Contrôle M2 Datascience
Optimisation avancée
06 mars 2017

Course book and written notes are authorized.
All electronic devices are prohibited.

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 sets $\Gamma_0(\mathcal{X})$ represents the set of proper, lower semi-continuous and convex functions on $\mathcal{X} \to (-\infty, +\infty]$. The spectral norm of a linear operator $A : \mathcal{X} \to \mathcal{Y}$ is defined by

$$|||A||| = \sup_{x \in \mathcal{X}, x \neq 0} \frac{\|Ax\|}{\|x\|} \, .$$

One has $|||A||| = \left|\left|\left|A^T\right|\right|\right|$. We recall the Moreau identity for every $f \in \Gamma_0(\mathcal{X})$, $x \in \mathcal{X}$ and $\tau > 0$:

$$\mathrm{prox}_{\tau f}(x) + \tau \, \mathrm{prox}_{\tau^{-1} f^*}\left(\frac{x}{\tau}\right) = x \, .$$

Exercice 1. Consider the problem

$$\min_{x \in \mathcal{X}} f(x) + g(Ax) \ , \ \text{where}$$

— $f \in \Gamma_0(\mathcal{X})$ is μ-strongly convex, for some $\mu > 0$;

— $g \in \Gamma_0(\mathcal{Y})$;

— $A : \mathcal{X} \to \mathcal{Y}$ is a linear operator.

We recall that since f is μ-strongly convex, then f^* is μ^{-1}-smooth *i.e.*, ∇f^* exists and is μ^{-1}-lipschitz continuous.

1. Recall the Lagrangian function and the expression of the dual problem. We assume from now on that the Lagrangian function has a saddle point.
2. Prove that the mapping $F : \lambda \mapsto f^*(-A^T\lambda)$ is L-smooth, for $L = \mu^{-1}|||A|||^2$.
3. Consider the iterates
$$\lambda^{k+1} = \mathrm{prox}_{\gamma g^*}\left(\lambda^k - \gamma \nabla F(\lambda^k)\right) \, .$$
Provide a sufficient condition under which the iterates converge. To which point does the algorithm converge ?

4. Prove that the algorithm rewrites

$$\lambda^{k+1} = \text{prox}_{\gamma g^*}\left(\lambda^k + \gamma A x^{k+1}\right) .$$

where $x^{k+1} = \arg\min_{x\in\mathcal{X}} f(x) + \langle \lambda^k, Ax\rangle$

5. Prove that the algorithm rewrites

$$\begin{aligned} x^{k+1} &= \arg\min_{x\in\mathcal{X}} f(x) + \langle \lambda^k, Ax\rangle \ , \\ z^{k+1} &= \arg\min_{z\in\mathcal{Y}} g(z) - \langle \lambda^k, z\rangle + \frac{\gamma}{2}\left\|z - Ax^{k+1}\right\|^2 \ , \\ \lambda^{k+1} &= \lambda^k + \gamma(Ax^{k+1} - z^{k+1})\,. \end{aligned}$$

6. Compare with the ADMM.
7. We consider the case where $f(x) = \|x-b\|^2$, where b is some fixed vector, and g is the indicator function of a non-empty closed convex set $\mathcal{C}$. Explicit the algorithm as a function of b, A and the projector onto $\mathcal{C}$.

Exercice 2. Let A be a $n\times p$ matrix and let b be a $n\times 1$ vector. Set $\eta > 0$. Consider the problem :

$$\min_{x\in\mathbb{R}^p} \eta\|x\|_1 + \frac{1}{2}\|Ax - b\|^2 \tag{1}$$

where $\|x\|_1 = \sum_{i=1}^p |x_i|$ if $x_1,\dots,x_p$ represent the components of x and where $\|\,.\,\|$ is the Euclidean norm.

1. What name is often given to Problem (1) ?

The above problem rewrites :

$$\min_{\substack{x\in\mathbb{R}^p, z\in\mathbb{R}^n \\ Ax=z}} \eta\|x\|_1 + \frac{1}{2}\|z - b\|^2 \,. \tag{2}$$

2. Write the Lagrangian $L((x,z);\lambda)$ associated to Problem (2).
3. Set $f(x) = \|x\|_1$. Prove that the saddle point $((x^*,z^*);\lambda^*)$ of the Lagrangian are characterized by :

$$\begin{aligned} &0 \in \partial f(x^*) + \frac{A^T\lambda^*}{\eta} \\ &\lambda^* = Ax^* - b \\ &z^* = Ax^*\,. \end{aligned}$$

We remind the reader that $\partial f(x) = \text{sign}(x_1)\times\cdots\times\text{sign}(x_n)$. This means that an element u satisfies $u\in\partial f(x)$ if and only if for every $i = 1,\dots,p$, $u_i \in \text{sign}(x_i)$.We note $A = (a_1,\dots,a_p)$ where $a_1,\dots,a_p$ are the columns of the matrix A.

4. Let $((x^*, z^*); \lambda^*)$ be a saddle point of the Lagrangian. Using Question 3, show that for every $i = 1, \ldots, p$,
$$|a_i^T \lambda^*| < \eta \implies x_i^* = 0\,.$$
5. Set $\eta_0 = \max_{i=1\ldots p} |a_i^T b|$. Verify that $((0,0); -b)$ is a saddle point of the Lagrangian when $\eta = \eta_0$.
6. When $\eta = \eta_0$, provide explicitely the expression of a minimizer of (1).

We now assume that $\eta < \eta_0$. We denote by Φ the dual function *i.e.*, $\Phi(\lambda) = \inf_{x,z} L((x,z);\lambda)$. We define
$$G(\lambda) = -\frac{\|\lambda + b\|^2}{2} + \frac{\|b\|^2}{2}\,.$$

7. Prove that
$$\Phi(\lambda) = \begin{cases} G(\lambda) & \text{if } \forall i, |a_i^T \lambda| \le \eta \\ -\infty & \text{otherwise.} \end{cases} \tag{3}$$
8. Define $\gamma = \Phi(-\frac{\eta}{\eta_0} b)$. Prove that $\gamma > -\infty$. Express γ as a function of η, η_0 and b.
9. Justify that any dual solution λ^* belongs to the set
$$S = \{\lambda \in \mathbb{R}^n \,:\, G(\lambda) \ge \gamma\}\,.$$

From now on, we set a fixed index $i \in \{1, \ldots, p\}$. We define $T_i = \max_{\lambda \in S} |a_i^T \lambda|$.

10. Using one of the previous questions, prove that if $T_i < \eta$, then every solution x to (1) satisfies $x_i = 0$.

A direct derivation shows that
$$T_i = |a_i^T b| + \|a_i\| \sqrt{\|b\|^2 - 2\gamma}\,. \tag{4}$$

11. Before solving numerically (1), an engineer decides, for every i, to compare the quantity (4) to the threshold η. What practical interest this test might have?
12. **[Bonus]** Prove equality (4).
Hint : write that $T_i = \max(P, Q)$ where $P = \max_{\lambda \in S} a_i^T \lambda$ and $Q = \max_{\lambda \in S} -a_i^T \lambda$, then calculate P and Q.