Convex optimization exercise sheet

Mathurin Massias

Notation

For a linear operator A, its adjoint is A^* and its Moore-Penrose pseudoinverse is $A^\dagger$.

1 Least squares from a linear algebra perspective

Exercise 1.1. *Let* $A \in \mathbb{R}^{n \times d}$*. Show that* $\operatorname{Ker} A = \operatorname{Ker} A^* A$*.*

Exercise 1.2. *Show that there always exist a solution to* $A^* A x = A^* b$*. PShow that this no longer holds in the infinite dimensional space (when* A *is a bounded linear operator between infinite dimensional Hilbert spaces.)*

Exercise 1.3. *Let* $A \in \mathbb{R}^{n \times d}$*,* $b \in \mathbb{R}^n$*. Show that solving Ordinary Least Squares:*

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

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

Show that the set of solutions is:

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

2 Gradient

Exercise 2.1. *Provide an example of setting where the gradient is not equal to the vector of partial derivatives.*

Exercise 2.2. *Show that*

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

Exercise 2.3. *Show that the gradient of a function is orthogonal to the level lines of that function.*

Exercise 2.4. *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*

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

Exercise 2.5. *Let* $A \in \mathbb{R}^{n \times d}$*. Provide an example where the gradient of* $x \mapsto \frac{1}{2} x^\top A x$ *is not equal to* Ax*.*

3 Convexity inequalities

Exercise 3.1. *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 3.2. *Let f be a L-smooth μ-strongly convex function. Show that for any x, y,*

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

4 Around fixed point schemes

Exercise 4.1. *Let $T : \mathcal{X} \to \mathcal{X}$ be a linear operator which is q-contractive, meaning that for all $x \in \mathcal{X}$,*

$$\|Tx\| \leq q\|x\| \ .$$

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