V. Schulz
M. Klar

SS 2022
Submission: 31.05.21, until 12:15

Numerical Optimization - Sheet 8

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 **Mathematics** (4 Points)

Let $p, q \in \mathbb{R}^n, p \neq 0$ and $B \in \mathbb{R}^{n \times n}$. Furthermore, we define $p^\perp := \{p' \in \mathbb{R}^n \,|\, p^\top p' = 0\}$. Please show, that the equation

$$B^+ = B + \frac{(q - Bp)p^\top}{p^\top p}$$

is the unique solution of the Broyden-conditions

$$\begin{aligned} B^+ p &= q && \text{(Secant Condition)} \\ B^+ p' &= Bp' \qquad \forall p' \in p^\perp. \end{aligned}$$

<u>Solution 1:</u>
We see that indeed B^+ solves (Secant Condition)

$$\begin{aligned} B^+ p &= Bp + \frac{(q - Bp)(p)^T p}{(p)^T p} \\ &= Bp + q - Bp = q. \end{aligned}$$

and also fulfills

$$\begin{aligned} B^+ p &= Bp + \frac{(q - Bp) \quad \overbrace{(p)^T p}^{=0,\text{due to } p \perp p}}{(p)^T p} \\ &= Bp. \end{aligned}$$

A linear mapping is uniquely defined by its values on a basis of $\mathbb{R}^n$. As $\{p\} \cup p^\perp$ contains a basis of $\mathbb{R}^n$, B^+ is uniquely defined by the above conditions.

Ex 2 (6 Points)

Let $A \in \mathbb{R}^{n \times n}$ be invertible and $x, y \in \mathbb{R}^n$. The Sherman-Morrison-Woodbury formula says that

(i) If $x^T A^{-1} y \neq -1$, then

$$(A + yx^T)^{-1} = A^{-1} - \frac{A^{-1} y x^T A^{-1}}{1 + x^T A^{-1} y}$$

(ii) If $x^T A^{-1} y = -1$, then $A + yx^T$ is singular.

Please show that the equation (i) and the assertion (ii) hold.

<u>Solution 2:</u>

(a) $x^T A^{-1} y \neq -1$:

$$\begin{aligned}
(A + yx^T)(A^{-1} - \frac{A^{-1}yx^TA^{-1}}{1 + x^TA^{-1}y}) &= I - \frac{yx^TA^{-1}}{1 + x^TA^{-1}y} + yx^TA^{-1} - \frac{yx^TA^{-1}yx^TA^{-1}}{1 + x^tA^{-1}y} \\
&= I - \frac{yx^TA^{-1}}{1 + x^TA^{-1}y} + \frac{yx^TA^{-1}(1 + x^TA^{-1}y)}{1 + x^TA^{-1}y} - \frac{yx^TA^{-1}yx^TA^{-1}}{1 + x^tA^{-1}y} \\
&= I + \frac{-yx^TA^{-1} + yx^TA^{-1} + yx^TA^{-1}(x^TA^{-1}y) - y(x^TA^{-1}y)x^TA^{-1}}{1 + x^TA^{-1}y} \\
&= I.
\end{aligned}$$

The conclusion follows due to the uniqueness of the inverse.

(b) Let $x, y \in \mathbb{R}^n$ with $x^T A^{-1} y = -1$. It follows that $y \neq 0$ and

$$(A + yx^T)A^{-1}y = y + y\ x^TA^{-1}y = y - y = 0.$$

Hence $A + yx^T$ is not injective.

Ex 3 **Data Science** (4 Points)

You are given the following matrices

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} 1 & 0 & -0.5 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}.$$

Use the Sherman-Morrison-Woodbury formula to compute the inverse B^{-1}.

Solution 3:

Set $v = (0, 0, 1)^\top$, and $u = (1, 0, 0)^\top$. Then we see that $B = A + vu^\top$ and we can compute

$$\begin{aligned}
B^{-1} &= (A + vu^\top)^{-1} = A^{-1} - \frac{A^{-1}vu^\top A^{-1}}{1 + u^\top A^{-1}v} \\
&= \begin{bmatrix} 1 & 0 & -0.5 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} - \frac{\begin{bmatrix} 1 & 0 & -0.5 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{bmatrix} 1 & 0 & -0.5 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}}{1 + \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{bmatrix} 1 & 0 & -0.5 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}} \\
&= \begin{bmatrix} 1 & 0 & -0.5 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} - \frac{\begin{pmatrix} -0.5 \\ 0 \\ 0.5 \end{pmatrix} (1, 0, -0.5)}{0.5} = \begin{bmatrix} 1 & 0 & -0.5 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} + \begin{bmatrix} 1 & 0 & -0.5 \\ 0 & 0 & 0 \\ -1 & 0 & 0.5 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 0 & -1 \\ 0 & 0.5 & 0 \\ -1 & 0 & 1 \end{bmatrix}
\end{aligned}$$

Ex 4 **Programming** (12 Points)

We define the sigmoidal function

$$\sigma(t) = \frac{1}{1+e^{-t}}.$$

It has the derivative $\sigma'(t) = \sigma(t)(1-\sigma(t))$. The module `gauss_newton` contains a function `generate_data(gamma=0)` which generates a data set (t_i, α_i) where $t_i \in \mathbb{R}$ and $\alpha_i \in \mathbb{R}$ with

$$\overline{\alpha}_i = \sigma(6t_i + 1) + \epsilon_i \gamma$$

for $i = 1, \ldots, 10$. The values $\epsilon_i \sim \mathcal{N}(0,1)$ are independently normally distributed and the real value $\gamma \in \mathbb{R}$ controls the influence of ϵ_i.

(i) Solve the problem

$$\min_{x\in\mathbb{R}^2} \frac{1}{2}\|F(x)\|^2,$$

with $F_i(x) = \sigma(x_1 t_i + x_2) - \overline{\alpha}_i$ for $i = 1, \ldots, 10$ and $\gamma = 0$ using the Gauss Newton algorithm as given in the lecture. Iterate until the size of the search direction is sufficiently small, i.e. until $\|\Delta x_k\| < \delta$ for some tolerance $\delta > 0$. The solution is of course $x^* = (6, 1)$.

(ii) Solve the above problem also for $\gamma =$ 5e-2, 1e-3.

(iii) Plot the norms of the search directions $\|\Delta x_k\|$ against the iteration count k and use a logarithmic scale in the $y-$axis. How do the cases $\gamma =$ 1e-1, 1e-3, 0 differ? Which behaviour should be observed for $\gamma = 0$?

Hint: The module `gauss_newton` contains the functions `armijo`, `sigmoidal()`, and `dsigmoidal()`, as well as two simple plot routines `plot(x, gamma)` and `convergence_plot(delta_x_list)`.