Exercise List: Properties and examples of convexity and smoothness

Robert M. Gower and Francis Bach.

September 13, 2019

Time to get familiarized with convexity, smoothness and a bit of strong convexity.

Notation: For every $x, y, \in \mathbb{R}^d$ let $\langle x, y\rangle \overset{\text{def}}{=} x^\top y$ and let $\|x\|_2 = \sqrt{\langle x, x\rangle}$.

Let $\sigma_{\min}(A)$ and $\sigma_{\max}(A)$ be the smallest and largest singular values of A defined by

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

Thus clearly

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

Let $\|A\|_F^2 \overset{\text{def}}{=} \mathrm{Tr}\left(A^\top A\right)$ 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 Convexity

We say that a twice differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y), \quad \forall x, y \in \mathbb{R}^d, \lambda \in [0,\, 1]. \tag{4}$$

or equivalently

$$v^\top \nabla^2 f(x) v \geq 0, \quad \forall x, v \in \mathbb{R}^d. \tag{5}$$

We say that f is μ–strongly convex if

$$v^\top \nabla^2 f(x) v \geq \mu \|v\|_2^2, \quad \forall x, v \in \mathbb{R}^d. \tag{6}$$

Ex. 1 — We say that $\|\cdot\| \to \mathbb{R}_+$ is a norm over $\mathbb{R}^d$ if it satisfies the following three properties

1. **Point separating:** $\|x\| = 0 \Leftrightarrow x = 0, \forall x \in \mathbb{R}^d$.
2. **Subadditive:** $\|x + y\| \leq \|x\| + \|y\|, \forall x, y \in \mathbb{R}^d$
3. **Homogeneous:** $\|ax\| = |a|\|x\|, \forall x \in \mathbb{R}^d, a \in \mathbb{R}$.

Part I

Prove that $x \mapsto \|x\|$ is a convex function.

Part II

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 $b \in \mathbb{R}^n$.

Part III

Let $f_i : \mathbb{R}^d \to \mathbb{R}$ be convex for $i = 1, \ldots, n$. Prove that $\sum_{i=1}^n f_i$ is convex.

Part IV

For 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 convex.

Part V

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

Part VI

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

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|y - x\|_2^2, \quad \forall x, y \in \mathbb{R}^d. \tag{7}$$

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

$$\|\nabla f(x)\|_2^2 \geq 2\mu(f(x) - f(x^*)), \quad \forall x. \tag{8}$$

2 Smoothness

We say that a function $f : \mathbb{R}^d \to \mathbb{R}$ is L–smooth if

$$\|\nabla f(x) - \nabla f(y)\| \quad \leq \quad L\|x - y\| \tag{11}$$

or equivalently if f is twice differentiable then

$$v^\top \nabla^2 f(x) v \leq L\|v\|_2^2, \quad \forall x, v \in \mathbb{R}^d. \tag{12}$$

Ex. 2 — *Part I*

Prove that $x \mapsto \frac{1}{2}\|x\|^2$ is 1–smooth.

Part II

Let $f : \mathbb{R}^d \to \mathbb{R}$ be twice differentiable and L–smooth. Show that

$$\sigma_{\max}(\nabla^2 f(x)) = \|\nabla^2 f(x)\|_2 \leq L.$$

Part III

For every twice differentiable L–smooth function $f : y \in \mathbb{R}^n \mapsto f(y)$, 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.

Part IV

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

Part V

For given scalars $y_i \in \mathbb{R}$ and vectors $a_i \in \mathbb{R}^d$ for $i = 1, \ldots, n$ 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!

Part VI

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

Part VII

Let $M > 0$ 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 $\phi_i''(t) \leq M$ for all $t \in \mathbb{R}$. Prove that $f(x)$ is $\frac{M}{n}\sigma_{\max}^2(A)$–smooth.

With this result, can you find a better estimate of the smoothness constant of the logistic regression loss?
Hint 1: ...

Part VIII

Co-coercivity. Let f be L-smooth, show that

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

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