Introduction to Optimization - Midterm 1

In what follows, $f : \mathbb{R}^N \to \mathbb{R}$ is L-smooth and μ-strongly convex, with $L, \mu > 0$.

Note: By Proposition 5.45 of the Lecture Notes (also Slide 6 of Lecture 6),[1] for every $x, y \in \mathbb{R}^N$, we have

$$f(x) \geq f(y) + \nabla f(y) \cdot (x - y) + \frac{\mu}{2}\|x - y\|^2. \tag{1}$$

1. Show that there exist constants $a > 0$ and $b, c \in \mathbb{R}$ such that, for every $x \in \mathbb{R}^N$, we have $f(x) \geq a\|x\|^2 + b\|x\| + c$.
   Fix $y = 0$ in (1), so that

   $$f(x) \geq f(0) + \nabla f(0) \cdot x + \frac{\mu}{2}\|x\|^2 \geq f(0) - \|\nabla f(0)\|\,\|x\| + \frac{\mu}{2}\|x\|^2,$$

   and then set $a = \frac{\mu}{2}$, $b = -\|\nabla f(0)\|$ and $c = f(0)$.

2. Prove that f has a exactly one minimizer.
   Since f is smooth, it is continuous. From 1, f is coercive:

   $$f(x) \geq \|x\|^2 \left[a + \frac{b}{\|x\|} + \frac{c}{\|x\|^2}\right] \to \infty, \quad \text{as } \|x\| \to \infty,$$

   because $a > 0$. By Weierstrass's Theorem, it has minimizers. There can be no more than one because strongly convex functions are strictly convex. Or, directly, using (2) with $x, y \in \operatorname{argmin}(f)$, we get $f(\frac{1}{2}x + \frac{1}{2}y) \leq \min(f) - \frac{\mu}{8}\|x - y\|^2 < \min(f)$ if $x \neq y$, yielding a contradiction.

Now, define the function $g : \mathbb{R}^N \to \mathbb{R}$, by $g(x) = f(x) - \frac{\mu}{2}\|x\|^2$ for $x \in \mathbb{R}^N$.

3. Verify that g is convex by using the definition.
   Given $x, y \in \mathbb{R}^N$, $\lambda \in (0, 1)$, the μ-strong convexity of f gives

   $$\begin{aligned} g\big(\lambda x + (1-\lambda)y\big) &= f\big(\lambda x + (1-\lambda)y\big) - \frac{\mu}{2}\,\|\lambda x + (1-\lambda)y\|^2 \\ &\leq \lambda f(x) + (1-\lambda) f(y) - \frac{\mu\lambda(1-\lambda)}{2}\|x - y\|^2 - \frac{\mu}{2}\,\|\lambda x + (1-\lambda)y\|^2 \\ &= \lambda g(x) + (1-\lambda) g(y) + \frac{\mu}{2} R(\lambda, x, y), \end{aligned}$$

   where $R(\lambda, x, y) = \lambda\|x\|^2 + (1-\lambda)\|y\|^2 - \lambda(1-\lambda)\|x - y\|^2 - \|\lambda x + (1-\lambda)y\|^2 \equiv 0$.[2]

[1] During Lecture 6, the students were encouraged to prove this by themselves. It is also Exercise 5.46 of the Lecture Notes. Anyway, this is done just as the implication $i) \Rightarrow ii)$ in Proposition 5.37: one rewrites the definition (2) as

$$f(x) - f(y) \geq \frac{f\big(y + \lambda(x - y)\big) - f(y)}{\lambda} + \frac{\mu(1-\lambda)}{2}\|x - y\|^2 \to \nabla f(y) \cdot (x - y) + \frac{\mu}{2}\|x - y\|^2, \quad \text{as } \lambda \to 0.$$

[2] Using 4 to prove 3 using the characterization of differentiable convex functions (so not the definition) gives half the mark.

4. Compute $\nabla g(x)$ in terms of $\nabla f(x)$, for $x \in \mathbb{R}^N$.
The gradient of the sum is the sum of the gradients: $\nabla g(x) = \nabla f(x) - \mu x$.

5. Prove that, for every $x, y \in \mathbb{R}^N$, we have $g(y) \leq g(x) + \nabla g(x) \cdot (y - x) + \dfrac{L-\mu}{2}\|y - x\|^2$.
The L-smoothness of f implies the formula in $ii)$ of the Baillon-Haddad Theorem, so

$$\begin{aligned} g(y) &= f(y) - \frac{\mu}{2}\|y\|^2 \\ &\leq \left[f(x) + \nabla f(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2\right] - \frac{\mu}{2}\|y\|^2 \\ &= g(x) + \frac{\mu}{2}\|x\|^2 + \big(\nabla g(x) + \mu x\big) \cdot (y - x) + \frac{L}{2}\|x - y\|^2 - \frac{\mu}{2}\|y\|^2 \\ &= g(x) + \nabla g(x) \cdot (y - x) + \frac{L}{2}\|x - y\|^2 + \frac{\mu}{2}\|x\|^2 + \mu x \cdot (y - x) - \frac{\mu}{2}\|y\|^2. \end{aligned}$$

The sum of the last three terms on the right-hand side is $-\frac{\mu}{2}\|x - y\|^2$.

6. What is the best Lipschitz constant you can establish for ∇g?
In view of 5, and the implication $ii) \Rightarrow i)$ in the Baillon-Haddad Theorem, g is $(L - \mu)$-smooth, which means that ∇g is $(L - \mu)$-Lipschitz. This is better than the $L + \mu$ that one obtains by just considering g as the sum of two smooth functions.
Bonus: It is also the best possible, since it is attained for $f(x) = \frac{L}{2}\|x\|^2$.

7. Deduce that, for every $x, y \in \mathbb{R}^N$, we have $(\nabla g(x) - \nabla g(y)) \cdot (x - y) \geq \dfrac{1}{L-\mu}\|\nabla g(x) - \nabla g(y)\|^2$.
We use 5 and the implication $ii) \Rightarrow iii)$ in the Baillon-Haddad Theorem.

8. Conclude that, for every $x, y \in \mathbb{R}^N$, we have $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \dfrac{\mu L}{L+\mu}\|x - y\|^2 + \dfrac{1}{L+\mu}\|\nabla f(x) - \nabla f(y)\|^2$.
Just replace $\nabla g(\cdot) = \nabla f(\cdot) - \mu(\cdot)$ in 7 to obtain

$$(\nabla f(x) - \nabla f(y)) \cdot (x - y) - \mu\|x - y\|^2 \geq \frac{1}{L-\mu}\left\|\nabla f(x) - \nabla f(y) - \mu(x - y)\right\|^2.$$

Then develop the square on the right-hand side, and gather the similar terms to obtain

$$\left[\frac{L+\mu}{L-\mu}\right](\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \frac{1}{L-\mu}\|\nabla f(x) - \nabla f(y)\|^2 + \frac{\mu L}{L-\mu}\|x - y\|^2.$$

Strong convexity. The function $f : \mathbb{R}^N \to \mathbb{R}$ is μ-strongly convex, for every $x, y \in \mathbb{R}^N$ and $\lambda \in (0, 1)$, one has

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\mu\lambda(1-\lambda)}{2}\|x - y\|^2. \tag{2}$$

Baillon-Haddad Theorem. Let $f : \mathbb{R}^N \to \mathbb{R}$ be convex and differentiable, and let $L > 0$. The following statements are equivalent:

i) f is L-smooth;

ii) $f(z) \leq f(x) + \nabla f(x) \cdot (z - x) + \frac{L}{2}\|z - x\|^2$, for all $x, z \in \mathbb{R}^N$;

iii) $f(x) \geq f(y) + \nabla f(y) \cdot (x - y) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2$, for all $x, y \in \mathbb{R}^N$; and

iv) $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2$, for all $x, y \in \mathbb{R}^N$.