V. Schulz
M. Klar

SS 2022
Submission: 19.04.21, until 12:15

Numerical Optimization - Sheet 2

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 (4 Points)

Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

(i) Show that
$$\mathrm{Ker}(A) = \mathrm{Im}(A^T)^\perp.$$

(ii) Show that the following assertions exclude each other:

- There exists an $x \in \mathbb{R}^n$ which fulfills $Ax = b$.
- There is a $y \in \mathbb{R}^m$ which fulfills $b^\top y = 1$ and $A^\top y = 0$.

<u>Solution 1:</u>

(i)
- "$\mathrm{Ker}(C) \subset \mathrm{Im}(C^\top)^\perp$"
  Let $v \in \mathrm{Ker}(C)$ be arbitrary $\Rightarrow \forall w \in \mathrm{Im}(C^\top)\ \exists a \in \mathbb{R}^m$ such that $w^\top v = [C^\top a]^\top v = a^\top C v = 0$.
- "$\mathrm{Ker}(C) \supset \mathrm{Im}(C^\top)^\perp$"
  Let $w \in \mathrm{Im}(C^\top)^\perp$ be arbitrary $\Rightarrow \forall a \in \mathbb{R}^m$ it holds that $a^\top C w = [C^\top a]^\top w = 0$ and therefore $Cw = 0$.

(ii) The above assertion implies
$$\mathrm{Ker}(A^\top) = \mathrm{Im}(A)^\perp \text{ and} \tag{1}$$
$$\mathrm{Ker}(A^\top)^\perp = \mathrm{Im}(A) \tag{2}$$

Additionaly we note that

(a) There exists an $x \in \mathbb{R}^n$ which fulfills $Ax = b \Leftrightarrow b \in \mathrm{Im}(A)$
(b) There is a $y \in \mathbb{R}^m$ which fulfills $b^\top y = 1$ and $A^\top y = 0 \Leftrightarrow y \in \mathrm{Ker}(A^\top)$ and $b^\top y = 1$

The connection between the assertions (a) and (b) is the vector b. We have that (a) excludes (b) because $b \in \mathrm{Im}(A) \Rightarrow b \in \mathrm{Ker}(A^\top)^\perp \Rightarrow b^\top y = 0$ for any $y \in \mathrm{Ker}(A^\top)$.
We have that (b) excludes (a) because $b^\top y = 1$ and $y \in \mathrm{Ker}(A^\top) \Leftrightarrow b^\top y = 1$ and $y \in \mathrm{Im}(A)^\perp$ $\Rightarrow b \notin \mathrm{Im}(A)$.

Ex 2 **Mathematics** (4 Points)

(Schur complement lemma) Let
$$D = \begin{bmatrix} A & b \\ b^\top & c \end{bmatrix},$$
where $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n, c \in \mathbb{R}$. Suppose that A is symmetric, positive definite. Prove that D is positive semidefinite if and only if $c - b^\top A^{-1} b \geq 0$.

<u>Solution 2:</u>

We show at first that

$$c - b^\top A^{-1} b \geq 0 \tag{3}$$

when D is positive semidefinite.
Let $v \in \mathbb{R}$ be arbitrary. As A is positive definite it is invertible and we have that for any $v \in \mathbb{R}$ there exists a $u \in \mathbb{R}^n$ such that

$$\begin{aligned} &Au + bv = 0 \\ &\Leftrightarrow u = -A^{-1} bv. \end{aligned} \tag{4}$$

For this pair we obtain

$$\left(u^\top, v\right) D \begin{pmatrix} u \\ v \end{pmatrix} = \left(u^\top, v\right) \begin{bmatrix} A & b \\ b^\top & c \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \left(u^\top, v\right) \begin{pmatrix} 0 \\ b^\top u + cv \end{pmatrix}.$$

Then (4) yields that

$$\left(u^\top, v\right) D \begin{pmatrix} u \\ v \end{pmatrix} = \left(u^\top, v\right) \begin{pmatrix} 0 \\ (c - b^\top A^{-1} b) v \end{pmatrix} = v^2 (c - b^\top A^{-1} b).$$

As D is positive semidefinite we conclude that

$$0 \leq \left(u^\top, v\right) D \begin{pmatrix} u \\ v \end{pmatrix} = v^2 (c - b^\top A^{-1} b).$$

To show the converse assertion let $u \in \mathbb{R}^n, v \in \mathbb{R}$ be arbitrary and choose the decomposition $u = u_1 + u_2$ and $v = v_1 + v_2$ given by

$$\begin{aligned} u_1 &= -A^{-1} bv \\ v_1 &= v, \end{aligned} \tag{5}$$

and

$$\begin{aligned} u_2 &= u - u_1 \\ v_2 &= 0. \end{aligned}$$

We then see that

$$\begin{aligned} &\left(u^\top, v\right) D \begin{pmatrix} u \\ v \end{pmatrix} = \left(u^\top, v\right) D \left(\begin{pmatrix} u_1 \\ v_1 \end{pmatrix} + \begin{pmatrix} u_2 \\ 0 \end{pmatrix} \right) = v^2 (c - b^\top A^{-1} b) + \left(u^\top, v\right) \begin{pmatrix} Au_2 \\ b^\top u_2 \end{pmatrix} \\ &= v^2 (c - b^\top A^{-1} b) + u^\top A u_2 + v b^\top u_2. \end{aligned}$$

We assume in the assertion that $v^2(c - b^\top A^{-1} b) \geq 0$, so it remains to show that $u^\top A u_2 + v b^\top u_2 \geq 0$. From (5) we obtain that

$$\begin{aligned} &u^\top A u_2 + v b^\top u_2 = u_1^\top A u_2 + u_2^\top A u_2 + v b^\top u_2 \\ &= - \left(A^{-1} bv\right)^\top A u_2 + u_2^\top A u_2 + v b^\top u_2 = -v b^\top u_2 + u_2^\top A u_2 + v b^\top u_2 \\ &= u_2^\top A u_2 \geq 0. \end{aligned} \tag{6}$$

Note that we cannot get a strinct inequality in (6) as u_2 could be 0. We can therefore conclude that

$$\left(u^\top, v\right) D \begin{pmatrix} u \\ v \end{pmatrix} \geq 0.$$

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

(i) Let $f : \mathbb{R}^n \to \mathbb{R}$, $x \mapsto b^\top x + a$ for $a \in \mathbb{R}$ and $b \in \mathbb{R}^n$ be an affine function. Show that f is convex.

(ii) Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and $c \in \mathbb{R}$. Show that the set

$$M := \{x \in \mathbb{R}^n \mid f(x) \leq c\}$$

is convex.

Solution 3:

(i) $f((1-\lambda)x + \lambda y) = b^\top((1-\lambda)x + \lambda y) + a = (1-\lambda)(b^\top x + a) + \lambda(b^\top y + a)$

(ii) Let $u, v \in M$. We have to show that $(1-\lambda)u + \lambda v \in M$ for all $\lambda \in [0,1]$
We immediately obtain the assertion from

$$f((1-\lambda)u + \lambda v) \leq (1-\lambda)f(u) + \lambda f(v) \leq f(a).$$

Ex 4 **Programming** (5 Points)

(i) Please implement a function `drv(f, x, h)` in Python 3, which returns the central difference for an arbitrary function $f : \mathbb{R} \to \mathbb{R}$ at point $x \in \mathbb{R}$ and step size h. The central difference of a function is given by

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}.$$

(ii) Implement a function `ddrv(f, x, h)` which computes the second derivative. Use the approximation

$$f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}.$$

(iii) Implement a function `pdrv(g, x, j, h)` which returns the partial derivative of a function $g : \mathbb{R}^n \to \mathbb{R}$ at point $x \in \mathbb{R}^n$ in direction of its j-th component, where $j = 0, \ldots, n-1$. The function g should take a `numpy`-array of length n as argument.

(iv) Implement a function `grad(g, x, h)` based on `pdrv(g, x, j, h)` which returns the gradient of g with respect to the standard Euclidean norm.

(v) Import the function `test` of the module `numdrv_test`. Execute it for each of your solutions via `test(drv)`, `test(ddrv)` and `test(pdrv)`. Your errors should be in the order of 10^{-7} or smaller.