V. Schulz
M. Klar

SS 2022
Submission: 26.04.21, until 12:15

Numerical Optimization - Sheet 3

If you are a student in mathematics please solve the exercises with no tag and the ones with the tag **Mathematics**. If you are a data science student please solve the problems with no tag and those with the tag **Data Science**. Submissions with tags other than your subject count as bonus points. The tag **Programming** marks programming exercises.

Ex 1 (6 Points)

The Rosenbrock function is defined by

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2.$$

Show that the point $x = (1, 1)$ is a local minimum of f and that the Hessian of f is positive definite at that point.

Solution 1:
The derivative $Df = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2})$ is given by

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= -2(1 - x_1) - 400(x_2 - x_1^2)x_1 \\ \frac{\partial f}{\partial x_2} &= 200(x_2 - x_1^2).\end{aligned}$$

Hence we see that the *first order necessary condition* for a local minimum

$$0 = Df((1, 1))$$

is fulfilled. The second derivative is given by

$$\begin{aligned}\frac{\partial^2 f}{\partial x_1^2} &= 2 - 400x_2 + 1200x_1^2 = 2 - 400(x_2 - 3x_1^2) \\ \frac{\partial^2 f}{\partial x_1 x_2} &= \frac{\partial^2 f}{\partial x_2 x_1} = -400x_1 \\ \frac{\partial^2 f}{\partial x_2^2} &= 200.\end{aligned}$$

In order to show that the *second order sufficient condition* for a local minimum is fulfilled at $(1, 1)$ we compute the determinant of $D^2 f - \lambda I$, which is given by

$$\begin{aligned}&(802 - \lambda)(200 - \lambda) - 160000 = 160400 - 1002\lambda + \lambda^2 - 160000 \\ &= 400 - 1002\lambda + \lambda^2.\end{aligned}$$

Claiming the determinant to be equal to 0 yields the solutions

$$\lambda_{1,2} = \{501 \pm \sqrt{501^2 - 400}\}.$$

We find $\sqrt{501^2 - 400} \approx 500,6$ which implies $\lambda_{1,2} > 0$ and hence that $D^2 f((1, 1))$ is positive definite.

Ex 2 (4 Points)

Let $H \in \mathbb{R}^{n\times n}$ be symmetric, $C \in \mathbb{R}^{m\times n}$ surjective, and $m \le n$. In addition, there is an $\alpha > 0$ such that

$$(v, Hv)_{\mathbb{R}^n} \ge \alpha \|v\|^2, \quad \forall v \in \operatorname{Kern} C$$

Show that

$$\begin{bmatrix} H & C^T \\ C & 0 \end{bmatrix} : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$$

is invertible.
Hint: Note the surjectivity of C implies that $\operatorname{Kern}(C^T) = \{0\}$.

<u>Solution 2:</u>
We know that for linear functions $f : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ injectivity, surjectivity and bijectivity are all equivalent.

It therefore suffices to show that

$$\begin{bmatrix} H & C^T \\ C & 0 \end{bmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = \mathbf{0} \Rightarrow \begin{pmatrix} v \\ w \end{pmatrix} = \mathbf{0}$$

which then implies invertibility.

So let as assume that we are given $\begin{pmatrix} v \\ w \end{pmatrix} \in \mathbb{R}^{n+m}, v \in \mathbb{R}^n, w \in \mathbb{R}^m$ such that

$$\begin{bmatrix} H & C^T \\ C & 0 \end{bmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = 0. \tag{1}$$

The second row in (1) says $Cv = 0$ which implies $v \in \operatorname{Kern}(C)$. If we multiply the first row by v from the left we obtain

$$\begin{aligned} Hv + C^T w &= 0 \, (\in \mathbb{R}^n) \\ \Leftrightarrow (v, Hv) + \underbrace{(v, C^T w)}_{=(Cv,w)=0} &= 0 \\ \Rightarrow 0 = (v, Hv) \ge \underbrace{\alpha}_{>0} \|v\|^2 &\Rightarrow \boxed{v = 0}. \end{aligned}$$

So v has to be the $0-$vector. Due to $\operatorname{Im}(C)^\perp = \operatorname{Ker}(C^T)$ the surjectivity of the linear function induced by C implies the injectivity of the linear function induced by C^T. However, as C^T induces an injective function we obtain that

$$\underbrace{Hv}_{=0(v=0)} + C^T w = C^T w = 0$$

$$C \text{ injective} \Rightarrow \boxed{w = 0}.$$

Hence $\begin{bmatrix} H & C^T \\ C & 0 \end{bmatrix}$ is invertible.

Ex 3 **Data Science, Programming** (6 Points)

(i) Implement the Rosenbrock function (see Ex. 1), its exact gradient and its exact Hessian w.r.t to the standard Euclidean scalar product.

(ii) Solve the optimization problem $\min f(x)$ using the function `minimize` of the module `scipy.optimize` with starting values $x_0 = (0, 0)$ and $x_0 = (0.99, 0.99)$.

- without any other parameters except from f and x_0.
- using of the parameter `jac`.
- using of the parameters `jac` and `hess` and `method="Newton-CG"`.

Print the solution of `minimize` into your iPython-Notebook.

Ex 4 **Mathematics** (4 Points)

A mapping $f \in C^1(S, \mathbb{R})$, where $S \subset \mathbb{R}^n$ is convex, is called strongly convex, if

$$(\nabla f(x) - \nabla f(y),\, x - y) \geq m\|x - y\|^2 \tag{2}$$

for some $m > 0$ and all $x, y \in S$. Show that f is strongly convex if, and only if

$$f(y) \geq f(x) + (\nabla f(x),\, y - x) + \frac{m}{2}\|x - y\|^2 \tag{3}$$

for $m > 0$ and all $x, y \in S$.

Solution 4:

(1) "⇒" (2)

In order to obtain an assertion about differences of function values from an assertion about differences of slopes we need to integrate. We choose $x, y \in S$ and define the suitable antiderivative $\phi : t \mapsto f(z_t), z_t := y + t(x - y)$, meaning the chain rule yields

$$\phi'(t) = (\nabla f(z_t), x - y).$$

The derivative of ϕ allows to exploit (1) for all values $t \in (0, 1)$

$$\phi'(t) - \phi'(0) = (\nabla f(z_t) - \nabla f(y), x - y) = \frac{1}{t}(\nabla f(z_t) - \nabla f(y), z_t - y) \geq \frac{1}{t}m\|z_t - y\|^2 = tm\|x - y\|^2$$

On top, integrating the differences of slopes for all values $t \in (0, 1)$ exaclty yields (2)

$$\begin{aligned}
f(x) - f(y) - (\nabla f(y), x - y) &= \int_0^1 (\nabla f(z_t) - \nabla f(y), x - y)dt \\
&= \phi(1) - \phi(0) - \phi'(0) = \int_0^1 \phi'(t) - \phi'(0)dt \\
&\geq m\|x - y\|^2 \int_0^1 t\,dt \;=\; \frac{m}{2}\|x - y\|^2.
\end{aligned}$$

(2) "⇒" (1)

First of all we can rewrite (2) as (2) $\Leftrightarrow (\nabla f(x), x - y) \geq f(x) - f(y) + \frac{m}{2}\|x - y\|^2$.
Interchanging the roles of x and y the yields

$$(2) \Rightarrow \begin{cases} \text{I)} \;\; (\nabla f(x), x - y) \geq f(x) - f(y) + \frac{m}{2}\|x - y\|^2 \\ \text{II)} \;\; (\nabla f(y), y - x) \geq f(y) - f(x) + \frac{m}{2}\|y - x\|^2 \end{cases}$$

From adding I) and II) we finally obtain

$$(\nabla f(x), x - y) + (\nabla f(y), y - x) = (\nabla f(x) - \nabla f(y), x - y) \geq m\|x - y\|^2.$$