Mathematical modelling

Lecture 5, March 15, 2022

Faculty of Computer and Information Science
University of Ljubljana

2021/22

Newton optimization method:

We would like to find the extrema of the function $F : \mathbb{R}^n \to \mathbb{R}$.

Since the extrema are *critical (or stationary) points*, the candidates are zeroes of the gradient, i.e.,

$$G(x) := \text{grad } F(x) = \begin{bmatrix} F_{x_1}(x) & \cdots & F_{x_n}(x) \end{bmatrix} = 0. \tag{1}$$

(1) is a system of n equations for n variables, the Jacobian of the vector function G is the so called Hessian of F:

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

If the sequence of iterates

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

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

Gradient descent

Optimization methods can also be used to ensure a sufficiently accurate starting approximation for the Newton-based techniques. (Like bisection does for a single one-variable equation.)

Finding solutions of the system $F(x) = 0$, where

$$F = [F_1, \ldots, F_n]^T : \mathbb{R}^n \to \mathbb{R}^n$$

is equivalent to finding global minima of

$$g(x) := \|F\|^2 = F_1(x)^2 + \ldots + F_n(x)^2 : \mathbb{R}^n \to \mathbb{R}.$$

We search for the local minima (which are not necessarily global minima!) of g as follows:

1. Choose x_0.
2. Determine the constant α in $x_r - \alpha \cdot \mathrm{grad}(g(x_r))$ which mimimizes

$$h(\alpha) = g(x_r - \alpha \cdot \mathrm{grad}(g(x_r))).$$

   (Or is significantly smaller than $h(0) = g(x_r)$.)
3. $x_{r+1} = x_r - \alpha \cdot \mathrm{grad}(g(x_r))$.

Quasi-Newtonov methods: Broyden's method

- For large n, the Newton's method is very expensive, since we need to evaluate n^2 partial derivatives at each step and use $\mathcal{O}(n^3)$ flops $(+, -, \cdot, :)$ to solve the linear system.
- Broyden's method avoids computing derivatives. For $n = m = 1$ it replaces the tangent by a secant throught the last two iterates. It mimicks this idea also for larger $n = m$.

Let B_r be an approximate for $J_f(x_r)$. Broyden's method works as follows:

1. Solve $B_r \Delta x_r = -f(x_r)$,
2. $x_{r+1} = x_r + \Delta x_r$,
3. Determine B_{r+1}.

The last step searches for a matrix B_{r+1}, which fulfils the secant condition:

$$B_{r+1}(x_{r+1} - x_r) = f(x_{r+1}) - f(x_r)$$

and is the closest to B_r in the spectral norm $\|\cdot\|_2$.

It turns out that

$$B_{r+1} = B_r + \frac{f(x_{r+1})(\Delta x_r)^T}{\|\Delta x_r\|_2^2}.$$

Recall from above the microwave oven example. The system of equations for the parameters α, a, b is:

$$\begin{aligned}
\frac{\alpha}{1+\sqrt{a^2+b^2}} - 0.27 &= 0 \\
\frac{\alpha}{1+\sqrt{(1-a)^2+(1-b)^2}} - 0.36 &= 0 \\
\frac{\alpha}{1+\sqrt{a^2+(2-b)^2}} - 0.3 &= 0.
\end{aligned}$$

https://zalara.github.io/Algoritmi/newtonsys.m
https://zalara.github.io/Algoritmi/broyden.m
https://zalara.github.io/Algoritmi/gradient_descent.m
https://zalara.github.io/Algoritmi/test_newtonsys_2.m

We have an overdetermined system

$$f : \mathbb{R}^n \to \mathbb{R}^m, \quad f(x) = (0, \dots, 0) \tag{2}$$

of m nonlinear equations for n unknowns, where $m > n$.

The system (2) generally does not have a solution, so we are looking for a solution of (2) by the least squares method, i.e., $\alpha \in \mathbb{R}^n$ such that the distance of $f(\alpha)$ from the origin is the smallest possible:

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

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

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

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 $\|f(x)\|^2$.

Example

We are given point $(x_i, y_i) \in \mathbb{R}^2$ for $i = 1, \ldots, m$ and are searching for the function

$$f(x, a, b) = ae^{bx}$$

which fits this data best by the method of least squares.

So we have the overdetermined system $F(a, b) = 0$, where

$$F : \mathbb{R}^2 \to \mathbb{R}^m, \quad F(a, b) = (y_1 - ae^{bx_1}, \ldots, y_m - ae^{bx_m}).$$

The Jacobian of F is

$$DF(a, b) = \begin{bmatrix} -e^{bx_1} & ax_1 e^{bx_1} \\ \vdots & \\ -e^{bx_m} & ax_m e^{bx_m} \end{bmatrix}.$$

Using the Gauss-Newton method:

- We choose initianl approximation (a_0, b_0),
- Calculate iterates

$$\begin{bmatrix} a_{r+1} \\ b_{r+1} \end{bmatrix} = \begin{bmatrix} a_r \\ b_r \end{bmatrix} - DF(a_r, b_r)^+ F(a_r, b_r)^T.$$

Chapter 4:

Curves and surfaces

- Curves
  - Definition and examples
  - Derivative
  - Arc length and the natural parametrization
  - Curvature
  - Plotting plane curves
  - Area bounded by plane curves
  - Curves in the polar form
  - Motion in $\mathbb{R}^3$
- Surfaces
  - Definition and examples
  - Cartesian, cylindrical and spherical coordinates
  - Surface of revolution
  - Tangent plane

Curves - definition and examples

A parametric curve (or parametrized curve) in $\mathbb{R}^m$ is a vector function

$$f : I \to \mathbb{R}^m, \qquad f(t) = \begin{bmatrix} f_1(t) \\ \vdots \\ f_m(t) \end{bmatrix},$$

where $I \subset \mathbb{R}$ is a bounded or unbounded interval.

The independent variable (in this case t) is the parameter of the curve.

For every value $t \in I$, $f(t)$ represents a point in $\mathbb{R}^m$.

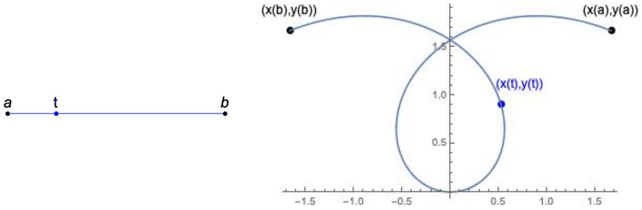
As t runs through I, $f(t)$ traces a path, or a curve, in $\mathbb{R}^m$.

If $m = 2$, then for every $t \in I$,

$$f(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \mathbf{r}(t)$$

is the position vector of a point in the plane $\mathbb{R}^2$.

All points $\{f(t), t \in I\}$ form a plane curve:

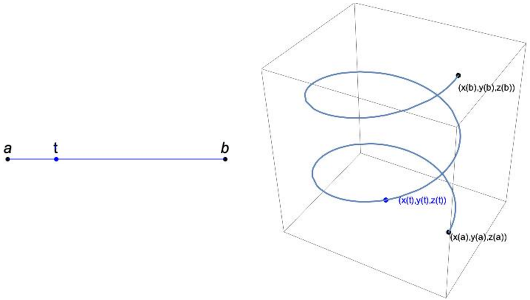


In this example $x(t) = t\cos t, y(t) = t\sin t, t \in [-3\pi/4, 3\pi/4]$

If $m = 3$, then

$$f(t) = \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} = \mathbf{r}(t)$$

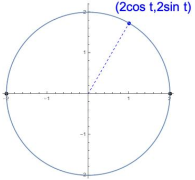
is the position vector of a point in $\mathbb{R}^3$ for every t, and $\{f(t), t \in I\}$ is a space curve:



In this example $x(t) = \cos t, y(t) = \sin t, z(t) = t/5, t \in [0, 4\pi]$

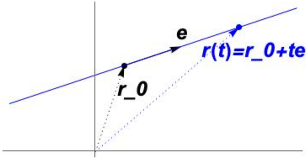
Example

$$f(t) = \begin{bmatrix} 2\cos t \\ 2\sin t \end{bmatrix}, t \in [0, 2\pi]$$



a circle with radius 2 and center $(0, 0)$

$f(t) = \mathbf{r}_0 + t\mathbf{e}, t \in \mathbb{R}$,
$\mathbf{r}_0, \mathbf{e} \in \mathbb{R}^m, \mathbf{e} \neq \mathbf{0}$

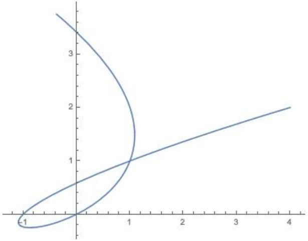


line through $\mathbf{r}_0$ in the direction of $\mathbf{e}$ in $\mathbb{R}^m$

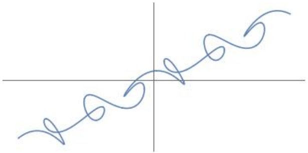
m=2:
slope $k = e_2/e_1$ if $e_1 \neq 0$
vertical if $\mathbf{e} = (0, e_2)$
horizontal if $\mathbf{e} = (e_1, 0)$

Example

$$f(t) = \begin{bmatrix} t^3 - 2t \\ t^2 - t \end{bmatrix}, t \in \mathbb{R}$$



$$f(t) = \begin{bmatrix} t + \sin(3t) \\ t + \cos(5t) \end{bmatrix}, t \in \mathbb{R}$$



A parametric curve $f(t), t \in [a, b]$ is closed if $f(a) = f(b)$.

Example

$$f(t) = \begin{bmatrix} \cos 3t \\ \sin 5t \end{bmatrix}, t \in [0, 2\pi]$$