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 $x \in \mathbb{R}^N$. Suggestion: How is V related to $\ker(A)$ $\big(= \operatorname{ran}(A^*)^\perp\big)$?

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?

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

$$\mathcal{L}(\hat{x}, y) \le \mathcal{L}(\hat{x}, \hat{y}) \le \mathcal{L}(x, \hat{y})$$

   for all $(x, y) \in \mathbb{R}^N \times \mathbb{R}^M$.

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

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

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

5. Prove that

$$\begin{aligned} 2\gamma\big(\mathcal{L}(x_{k+1}, p_{k+1}) - \mathcal{L}(\hat{x}, p_{k+1})\big) &\le \|x_k - \hat{x}\|^2 - \|x_{k+1} - \hat{x}\|^2 - \|x_{k+1} - x_k\|^2 \\ 2\gamma\big(\mathcal{L}(x_{k+1}, \hat{y}) - \mathcal{L}(x_{k+1}, y_{k+1})\big) &\le \|y_k - \hat{x}\|^2 - \|y_{k+1} - \hat{x}\|^2 - \|y_{k+1} - y_k\|^2 \\ 2\gamma\big(\mathcal{L}(x_{k+1}, y_{k+1}) - \mathcal{L}(x_{k+1}, p_{k+1})\big) &\le \delta\|y_{k+1} - p_{k+1}\|^2 + \tfrac{1}{\delta}\|y_{k+1} - y_k\|^2 \end{aligned}$$

   for every $k \ge 0$ and $\delta > 0$. Suggestion: Remember (1) the definition of subgradient, and (2) that $2ab \le \delta a^2 + \frac{1}{\delta}b^2$ for $a, b, \delta > 0$.

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

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

   for every $k \ge 0$.

7. Deduce that $\lim\limits_{k\to\infty} f(x_k) = f(\hat{x})$ and $\lim\limits_{k\to\infty} Ax_k = b$.

8. Prove that (x_k, y_k) converges to an optimal pair. Suggestion: Verify that for every optimal pair $(\hat{x}, \hat{y})$, $\lim\limits_{k\to\infty} \big[\|x_k - \hat{x}\|^2 + \|y_k - \hat{y}\|^2\big]$ exists.

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