, b, c) = ae^{-\left(\frac{x-c}{b}\right)^2}$
3. *Logistic model*: $F(x, a, b, c) = \frac{a}{(1+be^{-kx})}, k > 0$

Given the data points $\{(x_1, y_1), \dots, (x_m, y_m)\}$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$ we obtain a system of nonlinear equations for the parameters a_i:

$$f_i(a_1, \dots, a_p) = y_i - F(x_i, a_1, \dots, a_p) = 0, i = i, \dots, m.$$

Solving a system of nonlinear equations is a tough problem (even for $n = m = 1$) ...

Solutions are zeroes of a nonlinear vector function

$$f : \mathbb{R}^p \to \mathbb{R}^m$$

$$f(a_1, \dots, a_p) = (f_1(a_1, \dots, a_p), \dots, f_m(a_1 \dots, a_p)).$$

One strategy is to approximate these by zeroes of suitable linear approximations.

Example:
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 $z(0, 2) = 0.3$.

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

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

3.1 Vector functions of a vector variable

Let $f : D \to \mathbb{R}^m$, where $D \subset \mathbb{R}^n$:

$$f(x) = \begin{bmatrix} f_1(x)) \\ \vdots \\ f_m(x) \end{bmatrix} = \begin{bmatrix} f_1(x_1, \ldots, x_n)) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{bmatrix}$$

Special cases:

- $m = 1$: $f(x) = f(x_1, \ldots, x_n)$, is a scalar function of n variables
- $n = 1, m = 2$: $f(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix}$ represents a *parametric curve* in the plane $\mathbb{R}^2$
- $n = 1, m$ general: $f(t) = \begin{bmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{bmatrix}$ represents a *parametric curve* in $\mathbb{R}^n$
- $n = 2, m = 3$: f represents a *parametric surface* in $\mathbb{R}^3$.

The *derivative* of a vector function f is given by the *Jacobian matrix*:

$$J = Df = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \mathrm{grad} f_1 \\ \vdots \\ \mathrm{grad} f_m \end{bmatrix}$$

- $m = 1$, $n = 1$: $Df(x) = f'(x)$
- $m = 1, n$ general: $Df(x) = \mathrm{grad} f(x)$.
- n and m general: rows of Df contain gradients of the components f_i of f.

The *linear approximation* of f at the point a is the linear fuction that has the same value and the same derivative as f at a:

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

- $n = 1$, $m = 1$:

  $$L_a(x) = f(a) + f'(a)(x - a)$$

  is the linear approximation of a function of one variable (which you know from Calculus), its graph $y = L_a(x)$ is the tangent to the graph $y = f(x)$ at the point a,
- $n = 2$, $m = 1$, i.e. $f(x, y)$ is a function of two variables:

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

  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
Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$f(x, y, z) = \begin{bmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{bmatrix}.$$

At the point $(1, -1, 1)$

- the Jacobian matrix of f is

$$Df = \begin{bmatrix} 2x & 2y & 2z \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -2 & 2 \\ 1 & 1 & 1 \end{bmatrix}$$

- and the linear approximation to f is

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

Geometric picture:

every point (x_0, y_0, z_0) lies in the intersection of level surfaces

$$f_1(x, y, z) = f_1(x_0, y_0, z_0) \quad \text{in} \quad f_2(x, y, z) = f_2(x_0, y_0, z_0),$$

that is in our case

$$x^2 + y^2 + z^2 - 1 = x_0^2 + y_0^2 + z_0^2 - 1 \quad \text{in} \quad x + y + z = x_0 + y_0 + z_0.$$

The intersection of two surfaces in $\mathbb{R}^3$ determines an implicit curve in $\mathbb{R}^3$.

The vectors $\operatorname{grad} f_1(x_0, y_0, z_0)$ and $\operatorname{grad} f_2(x_0, y_0, z_0)$ (if they are nonzero) point each in the direction orthogonal to the corresponding level surface.

If nonzero, the vector

$$\operatorname{grad} f_1(x_0, y_0, z_0) \times \operatorname{grad} f_2(x_0, y_0, z_0)$$

points in the direction tangential to the implicit curve.

3.2 Solving systems of nonlinear equations

$f : D \to \mathbb{R}^m, D \subset \mathbb{R}^n$

We are looking for solutions of

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

In many cases an analytic solution does not even exist.

A number of *numerical methods* for approximate solutions is available. We will look at one, based on linear approximations.

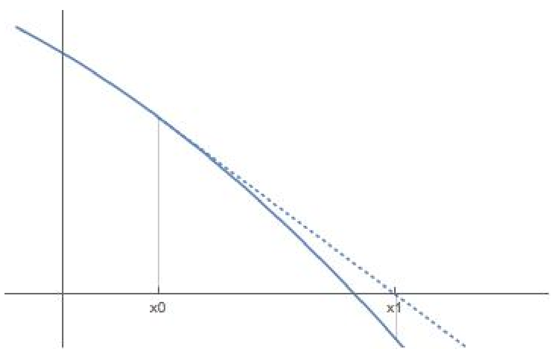
$n = 1, m = 1$: solving an equation $f(x) = 0$, $x \in \mathbb{R}$.

Newton's or *tangent method*:

We construct a recurssive sequence with

- x_0 initial term
- x_{k+1} solution of

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



so $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$

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

1. $f'(x) \neq 0$ for all $x \in I$, where I is an interval $[\alpha - r, \alpha + r]$ for some $r \geq |(\alpha - x_0)|$,
2. $f''(x)$ is continuous for all $x \in I$,
3. x_0 is close enough to the solution α.

Under these assumptions the convergence is *quadratic*:

if $\varepsilon_i = |x_i - \alpha|$ then $\varepsilon_{i+1} \leq M\varepsilon_i^2$,

where M is a constant bounded by $|f''(x)|/f'(x)$ on I.

$m = n > 1$:
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 the linear system for x_{k+1}

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

is solved at each step.

The sequence converges to a solution α if for some $r > 0$ the matrix $Df(x)$ is nonsingular for all x, $\|x - \alpha\| < r$, and $\|x_0 - \alpha\| < r$.

$m > n > 0$:
an overdetermined system $f(x) = 0$ of m nonlinear equations for n unknowns.

The system $f(x) = 0$ generally does not have a solution.

We are looking for a best fit to a solution, that is for α such that the distance of $f(\alpha)$ from 0 is the smallest possible:

$$\|f(\alpha)\|^2 = \min\{\|f(x)\|^2\}.$$

Gauss-Newtonova method is a generalization of the Newton method where instead of yhe inverse of the Jacobian its MP inverse is used at each step:

- x_0 initial approximation
- $x_{k+1} = x_k - Df(x_k)^+ f(x_k)$,

where $Df(x_k)^+$ is the MP inverse of $Df(x_k)$. If the matrix

$(Df(x_k)^T Df(x_k))$ is nonsingular at each step k then

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

At each step x_{k+1} is the least squares approximation to the solution of the overdetermined linear system $L_{x_k}(x) = 0$, that is,

$$\|L_{x_k}(x_{k+1})\|^2 = \min\{\|L_{x_k}(x)\|^2, x \in \mathbb{R}^n\}.$$

Convergence is not guaranteed, but:

- if the sequence x_k converges, the limit $x = \lim_k x_k$ is a local (but not nocessarily global) minimum of $\|f(x)\|^2$.

So: the Gauss-Newtonova method is an algorithm for the local minimu of $\|f(x)\|^2$.