NUMERICAL OPTIMIZATION

Sheet 7: Newton's method and more

Exercise one At the lecture, we have investigated what happens for Newton's method when we use a full rank matrix $A \in \mathbb{R}^{n\times n}$ to optimize $\min f(Ax)$ (compared to $\min f(x)$).

We have named the function we minimize $f_A(x) := f(Ax)$.

Then, we claimed the following equations for gradient and Hessian are true:

$$\nabla f_A(x) = A^T \nabla f(Ax).$$
$$H_{f_A}(x) = A^T H_f(Ax) A.$$

Using basic rules about partial derivatives, prove the two equations above.

Hint: Spend some time to get used to the notation. Even though they look the same, and $f_A(x)$ is defined to be $f(Ax)$, $\nabla f_A(x) \neq \nabla f(Ax)$. The vector $\nabla f_A(x)$ is the gradient of the function f_A evaluated at the point x, whereas $\nabla f(Ax)$ is the gradient of the function f evaluated at the point Ax.

Exercise two Perform one step of the Newton method for minimizing the function

$$f(x_1, x_2) = x_1^4 + 2x_1^3 + 2x_1^2 + x_2^2 - 2x_1 x_2,$$

starting in the initial point $(-1, 0)$.

We already knew from its original root-finding setting that the pure Newton method (Newton method with fixed step size $\alpha = 1$) can cycle or diverge if the initial point is not close to x^*, the local minimum. The following two exercises ask you to check two other cases where this happens:

Exercise three

1. Show that $f(x) = \log\left(e^x + e^{-x}\right)$ (a function of one variable) is convex.
2. Show that $f(x)$ has a unique global minimum at $x^* = 0$, but there exists t such that if $|x_n| > t$, then $|x_{n+1}| > |x_n|$, so the method diverges.

Exercise four

1. Show that $f(x) = -\log(x) + x$ is convex.
2. Show that $f(x)$ has a unique global minimum at $x^* = 1$. Next, show that there exists t such that for $x_0 < t$ the pure Newton method is convergent, but for $x_0 > t$ pure Newton method produces points not even in the domain of f.

Exercise five Let $f(x) = \sum_{i=1}^{n} \cosh(x_i)$. Suppose we are given a vector $a_1, \ldots, a_n$ and we investigate $f(x)$ in the bounded region $\forall i : -a_i \leq x_i \leq a_i$.

Find a simple formula (in terms of a_i) giving optimal values of m and M such that the largest eigenvalue of $H_f(x)$ is at most M and the lowest is at least m for all x within the bounds.