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 Continuous time

Exercise 3.1 (Preliminaries)**.** *Let $f : \mathcal{X} \to \mathbb{R}$ be twice differentiable. Let $\theta : \mathbb{R}_+ \to \mathcal{X}$ be differentiable. Show that*

$$\frac{d}{dt} f(\theta(t)) = \langle \nabla f(\theta(t)), \dot{\theta}(t) \rangle \ ,$$
$$\frac{d}{dt} \nabla f(\theta(t)) = \nabla^2 f(\theta(t)) \dot{\theta}(t) \ .$$

Exercise 3.2 (Everything decreases in gradient flow)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex differentiable function. Let $x(t)$ be the gradient flow on f, defined as a solution of $\dot{x}(t) = -\nabla f(x(t))$.*
Show that $f(x(t))$ decreases.
Show that $\|\nabla f(x(t))\|$ decreases.

4 Gradient

Exercise 4.1. *Provide an example of setting where the gradient is not equal to the vector of partial derivatives.*

Exercise 4.2. *Show that the gradient of a function is orthogonal to the level lines of that function.*

Exercise 4.3. *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 4.4 (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:*

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

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

Exercise 4.5. *Provide an example of matrix $A \in \mathbb{R}^{n \times d}$ such that the gradient of $x \mapsto \frac{1}{2} x^\top Ax$ is* not *equal to Ax.*

Exercise 4.6 (Exercise 3.2 in discrete time)**.** *Let f be a convex L-smooth function from $\mathbb{R}^d$ to $\mathbb{R}$. Show that the iterates of gradient descent on f with step size $0 < \alpha < 2/L$ have decreasing gradient norm.*
Is it still true when f is not convex?

5 Convexity inequalities

Exercise 5.1. *Let f be a convex and Gateaux-differentiable function. Let $L > 0$. Show that the following properties are equivalent:*

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

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

Exercise 5.2. *Let f be a L-smooth μ-strongly convex function. Show that for any x, y,*

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

Exercise 5.3. *Let $f : \mathbb{R}^d \to \mathbb{R}$, let $\mu > 0$. Show that f is μ-strongly convex function* with respect to the ℓ_2-norm *if and only if $f - \frac{\mu}{2}\|\cdot\|^2$ is convex.*
Provide a counter example when the underlying norm is not the Euclidean one.

6 Around fixed point schemes

Exercise 6.1. *Let $T : \mathcal{X} \to \mathcal{X}$ be a q-contractive operator, meaning that for all $x \in \mathcal{X}$,*

$$\|Tx\| \leq q\|x\| \ ,$$

with $q < 1$.
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 6.2. *Show that the results of the above exercise do not hold when $q = 1$.*

7 Constrained optimization

Exercise 7.1. *Show that the indicator function ι_C is convex (resp. lower semicontinuous, resp. proper) if C is convex (resp. closer, resp. nonempty).*

Exercise 7.2 (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*

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

8 Subdifferential and subgradients

Exercise 8.1. *Provide an example of convex function which has an empty subdifferential at some point of its domain.*

Exercise 8.2. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be separable: $f(x) = \sum_1^d f_i(x_i)$ where the f_i's are functions of the real variable.*
Show that $\partial f(x) = \partial f_1(x_1) \times \ldots \times f_d(x_d)$.
Compute the subdifferential of the ℓ_1-norm.

9 Fenchel transforms

Exercise 9.1. *Compute the Fenchel transform of the ℓ_p-norm for $p \in [1, +\infty]$.*

Exercise 9.2. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be separable: $f(x) = \sum_1^d f_i(x_i)$ where the f_i's are functions of the real variable.*
Show that $f^*(u) = (f_i^*(u_i))_{i \in [d]}$.

Exercise 9.3 ("Lipschitzing trick")**.** *Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be convex.*
Show that f *if* B*-Lipschitz if and only if the domain of* f^* *is included in the Euclidean ball of center 0 and radius* B*.*

Exercise 9.4 (Convolution smoothing)**.** *Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be convex.*
Show that $f \,\Box\, \frac{L}{2}\|\cdot\|^2$ *if* L*-smooth.*

10 The ℓ_1-norm

Exercise 10.1. *It is often claimed that the* ℓ_1*-norm is not differentiable at 0. Do you know a norm which is differentiable at 0?*

Exercise 10.2 (Huber function)**.** *Compute* $\|\cdot\|_1 \Box \frac{\varrho}{2}\|\cdot\|^2$*. In which sense is it a smooth approximation of the* ℓ_1*-norm?*

11 Logistic regression

Exercise 11.1. *Let* $b \in \{-1, 1\}^n, A \in \mathbb{R}^{n\times d}$*. Let* $f(x) = \sum_1^n \log(1 + \exp(-b_i a_i^\top x))$*.*
Compute ∇f*,* $\nabla^2 f$*.*
Is f *convex? Strictly convex? Strongly convex?*
Compute f^**.*