# Suggested Solutions to A.2

1. Consider the unconstrained optimization problem

$$\min f(x) \quad \text{s.t.} \quad x \in \mathbb{R}^n, \tag{P}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and $L_f$-smooth. Assume $X^* \subset \mathbb{R}^n$, the optimal set of (P), is nonempty. Let $f^*$ be the optimal value. Recall that the following useful details on GD method (convex case).

(1). the iterative process:

$$x^{k+1} = x^k - \frac{1}{L_k}\nabla f(x^k),$$

(2). $1/L_k$ is the $k$-th stepsize chosen by the constant-stepsize or the backtracking procedure.

(3). $L_k$ satisfies for any $x, y \in \mathbb{R}^n$,

$$f(x) - f\left(T_{L_k}(y)\right) \geq \frac{L_k}{2}\left\|x - T_{L_k}(y)\right\|^2 - \frac{L_k}{2}\|x-y\|^2 + f(x) - f(y) - \langle \nabla f(y), x-y\rangle.$$

(4). the bounds on $L_k$ (independent with $k$) is

$$\beta L_f \leq L_k \leq \alpha L_f.$$

(5). the theorem on sequence convergence under Fejér monotonicity.

Show that

(a) $\left\{f\left(x^k\right)\right\}_{k\geq 0}$ is nonincreasing.

(b) $\left\{\left\|x^k - x^*\right\|\right\}_{k\geq 0}$ is nonincreasing for any $x^* \in X^*$.

(c) $f(x^k) - f^* \leq \frac{\alpha L_f\left\|x^0 - x^*\right\|^2}{2k}$ for any $k \geq 1$ and $x^* \in X^*$.

(d) $\left\{x^k\right\}_{k\geq 0}$ converges to some optimal solution as $k \to \infty$.

(e) $\min\limits_{n=0,1,\cdots,k}\left\|\nabla f\left(x^n\right)\right\| \leq \frac{2\alpha^{1.5}L_f\left\|x^0-x^*\right\|}{\sqrt{\beta}k}$ for any $k \geq 1$ and $x^* \in X^*$.

(f) Under the constant stepsize rule in which $L_k \equiv L_f$ and $\alpha = \beta = 1$,

$$\left\|\nabla f\left(x^k\right)\right\| \leq \frac{2L_f\left\|x^0 - x^*\right\|}{k}$$

for any $k \geq 1$ and $x^* \in X^*$.

Hint: prove the norm of the gradient $\left\{\left\|\nabla f(x^k)\right\|\right\}_{k\geq 0}$ is nonincreasing.

*Proof.* (a) Let $x = y = x^k$ in (detail 3),

$$f(x^k) - f(x^{k+1}) \geq \frac{L_k}{2}\left\|x^k - x^{k+1}\right\|^2 = \frac{L_k}{2L_k^2}\left\|\nabla f(x^k)\right\|^2 \geq 0. \tag{Inq1}$$

(b) Let $x = x^*$, $y = x^k$ in (detail 3),

$$0 \geq f(x^*) - f(x^{k+1}) \geq \frac{L_k}{2} \left\|x^* - x^{k+1}\right\|^2 - \frac{L_k}{2} \left\|x^* - x^k\right\|^2 . \tag{Inq2.1}$$

Then

$$\left\|x^* - x^{k+1}\right\| \leq \left\|x^* - x^k\right\|^2 . \tag{Inq2.2}$$

(c) By (Inq2.1) and (detail 4),

$$f(x^k) - f^* \leq \frac{\alpha L_f}{2} \left\|x^* - x^k\right\|^2 - \frac{\alpha L_f}{2} \left\|x^* - x^{k-1}\right\|^2 \text{ for } k \geq 1.$$

Hence,

$$k\big[f(x^k) - f^*\big] \leq \sum_{n=1}^{k} \left[f\left(x^n\right) - f^*\right] \leq \frac{\alpha L_f}{2} \left\|x^* - x^k\right\|^2 \tag{Inq3}$$

where the first inequality in (Inq3) holds by (Inq1).

(d) $\cdots$

(e) By (Inq1) and (detail 4),

$$f(x^n) - f(x^{n+1}) \geq \frac{\beta}{2\alpha^2 L_f} \left\|\nabla f(x^n)\right\|^2 . \tag{Inq5.1}$$

Summing (Inq5.1) over $n = k, k+1, \cdots, 2k-1$ yields

$$f(x^k) - f^* \geq f(x^{2k}) - f^* + \frac{\beta}{2\alpha^2 L_f} \sum_{n=k}^{2k-1} \left\|\nabla f(x^n)\right\|^2 .$$

By (Inq3),

$$\frac{\alpha^3 L_f^2}{\beta k} \left\|x^* - x^k\right\| \geq \sum_{n=k}^{2k-1} \left\|\nabla f(x^n)\right\|^2 \geq k \min_{n=k,\cdots,2k-1} \left\|\nabla f(x^n)\right\|^2 \geq k \min_{n=0,\cdots,2k} \left\|\nabla f(x^n)\right\|^2 .$$

So

$$\min_{n=0,\cdots,2k-1} \left\|\nabla f(x^n)\right\|^2 \leq \frac{\alpha^3 L_f^2}{\beta k^2} \left\|x^* - x^k\right\|^2 . \tag{Inq5.2}$$

$$\min_{n=0,\cdots,2k} \left\|\nabla f(x^n)\right\|^2 \leq \frac{\alpha^3 L_f^2}{\beta k^2} \left\|x^* - x^k\right\|^2 .. \tag{Inq5.3}$$

For any $p = 1, 2, \cdots$, if $p$ is odd, then $\frac{p+1}{2}$ is an integer and

$$\min_{n=0,\cdots,p} \left\|\nabla f(x^n)\right\|^2 = \min_{n=0,\cdots,2\frac{p+1}{2}-1} \left\|\nabla f(x^n)\right\|^2 \leq \frac{\alpha^3 L_f^2}{\beta \left(\frac{p+1}{2}\right)^2} \left\|x^* - x^{\frac{p+1}{2}}\right\|^2 \leq \frac{4\alpha^3 L_f^2}{\beta p^2} \left\|x^* - x^p\right\|^2 .$$

If $p$ is even, then

$$\min_{n=0,\cdots,p} \left\|\nabla f(x^n)\right\|^2 = \min_{n=0,\cdots,2\frac{p}{2}} \left\|\nabla f(x^n)\right\|^2 \leq \frac{\alpha^3 L_f^2}{\beta \left(\frac{p}{2}\right)^2} \left\|x^* - x^{\frac{p}{2}}\right\|^2 \leq \frac{4\alpha^3 L_f^2}{\beta p^2} \left\|x^* - x^p\right\|^2 .$$

(f) Recall that

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

Let $x = x^k, y = x^{k+1}$,

$$\left\langle \nabla f(x^k) - \nabla f(x^{k+1}), \nabla f(x^k) \right\rangle \geq \left\| \nabla f(x^k) - \nabla f(x^{k+1}) \right\|^2 .$$

Set $a = \nabla f(x^k), b = \nabla f(x^{k+1})$, we have

$$(a-b)^T a \geq (a-b)^T (a-b).$$

Then

$$\begin{aligned} b^T b - a^T a =& (b - a + a)^T (b - a + a) - a^T a = (b-a)^T (b-a) - 2(a-b)^T a + a^T a - a^T a \\ \leq& (b-a)^T (b-a) - 2(a-b)^T (a-b) \leq 0. \end{aligned}$$

□

2.

3. (a) Notice that

$$f''(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ -2x & \text{if } x > 0. \end{cases}$$

Hence the Newton's method is not well-defined if intial point $x^0 \leq 0$. We set $\sqrt{c} \neq x^0 > 0$, the iterative formula is

$$x^{k+1} := x^k - \frac{f'(x^k)}{f''(x^k)} = \frac{1}{2} \left( x^k + \frac{c}{x^k} \right).$$

It's easy to show that $x^n \geq \sqrt{c}$ for $n = 1, 2, \cdots$ and $\{x^n\}_{n \geq 1}$ is decreasing, so $\{x^n\}_{n \geq 1}$ converges to $\sqrt{c}$ noticing the fixed points of the iterative equation are $\pm\sqrt{c}$.

(b) Let

$$f_i(x) = \frac{1}{1 + \exp\left(-b_i a_i^T x\right)}.$$

Then

$$\begin{aligned} \nabla l(x) &= -\frac{1}{m} \sum_{i=1}^{m} \left(1 - f_i(x)\right) b_i a_i + 2\lambda x \\ \nabla^2 l(x) &= -\frac{1}{m} \sum_{i=1}^{m} \left(1 - f_i(x)\right) f_i(x) b_i^2 a_i a_i^T + 2\lambda I. \end{aligned}$$

4. (a) Because $\Omega$ is bounded, there exists $M \in \mathbf{R}$ such that $\sup\{\|w\| \,|w \in \Omega\} < M$. Hence

$$\mu_\Omega(x) \leq \sup\{\|x\| + \|w\| \,|w \in \Omega\} = \|x\| + \sup\{\|w\| \,|w \in \Omega\} \leq \|x\| + M < \infty.$$

(b) Let $x, y \in \mathbf{R^n}$ and $0 \leq \lambda \leq 1$, set $z := \lambda x + (1-\lambda) y$. For any $\varepsilon > 0$, there exists $w_z \in \Omega$ such that

$$\mu_\Omega(z) - \varepsilon \leq \|z - \omega_z\| \leq \lambda \|x - \omega_z\| + (1-\lambda) \|y - \omega_z\| \leq \lambda \mu_\Omega(x) + (1-\lambda) \mu_\Omega(y).$$

The arbitrariness of $\epsilon$ implies that $\mu_\Omega$ is convex.

5. (a)

$$\begin{aligned}
g_1(\lambda x+(1-\lambda)y) &= f(\lambda x+(1-\lambda)y)-\frac{\sigma}{2}\|\lambda x+(1-\lambda)y\|^2\\
&\le \lambda f(x)+(1-\lambda)f(y)-\frac{\sigma}{2}\left(\lambda(1-\lambda)\|x-y\|^2+\|\lambda x+(1-\lambda)y\|^2\right)\\
&= \lambda f(x)+(1-\lambda)f(y)-\lambda\frac{\sigma}{2}\|x\|^2-(1-\lambda)\|y\|^2\\
&= \lambda g_1(x)+(1-\lambda)g_1(y)
\end{aligned}$$

   (b) Notice that $g_2(x)=g_1(x)+\frac{\sigma}{2}\left(\|x^*\|^2-2x^{*T}x\right)$. The result can be obtained by using the definition of convex functions.

   (c) The result can be obtained by using the definition of strongly convex functions.

6. For any $x\in\mathbf{R^n}$, we have

   (a) $d_\Omega(x)$ is well-defined.
   Just notice that the real subset $\{\|x-z\|\,|z\in\Omega\}$ is lower bounded.

   (b) $P_\Omega(x)$ is nonempty.
   By the definition of $d_\Omega(x)$, we have $z_n\in\Omega$ satisfying $\|x-z^n\|\le d_\Omega(x)+\frac{1}{n}$. On the other hand, the boundedness and closeness of $\Omega$ implies that there exists a limiting point $w\in\Omega$ of $(z_n)_{n\ge1}$. Hence $\|x-w\|=d_\Omega(x)$.
   In fact, just notice that the distance operator $\|x-\cdot\|$ is continuous and $\Omega$ is compact for any fixed $x\in\mathbf{R^n}$. Hence the distance operator attains it's infimum over $\Omega$ at some point.

   (c) $P_\Omega(x)$ is a singleton.
   Suppose there exist $w_1\ne w_2$ with $w_1,w_2\in P_\Omega(x)$, noticing $\frac{w_1+w_2}{2}\in\Omega$ since that $\Omega$ is convex. Then

$$d_\Omega(x)\le\left\|x-\frac{w_1+w_2}{2}\right\|<\frac{1}{2}\|x-w_1\|+\frac{1}{2}\|x-w_1\|=d_\Omega(x).$$

   It's a contradiction noticing that the second inequality is strict.
   In fact,

$$\left\|x-\frac{w_1+w_2}{2}\right\|^2=\left(\frac{1}{2}\|x-w_1\|+\frac{1}{2}\|x-w_2\|\right)^2\Rightarrow\|w_1-w_2\|^2=0.$$

7. *Proof.* (a) Pick $\bar v\in\partial f(\bar x)$. (By the way, Theorem 5.2.4 guarantees the nonemptiness of $\partial f(\bar x)$). Then we have

$$f(x)\ge f(\bar x)+\langle\bar v,x-\bar x\rangle\text{ for all }x\in\mathbb{R}^n.\tag{1}$$

   Moreover, the conclusion condition $0\in\partial f(\bar x)+N_\Omega(\bar x)$ shows that $-v\in N_\Omega(\bar x)$, or equivalently,

$$\langle-\bar v,x-\bar x\rangle\le0\text{ for all }x\in\Omega.\tag{2}$$

   Combining (1) and (2), we derive

$$f(x)\ge f(\bar x)+\langle\bar v,x-\bar x\rangle\text{ for all }x\in\Omega.$$

   (b) By second-order Taylor expansion, we have

$$f(\bar x+\alpha(x-\bar x))=f(\bar x)+\alpha(x-\bar x)^T\nabla f(\bar x)+\frac{(x-\bar x)^T\nabla f(\bar x)(x-\bar x)}{2}\alpha^2+o(\alpha^2).$$

   If we choose $x\in\Omega$ and $\alpha>0$ sufficiently small, then the convexity of $\Omega$ implies that $\bar x+\alpha(x-\bar x)\in\Omega$. Considering Condition (a) and (b), we have

$$f(\bar x+\alpha(x-\bar x))\ge f(\bar x)+\frac{(x-\bar x)^T\nabla f(\bar x)(x-\bar x)}{2}\alpha^2>f(\bar x)$$

   for sufficiently small $\alpha>0$. Hence $\bar x$ is a strictly minimizer of $f$ over $\Omega$.

□

8. *Proof.* Define

$$\mathcal{D}(\bar{x}) = \left\{ \alpha (x - \bar{x}) \middle| \forall \alpha \geq 0 \text{ and } x \in \Omega \right\}.$$

First, we show $\mathcal{D}(\bar{x}) \subset T(\bar{x})$. For any $0 \neq d \in \mathcal{D}(\bar{x})$, i.e, $\exists \alpha_d > 0, x_d \in \Omega$ with $d = \alpha_d (x_d - \bar{x})$, then $d$ is a feasible direction of $\Omega$ at $\bar{x}$ (think of why) and hence $d \in T(\bar{x})$ (think of why).

Secondly, we show $T(\bar{x}) \subset \mathrm{cl}(\mathcal{D})$. Actually, for any $d \in T(\bar{x})$, $\exists t_k \downarrow 0, d^k \to d$ with $\bar{x} + t_k d^k \in \Omega$. Notice that

$$d^k = \frac{1}{t_k} \left( \bar{x} + t_k d^k - \bar{x} \right),$$

i.e., $d^k \in \mathcal{D}(\bar{x})$. Then $d \in \mathrm{cl}(\mathcal{D})$. □

**more details:** Is $\mathcal{D}(\bar{x}) / \{0\}$ the collection of feasible directions to $\Omega$ at $\bar{x}$? Is $\mathcal{D}(\bar{x})$ closed?

9. *Proof.* $\supset$: For any $d$ satisfying $\langle d, x - \bar{x} \rangle \leq 0, \ \forall x \in \Omega$, we have

$$\limsup_{x \xrightarrow{\Omega} \bar{x}} \frac{\langle d, x - \bar{x} \rangle}{\|x - \bar{x}\|} \leq 0,$$

i.e., $d \in N(\bar{x})$.

$\subset$: Pick $d \in N(\bar{x})$ arbitrarily, for any $\bar{x} \neq x \in \Omega$, define

$$x^k = \frac{1}{k} x + \left( 1 - \frac{1}{k} \right) \bar{x},$$

noticing that $x^k \in \Omega$, (the convexity of $\Omega$), $x^k \to \bar{x}$ and $d \in N(\bar{x})$, we have

$$0 \geq \limsup_{k \to \infty} \frac{\langle d, x^k - \bar{x} \rangle}{\|x^k - \bar{x}\|} = \frac{\langle d, x - \bar{x} \rangle}{\|x - \bar{x}\|},$$

that is to say, $\langle d, x - \bar{x} \rangle \leq 0$. □

10. *Proof.* (a)

(b)

(c) Let

$$\Omega = \left\{ (x_1, x_2) \in \mathbb{R}^2 \middle| \forall x_1 \in \mathbb{R}^n, x_2 \geq e^{x_1} \right\}$$

and

$$D = \mathbb{R} \times \{0\}.$$

Then for any $p = (p_1, p_2) \in \mathbb{R}^2$, we have

$$\inf_{x \in D} \langle p, x \rangle = \inf_{x_1 \in \mathbb{R}} p_1 x_1 = -\infty.$$

□