Convex optimization exercise sheet

Mathurin Massias

Notation

For a linear operator A, its adjoint is A^* and its Moore-Penrose pseudoinverse is $A^\dagger$. $\mathcal{X}$ and $\mathcal{Y}$ are two Hilbert spaces.

1 Convexity

Exercise 1.1 (Pointwise sup preserves convexity)**.** *Let $(f_i)_I$ be a family of convex functions (not necessarily countable). Show that $x \mapsto \sup_{i \in I} f_i(x)$ is convex.*

Exercise 1.2 (Precomposition by linear operator preserves convexity)**.** *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function and $A : \mathcal{X} \to \mathcal{Y}$ a linear operator. Show that $f(A\cdot)$ is convex (on $\mathcal{Y}$).*

Exercise 1.3 (Misconceptions on existence of minima)**.** *Provide an example of convex function which does not admit a minimizer.*
What if the function is continuous and lower bounded ?

2 Least squares

2.1 From a linear algebra perspective

Exercise 2.1. *Let $A \in \mathbb{R}^{n \times d}$. Show that* $\operatorname{Ker} A = \operatorname{Ker} A^* A$.

Exercise 2.2. *Show that there always exist a solution to $A^*Ax = A^*b$.*
Show that this no longer holds in the infinite dimensional space (when A is a bounded linear operator between infinite dimensional Hilbert spaces.)

Exercise 2.3. *Let $A \in \mathbb{R}^{n \times d}$, $b \in \mathbb{R}^n$. Show that solving Ordinary Least Squares:*

$$\min \frac{1}{2} \|Ax - b\|^2 \ ,$$

amounts to solving $A^*Ax = A^*b$ (aka the normal equations*).*

Show that the set of solutions is:

$$(A^*A)^\dagger A^* b + \operatorname{Ker} A \ .$$

2.2 Gradient descent on least squares

Exercise 2.4 (Gradient descent on isotropic parabola)**.** *Let $A \in \mathbb{R}^{n \times d}$ be such that its condition number*[1] *$\kappa(A)$ is equal to 1. Show that gradient descent with stepsize $1/L$ converges in a single iteration for the problem* $\min \frac{1}{2}\|Ax - b\|^2$.

[1] i.e. the ratio between the largest and the smallest eigenvalues of A^*A.

3 Gradient

Exercise 3.1. *Provide an example of setting where the gradient is not equal to the vector of partial derivatives.*

Exercise 3.2. *Let $f : \mathcal{X} \to \mathbb{R}$ and $\theta : \mathbb{R}_+ \to \mathcal{X}$ be differentiable. Show that*

$$\frac{d}{dt} f(\theta(t)) = \langle \nabla f(\theta(t)), \dot{\theta}(t) \rangle \ .$$

Exercise 3.3. *Show that the gradient of a function is orthogonal to the level lines of that function.*

Exercise 3.4. *Let $A \in \mathbb{R}^{n \times d}$, $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ defined as $g(x) = f(Ax)$ for all $x \in \mathbb{R}^d$. Show that*

$$\nabla g(x) = A^* \nabla f(Ax) \ ,$$
$$\nabla^2 g(x) = A^* \nabla^2 f(Ax) A \ .$$

Exercise 3.5 (Polyak-Łojasiewicz inequality)**.** *Let f be a μ-strongly-convex and differentiable function. Let $x^* = \operatorname{argmin} f(x)$. Show that f satisfies the Polyak-Łojasiewicz inequality:*

$$\mu(f(x) - f(x^*)) \le \frac{1}{2} \|\nabla f(x)\|^2 \ . \tag{1}$$

Provide an example of function which is not strongly convex, but satisfies the inequality.

Exercise 3.6. *Provide an example of matrix $A \in \mathbb{R}^{n \times d}$ such that the gradient of $x \mapsto \frac{1}{2} x^\top A x$ is* not *equal to Ax.*

4 Convexity inequalities

Exercise 4.1. *Let f be a convex and Gateaux-differentiable function. Let $L > 0$. Show that the following properties are equivalent:*

1. $\forall x, y,\ \|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|$
2. $\forall x, y,\ f(x) \le f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2}\|x - y\|^2$
3. $\forall x, y,\ \frac{1}{L}\|\nabla f(x) - \nabla f(y)\| \le \langle x - y, \nabla f(x) - \nabla f(y) \rangle$

Note: 2 is known as the descent lemma; 3 is known as the Baillon-Haddad theorem, or cocoercivity of the gradient.

Exercise 4.2. *Let f be a L-smooth μ-strongly convex function. Show that for any x, y,*

$$\frac{\mu L}{\mu + L}\|x - y\|^2 + \frac{1}{L + \mu}\|\nabla f(x) - \nabla f(y)\|^2 \le \langle x - y, \nabla f(x) - \nabla f(y) \rangle \ .$$

5 Around fixed point schemes

Exercise 5.1. *Let $T : \mathcal{X} \to \mathcal{X}$ be a linear operator which is q-contractive, meaning that for all $x \in \mathcal{X}$,*

$$\|Tx\| \le q\|x\| \ .$$

Show that T admits at most one fixed point.
Show that the sequence defined by $x_0 \in \mathcal{X}$, $x_{k+1} = Tx_k$ converges, and that the limit is a fixed point.
Show that, denoting x^* this fixed point, the sequence x_k converges to x^* at linear speed.

Exercise 5.2. *Show that the results of the above exercise do not hold when $q = 1$.*

6 Constrained optimization

Exercise 6.1. *Show that the indicator function ι_C is convex (resp. lower semicontinuous, resp. proper) if C is convex (resp. closer, resp. nonempty).*

Exercise 6.2 (Global optimality condition for constrained convex optimisation)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex differentiable function, let C be a convex subset of $\mathbb{R}^d$. Show that $x^* \in \operatorname{argmin}_{x \in C} f(x)$ if and only if*

$$\forall x \in \mathbb{R}^d,\ \langle \nabla f(x^*), x - x^* \rangle \geq 0 \ \ .$$