Introduction to Optimization. Mock Exam 2022-2023.

Consider the problem $(\mathcal{P})$ of minimizing a continuous convex function $f : \mathbb{R}^N \to \mathbb{R}$ over the affine subspace $V = \{x \in \mathbb{R}^N : Ax = b\}$, for given $A \in \mathbb{R}^{M\times N}$ and $b \in \mathbb{R}^M$.

1. If ι_V denotes the indicator function of V, prove that $\partial\iota_V(x) = \operatorname{ran}(A^*)$ for each $x \in V$. Suggestion: How is V related to $\ker(A)$ $\big(= \operatorname{ran}(A^*)^\perp\big)$?
For any $x_0 \in V$, we have $V = x_0 + \ker(A)$. As seen in class, $\partial\iota_V(x) = \emptyset$ if $x \notin V$ and $\partial\iota_V(x) = \{z : z\cdot(v-x) \leq 0\ \forall v \in V\} = \{z : z\cdot v \leq 0\ \forall v \in \ker(A)\} = \ker(A)^\perp = \operatorname{ran}(A^*)$ if $x \in V$.

2. Use the first order optimality condition for $(\mathcal{P})$, obtained from Fermat's Rule, to show that $\hat{x}$ is a solution of $(\mathcal{P})$ if, and only if, $A\hat{x} = b$ and there exists $\hat{y} \in \mathbb{R}^M$ such that $-A^*\hat{y} \in \partial f(\hat{x})$.[1] We say $(\hat{x}, \hat{y})$ is an *optimal pair*. Is this related to Lagrange multipliers?
The problem is to minimize $f + \iota_V$ over $\mathbb{R}^N$. By convexity, $\hat{x}$ is a solution if, and only if, $0 \in \partial f(\hat{x}) + \partial\iota_V(\hat{x})$, which means that $\hat{x} \in V$ and $0 \in \partial f(\hat{x}) + \hat{z}$ for some $\hat{z} \in \operatorname{ran}(A^*)$. This is equivalent to saying that there is $\hat{y} \in \mathbb{R}^M$ such that $-A^*\hat{y} \in \partial f(\hat{x})$. The Lagrange multiplier theorem seen in class for the differentiable case (Lecture 9) gives $-A^*\hat{y} \in \nabla f(\hat{x})$ (see also Lecture 11).

3. Define the *Lagrangian* of the problem by $\mathcal{L}(x, y) = f(x) + y\cdot(Ax - b)$, for $(x, y) \in \mathbb{R}^N \times \mathbb{R}^M$. Show that if $(\hat{x}, \hat{y})$ is an optimal pair, then
$$\mathcal{L}(\hat{x}, y) \leq \mathcal{L}(\hat{x}, \hat{y}) \leq \mathcal{L}(x, \hat{y})$$
for all $(x, y) \in \mathbb{R}^N \times \mathbb{R}^M$.
The first inequality is an equality because $A\hat{x} = b$. Now, since $-A^*\hat{y} \in \nabla f(\hat{x})$, the subgradient inequality gives $f(x) \geq f(\hat{x}) - A^*\hat{y}\cdot(x - \hat{x})$. It suffices to add $y\cdot(Ax - b)$ on both sides (and use the property of the transpose) to get the second inequality.

In what follows, we establish the convergence of the algorithm given by

$$\begin{cases} p_{k+1} &= \operatorname{argmax}\Big\{\mathcal{L}(x_k, y) - \frac{1}{2\gamma}\|y - y_k\|^2 \ : \ y \in \mathbb{R}^M\Big\} \\ x_{k+1} &= \operatorname{argmin}\Big\{\mathcal{L}(x, p_{k+1}) + \frac{1}{2\gamma}\|x - x_k\|^2 \ : \ x \in \mathbb{R}^N\Big\} \\ y_{k+1} &= \operatorname{argmax}\Big\{\mathcal{L}(x_{k+1}, y) - \frac{1}{2\gamma}\|y - y_k\|^2 \ : \ y \in \mathbb{R}^M\Big\}, \end{cases}$$

with $\gamma > 0$, and starting from an initial point $(x_0, y_0) \in \mathbb{R}^N \times \mathbb{R}^M$.

4. Write the optimality conditions corresponding to the three subiterations, in order to find closed formulas for p_{k+1} and y_{k+1}, and to express x_{k+1} in terms of a proximal step.
In the first, the functions involved are differntiable, so the optimality condition is $Ax_k - b + \frac{1}{\gamma}(p_{k+1} - y_k)$. In other words, $p_{k+1} = y_k + \gamma(Ax_k - b)$. Similarly, $y_{k+1} = y_k + \gamma(Ax_{k+1} - b)$. For the second inequality, we have $0 \in \partial f(x_{k+1}) + A^*p_{k+1} + \frac{1}{\gamma}(x_{k+1} - x_k)$, which is the same as $x_{k+1} = (I + \gamma\partial f)^{-1}\big(x_k - \gamma A^*p_{k+1}\big)$.

In parts 5, 6 and 7, $(\hat{x}, \hat{y})$ is any optimal pair.

5. Prove that
$$\begin{aligned} 2\gamma\big(\mathcal{L}(x_{k+1}, p_{k+1}) - \mathcal{L}(\hat{x}, p_{k+1})\big) &\leq \|x_k - \hat{x}\|^2 - \|x_{k+1} - \hat{x}\|^2 - \|x_{k+1} - x_k\|^2 \\ 2\gamma\big(\mathcal{L}(x_{k+1}, \hat{y}) - \mathcal{L}(x_{k+1}, y_{k+1})\big) &\leq \|y_k - \hat{y}\|^2 - \|y_{k+1} - \hat{y}\|^2 - \|y_{k+1} - y_k\|^2 \\ 2\gamma\big(\mathcal{L}(x_{k+1}, y_{k+1}) - \mathcal{L}(x_{k+1}, p_{k+1})\big) &\leq \delta\|y_{k+1} - p_{k+1}\|^2 + \frac{1}{\delta}\|y_{k+1} - y_k\|^2 \end{aligned}$$

[1] Since f is continuous, we have $\partial(f + \iota_V) = \partial f + \partial\iota_V$. You do not need to prove this.

for every $k \geq 0$ and $\delta > 0$. Suggestion: Remember (1) the definition of subgradient, and (2) that $2ab \leq \delta a^2 + \frac{1}{\delta} b^2$ for $a, b, \delta > 0$.
By 4, $-A^* p_{k+1} - \frac{1}{\gamma}(x_{k+1} - x_k) \in \partial f(x_{k+1})$. Using the subgradient inequality, we obtain

$$\begin{aligned} 2\gamma \mathcal{L}(\hat{x}, p_{k+1}) &= 2\gamma f(\hat{x}) \\ &\geq 2\gamma f(x_{k+1}) - 2\gamma\big(A^* p_{k+1} + \tfrac{1}{\gamma}(x_{k+1} - x_k)\big) \cdot (\hat{x} - x_{k+1}) \\ &= 2\gamma f(x_{k+1}) - 2\gamma A^* p_{k+1} \cdot (\hat{x} - x_{k+1}) + 2(x_{k+1} - x_k) \cdot (\hat{x} - x_{k+1}) \\ &= 2\gamma \mathcal{L}(x_{k+1}, p_{k+1}) + 2(x_{k+1} - x_k)) \cdot (\hat{x} - x_{k+1}). \end{aligned}$$

We use $\|a \pm b\|^2 = \|a\|^2 + \|b\|^2 \pm 2a \cdot b$ conveniently to obtain the first inequality. The second inequality is obtained by rewriting the left-hand side as $2\gamma(\hat{y} - y_{k+1}) \cdot (Ax_{k+1} - b) = 2(\hat{y} - y_{k+1}) \cdot (y_{k+1} - y_k)$. For the last one, we use Cauchy-Schwarz inequality and the second suggestion.

6. Show that if $\gamma\|A\| < 1$, there is $\varepsilon > 0$ such that

$$\|x_{k+1} - \hat{x}\|^2 + \|y_{k+1} - \hat{y}\|^2 + 2\gamma\big(\mathcal{L}(x_{k+1}, \hat{y}) - \mathcal{L}(\hat{x}, p_{k+1})\big) + \varepsilon\|Ax_{k+1} - b\|^2 \leq \|x_k - \hat{x}\|^2 + \|y_k - \hat{y}\|^2$$

for every $k \geq 0$.
We sum the three inequalities in 5 and choose δ appropriately to cancel the remaining terms: First observe that $\|y_{k+1} - p_{k+1}\| = \|\gamma A(x_{k+1} - x_k)\| \leq \gamma\|A\|\|x_{k+1} - x_k\|$. If we take $\delta \in (1, (\gamma\|A\|)^{-2})$ both $\|x_{k+1} - x_k\|^2$ and $\|y_{k+1} - y_k\|^2$ remain with negative coefficients. Actually, we have proved that there exist $\varepsilon_1, \varepsilon_2 > 0$ such that

$$\|x_{k+1} - \hat{x}\|^2 + \|y_{k+1} - \hat{y}\|^2 + 2\gamma\big(\mathcal{L}(x_{k+1}, \hat{y}) - \mathcal{L}(\hat{x}, p_{k+1})\big) + \varepsilon_1\|Ax_{k+1} - b\|^2 + \varepsilon_2\|x_{k+1} - x_k\|^2$$
$$\leq \|x_k - \hat{x}\|^2 + \|y_k - \hat{y}\|^2$$

for every $k \geq 0$.

7. Deduce that $\lim_{k\to\infty} f(x_k) = f(\hat{x})$ and $\lim_{k\to\infty} Ax_k = b$.
Using the telescopic property, we see that the nonnegative series $\sum\big(\mathcal{L}(x_{k+1}, \hat{y}) - \mathcal{L}(\hat{x}, p_{k+1})\big)$ and $\sum \|Ax_{k+1} - b\|^2$ are convergent, so their general terms go to 0.

8. Prove that (x_k, y_k) converges to an optimal pair. Suggestion: Verify that for every optimal pair $(\hat{x}, \hat{y})$, $\lim_{k\to\infty} \big[\|x_k - \hat{x}\|^2 + \|y_k - \hat{y}\|^2\big]$ exists.
From 6, the nonnegative sequence $\|x_k - \hat{x}\|^2 + \|y_k - \hat{y}\|^2$ is nonincreasing. Therefore $\lim_{k\to\infty} \big[\|x_k - \hat{x}\|^2 + \|y_k - \hat{y}\|^2\big]$ exists, and the sequence (x_k, y_k) is bounded. Suppose (x_{n_k}, y_{n_k}) converges to some (x_∞, y_∞). From 7, x_∞ is a solution of $(\mathcal{P})$. On the other hand, by 4 (with $k+1$ replaced by k), we have

$$f(z) \geq f(x_k) + \big(-A^* p_k - \tfrac{1}{\gamma}(x_k - x_{k-1})\big) \cdot (z - x_k)$$

for all z. Since $\lim_{k\to\infty} \|p_k - y_k\| = \lim_{k\to\infty} \|x_k - x_{k-1}\| = 0$, we can pass to the limit in the inequality above to obtain

$$f(z) \geq f(x_\infty) + (-A^* y_\infty) \cdot (z - x_\infty),$$

for all z, and conclude that $-A^* y_\infty \in \partial f(x_\infty)$. It follows that (x_∞, y_∞) is an optimal pair, and so $\lim_{k\to\infty} \big[\|x_k - x_\infty\|^2 + \|y_k - y_\infty\|^2\big]$ must be 0.