Exercise Sheet 1

Lionel Tondji

Strong convexity and smoothness

July 17, 2024

Notation: For every $x, y \in \mathbb{R}^d$, let $\langle x, y\rangle = x^\top y$ and $\|x\|_2 = \sqrt{\langle x, x\rangle}$.
Let $\sigma_{min}(A)$ and $\sigma_{max}(A)$ be the smallest and largest singular values of A defined by

$$\sigma_{min}(A) = \min_{x\in\mathbb{R}^d} \frac{\|Ax\|_2}{\|x\|_2} \quad \text{and} \quad \sigma_{max}(A) = \max_{x\in\mathbb{R}^d} \frac{\|Ax\|_2}{\|x\|_2}. \tag{1}$$

Thus clearly

$$\frac{\|Ax\|_2^2}{\|x\|_2^2} \leq \sigma_{max}(A)^2, \quad \forall\, x \in \mathbb{R}^d. \tag{2}$$

Let $\|A\|_F^2 = \text{Trace}(A^\top A)$ denote the Frobenius norm of A. Finally, a result you will need, for every symmetric matrix G the $L2$ induced matrix norm can be equivalently defined by

$$\|G\|_2 = \sigma_{max}(G) = \sup_{x\in\mathbb{R}^d, x\neq 0} \frac{|\langle Gx, x\rangle|}{\|x\|_2^2} = \max_{x\in\mathbb{R}^d, x\neq 0} \frac{\|Gx\|_2}{\|x\|_2} \tag{3}$$

1 The Prox Operator

Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex and lower semi-continuous and define, for $\lambda > 0$ the proximal mapping $prox_{\lambda f} : \mathbb{R}^d \to \mathbb{R}^d$

$$prox_{\lambda f}(x) = arg\min_{y\in\mathbb{R}^d} \left\{ \frac{1}{2}\|y - x\|_2^2 + \lambda f(y) \right\} \tag{4}$$

1. Show that $prox_{\lambda f}$ is well defined, i.e. that the minimization problem has a unique solution for all x.

2. Prove that $prox_{\lambda f}$ is non-expansive, i.e. that for every $x, y \in \mathbb{R}^d$ it holds that

$$\|prox_{\lambda f}(x) - prox_{\lambda f}(y)\|_2 \leq \|x - y\|_2, \quad \forall x, y \in \mathbb{R}^d$$

2 Convexity

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice continuously differentiable function.

(i) Show that the following assertions are equivalent :

  a) $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \quad \forall\, x, y \in \mathbb{R}^d,\ \lambda \in [0,1].$
  b) $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle, \quad \forall\, x, y \in \mathbb{R}^d.$
  c) $\langle \nabla^2 f(x)v, v\rangle \geq 0, \quad \forall\, x, v \in \mathbb{R}^d$

  When f verifies one of the above inequalities we say that f is convex.

(ii) a) Let $g : \mathbb{R}^d \to \mathbb{R}$ be convex and $f : \mathbb{R} \to \mathbb{R}$ a convex nondecreasing function. Show that $f \circ g$ is a convex function.
  b) Is the composition of 2 convex functions itself a convex function?

(iii) Are the following applications convex functions ? Justify your answer.

  a) $x \mapsto \|x\|$
  b) $x \mapsto \|x\|^2$
  c) $x \mapsto x^2$
  d) $x \mapsto e^{-e^{-x}}, x \in [0, +\infty)$
  e) $x \mapsto (-x^{\frac{1}{3}})^2, x \in [0, +\infty)$
  f) $x \mapsto (-log(x))^2$

(iv) For every convex function $f : y \in \mathbb{R}^m \mapsto f(y)$, prove that $g : x \in \mathbb{R}^d \mapsto f(Ax - b)$ is a convex function, where $A \in \mathbb{R}^{n\times d}$ and $b \in \mathbb{R}^n$.

(v) Let $f_i : \mathbb{R}^d \mapsto \mathbb{R}$ be convex for $i = \{1, \dots, n\}$. Prove that $\sum_{i=1}^n f_i$ is convex.

(vi) For a given scalars $y_i \in \mathbb{R}$ and vectors $a_i \in \mathbb{R}^d$ for $i = \{1, \dots, m\}$ prove that the logistic regression function $f(x) = \frac{1}{n}\sum_{i=1}^n ln(1 + e^{-y_i\langle x, a_i\rangle})$ is convex.

(vii) Let $A \in \mathbb{R}^{n\times d}$ have full column rank. Prove that $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ is $\sigma_{min}^2(A)$–strongly convex.

(viii) Now suppose that the function $f(x)$ is μ–strongly convex, that is, it satisfies

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|y - x\|_2^2, \quad \forall\, x, y \in \mathbb{R}^d. \tag{5}$$

  Prove that $f(x)$ satisfies the Polyak–Lojasiewicz condition, that is

$$\|\nabla f(x)\|_2^2 \geq 2\mu(f(x) - f(x^*)), \quad \forall x \tag{6}$$

3 Smoothness

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice continuously differentiable function.

(i) Show that the following assertions are equivalent :

a) $\|\nabla f(x) - \nabla f(y)\| \leq \|x - y\|, \quad \forall\, x, y \in \mathbb{R}^d$

b) $f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|_2^2, \quad \forall\; x, y \in \mathbb{R}^d$.

c) $\langle \nabla^2 f(x) v, v \rangle \leq L\|v\|_2^2, \quad \forall\; x, v \in \mathbb{R}^d$

When f verifies one of the above inequalities we say that f is L-smooth or that the gradients of f are Lipschitz continuous.

(ii) Prove that $x \mapsto \frac{1}{2}\|x\|_2^2$ is 1-smooth.

(iii) For every twice differentiable L–smooth function $f : y \in \mathbb{R}^m \mapsto f(y)$, prove that $g : x \in \mathbb{R}^d \mapsto f(Ax - b)$ is a smooth function, where $A \in \mathbb{R}^{n \times d}$ and $b \in \mathbb{R}^n$. Find the smoothness constant of g.

(iv) Let $f_i : \mathbb{R}^d \mapsto \mathbb{R}$ be be a twice differentiable and L_i–smooth for $i = \{1, \ldots, n\}$. Prove that $\frac{1}{n}\sum_{i=1}^n f_i$ is $\left(\frac{1}{n}\sum_{i=1}^n L_i\right)$-smooth.

(v) For a given scalars $y_i \in \mathbb{R}$ and vectors $a_i \in \mathbb{R}^d$ for $i = \{1, \ldots, m\}$ prove that the logistic regression function $f(x) = \frac{1}{n}\sum_{i=1}^n ln(1 + e^{-y_i\langle x, a_i \rangle})$ is smooth. Find the smoothness constant!

(vi) Let $A \in \mathbb{R}^{n \times d}$ be any matrix. Prove that $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ is $\sigma_{max}^2(A)$–smooth.

(vii) Let $C > 0$ be a positive constant. Let $f(x) = \frac{1}{n}\sum_{i=1}^n \phi_i(a_i^\top x)$ where $\phi_i : \mathbb{R} \to \mathbb{R}$ is a scalar function such that $\phi_i^{''}(t) \leq C, \ \forall\, t \in \mathbb{R}$. Prove that $f(x)$ is $\frac{C}{n}\sigma_{max}^2(A)$-smooth. With this result, can you find a better estimate of the smoothness constant of the logistic regression loss ?
Hint: Show that $-\nabla f^2(x) + \frac{C}{n}A^\top A$ is positive semidefinite.

(viii) *Co-coercivity.* Let f be L-smooth, show that

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \frac{1}{L}\|\nabla f(y) - \nabla f(x)\|_2^2$$

Hint: Show that $f(y) - f(x) \leq \langle \nabla f(y), y - x \rangle - \frac{1}{2L}\nabla f(x) - \nabla f(y)\|_2^2$.