Introduction to Optimization - Midterm 2

Consider the problem

$$\min_{x \in C} f(x),$$

where $f : \mathbb{R}^N \to \mathbb{R}$ is L-smooth and convex, and $C \subset \mathbb{R}^N$ is nonempty, closed and convex. We denote by $\mathrm{Proj}_C(x)$ the projection of $x \in \mathbb{R}^N$ onto C. The set of solutions, which we assume to be nonempty, is denoted by S, and the optimal value by f^*. This problem can be solved by means of the *Projected Gradient Method*, which starts from some $x_0 \in \mathbb{R}^N$, picks $\gamma \in (0, \frac{1}{L}]$, and then iterates

$$x_{k+1} = \mathrm{Proj}_C\left(x_k - \gamma \nabla f(x_k)\right), \tag{1}$$

for $k \geq 0$. In other words, at each iteration, we perform a gradient step and then project onto C. The purpose of this assignment is to prove that the Projected Gradient Methods converges. The purpose of this exercise is to prove the convergence of this algorithm. To this end, do as follows:

1. Let $w \in \mathbb{R}^N$ and let $\overline{w} = \mathrm{Proj}_C(w)$. Show that for all $z \in C$, we have

   $$(\overline{w} - w) \cdot (\overline{w} - z) \leq 0.$$

   This is Corollary 5.20 of the Lecture Notes. Write $\phi(z) = \frac{1}{2}\|z - w\|^2$, so that $\overline{w}$ is the unique solution to $\min_{z \in C} \phi(w)$ (recall the ϕ is strongly convex). By Fermat's Rule II (Slide 9 of Lecture 5, or Theorem 5.17 of the Lecture Notes), we have $0 \leq \nabla\phi(\overline{w}) \cdot (z - \overline{w}) = (\overline{w} - w) \cdot (z - \overline{w})$ for every $z \in C$.

2. Deduce that, for every $z \in C$ and $k \geq 0$, we have

   $$\gamma \nabla f(x_k) \cdot (x_{k+1} - z) \leq (x_{k+1} - x_k) \cdot (z - x_{k+1}).$$

   Use 1 with $w = x_k - \gamma \nabla f(x_k)$ and $\overline{w} = x_{k+1}$ and rearrange the terms.

3. Infer that, for every $k \geq 1$, we have

   $$\gamma \nabla f(x_k) \cdot (x_{k+1} - x_k) \leq -\|x_{k+1} - x_k\|^2.$$

   Use 2 with $z = x_k$, which belongs to C by definition.

4. Verify that, for every $k \geq 1$, we have

   $$f(x_{k+1}) \leq f(x_k) + (x_{k+1} - x_k) \cdot \nabla f(x_k) + \frac{L}{2}\|x_{k+1} - x_k\|^2.$$

   Since f is L-smooth, we have $f(y) \leq f(x) + \nabla f(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2$ for every x, y (Slide 9 of Lecture 4 or Proposition 4.11 of Lecture Notes). Use this with $x = x_k$ and $y = x_{k+1}$.

5. Show that, for every $k \geq 1$, we have

   $$f(x_{k+1}) \leq f(x_k) - \frac{L}{2}\|x_{k+1} - x_k\|^2.$$

   Combine 3 and 4, and use the fact that $\gamma L \leq 1$.

6. Deduce that $\lim_{k \to \infty} f(x_k)$ exists, and that $\lim_{k \to \infty} \|x_{k+1} - x_k\| = 0$.
   Since $x_k \in C$ for $k \geq 1$, $f(x_k) \geq f^*$. The sequence $\big(f(x_k)\big)$ is nonincreasing and bounded from below, thus $\lim_{k \to \infty} f(x_k)$ exists. Next, $0 \leq \|x_{k+1} - x_k\| \leq \sqrt{f(x_k) - f(x_{k+1})} \to 0$ as $k \to \infty$.

7. Prove that, for every $\hat{x} \in S$ and $k \geq 1$, we have

$$2\gamma\big(f(x_{k+1}) - f^*\big) \leq \|x_k - \hat{x}\|^2 - \|x_{k+1} - \hat{x}\|^2.$$

By convexity, $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$ for every x, y. Use this with $x = x_k$ and $y = \hat{x}$ to obtain $f(x_k) \leq f^* + \nabla f(x_k) \cdot (x_k - x^*)$. Combine with 4 and multiply by 2γ to get

$$2\gamma\big(f(x_{k+1}) - f^*\big) \leq 2\gamma \nabla f(x_k) \cdot (x_{k+1} - \hat{x}) + \gamma L \|x_{k+1} - x_k\|^2.$$

Then use 2 with $z = \hat{x}$ and rewrite the right-hand side (using $\gamma L \leq 1$ and $\|a \pm b\|^2 = \|a\|^2 + \|b\|^2 \pm 2a \cdot b$) to conclude.

8. Conclude that, for every $k \geq 1$, the following convergence rate holds:

$$f(x_k) - f^* \leq \frac{\operatorname{dist}(x_0, S)^2}{2\gamma(k+1)}.$$

The sequence $\big(f(x_k)\big)$ is nonincreasing, whence

$$2\gamma(k+1)\big(f(x_k) - f^*\big) \leq \sum_{j=0}^{k} \big(f(x_j) - f^*\big) \leq \|x_0 - \hat{x}\|^2 - \|x_{k+1} - \hat{x}\|^2 \leq \|x_0 - \hat{x}\|^2.$$

Pick $\hat{x} = \operatorname{Proj}_S(x_0)$ (or take the infimum over all $\hat{x} \in S$ on the right-hand side) to conclude.

9. Finally, prove that x_k converges, as $k \to \infty$, to a point in S.
By 7, the sequence (x_k) is bounded. Therefore, it must have a limit point $\bar{x}$, so that $\|x_{k_n} - \bar{x}\| \to 0$ as $n \to \infty$. Also by 7, the sequence $(\|x_k - \bar{x}\|)$ is nonincreasing, and so has a limit as $k \to 0$. This limit must be zero: $\lim_{k\to\infty} \|x_k - \bar{x}\| = \lim_{n\to\infty} \|x_{k_n} - \bar{x}\| = 0$, whence x_k converges to $\bar{x}$.