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 if 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

\[ y_i = F(x_i, a_1, \ldots, a_p), \quad i = 1, \ldots, m \]

for the parameters \( a_1, \ldots, a_p \).

Examples of nonlinear models:

1. **Exponential decay** \( F(x, a, k) = ae^{-kx} \) ali rast
   \( F(x, a, k) = a(1 - e^{-kx}), \quad k > 0 \)
2. **Gaussian model**: \( F(x, a, b, c) = ae^{-\left(\frac{x-c}{b}\right)^2} \)
3. **Logistic model**: \( F(x, a, b, c) = \frac{a}{1 + be^{-kx}}, \quad k > 0 \)
Given the data points \( \{ (x_1, y_1), \ldots, (x_m, y_m) \} \), \( x_i \in \mathbb{R}^n, y_i \in \mathbb{R} \) we obtain a system of nonlinear equations for the parameters \( a_i \):

\[
f_i(a_1, \ldots, a_p) = y_i - F(x_i, a_1, \ldots, a_p) = 0, \quad i = i, \ldots, m.
\]

Solutions are zeroes of a nonlinear vector function

\[
f : \mathbb{R}^p \rightarrow \mathbb{R}^m
\]

\[
f(a_1, \ldots, a_p) = (f_1(a_1, \ldots, a_p), \ldots, f_m(a_1, \ldots, a_p)).
\]

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

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

Example of a nonlinear model:
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 \( u(r) = \frac{\alpha}{1 + r} \).

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 \( \alpha, a, b \):

\[
\frac{\alpha}{1 + \sqrt{a^2 + b^2}} = 0.27
\]

\[
\frac{\alpha}{1 + \sqrt{(1 - a)^2 + (1 - b)^2}} = 0.36
\]

\[
\frac{\alpha}{1 + \sqrt{a^2 + (2 - b)^2}} = 0.3
\]
3.1 Vector functions of a vector variable

Let $f$ be a function from $D \subset \mathbb{R}^n$ to $\mathbb{R}^m$,

$$f\text{ maps a vector } x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in D \text{ to a vector } f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}$$

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

Examples:

1. A linear vector function $f : \mathbb{R}^n \to \mathbb{R}^m$ is given by
   $f : x \mapsto Ax + b$, where $A$ is a matrix of order $m \times n$ and $b \in \mathbb{R}^m$.

2. A nonlinear vector function $f : \mathbb{R}^3 \to \mathbb{R}^2$ could, for example, be given by
   $$f : \begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{bmatrix}.$$
The derivative of a vector function \( f \) is given by the Jacobian matrix:

\[
J = Df = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n}
\end{bmatrix}
\]

- If \( m = 1, n = 1 \), \( f \) is a function from \( D \subset \mathbb{R} \) to \( \mathbb{R} \) and \( Df(x) = f'(x) \).
- For \( m = 1 \) and \( n \) general \( f \) is a function of \( n \) variables and \( Df(x) = \text{grad} f(x) \).
- For \( n \) and \( m \) general \( Df(x) = \begin{bmatrix} \text{grad} f_1 \\ \vdots \\ \text{grad} f_m \end{bmatrix} \).

Examples

1. If \( f : \mathbb{R}^n \to \mathbb{R}^m \) is the linear function \( x \mapsto Ax + b \), then \( Df(x) = A \).
2. If \( f : \mathbb{R}^3 \to \mathbb{R}^2 \) is given by

\[
f : \begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{bmatrix}
\]

then

\[
Df(x) = \begin{bmatrix} 2x & 2y & 2z \\ 1 & 1 & 1 \end{bmatrix}
\]
The *linear approximation* of $f$ at the point $a$ is the linear function 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:

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

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:

The linear approximation to the function

$$f : \begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{bmatrix}$$

at $a = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

is the linear function

$$L_a(x, y, z) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & -2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x - 1 \\ y + 1 \\ z - 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 + 2(x - 1) - 2(y + 1) + 2(z - 1) \\ 1 + (x - 1) + (y + 1) + (z - 2) \end{bmatrix} = \begin{bmatrix} 2 & -2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 4 \\ 0 \end{bmatrix}.$$
Geometric picture:

Given a vector function \( f : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \), every point \((x_0, y_0, z_0)\) lies in the intersection of level surfaces

\[
f_1(x, y, z) = c_1 \quad \text{in} \quad f_2(x, y, z) = c_2,
\]

where \( c_1 = f_1(x_0, y_0, z_0) \) and \( c_2 = f_2(x_0, y_0, z_0) \).

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

If they are nonzero, the vectors \( \text{grad} f_1(x_0, y_0, z_0) \) and \( \text{grad} f_2(x_0, y_0, z_0) \) are normal vectors of the two level surfaces, and

\[
\text{grad} f_1(x_0, y_0, z_0) \times \text{grad} f_2(x_0, y_0, z_0)
\]

is tangential to the implicit curve.

For the function

\[
f : \begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{bmatrix}
\]

the implicit curve through \((1, -1, 1)\) is given by

\[
x^2 + y^2 + z^2 - 1 = 2 \quad \text{in} \quad x + y + z = 1,
\]

and the tangent vector is

\[
\text{grad} f_1(1, -1, 1) \times \text{grad} f_2(1, -1, 1) = \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ 0 \\ 4 \end{bmatrix}.
\]
3.2 Solving systems of nonlinear equations

\( f : D \rightarrow \mathbb{R}^m, D \subset \mathbb{R}^n \)

We are looking for solutions of

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

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.

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

**Newton's or tangent method:**

We construct a recursive 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 $\alpha$, $f(\alpha) = 0$, 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 $\alpha$.

Under these assumptions the convergence is \textit{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 $\alpha$ if for some $r > 0$ the matrix $Df(x)$ is nonsingular for all $x$, $\|x - \alpha\| < r$, and $\|x_0 - \alpha\| < r$. 
Application to optimization: **Newton optimization method**

Let $F : \mathbb{R}^n \to \mathbb{R}$, we are looking for the minimum (or maximum) of $F$.

The first step is to find the critical points, i.e. solution of

$$f = \text{grad} F = \begin{bmatrix} F_{x_1} \\ \vdots \\ F_{x_n} \end{bmatrix} = 0.$$

This is a system of $n$ equations for $n$ variables, the Jacobian of the vector function $f$ is the Hessian of $F$:

$$Df(x) = H(x) = \begin{bmatrix} F_{x_1 x_1} & \cdots & F_{x_1 x_n} \\ \vdots & \ddots & \vdots \\ F_{x_n x_1} & \cdots & F_{x_n x_n} \end{bmatrix}.$$

If the sequence of iterates

$$x_0, \quad x_{k+1} = x_k - H^{-1}(x_k)\text{grad} F(x_k)$$

converges, the limit is a critical point of $F$, i.e. a candidate for the minimum (or maximum).

$m > n > 0$:
We have 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 $\alpha$ such that the distance of $f(\alpha)$ from 0 is the smallest possible:

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

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

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

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 necessarily global) minimum of $\|f(x)\|^2$.

It follows that the Gauss-Newton method is an algorithm for the local minimum of $\|f(x)\|^2$. 