Introduction to Optimization - Exam

Let $g : \mathbb{R}^N \to \mathbb{R}$ be convex and continuous, and let $h : \mathbb{R}^N \to \mathbb{R}$ be convex and continuously differentiable. Define $F := g - h$, and assume it is bounded from below. Note that F may not be convex. We say $\hat{x} \in \mathbb{R}^N$ is a *critical point* of F if $\nabla h(\hat{x}) \in \partial g(\hat{x})$.

1. Prove that every minimizer of F is a critical point of F.
   If $\bar{x}$ is a minimizer of F, then $g(\bar{x}) - h(\bar{x}) \leq g(x) - h(x)$ for every $x \in \mathbb{R}^N$. The convexity of h gives

   $$g(x) \geq g(\bar{x}) + h(x) - h(\bar{x}) \geq g(\bar{x}) + \nabla h(\bar{x}) \cdot (x - \bar{x}).$$

   Since this holds for every $x \in \mathbb{R}^N$, it means that $\nabla h(\bar{x}) \in \partial g(\bar{x})$, by definition.

In what follows, we fix $\gamma > 0$. Given $z \in \mathbb{R}^N$, define $Q_z : \mathbb{R}^N \to \mathbb{R}$ by

$$Q_z(x) := g(x) - \nabla h(z) \cdot (x - z) + \frac{1}{2\gamma}\|x - z\|^2.$$

2. Show that, for every $z \in \mathbb{R}^N$, Q_z has a unique minimizer.
   The function Q_z is continuous (as the sum of three continuous functions) and strongly convex (as the sum of a convex one, an affine one, and a strongly convex one). Therefore, it has a unique minimizer.

Starting from $x_0 \in \mathbb{R}^N$, we construct a sequence (x_k) by iterating

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^N} \left\{ g(x) - \nabla h(x_k) \cdot (x - x_k) + \frac{1}{2\gamma}\|x - x_k\|^2 \right\}, \quad \text{for } k \geq 0. \tag{1}$$

3. Explain why the equality sign in (1) is a slight abuse of notation.
   The left-hand side is a point and the right-hand side is a set. Since the latter is a singleton (by 2), there is no ambiguity.

4. Write the first order optimality condition for the optimization problem given in (1).
   By Fermat's Rule (and continuity), we have $0 \in \partial g(x_{k+1}) - \nabla h(x_k) + \frac{1}{\gamma}(x_{k+1} - x_k)$.

5. Verify that, for every $k \geq 0$, we have

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

   By the definition of x_{k+1}, we have $Q_{x_k}(x_{k+1}) \leq Q_{x_k}(x_k)$, which is exactly what we want.

6. Prove that, for every $k \geq 0$, we have

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

   By the convexity of h, we have $h(x_{k+1}) \geq h(x_k) + \nabla h(x_k) \cdot (x_{k+1} - x_k)$. Combining this with 5, and rearranging the terms, gives the result.

7. Show that the sequence $\big(F(x_k)\big)$ is convergent, and that $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$.
   The sequence $\big(F(x_k)\big)$ is nonincreasing (by 6) and bounded from below (because F is bounded from below), so it converges. From 6, we also get $0 \leq \|x_{k+1} - x_k\|^2 \leq F(x_k) - F(x_{k+1}) \to 0$ as $k \to \infty$.

8. Prove that, if (x_k) has a convergent subsequence, its limit must be a critical point of F.
   Suppose $x_{k_n} \to \hat{x}$ as $n \to \infty$. Then, $\nabla h(x_{k_n}) \to \nabla h(\hat{x})$, because h is continuously differentiable. Since $\|x_{k_n+1} - x_{k_n}\| \to 0$ as $n \to \infty$, also x_{k_n+1} converges to $\hat{x}$. Since $\nabla h(x_k) - \frac{1}{\gamma}(x_{k_n+1} - x_{k_n}) \in \partial g(x_{k_n+1})$, we have

$$g(y) \geq g(x_{k_n+1}) + \left(\nabla h(x_k) - \frac{1}{\gamma}(x_{k_n+1} - x_{k_n})\right) \cdot (y - x_{k_n+1}), \quad \text{for every } y \in \mathbb{R}^N.$$

   Passing to the limit gives $g(y) \geq g(\hat{x}) + \nabla h(\hat{x}) \cdot (y - \hat{x})$ for every $y \in \mathbb{R}^N$, and so $\nabla h(\hat{x}) \in \partial g(\hat{x})$.

In what follows, suppose F is coercive[1], and has a unique critical point $\bar{x}$.

9. Show that $\bar{x}$ is the unique minimizer of F.
   Since F is continuous and coercive, it has minimizers. Every minimizer must be a critical point (by 1). Since F only has only one critical point, it cannot have more than one minimizer.

10. Show that x_k converges to $\bar{x}$ as $k \to \infty$.
    Since $\big(F(x_k)\big)$ is nonincreasing, $x_k \in \{z : F(z) \leq F(x_0)\}$. This set is bounded because F is coercive. It follows that (x_k) is bounded. By 8, every convergent subsequence must converge to $\bar{x}$, and this implies that the whole sequence converges to that limit.

[1] $\lim_{\|x\| \to \infty} F(x) = +\infty$