Convex analysis and optimisation

2020, November 18
9h30 – 12h

You may use your courses notes during the exam, but you shouldn't use any online resource nor communicate with each other. A fraud could lead to a ban to pass university exams in France for up to five years. At the end of the exams, you scan your copy and send it, preferably in a unique file (e.g. zip) to `quentin.merigot@universite-paris-saclay.fr`

The exam has two parts: Part I corresponds to the algorithmic part by J.C. Pesquet and Part II to the convex analysis part by Q. Mérigot. They are graded independently and account each for half of the final grade.

Optimisation exam

Part 1

Answer the following questions concerning the article : "Convergence of Proximal Gradient Algorithm in the Presence of Adjoint Mismatch" by E. Chouzenoux, J.-C. Pesquet, C. Riddell, M. Savanier, and Y. Trousset [1]. All the provided answers should be properly justified.

1. What is the subdifferential of the objective function minimized in (10) ?
2. Give the expression of the Douglas-Rachford algorithm when applied to (10).
3. Specify convergence conditions for the obtained algorithm.
4. Assume that $\kappa = 0$ and
$$\overline{K}H = \begin{bmatrix} 4 & 4 \\ 0 & 1 \end{bmatrix}.$$
Is L cocoercive ? Same question if $\kappa = 1$.
5. Assume that L is cocoercive and $(\forall x \in \mathcal{H})\ g(x) = \|x\|^{3/2}$.
What can be said about $\mathrm{Fix} T_\gamma$ with $\gamma \in\ ]0, +\infty[$?
6. What is the applicative interest of Theorem 3.11 ?
7. Are the conditions provided in Proposition 3.15 necessary and sufficient for the weak convergence of the proximal gradient algorithm in the presence of adjoint mismatch ? Why ?
8. Is δ defined in Section 3.4 the same as δ defined in Section 4 ?
9. What does the red plot in dashed line in Figure 3 illustrate ?
10. Which criticism does the sentence just before Figure 4 inspire you ?

1. http ://www.optimization-online.org/DB_FILE/2020/10/8055.pdf

Part II

The two exercises are independent. If you do not find the solution to a question, you can admit the answer and skip it. Given the length of this part, it will be possible to get the maximal grade without answering all the questions.

Exercise 1. We consider an optimization problem over the space of square N-by-N matrices $\mathcal{M}_N(\mathbb{R})$, endowed with the scalar product $\langle X|Y\rangle = \sum_{1\le i,j\le N} X_{ij}Y_{ij}$ and the Euclidean norm $\|X\| = \sqrt{\langle X|X\rangle}$. The space $\mathbb{R}^N$ is endowed with the standard scalar product and Euclidean norm. We consider the function $e : \mathbb{R} \to \overline{\mathbb{R}}$ defined by

$$e(r) = \begin{cases} r\log r - r & \text{if } r > 0, \\ 0 & \text{if } r = 0, , \\ +\infty & \text{if } r < 0. \end{cases}$$

and $E : \mathcal{M}_N(\mathbb{R}) \to \overline{\mathbb{R}}$ defined by $E(X) = \sum_{1\le i,j\le N} e(X_{ij})$.
We consider the set $\Gamma_N \subseteq \mathcal{M}_N(\mathbb{R})$ of matrices whose rows and columns sum to one, i.e.

$$X \in \Gamma_N \iff \begin{cases} \forall i \in \{1,\dots,N\}, \sum_{1\le j\le N} X_{ij} = 1 \\ \forall j \in \{1,\dots,N\}, \sum_{1\le i\le N} X_{ij} = 1 \end{cases}$$

We finally fix a matrix $C \in \mathcal{M}_N(\mathbb{R})$, and we consider the minimisation problem

$$P = \min_{X\in\Gamma_N} \langle C|X\rangle + E(X)$$

1. Prove that e (resp. E) is strictly convex and lower semi-continuous, and compute its convex conjugate e^* (resp. E^*).
2. Prove that the minimization problem P has a unique solution.

   We introduce a few notations:
   - $\mathbf{1}_N = (1,\dots,1)^T$ *is the column vector of length N containing only* 1s.
   - *We define a linear operator* $A : \mathcal{M}_N(\mathbb{R}) \to \mathbb{R}^N \times \mathbb{R}^N$ *by* $AX = (X\mathbf{1}_N, (X^T)\mathbf{1}_N)$.
   - *The adjoint of A is the linear operator* $A^* : \mathbb{R}^N \times \mathbb{R}^N \to \mathcal{M}_N(\mathbb{R})$ *characterized by* $\langle A(X)|(\Phi,\Psi)\rangle = \langle X|A^*(\Phi,\Psi)\rangle$ *for any* $X \in \mathcal{M}_N(\mathbb{R})$ *and any* $\Phi,\Psi \in \mathbb{R}^N$.
   - $G(X) = \langle C|X\rangle + E(X)$,
   - $H : \mathbb{R}^N\times\mathbb{R}^N \to \overline{\mathbb{R}}$ *is the convex indicator function of* $(\mathbf{1}_N, \mathbf{1}_N)$, *i.e.* $H(U,V) = 0$ *if* $U = V = \mathbf{1}_N$ *and* $+\infty$ *otherwise.*
3. Show that $P = \min_{X\in\mathcal{M}_N(\mathbb{R})} G(X) + H(AX)$.
4. Using Fenchel-Young's inequality, prove that $P \ge D$, where

$$D = \sup_{(\Phi,\Psi)\in\mathbb{R}^N\times\mathbb{R}^N} -G^*(A^*(\Phi,\Psi)) - H^*(-\Phi,-\Psi).$$

5. Compute A^* (you should find $A^*(\Phi,\Psi)_{ij} = \Phi_i + \Psi_j$), G^* and H^*. Deduce that the dual problem can be rewritten as

$$D = \sup_{(\Phi,\Psi)\in\mathbb{R}^N\times\mathbb{R}^N} \langle\Phi|\mathbf{1}_N\rangle + \langle\Psi|\mathbf{1}_N\rangle - \sum_{1\le i,j\le N} \exp(\Phi_i + \Psi_j - C_{ij}).$$

6. Assume that (Φ^0, Ψ^0) is a maximizer of D and consider the matrix $X^0 \in \mathcal{M}_N(\mathbb{R})$ defined by $X^0_{ij} = \exp(\Phi^0_i + \Psi^0_j - C_{ij})$.

a. Write the optimality condition for (Φ^0, Ψ^0) and deduce that X^0 belongs to Γ_N.

b. Prove that $\langle \Phi^0 | \mathbf{1}_N\rangle + \langle \Psi^0 | \mathbf{1}_N\rangle - \sum_{1\leq i,j\leq N} \exp(\Phi^0_i + \Psi^0_j - C_{ij}) = \langle C | X^0\rangle + E(X^0)$.

c. Conclude that $P = D$, and that X^0 is a minimizer of P.

Exercise 2. Let $f \in \mathcal{C}^0(\mathbb{R}^n)$ be convex.

1. Fix a point $x \in \mathbb{R}^n$, and define $K = \overline{\mathrm{conv}} S$ where

$$S = \left\{ s \in \mathbb{R}^n \mid \exists x_n \to x, \text{ s.t. } f \text{ is differentiable at } x_n \text{ and } \lim_{n\to\infty} \nabla f(x_n) = s \right\}.$$

a. Prove that $S \subseteq \partial f(x)$ and $K \subseteq \partial f(x)$.

The goal of the next questions is to prove the converse inclusion.

b. Fix some vector $v \in \mathbb{R}^n$. Prove that for all $t_n = 1/n$, there exists $v_n \in \mathbb{R}^n$ such that $\|v_n - v\| \leq t_n$ and such that f is differentiable at $x_n = x + t_n v_n$.

c. Prove that, taking a subsequence if necessary, one can assume that $\nabla f(x_n)$ converges to a vector $s \in S$. Show that $f^+(x, v) \leq \langle s|v\rangle$.

d. Deduce that $f^+(x, v) \leq \sigma_K(v)$ where σ_K is the support function of K. Conclude that $\partial f(x) \subseteq K$.

2. *Application.* Assume there exists $G \in \mathcal{C}^0(\mathbb{R}^d, \mathbb{R}^d)$ such that $\forall x \in \mathbb{R}^n, G(x) \in \partial f(x)$. Prove that f belongs to $\mathcal{C}^1(\mathbb{R}^n)$ and that $\nabla f = G$.