Continuous Optimization **Spring 2025**

Exercise 3: Subgradient Method

Lecturer: Aurelien Lucchi

Problem 1 (Properties subdifferential):
Prove that:

1. The subdifferential is always a closed convex set.
2. The subdifferential is not always non-empty for non-convex functions.

Problem 2 (Convergence rate for strongly-convex functions):
Prove that if $\|\mathbf{g}_i\| \le L \ \forall i$,

$$f\left(\frac{2}{K(K+1)}\sum_{k=0}^{K-1}(k+1)\mathbf{x}_k\right) - f^* \le \frac{2L^2}{\mu(K+1)}.$$

1. Prove that $\|\mathbf{x}_{k+1}-\mathbf{x}^*\|_2^2 \le (1-\mu\eta_k)\|\mathbf{x}_k-\mathbf{x}^*\|_2^2 - 2\eta_k(f(\mathbf{x}_k)-f(\mathbf{x}^*)) + \eta_k^2L^2$
2. Rewrite the above expression using $\eta_k = \frac{2}{\mu(k+1)}$ and calculate $k(f(\mathbf{x}_k)-f(\mathbf{x}^*))$
3. Observe that $k(f(\mathbf{x}_k)-f(\mathbf{x}^*)) \le \frac{\mu}{4}\left(k(k-1)\|\mathbf{x}_k-\mathbf{x}^*\|_2^2 - k(k+1)\|\mathbf{x}_{k+1}-\mathbf{x}^*\|_2^2\right) + \frac{1}{\mu}L^2$
4. Sum the above expression for $k \in \{1, \cdots, K\}$. Hint: The series is telescopic.
5. Conclude using convexity.

Problem 3 (Polyak's step size):
In the subgradient lecture, we have seen that

$$\begin{aligned}\|\mathbf{x}_{k+1}-\mathbf{x}^*\|_2^2 &= \|\mathbf{x}_k-\eta_k\mathbf{g}_k-\mathbf{x}^*\|_2^2\\ &= \|\mathbf{x}_k-\mathbf{x}^*\|_2^2 - 2\eta_k\langle\mathbf{x}_k-\mathbf{x}^*,\mathbf{g}_k\rangle + \eta_k^2\|\mathbf{g}_k\|_2^2\\ &\le \|\mathbf{x}_k-\mathbf{x}^*\|_2^2 - 2\eta_k(f(\mathbf{x}_k)-f(\mathbf{x}^*)) + \eta_k^2\|\mathbf{g}_k\|_2^2. \qquad (1)\end{aligned}$$

1. Prove that the optimal step size is equal to $\eta_k = \frac{(f(\mathbf{x}_k)-f(\mathbf{x}^*))}{\|\mathbf{g}_k\|_2^2}$ (this step size is known as Polyak's step size).
2. Does this step size guarantee convergence?
3. What's the drawback of this approach?

Problem 4 (Programming: regularized logistic regression):
Complete TODOs in the Jupyter Notebook provided to implement the (sub)-gradient method for the ℓ^1 and ℓ^2-regularized Logistic Regression.