MGMT 690 - Pset 3

Spring 2024

Instructions:

- This pset is due on Sunday, April 21 at 11:59pm.
- Completed psets should be submitted to Gradescope.
- **Exercises** are for your own review only. They do not need to be submitted and will not be graded.
- **Complete all problems 1–4.**

Exercises

- Consider $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x) = |x_1| + 2\,|x_2|\,.$$

Show that $\partial f(1,0) = \{(1,y) : |y| \le 2\}$. Thus, $(1,2) \in \partial f(1,0)$. Next, show that $f((1,0) - t(1,2)) > f(1,0)$ for all $t > 0$. This example shows that a negative subgradient is not necessarily a descent direction.

- Fill in the missing details in the proof of Theorem 21 in the notes.

Problems

1. [15pts] Let $f_i : \mathbb{R}^d \to \mathbb{R}$ be convex and differentiable for $i = 1, \dots, n$. Let $F(x) \coloneqq \max_i f_i(x)$. Show that

$$\partial F(x) = \mathrm{conv}(\{\nabla f_i(x) : f_i(x) = F(x)\}).$$

2. [25pts] Let $\gamma > 1$ and consider the following function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x) = \begin{cases} \sqrt{x_1^2 + \gamma x_2^2} & \text{if } |x_2| \le x_1 \\ \frac{x_1 + \gamma |x_2|}{\sqrt{1+\gamma}} & \text{else} \end{cases}$$

This function is convex and $\sqrt{\gamma}$-Lipschitz (you do not need to prove this).

Consider the subgradient method with *exact* line-search initialized at $x^{(0)} = (\gamma, 1)$, i.e., for $t \ge 1$, let $g \in \partial f(x^{(t-1)})$ and set

$$x^{(t)} = \operatorname*{arg\,min}_{x \in x^{(t-1)} - \mathbb{R}_+ g} f(x)$$

(a) [15pts] Prove that for a general convex function $f : \mathbb{R}^n \to \mathbb{R}$, if f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$. Recall, if f is differentiable at x, then $\nabla f(x)$ is defined to be the unique vector in $\mathbb{R}^n$ so that for all $u \in \mathbb{R}^n$,

$$\frac{d}{dt} f(x + tu) = \langle \nabla f(x), u \rangle .$$

(b) [10pts] Prove by induction that $x^{(t)} = \left(\gamma \left(\frac{\gamma-1}{\gamma+1} \right)^t , \left(\frac{1-\gamma}{\gamma+1} \right)^t \right)$ for all $t \geq 0$.

This shows that the subgradient method with exact line-search converges to the origin where $f(0) = 0$. On the other hand, f can be made arbitrarily negative by sending $x_1 \to -\infty$.

3. [30pts] Let $f : \mathbb{R}^n \to \mathbb{R}$ be a L-Lipschitz convex function with minimizer $x^\star$ and minimum value f^*. Suppose that f satisfies the following growth condition parameterized by $\delta > 0, \alpha > 0$:

$$f(x) - f^\star \leq \delta \qquad \Longrightarrow \qquad f(x) - f^* \geq \alpha \left\| x - x^\star \right\|^2 .$$

Suppose we are given $x_0 \in \mathbb{R}^n$ with $\|x_0 - x^\star\| \leq R$.

Fill in the missing details (i.e., replace the ?s) in the following *restarted* subgradient method. Consider the following algorithm:

Algorithm 1 Restarted subgradient method

Given: $L, R, \alpha, \delta, x_0$

- For each $k = 0, \ldots$
  - Run the subgradient method with constant stepsizes (see Corollary 7 in the notes) with initial iterate x_k for

    $$T_k = ?$$

    iterations. Let x_{k+1} to be the output of the subgradient method.

By setting $T_0 =?$, we can ensure the following property:

Lemma 1. *It holds that $f(x_1) - f^\star \leq \delta$.*

Proof. ? ∎

For $k \geq 1$, define $\delta_k = \frac{2}{2^k}\delta \leq \delta$. By setting $T_k =?$ for $k \geq 1$, we can ensure the following property:

Lemma 2. *It holds that $f(x_k) - f^\star \leq \delta_k$.*

Proof. ? ∎

We conclude that:

Proposition 1. *The restarted subgradient method with constant stepsizes and horizons* $T_0 =?$ *and* $T_k =?$ *for all* $k \geq 1$ *achieves a gap* $f(x) - f^\star \leq \epsilon$ *after at most*

$$O\left(\frac{L^2 R^2}{\delta^2} + \frac{L^2}{\alpha\epsilon}\right)$$

total (inner) iterations. Thus for $\epsilon \ll \frac{\delta^2}{\alpha R^2}$*, this convergence rate is* $O\left(\frac{L^2}{\alpha\epsilon}\right)$.

Compare this rate with Corollary 7 in the notes.

4. [30pts] This problem extends the accelerated gradient descent method for L-smooth convex functions and its analysis to other "smoothly-proxable" convex problems.

   Formally, consider a minimization problem of the form

$$\min_{x \in \Omega} F(x)$$

   where $F : \mathbb{R}^n \to \mathbb{R}$ is an arbitrary function and $\Omega \subseteq \mathbb{R}^n$ is an arbitrary set. We say that

$$\texttt{prox} : \mathbb{R}^n \to \Omega$$

   is a *smooth prox-oracle* for this problem if `prox` satisfies the following property: Given $y \in \mathbb{R}^n$, define $g(y) \coloneqq L(y - \texttt{prox}(y))$. Then, for all $z \in \Omega$, it holds that

$$F(\texttt{prox}(y)) \leq f(z) + \langle g(y), y - z \rangle - \frac{\|g(y)\|^2}{2L}. \tag{1}$$

   We will replace the gradient step in accelerated gradient descent with the prox oracle:

Algorithm 2 Accelerated gradient descent for smoothly proxable problems

Given $x_0 \in \mathbb{R}^d$, $F : \mathbb{R}^n \to \mathbb{R}$ and $\texttt{prox} : \mathbb{R}^n \to \Omega$

- Set $y_0 = x_0$ and $\lambda_{-1} = 1$
- For $t = 0, \ldots$

$$\lambda_t = \frac{1 + \sqrt{1 + 4\lambda_{t-1}^2}}{2}$$
$$\gamma_t = \frac{\lambda_{t-1} - 1}{\lambda_t}$$
$$x_{t+1} = \texttt{prox}(y_t) = y_t - \frac{1}{L} g(y_t)$$
$$y_{t+1} = x_{t+1} + \gamma_t (x_{t+1} - x_t)$$

(a) [10pts] Modify the analysis of **??** to show that:

Theorem 1. *Suppose $F : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subseteq \mathbb{R}^n$ and suppose $\texttt{prox} : \mathbb{R}^n \to \Omega$ is a smooth prox-oracle for $\min_{x\in\Omega} F(x)$. Furthermore, suppose F has minimizer $x^\star$ with minimum value $F^\star$. Then, it holds that*

$$F(x_T) - F^\star = O\left(\frac{L\left\|x_0 - x^\star\right\|^2 + F(x_0) - F(x^\star)}{T^2}\right).$$

(b) [10pts] Suppose $F : \mathbb{R}^n \to \mathbb{R}$ is an L-smooth convex function and $\Omega \subseteq \mathbb{R}^n$ is nonempty, closed, and convex. Define

$$\texttt{prox}(y) \coloneqq \arg\min_{x\in\Omega} \left\{F(y) + \langle \nabla F(y), x - y\rangle + \frac{L}{2}\left\|x - y\right\|^2\right\}.$$

Prove that this map is well-defined, is equal to

$$\texttt{prox}(y) = \Pi_\Omega\left(y - \frac{1}{L}\nabla F(y)\right),$$

and is a smooth prox-oracle for $\min_{x\in\Omega} F(x)$.

(c) [10pts] Suppose $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ are L-smooth convex functions and define

$$F(x) \coloneqq \max_{i\in[k]} f_i(x).$$

Define

$$\texttt{prox}(y) \coloneqq \arg\min_{x\in\mathbb{R}^n} \max_{i\in[k]} \left\{f_i(y) + \langle \nabla f_i(y), x - y\rangle + \frac{L}{2}\left\|x - y\right\|^2\right\}.$$

Prove that this map is well-defined and is a smooth prox-oracle for $\min_{x\in\Omega} F(x)$.