Exam M2 Datascience
Convex analysis, monotone operators and optimization
18 January 2018

Course book and written notes are authorized.
All electronic devices are prohibited.
Questions can be answered either in French or in English.

The sets $\mathcal{X}$ and $\mathcal{Y}$ respectively represent the Euclidean spaces $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively, where d, m are positive integers. The set $\Gamma_0(\mathcal{X})$ represents the set of proper, lower semi-continuous and convex functions on $\mathcal{X} \to (-\infty, +\infty]$.

Exercise 1 (Relaxed iterations involving α-averaged operators)**.** Let $T : \mathcal{X} \to \mathcal{X}$ be an α-averaged operator, where $0 < \alpha < 1$. Denote by $\mathcal{S}$ the set of fixed points of T and assume that $\mathcal{S} \neq \emptyset$. Let (λ_k) be a sequence of real numbers s.t. for every k, $0 < \lambda_k < 1/\alpha$, and

$$\sum_{k=0}^{\infty} \lambda_k(1 - \alpha\lambda_k) = +\infty\,.$$

Let $x_0 \in \mathcal{X}$ and consider the sequence (x_k) given by

$$x_{k+1} = (1 - \lambda_k)x_k + \lambda_k T(x_k)\,. \tag{1.1}$$

1. Using the definition of α-averaged operators, prove that there exists a non-expansive operator $R : \mathcal{X} \to \mathcal{X}$ s.t. for all k,
$$x_{k+1} = (1 - \alpha\lambda_k)x_k + \alpha\lambda_k R(x_k)\,.$$
2. What is the set of fixed points of R?
3. Justify that for every k, there exists $0 < \beta_k < 1$ and a β_k-averaged operator S_k s.t.
$$x_{k+1} = S_k(x_k)\,.$$
Provide the value of β_k as a function of α, λ_k, and provide the set of fixed points of S_k.
4. For some constant $\delta_k > 0$ to be provided, show that for every $x^* \in \mathcal{S}$,
$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \delta_k\|(I - R)(x_k)\|^2\,,$$
where $I : \mathcal{X} \to \mathcal{X}$ stands for the identity.
5. Show that
$$\sum_{k=1}^{+\infty} \delta_k\|(I - R)(x_k)\|^2 < \infty$$

6. Deduce that $\liminf_{k\to\infty} \|(I-R)(x_k)\| = 0$.
Hint : Use that for every non-negative sequences (a_k), (b_k) such $\sum_k a_k = +\infty$, the fact that $\sum_k a_k b_k < \infty$ implies that $\liminf b_k = 0$. (You don't have to show this).
7. Show the identity

$$(I-R)(x_{k+1}) = (1-\beta_k)(I-R)(x_k) + R(x_k) - R(x_{k+1})\,.$$

8. Prove that the sequence $(\|(I-R)(x_k)\|)$ is decreasing.
9. Deduce that $\lim_{k\to\infty} \|(I-R)(x_k)\| = 0$ (Justification in one or two lines).
10. The sequence (x_k) has cluster points [1]. Why ?
11. Let $\bar{x}$ be a cluster point of (x_k). Prove that $\bar{x} \in \mathcal{S}$.
12. Prove that (x_k) admits a single cluster point.
Hint : Use Question 4.
13. Conclude about the convergence of the sequence (x_k).

Let f, g be two functions in $\Gamma_0(\mathcal{X})$. Denote by $C_{\partial f}$ and $C_{\partial g}$ the Cayley transforms of ∂f and ∂g respectively. Denote by $T = \frac{1}{2}(I + C_{\partial f} C_{\partial g})$ the Douglas-Rachford operator. We assume that T admits a fixed point.

14. Explain how to recover the minimizers of $f+g$ from the fixed points of T (use a result from the course).
15. Provide a value of α s.t. T is α-averaged (no justification is required, just provide the value).
16. Write the algorithm corresponding to the iterations (1.1) as a function of (λ_k and the proximity operators of f and g.
17. Characterize a sequence which converges to a minimizer of $f+g$?

Exercise 2 (ADMM for solving a consensus problem)**.** Consider a set of N agents (computing devices). Assume that for $n \in \{0, \ldots, N-1\}$, Agent n has a private cost function $f_n \in \Gamma_0(\mathbb{R})$ with the domain $\operatorname{dom}(f_n) = \mathbb{R}$. Our purpose is to solve the problem

$$\min_{v\in\mathbb{R}} \sum_{n=0}^{N-1} f_n(v), \tag{2.1}$$

where this minimum is assumed to exist. The difficulty is that each agent can perform operations on this agent's private function only.

1. Defining the function

$$F : \mathbb{R}^N \to \mathbb{R}, \quad x = (x_0, \ldots, x_{N-1}) \mapsto F(x) = \sum_{n=0}^{N-1} f_n(x_n),$$

1. Cluster point = valeur d'adhérence.

show that Problem (2.1) has the same set of minima as the *consensus problem*

$$\min_{x \in \mathbb{R}^N} F(x) + \iota_{\mathcal{C}}(x), \tag{2.2}$$

where $\iota_{\mathcal{C}}$ is the indicator function on a closed and convex set $\mathcal{C}$ to be defined.

Our purpose is to solve Problem (2.2) using ADMM.

2. Show that the set of saddle points of Problem (2.2) is not empty.
3. Provide the expression of $\text{prox}_{\iota_{\mathcal{C}}}(x)$.
4. Given a step size $\gamma > 0$ provide the expressions of ADMM iterations. We denote as x^k the vector of primal variables and as ϕ^k the vector of dual variables at Iteration k, as defined in the lecture notes, and we write $x^k = (x_0^k, \ldots, x_{N-1}^k)$ and $\phi^k = (\phi_0^k, \ldots, \phi_{N-1}^k)$. It is assumed that the initial value ϕ^0 satisfies $\sum_{n=0}^{N-1} \phi_n^0 = 0$. Take this last observation into account to simplify the expressions of ADMM iterations.
5. Assuming that the N agents are connected to a *fusion center* through a communication network, suggest a pseudocode for describing the algorithm. The pattern of this pseudocode should be as follows :

At Iteration $k+1$,

`For` $n = 0, \ldots, N-1$,
$\{$
`/* Computation performed by Agent` n
`/* Data sent by Agent` n `to Fusion Center`
$\}$

`/* Operation performed by Fusion Center`
`/* Data broadcasted to all agents`

`For` $n = 0, \ldots, N-1$,
$\{$
`/* Computation performed by Agent` n
$\}$

Exercise 3 (Dual of the square-root Lasso problem)**.** We consider in this exercise the square-root Lasso problem given by

$$\min_{x \in \mathcal{X}} \|Ax - b\|_2 + \lambda \|x\|_1 \tag{$\sqrt{\text{Lasso}}\ \lambda$}$$

where $\lambda > 0$, $A : \mathcal{X} \to \mathcal{Y}$ is a linear operator and $b \in \mathcal{Y}$. The advantage of this problem as compared to the Lasso problem is that one can get statistical guarantees by choosing a regularization parameter λ that is independent of the noise level of the model.

1. Show that there exists at least one minimizer $x_\star^{(\lambda)}$ to the square-root Lasso problem ($\sqrt{\text{Lasso}}\ \lambda$).
2. Let $x_\star^{(\lambda)}$ be a solution to the square-root Lasso problem ($\sqrt{\text{Lasso}}\ \lambda$). We assume that $Ax_\star^{(\lambda)} - b \neq 0$.
   Using Fermat's rule, show that there exists $\alpha > 0$ such that $x_\star^{(\lambda)}$ is also solution to the Lasso problem
   $$\min_{x \in \mathcal{X}} \frac{1}{2} \|Ax - b\|_2^2 + \alpha \|x\|_1 \ .$$
3. What is the Fenchel-Legendre transform of $g : x \mapsto \lambda \|x\|_1$? (you can make the computations or provide directly the answer using the course).
4. Compute the Fenchel-Legendre transform of $f : z \mapsto \|z - y\|_2$?
5. Compute a dual problem to the square-root Lasso problem ($\sqrt{\text{Lasso}}\ \lambda$).
6. Does strong duality hold ?
7. Propose an algorithm for the resolution of the square-root lasso problem. After how many iterations are we guaranteed to have an ϵ-solution ? Comment of the advantages and the limits of your algorithmic choices.