Exercise List: Properties and examples of convexity and smoothness

Robert M. Gower.

October 2, 2017

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

\[
\sigma_{\text{min}}(A) \overset{\text{def}}{=} \min_{x \in \mathbb{R}^d} \frac{\|Ax\|_2}{\|x\|_2}, \quad \sigma_{\text{max}}(A) \overset{\text{def}}{=} \max_{x \in \mathbb{R}^d} \frac{\|Ax\|_2}{\|x\|_2}.
\]

(1)

Thus clearly

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

(2)

Let $\|A\|_F^2 \overset{\text{def}}{=} \text{Tr}(A^\top A)$ denote the Frobenius norm of $A$. Finally, a result you will need, for every symmetric positive semi-definite matrix $G$ the $L^2$ induced matrix norm can be equivalently defined by

\[
\|G\|_2 = \sigma_{\text{max}}(G) = \sup_{x \in \mathbb{R}^d, x \neq 0} \frac{\langle Gx, x \rangle_2}{\|x\|_2^2} = \max_{x \in \mathbb{R}^d, x \neq 0} \frac{\|Gx\|_2}{\|x\|_2}.
\]

(3)

1 Convexity

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

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

(4)

or equivalently

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

(5)

We say that $f$ is $\mu$–strongly convex if

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

(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:** $\|x\| = 0 \iff x = 0, \forall x \in \mathbb{R}^d$.
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 $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$.

**Part III**
Let $f_i : \mathbb{R}^d \to \mathbb{R}$ be convex for $i = 1, \ldots, m$. Prove that $\sum_{i=1}^m f_i$ 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 $f(x) = \sum_{i=1}^m \ln(1 + e^{-y_i \langle x, a_i \rangle})$ is convex.

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

**Answer (Ex. I)** — Let $x, y \in \mathbb{R}^d$ and $\lambda \in [0, 1]$. It follows that

$$\|\lambda x + (1 - \lambda)y\| \overset{item 2}{\leq} \|\lambda x\| + \|(1 - \lambda)y\|$$

$$\overset{item 3}{\leq} \lambda\|x\| + (1 - \lambda)\|y\|. \quad \blacksquare$$

**Answer (Ex. III)** — Let $x, y \in \mathbb{R}^d$ and $\lambda \in [0, 1]$. It follows that

$$g(\lambda x + (1 - \lambda)y) = f(A(\lambda x + (1 - \lambda))y - b)$$

$$= f(\lambda(Ax - b) + (1 - \lambda)(Ay - b))$$

$$= \lambda f(Ax - b) + (1 - \lambda)f(Ay - b) \quad (7)$$

$f$ is convex. $\lambda f(Ax - b) + (1 - \lambda)f(Ay - b). \quad \blacksquare$

**Answer (Ex. VI)** — Immediate through either definition.
**Answer (Ex. IV)** — From exercise VI we need only prove that \( f(x) = \ln(1 + e^{-y(x,w)}) \) is convex for a given \( y \in \mathbb{R} \) and \( w \in \mathbb{R}^d \). From exercise III we need only prove that \( \phi(\alpha) = \ln(1 + e^\alpha) \) is convex, since \( x \mapsto -y(x,w) \) is a linear function. The convexity of \( f(\alpha) \) now follows by differentiating once

\[
\phi'(\alpha) = \frac{e^\alpha}{1 + e^\alpha},
\]

then differentiating again

\[
\phi''(\alpha) = \frac{e^\alpha}{(1 + e^\alpha)^2} - \frac{e^{2\alpha}}{(1 + e^\alpha)^2} = \frac{e^\alpha}{(1 + e^\alpha)^2} \geq 0, \quad \forall \alpha.
\]

We can now call upon the definition (5), but since \( \alpha \in \mathbb{R} \) is a scalar, the above already proves that \( \phi(\alpha) \) is convex.

**Answer (Ex. V)** — Differentiating twice we have that

\[
\nabla^2 f(x) = A^\top A.
\]

Consequently

\[
v^\top \nabla^2 f(x)v = v^\top A^\top A v = \|Av\|^2_2 \geq \sigma_{\text{min}}(A)^2 \|v\|^2_2.
\]

## 2 Smoothness

We say that a function \( f : \mathbb{R}^d \to \mathbb{R} \) is \( L \)-smooth if

\[
\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \quad (9)
\]

or equivalently if \( f \) is twice differentiable then

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

**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

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

Part III

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

Part IV

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

Part V

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

Part VI

Let \( A \in \mathbb{R}^{m \times d} \) be any matrix. Prove that \( \|Ax - b\|^2_2 \) is \( \sigma_{\text{max}}^2(A) \)-smooth. Hint 1: ...

Answer (Ex. I) — Clearly \( \nabla^2 \frac{1}{2} \|x\|^2 = I \) and thus follows from definition (9).

Answer (Ex. II) — Using the definition of the induced norm we have that
\[
\|\nabla^2 f(x)\|_2^2 = \sup_{v \neq 0} \frac{v^\top \nabla^2 f(x)v}{\|v\|_2^2} \overset{(10)}{\leq} \sup_{v \neq 0} \frac{L\|v\|_2^2}{\|v\|_2^2} = L.
\]

Answer (Ex. III) — Differentiating \( g(x) \) once gives
\[
\nabla g(x) = A^\top \nabla f(Ax - b).
\]

First we prove the claim using the definition (9). Indeed note that
\[
\|\nabla g(x) - \nabla g(y)\|_2 \leq \|A^\top (\nabla f(Ax - b) - \nabla f(Ay - b))\|_2 \\
\leq \|A^\top\|_2\|\nabla f(Ax - b) - \nabla f(Ay - b)\|_2 \\
\leq L\|A^\top\|_2\|Ax - b - (Ay - b)\|_2 \\
\leq L\|A^\top\|_2\|A\|_2\|x - y\|_2.
\]

This the smoothness parameter is given by \( L\|A\|_2^2 \) where we used that \( \|A^\top\|_2 = \|A\|_2 \). This completes the proof.
We can also prove the claim using (10). Differentiating again we have that
\[ \nabla^2 g(x) = A^\top \nabla^2 f(Ax - b) A. \]
Consequently
\[ \| \nabla^2 g(x) \|^2_2 \leq \| A \|^2_2 \| \nabla^2 f(Ax - b) \|^2_2 \leq L \| A \|^2_2. \]
We could further tighten this by considering the smoothness constant of \( f \) restricted to the set \( \{ x | Ax - b \} \) which might be smaller than \( \mathbb{R}^d \).

**Answer (Ex. IV)** — Clearly
\[ \nabla^2 \left( \frac{1}{n} \sum_{i=1}^{n} f_i(x) \right) = \frac{1}{n} \sum_{i=1}^{n} \nabla^2 f_i(x) \leq \frac{1}{n} \sum_{i=1}^{n} L_i I. \]
You can also prove this using the definition (9) and applying repeatedly the subadditivity of the norm.

**Answer (Ex. V)** — First note that from (11) the function \( \phi(\alpha) = \ln(1 + e^\alpha) \) is 1–smooth. Consequently from exercise III the function \( f_i(x) = \ln(1 + e^{-y_i \langle x, a_i \rangle}) \) is \( y_i^2 \| a_i \|^2_2 \)–smooth.

Finally from exercise IV the logistic regression function is \( \sum_{i=1}^{m} \frac{y_i^2 \| a_i \|^2_2}{m} \)–smooth. But this is not the tightest smoothness constant. Indeed, first it is not hard to show that
\[ \phi''(\alpha) = \frac{e^\alpha}{(1 + e^\alpha)^2} \leq \frac{1}{4}, \quad \forall \alpha. \tag{11} \]
Furthermore, by analysing directly the Hessian of \( f(x) - \sum_{i=1}^{m} f_i(x) \) we see that
\[ \nabla^2 f(x) = A^\top \Phi(x) A, \]
where \( \Phi(x) = \text{diag} \left( \frac{e^{\alpha_i}}{(1 + e^{\alpha_i})^2} \right) \), where \( \alpha_i = -y_i \langle a_i, x \rangle \). Consequently
\[ \| \nabla^2 f(x) \|_2 = \| A^\top \Phi(x) A \|_2 \leq \| A \|^2_2 \| \Phi(x) \|_2 \leq \frac{\| A \|^2_2}{4}. \]
This is a much tighter smoothness constant.

**Answer (Ex. VI)** — Differentiating twice we have that
\[ \nabla^2 f(x) = A^\top A. \]
Consequently
\[ v^\top \nabla^2 f(x) v = v^\top A^\top A v \leq \| A v \|^2_2 \leq \| v \|^2_2. \]