Continuous Optimization **Spring 2025**

Exercise 6: Newton's method

Lecturer: Aurelien Lucchi

Problem 1 (Affine invariance property of Newton's method):

Consider a function $f : \mathbb{R}^d \to \mathbb{R}$ and a non-singular matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$. Let $\mathbf{x} = \mathbf{A}\mathbf{y}$ and $g(\mathbf{y}) = f(\mathbf{A}\mathbf{y})$.

1. Show that the per-step update of Newton's method that minimizes $g(\mathbf{y})$ is equal to
$$\mathbf{y}^+ = \mathbf{y} - \mathbf{A}^{-1}(\nabla^2 f(\mathbf{A}\mathbf{y}))^{-1}\nabla f(\mathbf{A}\mathbf{y}).$$
2. Show that
$$\mathbf{x}^+ = \mathbf{x} - (\nabla^2 f(\mathbf{x}))^{-1}\nabla f(\mathbf{x}).$$
3. What conclusion can you draw from the previous equation?

Hint: Given a composition of functions $g = f(h(\mathbf{y}))$, where $h : \mathbb{R}^d \to \mathbb{R}^p$ and $f : \mathbb{R}^p \to \mathbb{R}$, and such that $\frac{\partial^2 h}{\partial y_i \partial y_j} = 0$, then
$$\nabla^2 g = \mathbf{J}_h^\top \mathbf{H}_f \mathbf{J}_h, \tag{1}$$
where $\mathbf{J}_h$ is the Jacobian matrix of h, and $\mathbf{H}_g$ is the Hessian matrix of g.

Problem 2 (Quadratic convergence of Newton's method):

1. Recall the following theorem derived in class.

> **Theorem 1** (Undamped)**.** *Assume that $f(\cdot)$ satisfies the Assumptions seen in class and that $\|\mathbf{x}_k - \mathbf{x}^*\| \leq \frac{2\mu}{3L}$, then*
> $$\|\mathbf{x}_{k+1} - \mathbf{x}^*\| \leq \frac{3L}{2\mu}\|\mathbf{x}_k - \mathbf{x}^*\|^2. \tag{2}$$

   Using the above theorem, provide a bound on $\|\mathbf{x}_{k+s} - \mathbf{x}^*\|$.

2. Assuming $\|\mathbf{x}_0 - \mathbf{x}^*\| \leq \frac{\mu}{3L}$, prove that $\|\mathbf{x}_k - \mathbf{x}^*\| \leq \left(\frac{1}{2}\right)^{2^k - 1} \cdot \frac{\mu}{3L}$.
3. Assuming $\|\mathbf{x}_0 - \mathbf{x}^*\| \leq \frac{\mu}{3L}$, prove that the Hessian satisfies the following relative error bound:
$$\frac{\|\nabla f^2(\mathbf{x}_k) - \nabla f^2(\mathbf{x}^*)\|}{\|\nabla f^2(\mathbf{x}^*)\|} \leq 2\left(\frac{1}{3}\right)^{2^k - 1}.$$

Problem 3 (Convergence in terms of gradient norm):

We optimize a function $f : \mathbb{R}^d \to \mathbb{R}$ using Newton's method that produces the iterates:
$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{p}_k, \tag{3}$$
where $\mathbf{p}_k := [\nabla^2 f(\mathbf{x}_k)]^{-1}\nabla f(\mathbf{x}_k)$.

1. Using the relation $\nabla f(\mathbf{x}_k) + \nabla^2 f(\mathbf{x}_k)\mathbf{p}_k = 0$, prove that
$$\|\nabla f(\mathbf{x}_{k+1})\| \leq \int_0^1 \|\nabla^2 f(\mathbf{x}_k + t\mathbf{p}_k) - \nabla^2 f(\mathbf{x}_k)\|\|\mathbf{p}_k\|\, \mathrm{d}t.$$
2. Since $\nabla^2 f$ is non-singular and Lipschitz continuous, there is a radius $r > 0$ such that $\|\nabla^2 f(\mathbf{x}_k)^{-1}\| \leq 2\|\nabla^2 f(\mathbf{x}^*)^{-1}\|$ for all $\mathbf{x}_k$ such that $\|\mathbf{x}_k - \mathbf{x}^*\| \leq r$. Use this result to prove that
$$\|\nabla f(\mathbf{x}_{k+1})\| \leq 2L\|[\nabla^2 f(\mathbf{x}^*)]^{-1}\|^2\|\nabla f(\mathbf{x}_{k+1})\|^2 \quad \text{if } \|\mathbf{x}_k - \mathbf{x}^*\| \leq r,$$
   i.e. the gradient norm converges to zero quadratically.