Optimization in Machine Learning | **Universität Mannheim**
HWS 2024 | Prof. Simon Weißmann, Felix Benning

Sheet 3

For the exercise class on the 17.10.2023 at 12:00.
Hand in your solutions by 10:15 in the lecture on Tuesday 15.10.2024.

Exercise 1 (Convergence Speed). **(3 Points)**

Proof that

(i) if we have

$$\limsup_{k\to\infty} \frac{e(x_{k+1})}{e(x_k)} = 0,$$

then $e(x_k)$ converges super-linearly. (1 pt)

(ii) If for $c \in (0,1)$ we have

$$\limsup_{k\to\infty} \frac{e(x_{k+1})}{e(x_k)} < c,$$

then $e(x_k)$ converges linearly with rate c. (1 pt)

(iii) If for $c \in (0,1)$ we have

$$\limsup_{k\to\infty} \frac{e(x_{k+1})}{e(x_k)^2} < c,$$

then $e(x_k)$ converges super-linearly with rate c. (1 pt)

Exercise 2 (Sub-gradients). **(2 Points)**

Let $f, g : \mathbb{R}^d \to \mathbb{R}$ be convex functions.

(i) Prove that $\partial f(x)$ is a convex set for any $x \in \mathbb{R}^d$. (1 pt)

(ii) For $h(x) = f(Ax+b)$ prove $\partial h(x) \supseteq A^T \partial f(Ax+b)$. Prove equality for invertible A. (1 pt)

Exercise 3 (Lasso). **(6 Points)**

Let

$$f(x) = \tfrac{1}{2}\|x-y\|^2 + \lambda\|x\|_1$$

for $x \in \mathbb{R}^d$ be the Lagrangian form of the least squares LASSO method.

(i) Compute a sub-gradient of f. (2 pts)

(ii) Prove that f is convex. (1 pt)

(iii) Find a global minimum of f. (1 pt)

(iv) Implement f as a sub-type of "DifferentiableFunction" (even though it is not) by returning a single sub-gradient and apply gradient descent to verify the global minimum `https://classroom.github.com/a/Bm7FMb12` (2 pts).

Exercise 4 (Momentum Matrix). **(2 Points)**

let $D = \operatorname{diag}(\lambda_1, \dots, \lambda_d)$, $\alpha, \beta > 0$ and define

$$T = \begin{pmatrix} (1+\beta)\mathbb{I} - \alpha D & -\beta\mathbb{I} \\ \mathbb{I} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{2d\times 2d}$$

Prove there exists a regular $S \in \mathbb{R}^{2d\times 2d}$ such that

$$S^{-1}TS = \hat{T} = \begin{pmatrix} T_1 & & \\ & \ddots & \\ & & T_d \end{pmatrix}$$

with

$$T_i = \begin{pmatrix} 1+\beta-\alpha\lambda_i & -\beta \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}.$$

Exercise 5 (PL-Inequality). **(6 Points)**

Assume $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth and satisfies the Polyak-Łojasiewicz inequality

$$\|\nabla f(x)\|^2 \geq 2c(f(x) - f_*) \tag{PL}$$

for some $c > 0$ and all $x \in \mathbb{R}^d$ with $f_* = \min_x f(x) > -\infty$.

(i) Prove that gradient descent with fixed step size $\alpha_k = \frac{1}{L}$ converges linearly in the sense

$$f(x_k) - f_* \leq (1 - \tfrac{c}{L})^k (f(x_0) - f_*). \qquad \text{(2 pts)}$$

(ii) Prove that μ-strong-convexity and L-smoothness imply the PL-inequality. (2 pts)

(iii) Use a graphing calculator to find c such that $f(x) = x^2 + 3\sin^2(x)$ satisfies the PL-condition (argue why $x \to \infty$ is not a problem) and prove it is not convex. (2 pts)

Exercise 6 (Weak PL-Inequality). **(5 Points)**

Assume $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth and satisfies the "weak PL inequality"

$$\|\nabla f(x)\| \geq 2c(f(x) - f_*)$$

for some $c > 0$ and all $x \in \mathbb{R}^d$ with $f_* = \min_x f(x) > -\infty$.

(i) Let $a_0 \in [0, \frac{1}{q}]$ for some $q > 0$ and assume for the sequence $(a_n)_{n\in\mathbb{N}}$ that it is positive and satisfies a diminishing contraction

$$0 \leq a_{n+1} \leq (1 - q a_n) a_n \qquad \forall n \geq 0.$$

Prove the convergence rate

$$a_n \leq \frac{1}{nq + 1/a_0} \leq \frac{1}{(n+1)q}. \qquad \text{(2 pts)}$$

Hint. *A useful checkpoint might be the telescoping sum of*

$$\frac{1}{a_{n+1}} - \frac{1}{a_n} \geq q.$$

(ii) Prove that f is bounded. More specifically $e(x) := f(x) - f_* \leq \frac{L}{2c^2}$ for all x. (1 pt)

Hint. *Use Sheet 1 Exercise 1 (i).*

(iii) For gradient descent $x_{n+1} - x_n = -\alpha_n \nabla f(x_n)$ with constant step size $\alpha_k = \frac{1}{L}$ prove the convergence rate

$$f(x_n) - f_* \leq \frac{L}{2c^2(n+1)}.$$ (2 pts)