Exam — Introduction to Optimization (WBMA054-05)

Friday, November 10, 2023, 08.30h–10.30h CET

University of Groningen

Instructions

1. Except for the official *cheat sheet*, the use of books or notes is not allowed.
2. Justify all your answers.
3. Write both your last name and student number on the answer sheets.
4. Each question is worth 1 point. Question 10 is optional.

---

1. Let $\phi : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ be closed and $\mu-$strongly convex. Use the Reciprocity formula to show why ϕ^* (the Fenchel conjugate of ϕ) must be $\frac{1}{\mu}-$smooth.
Take $z_1^* \in \partial\phi(z_1)$ and $z_2^* \in \partial\phi(z_2)$. By strong convexity, we have
$$(z_1^* - z_2^*) \cdot (z_1 - z_2) \geq \mu\|z_1 - z_2\|^2.$$
The reciprocity formula says that $z_i \in \partial\phi^*(z_i^*)$, $i = 1, 2$. The Cauchy-Schwarz inequality gives $\mu\|z_1 - z_2\| \leq \|z_1^* - z_2^*\|$, which implies $\partial\phi^*(z_i^*) = \{z_i\}$, whence $\nabla\phi^*(z_i^*) = z_i$. It follows that $\|\nabla\phi^*(z_1^*) - \nabla\phi^*(z_2^*)\| \leq \frac{1}{\mu}\|z_1 - z_2\|$, as we wanted.

In all that follows, we consider the problem $(\mathcal{P})$ of minimizing a continuous and $\mu-$strongly convex function $f : \mathbb{R}^N \to \mathbb{R}$ on $V = \{x \in \mathbb{R}^N : Ax = b\}$, where $A \in \mathbb{R}^{M\times N}$ and $b \in \mathbb{R}^M$.

2. Why can we assure that problem $(\mathcal{P})$ has a unique solution?
The problem is to minimize $f + \iota_V$, which is closed (sum of closed and continuous) and strongly convex (sum of convex and strongly convex).

3. Use the first order optimality condition to show that $\hat{x}$ is a solution of $(\mathcal{P})$ if, and only if, $A\hat{x} = b$ and there exists $\hat{y} \in \mathbb{R}^M$ such that $-A^T\hat{y} \in \partial f(\hat{x})$.[1] We say $(\hat{x}, \hat{y})$ is an *optimal pair*.
Following the suggestion, the optimality condition gives $0 \in \partial f(\hat{x}) + \mathrm{ran}(A^T)$, which says that $\hat{x} \in V$ $(A\hat{x} = b)$ and $-A^T\hat{y} \in \partial f(\hat{x})$ for some $\hat{y} \in \mathbb{R}^M$.

In the rest of the exam, we shall establish the convergence of a sequence (x_k, y_k), constructed from an initial point $(x_0, y_0) \in \mathbb{R}^N \times \mathbb{R}^M$, by iterating
$$\begin{cases} x_{k+1} &= \operatorname{argmin}\left\{L(x, y_k) \;:\; x \in \mathbb{R}^N\right\} \\ y_{k+1} &= y_k + \alpha(Ax_{k+1} - b), \end{cases}$$
with $\alpha > 0$ and $L(x, y) = f(x) + y \cdot (Ax - b) = f(x) + (A^Ty) \cdot x - y \cdot b$, for each $(x, y) \in \mathbb{R}^N \times \mathbb{R}^M$.

4. Why is x_{k+1} well defined?
The function $x \mapsto L(x, y_k)$ is closed and strongly convex, because it is the sum of the closed and strongly convex function f, plus an affine function. Hence, it has a unique minimizer.

[1]Since f is continuous, we have $\partial(f + \iota_V) = \partial f + \partial\iota_V = \partial f + \mathrm{ran}(A^T)$. You do not need to prove this.

5. Write the optimality condition satisfied by x_{k+1} (this comes from the first subiteration).
The optimality condition says that $0 \in \partial L(x_{k+1}, y_k)$. If we show that $\partial L(x_{k+1}, y_k) = \partial f(x_{k+1}) + A^T y_k$, then $0 \in \partial f(x_{k+1}) + A^T y_k$, and thus $-A^T y_k \in \partial f(x_{k+1})$. We shall see that $z^* \in \partial L(x_{k+1}, y_k)$ if, and only if, $z^* - A^T y_k \in \partial f(x_{k+1})$. Indeed,

$$\begin{aligned} z^* \in \partial L(x_{k+1}, y_k) &\Leftrightarrow L(z, y_k) \geq L(x_{k+1}, y_k) + z^* \cdot (z - x_{k+1}) \quad \forall z \\ &\Leftrightarrow f(z) + (A^T y_k) \cdot z \geq f(x_{k+1}) + (A^T y_k) \cdot x_{k+1} + z^* \cdot (z - x_{k+1}) \quad \forall z \\ &\Leftrightarrow f(z) \geq f(x_{k+1}) + (z^* - A^T y_k) \cdot (z - x_{k+1}) \quad \forall z \\ &\Leftrightarrow z^* - A^T y_k \in \partial f(x_{k+1}), \end{aligned}$$

as required.

6. Show that the dual $(\mathcal{D})$ of problem $(\mathcal{P})$ is $\min\{h(y) : y \in \mathbb{R}^M\}$, where $h(y) = f^*(-A^T y) + b \cdot y$.
In the notation of the lectures, $g(z) = \iota_{\{b\}}(z)$, and so $g^*(z) = b \cdot z$ (seen in class).

7. Compute ∇h, and verify that h is $\ell-$smooth, with $\ell = \frac{\|A\|^2}{\mu}$.
Using the Chain Rule, we have $\nabla h(y) = -A \nabla f^*(-A^T y) + b$. Question 1 shows that f^* is $\frac{1}{\mu}-$smooth. By composition,

$$\begin{aligned} \|\nabla h(y_1) - \nabla h(y_2)\| &= \left\|A\big(\nabla f^*(-A^T y_1) - \nabla f^*(-A^T y_2)\big)\right\| \\ &\leq \|A\| \left\|\nabla f^*(-A^T y_1) - \nabla f^*(-A^T y_2)\right\| \\ &\leq \|A\| \tfrac{1}{\mu} \left\|A^T\big(y_1 - y_2\big)\right\| \\ &\leq \|A\| \tfrac{1}{\mu} \|A^T\| \, \|y_1 - y_2\|. \end{aligned}$$

Since $\|A^T\| = \|A\|$, the conclusion follows.

8. Show that the sequence (y_k) satisfies $y_{k+1} = y_k - \alpha \nabla h(y_k)$.
Question 5 implies that $x_{k+1} = \nabla f^*(-A^T y_k)$, by reciprocity. Question 7 then gives $\nabla h(y_k) = -A x_{k+1} + b$, and so the second subiteration is equivalent to $y_{k+1} = y_k - \alpha \nabla h(y_k)$.

9. For which values of α can we guarantee that the sequence (x_k, y_k) converges to an optimal pair $(\hat{x}, \hat{y})$ as $k \to \infty$ (express the result in terms of μ and $\|A\|$).
First, h attains its minimum at $\hat{y}$, because it is convex and $\nabla h(\hat{y}) = -A \nabla f^*(-A^T \hat{y}) + b = -A\hat{x} + b = 0$ (we have used reciprocity again). Since h is also $\ell-$smooth, the convergence of the gradient method holds for all $\alpha < \frac{2}{\ell} = \frac{2\mu}{\|A\|^2}$.

10. (Bonus) What can you say about the convergence rate of this algorithm?
In general, $h(y_k) - v^* = \mathcal{O}(1/k)$, where v^* is the dual value, which coincides with the primal value because there is no duality gap. The rate becomes linear if h is strongly convex or, more generally, if it has quadratic growth. The latter happens, for example, if f is smooth.