

# 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. Number of coffee cups (☕) indicates difficulty of exercise.

## 1 Convexity

**Exercise 1.1.** ☕ Show that local minimizers of convex functions are global minimizers.

**Exercise 1.2** (Pointwise supremum 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 1.3** (Precomposition by linear operator preserves convexity). ☕ Let  $f : \mathcal{Y} \rightarrow \mathbb{R}$  be a convex function and  $A : \mathcal{X} \rightarrow \mathcal{Y}$  a linear operator. Show that  $f(A \cdot)$  is convex (on  $\mathcal{X}$ ).

**Exercise 1.4** (Misconceptions on existence of minimizers). ☕ Provide an example of convex function which does not admit a minimizer.

What if the function is continuous and lower bounded?

**Exercise 1.5.** ☕ Show that a strictly convex function has at most one minimizer.

**Exercise 1.6** (Jensen's inequality). ☕ Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be convex. Let  $n \in \mathbb{N}$ ,  $x_1, \dots, x_n \in \mathbb{R}^d$ , and let  $\lambda_1, \dots, \lambda_n$  be positive scalars summing to 1.

Show that  $f(\sum_{i=1}^n \lambda_i x_i) \leq \sum_{i=1}^n \lambda_i f(x_i)$ .

**Exercise 1.7.** ☕ Show that a function is lower semicontinuous if its sublevel sets are closed.

**Exercise 1.8.** ☕ Show that the sublevel sets of a convex function are convex. Find a function with convex sublevel sets which is not convex.

**Exercise 1.9** (First order characterization of convex functions). ☕☕ Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a differentiable function. Show that the following are equivalent:

1.  $f$  is convex
2.  $f$  lies above its tangents:  $\forall x, y \in \mathbb{R}^d, f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$
3.  $\nabla f$  is monotone:  $\forall x, y \in \mathbb{R}^d, \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$

**Exercise 1.10** (Continuity). ☕☕☕ Show that a convex function is locally Lipschitz (hence continuous) on the interior of its domain.

**Exercise 1.11** (Characterization of strongly convex functions in the Euclidean case). ☕ Show that  $f$  is  $\mu$ -strongly convex with respect to the Euclidean norm if and only if  $f - \frac{\mu}{2} \|\cdot\|^2$  is convex.

**Exercise 1.12.** ☕ Show that a strongly convex function admits exactly one minimizer.

## 2 Least squares

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

Show that for any  $b \in \mathbb{R}^n$  there exist a solution to  $A^*Ax = A^*b$ .

☹☹ Show that there does not always exist a solution to  $A^*Ax = A^*b$  in the infinite dimensional case (when  $A$  is a bounded linear operator between infinite dimensional Hilbert spaces.)

**Exercise 2.2.** ☹ Let  $A \in \mathbb{R}^{n \times d}$ ,  $b \in \mathbb{R}^n$ . Show that solving Ordinary Least Squares:

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

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

Show that the set of solutions is:

$$A^\dagger b + \text{Ker } A.$$

**Exercise 2.3.** ☹ When is  $x \mapsto \|Ax - b\|^2$  strictly convex? Strongly convex?

**Exercise 2.4** (Least squares with intercept). ☹☹ An intercept  $x_0$  is a constant scalar term in the linear prediction function, that becomes  $a \mapsto a^\top x + x_0$ . Fitting an intercept can be done by adding a column of 1s to  $A$ . Alternatively, show that the solution of least squares with intercept,

$$(\hat{x}, \hat{x}_0) \in \underset{x \in \mathbb{R}^d, x_0 \in \mathbb{R}}{\text{argmin}} \frac{1}{2} \|Ax - b - x_0 \mathbf{1}\|^2 \quad (2)$$

is given by:

$$\hat{x} = \hat{x}_c, \quad (3)$$

$$\hat{x}_0 = \frac{1}{n} \sum_{i=1}^n (a_i^\top \hat{x} - b_i), \quad (4)$$

where  $\hat{x}_c$  is the solution of least squares without intercept on centered data  $A_c$  and  $b_c$  (versions of  $A$  and  $b$  where the rowwise mean has been subtracted).

**Exercise 2.5** (Gradient descent on isotropic parabola). ☹ Let  $A \in \mathbb{R}^{n \times d}$  be such that the condition number of  $A^\top A$  is equal to 1. Show that gradient descent with stepsize  $1/L$  converges in a single iteration for the problem  $\min \frac{1}{2} \|Ax - b\|^2$ .

## 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.** ☹ Show that the gradient of a function is orthogonal to the level lines of that function.

**Exercise 3.3.** ☹ Compute the gradients and Hessians of  $x \mapsto \|x\|^2$ ,  $x \mapsto \|x\|$ ,  $x \mapsto a^\top x$ . Is it true that the gradient of  $x \mapsto \frac{1}{2} x^\top Ax$  is equal to  $Ax$ ?

**Exercise 3.4.** ☹☹ Compute the gradient of the logdet function,  $M \mapsto \log \det(M)$ .

**Exercise 3.5.** ☹ Let  $A \in \mathbb{R}^{n \times d}$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  defined as  $g(x) = f(Ax)$  for all  $x \in \mathbb{R}^d$ . Show that

$$\begin{aligned} \nabla g(x) &= A^* \nabla f(Ax), \\ \nabla^2 g(x) &= A^* \nabla^2 f(Ax) A. \end{aligned}$$

---

<sup>1</sup>i.e. the ratio between the largest and the smallest eigenvalues of  $A^\top A$ .

## 4 Convexity inequalities

**Exercise 4.1.** 🐼🐼 Let  $f$  be a convex and 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.  $\forall x, y, \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \leq \langle x - y, \nabla f(x) - \nabla f(y) \rangle$

**Exercise 4.2.** 🐼 Let  $f$  be a twice differentiable  $L$ -smooth function. Show that for all  $x \in \mathbb{R}^d$ ,  $\nabla^2 f(x) \preceq L \text{Id}$ .

**Exercise 4.3.** 🐼 Let  $f$  be a differentiable function. Show that the following properties are equivalent:

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

**Exercise 4.4.** 🐼 Let  $f$  be a twice differentiable  $\mu$ -strongly convex function. Show that for all  $x \in \mathbb{R}^d$ ,  $\mu \text{Id} \preceq \nabla^2 f(x)$ .

**Exercise 4.5** (Polyak-Łojasiewicz inequality). 🐼🐼 Let  $f$  be a  $\mu$ -strongly-convex and differentiable function. Let  $x^* = \text{argmin } f(x)$ . Show that  $f$  satisfies the Polyak-Łojasiewicz inequality:

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

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

**Exercise 4.6.** 🐼🐼 Let  $f$  be a  $L$ -smooth  $\mu$ -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 4.7** (3 point descent lemma). 🐼 Let  $f$  be convex and  $L$ -smooth. Show that for any triplet  $(x, y, z)$ ,

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

## 5 Gradient descent

**Exercise 5.1** (Exercise 8.2 in discrete time). 🐼🐼 Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a convex, twice differentiable  $L$ -smooth function.

Show that the iterates of gradient descent on  $f$  with step size  $0 < \alpha < 2/L$  have decreasing gradient norm.

Can you show it if  $f$  is not twice differentiable?

Is it still true when  $f$  is not convex?

**Exercise 5.2.** 🐼 Provide a finite dimensional example of convex  $L$ -smooth function  $f$  such that gradient descent with stepsize  $< 2/L$  diverges.

**Exercise 5.3** (Gradient descent on Lipschitz function without knowing the horizon). 🐼🐼 For gradient descent on a Lipschitz differentiable objective, the classical proof assumes that the total number of iterations  $t$  is known in advance, to set the fixed stepsize  $\eta \propto 1/\sqrt{t}$ .

Show that using a decreasing stepsize  $\eta_k \propto 1/\sqrt{k}$  leads to a rate of order  $\log k/\sqrt{k}$  on the ergodic or best iterate.

## 6 Constrained optimization

**Exercise 6.1.** ☛ Show that the indicator function  $\iota_C$  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  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  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

$$\forall x \in C, \langle \nabla f(x^*), x - x^* \rangle \geq 0 .$$

**Exercise 6.3.** ☛☛ Let  $C$  be a non-empty closed convex set. Show that for all  $x \in C, y \in \mathbb{R}^d$ ,

$$\|x - y\|^2 \geq \|x - \Pi_C(y)\|^2 + \|y - \Pi_C(y)\|^2 .$$

In particular this shows that  $\|y - x\| \geq \|x - \Pi_C(y)\|$ , which says that projection can only get you closer to optimum (taking  $x = x^*$ , if your current iterate is  $y$ ).

## 7 Subdifferential and subgradients

**Exercise 7.1.** ☛ Provide an example of convex function which has an empty subdifferential at some point of its domain.

**Exercise 7.2.** ☛ Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be separable:  $f(x) = \sum_1^d f_i(x_i)$  where the  $f_i$ 's are functions of the real variable.

Show that  $\partial f(x) = \partial f_1(x_1) \times \dots \times \partial f_d(x_d)$ .

Compute the subdifferential of the  $\ell_1$ -norm.

**Exercise 7.3.** ☛ Show that the subdifferential of a function at a point writes as an intersection of half-spaces. Show that it is convex and closed.

**Exercise 7.4.** ☛ Show that the subdifferential of  $\iota_C$  at  $x \in C$  is equal to the normal cone of  $C$  at  $x$ , that is:

$$\mathcal{N}_C(x) = \{u : \forall z \in C, \langle u, z - x \rangle \leq 0\}$$

**Exercise 7.5.** ☛☛ Compute the subdifferential of the Euclidean norm at any point in  $\mathbb{R}^d$ .

**Exercise 7.6** (Exercise 6.2 revisited). ☛☛ Show that the first order optimality condition for differentiable convex constrained optimization rewrites:

$$x^* \in \operatorname{argmin}_C f(x) \Leftrightarrow -\nabla f(x^*) \in \mathcal{N}_C(x^*) .$$

**Exercise 7.7** (Non emptiness of subdifferential). ☛☛☛ Let  $f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$  be convex. Show that if  $x \in \operatorname{int}(\operatorname{dom} f)$ ,  $\partial f(x)$  is non empty and compact.

Show that if  $x$  lies on the boundary of  $\operatorname{dom} f$ ,  $\partial f(x)$  is either empty or unbounded.

## 8 Continuous time

**Exercise 8.1** (Preliminaries). ☛ Let  $f : \mathcal{X} \rightarrow \mathbb{R}$  be twice differentiable. Let  $\theta : \mathbb{R}_+ \rightarrow \mathcal{X}$  be differentiable. Show that

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

**Exercise 8.2** (Everything decreases in gradient flow). 🐢 Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a convex twice 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.

## 9 Bregmaneries

**Exercise 9.1** (Some misconceptions). 🐢 Why is a Bregman divergence not “like a distance” in general? Is it “like a distance” in the Euclidean case?

**Exercise 9.2** (KL is Bregman divergence of negative entropy). 🐢🐢 Show that the Bregman divergence associated to negative entropy  $x \mapsto \sum_1^d x_i \log(x_i)$  is the Kullback-Leibler divergence:

$$D(x, y) = \sum_1^d y_i \log \left( \frac{x_i}{y_i} \right) .$$

**Exercise 9.3.** 🐢 Let  $x, y, z \in \mathbb{R}^d$  and  $D$  be a Bregman divergence associated to a differentiable convex function  $\phi$ . Show that

$$D(x, y) + D(z, x) - D(z, y) = \langle \nabla \phi(x) - \nabla \phi(y), x - z \rangle .$$

How is this a generalization of Pythagoras theorem?

**Exercise 9.4** (Exercise 6.3 for Bregman projection). 🐢🐢 Let  $\phi$  a differentiable convex function. Let  $D$  be the associated Bregman divergence. Let  $C$  be a non-empty closed and convex subset of  $\mathbb{R}^d$ , let  $z \in \mathbb{R}^d$  and  $y = \operatorname{argmin}_{y' \in C} D(y', z)$  its Bregman projection onto  $C$ . Show that for all  $x \in C$ :

$$D(x, z) \geq D(x, y) + D(y, z) .$$

This definitely shows that  $D$  is not “like a distance”.

**Exercise 9.5** (Bregman projection onto the simplex). 🐢🐢 Show that the Bregman projection of  $x \in \mathbb{R}_+^d$  onto the simplex  $\Delta$ , when the potential is the negative entropy, is  $(x_i / \sum_1^d x_i)_i$ . Compare to the Euclidean projection.

## 10 Sparsity

**Exercise 10.1.** 🐢 It is often claimed that the  $\ell_1$ -norm is not differentiable at 0. Do you know a norm which is differentiable at 0?

**Exercise 10.2** (Huber function). 🐢🐢 Compute  $\|\cdot\|_1 \square \frac{\rho}{2} \|\cdot\|^2$ . In which sense is it a smooth approximation of the  $\ell_1$ -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  $\nabla f$ ,  $\nabla^2 f$ .

Is  $f$  convex? Strictly convex? Strongly convex?

🐢🐢 Compute  $f^*$ .

## 12 Fixed point schemes

**Exercise 12.1.** ☹☹ Let  $T : \mathcal{X} \rightarrow \mathcal{X}$  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 12.2.** ☹ Show that the results of [Exercise 12.1](#) do not hold when  $q = 1$ .