Exam — Introduction to Optimization (WMMA054-05)

Thursday, November 11, 2022, 08.30h–10.30h

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. The exam grade is the number of marks plus 1. Use your time wisely.

Let let $C \subset \mathbb{R}^N$ be nonempty, convex and compact (closed and bounded). This part concerns

$$(\mathcal{P}_1) \qquad \min \{ a \cdot z \, : \, z \in C \},$$

where $a \in \mathbb{R}^N$.

1. Show that, for every $a \in \mathbb{R}^N$, $(\mathcal{P}_1)$ has at least one solution. [1p]

   The function $\phi : \mathbb{R}^N \to \mathbb{R}$, defined by $\phi(z) = a \cdot z$, is continuous, and C is compact.

2. Set $f(z) = \iota_C(z)$ and $g(z) = a \cdot z$.

   (a) Compute f^* and g^*, and determine the (Fenchel-Rockafellar) dual of $(\mathcal{P}_1)$. Observe that there is no duality gap. [1p]

   We saw in class that $f^*(y) = \sup\{y \cdot z : z \in C\}$ and $g^*(y) = \iota_{\{a\}}(y)$. The dual is

   $$\min_{y \in \mathbb{R}^N} \left\{ f^*(-y) + g^*(y) \right\} = \min_{y \in \mathbb{R}^N} \left\{ \sup\{-y \cdot z : z \in C\} : y = a \right\} = \sup\{-y \cdot z : z \in C\}.$$

   Changing sup to inf and using 1, this is $-\min\{y \cdot z : z \in C\}$. Clearly, $\alpha + \alpha^* = 0$.

   (b) Show that the primal-dual method is reduced to the projected gradient method. [1p]

   Each iteration of the primal-dual method is given by $x_{k+1} = \mathrm{Proj}_C(x_k - \tau y_k)$ and $y_{k+1} = \mathrm{Proj}_{\{a\}}\big(y_k + \sigma(2x_{k+1} - x_k)\big) = a$, which gives $x_{k+1} = \mathrm{Proj}_C(x_k - \tau a)$. This is the same as the projected gradient method because $\nabla g(x) \equiv a$.

3. Write the first-order optimality condition for $(\mathcal{P}_1)$. [1p]

   Since ϕ is convex and $\nabla \phi(z) \equiv a$, Fermat's Rule says that $\hat{x}$ is a solution if, and only if, $a \cdot (c - \hat{x}) \leq 0$ for all $c \in C$.

Now let $f : \mathbb{R}^N \to \mathbb{R}$ be convex and L-smooth, and let us analyze the convergence of the following algorithm to approximate

$$(\mathcal{P}_2) \qquad \hat{f} = \min\{\, f(z) \,:\, z \in C\,\}.$$

Starting with $x_0 \in C$, define a sequence (x_k, y_k) inductively by

$$\begin{cases} y_k & \in & \operatorname{argmin}\{\, \nabla f(x_k) \cdot z \,:\, z \in C\,\} \\ \gamma_k & = & \operatorname{argmin}\{\, f\big(x_k + \gamma(y_k - x_k)\big) \,:\, \gamma \in [0,1]\,\} \\ x_{k+1} & = & x_k + \gamma_k(y_k - x_k) \end{cases}$$

for $k \geq 1$.

4. Verify that $x_k \in C$ for all $k \geq 0$. [1p]

Induction: $x_0 \in C$. Now, suppose $x_k \in C$. Since $y_k \in C$, C is convex and x_{k+1} is a convex combination of x_k and y_k, $x_{k+1} \in C$.

5. For each $k \geq 1$ and $\gamma \in [0,1]$, set $z_k(\gamma) = x_k + \gamma(y_k - x_k)$. Show that $f(x_{k+1}) \leq f\big(z_k(\gamma)\big)$. Deduce that $f(x_{k+1}) \leq f(x_k)$. [1p]

It is immediate from the definitions.

6. Prove that there is a constant $D \geq 0$ such that

$$f(x_{k+1}) - \hat{f} \leq (1-\gamma)\big(f(x_k) - \hat{f}\big) + \gamma^2 D$$

for all $\gamma \in [0,1]$ and $k \geq 1$. [1p]

Suggestion: (1) Use question 5 and the Descent Lemma; (2) Which sets have finite diameter?

From the Descent Lemma, we have

$$\begin{aligned} f(z_k(\gamma)) &\leq f(x_k) + \nabla f(x_k) \cdot (z_k(\gamma) - x_k) + \frac{L}{2}\|z_k(\gamma) - x_k\|^2 \\ &= f(x_k) + \gamma\, \nabla f(x_k) \cdot (y_k - x_k) + \frac{\gamma^2 L}{2}\|y_k - x_k\|^2 \end{aligned}$$

Let $\hat{x}$ minimize f on C (f is continuous and C is compact). By the definition of y_k, $\nabla f(x_k) \cdot y_k \leq \nabla f(x_k) \cdot \hat{x}$. Using this, along with 5 and the convexity of f, we obtain

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \gamma\, \nabla f(x_k) \cdot (\hat{x} - x_k) + \frac{\gamma^2 L}{2}\|y_k - x_k\|^2 \\ &\leq f(x_k) + \gamma\left(\hat{f} - f(x_k)\right) + \frac{\gamma^2 L}{2}\|y_k - x_k\|^2. \end{aligned}$$

Finally, since y_k and x_k belong to C, which is bounded, we have

$$\|y_k - x_k\| \leq \operatorname{diam}(C) := \sup\big\{\|u - v\| : u, v \in C\big\} < \infty.$$

Writing $D = \frac{L}{2}\operatorname{diam}(C)^2$, we obtain

$$f(x_{k+1}) \leq f(x_k) + \gamma\left(\hat{f} - f(x_k)\right) + \gamma^2 D.$$

It suffices to subtract $\hat{f}$ on both sides and rearrange the terms to conclude.

7. For each k, substitute $\gamma = \frac{2}{k+2}$ in the preceding inequality to deduce that

$$(k+2)^2\big(f(x_{k+1}) - \hat{f}\big) \leq (k+1)^2\big(f(x_k) - \hat{f}\big) + 4D.$$

Do this for each k to show that

$$f(x_k) - \hat{f} \leq \frac{4D}{k+1}$$

for all $k \geq 1$. [1p]

With the substitution, and multiplying by $(k+1)^2$, we obtain

$$\begin{aligned} (k+2)^2\big(f(x_{k+1}) - \hat{f}\big) &\leq k(k+2)\big(f(x_k) - \hat{f}\big) + 4D \\ &\leq (k+1)^2\big(f(x_k) - \hat{f}\big) + 4D. \end{aligned}$$

Summing from 0 to $k-1$, and using the telescopic property, we deduce that

$$(k+1)^2\big(f(x_k) - \hat{f}\big) \leq f(x_0) - \hat{f} + 4Dk.$$

Since $f(x_0) - \hat{f} \leq D$, a rough bound for the right-hand side is $4D(k+1)$.

8. Given $\varepsilon > 0$, how many iterations of this method give you the certainty that you have found a point $\hat{x} \in C$ such that $f(\hat{x}) \leq \hat{f} + \varepsilon$? [1p]

Let K be the first positive integer greater than, or equal to, $4D\varepsilon^{-1} - 1$. Then

$$f(x_K) - \hat{f} \leq \frac{4D}{K+1} \leq \varepsilon.$$

Of course, any integer greater than K would do as well.