Optimization Algorithms
Exercise 1

Marc Toussaint
Learning & Intelligent Systems Lab, TU Berlin
Marchstr. 23, 10587 Berlin, Germany

Winter 2020/21

- During the tutorials the tutor will ask students randomly to present their solutions or questions. What you present does not have to be a perfect solution, but please try to prepare something.
- You will need to present your solutions via zoom by sharing the screen, e.g. by running code online, showing a picture of a hand-written solution, or a LaTeX'ed solution. Please be prepared for this.

1 Boyd & Vandenberghe

Read sections 1.1, 1.3 & 1.4 of Boyd & Vandenberghe "Convex Optimization". This is for you to get an impression of the book. Learn in particular about their categories of convex and non-linear optimization problems.

2 Your first gradient descent

Consider the following functions over $x \in \mathbb{R}^n$:

$$f_{\text{sq}}(x) = x^\top C x \ , \tag{1}$$

$$f_{\text{hole}}(x) = \frac{x^\top C x}{a^2 + x^\top C x} \ . \tag{2}$$

For $C = \mathbf{I}$ (identity matrix) the first would be fairly simple to optimize. The C matrix changes the *conditioning* ("skewedness of the Hessian") of these functions to make them a bit more interesting. We assume that C is a diagonal matrix with entries $C(i,i) = c^{\frac{i-1}{n-1}}$. We choose a conditioning[1] $c = 10$, and width $a = .1$.
In the previous exercise you learned to code and display these functions. You now test simple optimization methods.

a) What are the gradients $\nabla f_{\text{sq}}(x)$ and $\nabla f_{\text{hole}}(x)$?

b) What are the Hessians $\nabla^2 f_{\text{sq}}(x)$ and $\nabla^2 f_{\text{hole}}(x)$?

c) Implement a simple fixed stepsize gradient descent, iterating $x_{k+1} = x_k - \alpha \nabla f(x_k)$, with start point $x_0 = (1, 1)$, $c = 10$, $a = 0.1$, and heuristically chosen α. In each iteration you should ouput basic information (#iteration, current-cost) and store the current x as a column to a large matrix (the "path"). After the run, display the function together with this path. (The previous exercise already had the option for this: passing the plotFunc the `trace_xy`, argument.)

d) Implement gradient descent with backtracking, as described on slide 02:05 (with default parameters ϱ). Test the algorithm on $f_{\text{sq}}(x)$ and $f_{\text{hole}}(x)$ with start point $x_0 = (1, 1)$.

[1] The word "conditioning" generally denotes the ratio of the largest and smallest Eigenvalue of the Hessian.