V. Schulz
M. Klar

SS 2022
Submission: 12.04.22, until 12:15

Numerical Optimization - Sheet 1

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 $f : \mathbb{R}^n \to \mathbb{R}$ be convex, that means

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda) \cdot f(x) + \lambda \cdot f(y), \quad \forall x, y \in \mathbb{R}^n,\ \forall \lambda \in [0,1].$$

Show that any local minimum of f is already a global minimum.

Solution 1:
Let $\hat{x}$ be a local minimum in an arbitrary, fixed (open) neighbourhood $\hat{x} \in U \in \mathbb{R}^n$, i.e.

$$f(\hat{x}) \leq f(x)$$

for all $x \in U$. Let us assume there is a $\hat{y} \in \mathbb{R}^n$ such that

$$f(\hat{y}) < f(\hat{x}).$$

We obtain, that for all $\lambda \in (0,1)$

$$f(\lambda\hat{x} + (1-\lambda)\hat{y}) \leq \lambda f(\hat{x}) + (1-\lambda) f(\hat{y}) < f(\hat{x}) + (1-\lambda) f(\hat{x}) = f(\hat{x}).$$

As U is open, there is a $\hat{\lambda} \in (0,1)$ such that $\hat{\lambda}\hat{x} + (1-\hat{\lambda})\hat{y} \in U$ which yields a contradiction to the assumption of local optimality of $\hat{x}$.

Ex 2 (4 Points)

Consider a linear mapping $f : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto Cx$, defined by a matrix $C \in \mathbb{R}^{m \times n}$ with $m \leq n$. Show that f being surjective is equivalent to the matrix C having full rank.

Solution 2:
Definition rank (see Elements of Mathematics)
Let $A \in \mathbb{R}^{m \times n}$ be a real matrix with SVD $A = U\Sigma V^\top$. Then the rank of A is the number of positive singular values.

Definition pseudoinverse of A (see Elements of Mathematics)
Let $A \in \mathbb{R}^{m \times n}$ be a real matrix with SVD $A = U\Sigma V^\top$. Then the pseudoinverse of A is defined by

$$A^+ := V\Sigma^+ U^\top := V \left[\begin{array}{ccc|ccc} \sigma_1^{-1} & & & & \vdots & \\ & \ddots & & \cdots & 0 & \cdots \\ & & \sigma_k^{-1} & & \vdots & \\ \hline & \vdots & & & \vdots & \\ \cdots & 0 & \cdots & \cdots & 0 & \cdots \\ & \vdots & & & \vdots & \end{array}\right] U^\top.$$

Definition surjective (see Elements of Mathematics)
Let $f : X \to Y$ be a function. Then f is called surjective if for any $y \in Y$ there exists an $x \in X$ such that $f(x) = y$.

Proof:
Full Rank $\Rightarrow$ Surjective

Let $y \in \mathbb{R}^m$ be arbitrary and set $x := C^+ y$. Because C is of full rank we obtain

$$\Sigma\Sigma^+ = \mathrm{id}_m\,.$$

Hence

$$Cx = CC^+ y = U\Sigma V V^\top \Sigma^+ U^\top y = U\Sigma\,\mathrm{id}_n\,\Sigma^+ U^\top y = U\,\mathrm{id}_m\,U^\top y = \mathrm{id}_m\,y = y.$$

Surjective $\Rightarrow$ Full Rank

As f is surjective we know that for all $k \in \{1, \ldots, m\}$ there are x_k such that

$$u_k = Cx_k$$

for each column u_k of $U \in \mathbb{R}^{m\times m}$. The alternative representation

$$C = \sum_{i=1}^{m} \sigma_i u_i v_i^\top$$

yields

$$u_k = \sum_{i=1}^{m} \sigma_i u_i v_i^\top x_k = \sum_{i=1}^{m} \sigma_i u_i \alpha_i.$$

As the columns of U are linearly independent we obtain that $\sigma_k \neq 0$ for all $k \in \{1, \ldots, m\}$.

Ex 3 (4 Points)

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as

$$f(x) = \frac{1}{2}\left[x_1^2 + 2x_1x_2 + 4x_2^2\right].$$

Find a scalar product such that $\nabla f(x) = x\ \forall x \in \mathbb{R}^2$.

<u>Solution 3:</u>

$$(x, y)_A := x^\top \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix} y$$

yields

$$\frac{\partial f}{\partial x} = (x_1 + x_2, 4x_2 + x_1).$$

The gradient is a vector $g(x)$ sucht that

$$(g(x), v)_A = \frac{\partial f}{\partial x} v \ \ \forall v \in \mathbb{R}^2.$$

We set $g(x) := x$ and find

$$(x, e_1)_A = x_1 + x_2 = \frac{\partial f}{\partial x} e_1,$$

$$(x, e_2)_A = 4x_2 + x_1 = \frac{\partial f}{\partial x} e_2.$$

Ex 4 **Programming** (3 Points)

Create a contour plot of the Rosenbrock function

$$f(x, y) = (a - x)^2 + b(y - x^2)^2,$$

for $a = 1$ and $b = 100$. Try to approximately find the region of its global minimum in the plot.
Hint:

- You can obtain a logarithmic scaling of the contour lines in `plt.contour()` via the option `locator=ticker.LogLocator()`.
- Please submit the solution as iPython Notebook. Read the information page on Olat if you need information about Python 3 or iPython Notebooks.