Biomath 210 Homework 6

Huwenbo Shi (603-778-363) shihuwenbo@ucla.edu

December 9, 2015

Problem 7.1

To prove the first equality, we notice that

$$\begin{aligned} f \circ P_C(\mathbf{y} + t\mathbf{v}) - f \circ P_C(\mathbf{y}) &= s_f[P_C(\mathbf{y} + t\mathbf{v}), P_C(\mathbf{y})][P_C(\mathbf{y} + t\mathbf{v}) - P_C(\mathbf{y})]_7 \\ &= s_f[P_C(\mathbf{y} + t\mathbf{v}), P_C(\mathbf{y})]s_{P_C}(\mathbf{y} + t\mathbf{v}, \mathbf{y})t\mathbf{v}, \end{aligned} \tag{1}$$

which implies

$$\frac{f \circ P_C(\mathbf{y} + t\mathbf{v}) - f \circ P_C(\mathbf{y})}{t} = s_f[P_C(\mathbf{y} + t\mathbf{v}), P_C(\mathbf{y})]s_{P_C}(\mathbf{y} + t\mathbf{v}, \mathbf{y})\mathbf{v}. \tag{2}$$

Sending t to 0, yields $d_{\mathbf{v}} f \circ P_C(\mathbf{y}) = df(P_C(\mathbf{y}))dP_C(\mathbf{y})\mathbf{v} = df(P_C(\mathbf{y}))d_{\mathbf{v}}P_C(\mathbf{y})$. For the boundary point $\mathbf{y}$, $P_C(\mathbf{y}) = \mathbf{y}$, therefore, we have

$$d_{\mathbf{v}} f \circ P_C(\mathbf{y}) = df(\mathbf{y})d_{\mathbf{v}}P_C(\mathbf{y}) = -\mathbf{g}^* d_{\mathbf{v}}P_C(\mathbf{y}). \tag{3}$$

To prove the second equality, we show that $d_{\mathbf{v}}P_C(\mathbf{y}) = P_T(\mathbf{v})$. From Equation (3.5) of chapter 3, and Equation (7.8) of appendix, we know that

$$d_{\mathbf{v}}\mathrm{dist}(\mathbf{y}, C) = \|\mathbf{v} - d_{\mathbf{v}}P_C(\mathbf{y})\| = \mathrm{dist}(\mathbf{v}, T) = \inf_{\mathbf{z}\in T} \|\mathbf{v} - \mathbf{z}\| = \|\mathbf{v} - P_T(\mathbf{v})\|. \tag{4}$$

Therefore, $d_{\mathbf{v}}P_C(\mathbf{y}) = P_T(\mathbf{v})$. And we conclude that

$$d_{\mathbf{v}} f \circ P_C(\mathbf{y}) = df(\mathbf{y})d_{\mathbf{v}}P_C(\mathbf{y}) = -\mathbf{g}^* d_{\mathbf{v}}P_C(\mathbf{y}) = -\mathbf{g}^* P_T(\mathbf{v}). \tag{5}$$

From proposition 5.2.4, $\mathbf{g} = P_T(\mathbf{g}) + P_{T^\circ}(\mathbf{g})$. Therefore, $d_{\mathbf{v}} f \circ P_C(\mathbf{y}) = -P_T(\mathbf{g})^* P_T(\mathbf{v}) - P_{T^\circ}(\mathbf{g})^* P_T(\mathbf{v})$. Restricting $t \geqslant 0$ in Equation (2), if $f(P_C(\mathbf{y} + t\mathbf{v})) < f(P_C(\mathbf{y}))$, then $f(P_C(\mathbf{y}+t\mathbf{v})) - f(P_C(\mathbf{y})) < 0$ and $d_{\mathbf{v}} f \circ P_C(\mathbf{y}) < 0$. On the other hand, if $d_{\mathbf{v}} f \circ P_C(\mathbf{y}) < 0$, we have $f(P_C(\mathbf{y} + t\mathbf{v})) < f(P_C(\mathbf{y}))$. Thus, $-P_T(\mathbf{g})^* P_T(\mathbf{v}) - P_{T^\circ}(\mathbf{g})^* P_T(\mathbf{v})$ is necessary and sufficient for $\mathbf{v}$ to be a descent direction.

When $\mathbf{g} = \mathbf{v}$, we have

$$\begin{aligned} d_{\mathbf{v}} f \circ P_C(\mathbf{y}) &= -P_T(\mathbf{v})^* P_T(\mathbf{v}) - P_{T^\circ}(\mathbf{v})^* P_T(\mathbf{v}) \\ &= -P_T(\mathbf{v})^* P_T(\mathbf{v}) = -\mathbf{v}^* P_T(\mathbf{v}), \end{aligned} \tag{6}$$

which implies $[P_T(\mathbf{v}) - \mathbf{v}]^* P_T(\mathbf{v}) = 0$. Since $\mathbf{v}$ is arbitrary, $P_T(\mathbf{v})$ must be $\mathbf{0}$ for the equality to hold.

When $\mathbf{g} \in T^\circ$, we have $\mathbf{g} = P_{T^\circ}(\mathbf{g})$, which entails $-P_T(\mathbf{g})^* P_T(\mathbf{v}) - \mathbf{g}^* P_T(\mathbf{v}) = -\mathbf{g}^* P_T(\mathbf{v})$ and $-P_T(\mathbf{g})^* P_T(\mathbf{v}) = 0$. Since $\mathbf{g}$ is arbitrary, $P_T(\mathbf{v})$ must be $\mathbf{0}$. When $P_T(\mathbf{v}) = \mathbf{0}$, we have $\mathbf{g} = P_T(\mathbf{g}) + P_{T^\circ}(\mathbf{g}) = P_T(\mathbf{v}) + P_{T^\circ}(\mathbf{g}) = P_{T^\circ}(\mathbf{g})$. Therefore, $\mathbf{g} \in T^\circ$.

Problem 7.2

We first derive the semidifferential of $P_S(\mathbf{x})$. We note that

$$\frac{P_S(\mathbf{x} + t\mathbf{v}) - P_S(\mathbf{x})}{t} = s_{P_S}(\mathbf{x} + t\mathbf{v}, \mathbf{x})\mathbf{v}. \tag{7}$$

Since $\mathbf{x} \in S$, we have $P_S(\mathbf{x}) = \mathbf{x}$. For t approaching 0, i.e. the perturbation in $\mathbf{x}$ by $t\mathbf{v}$ doesn't change the relative rank order of magnitude of $\mathbf{x}_i$, we have

$$P_S(\mathbf{x} + t\mathbf{v})_i = \begin{cases} \mathbf{x}_i + t\mathbf{v}_i & \text{if } \mathbf{x}_i \neq 0 \\ 0 & \text{if } \mathbf{x}_i = 0 \end{cases} \tag{8}$$

Therefore,

$$\frac{P_S(\mathbf{x} + t\mathbf{v}) - P_S(\mathbf{x})}{t} = \begin{cases} \mathbf{v}_i & \text{if } \mathbf{x}_i \neq 0 \\ 0 & \text{if } \mathbf{x}_i = 0 \end{cases} \tag{9}$$

Sending t to 0, yields the conclusion stated in the problem.

For the second part of the problem, from problem 7.1 we know that

$$d_{\mathbf{v}} f \circ P_S(\mathbf{y}) = \nabla f(\mathbf{y})^* d_{\mathbf{v}} P_S(\mathbf{y}) = \sum_{\mathbf{y}_i \neq 0} \nabla f(\mathbf{y})_i d_{\mathbf{v}} P_S(\mathbf{y})_i. \tag{10}$$

For $\mathbf{v} = -\nabla f(\mathbf{y})$, we have

$$d_{\mathbf{v}} f \circ P_S(\mathbf{y}) = -\sum_{\mathbf{x}_i \neq 0} \frac{\partial}{\partial \mathbf{y}_i} f(\mathbf{y})^2 < 0, \tag{11}$$

unless $\frac{\partial}{\partial \mathbf{y}_i} f(\mathbf{y}) = 0$ for all $\mathbf{y}_i \neq 0$.

For the third part of the problem, we prove by finding counterexample,

$$d_{-\mathbf{Au}} f \circ P_S(\mathbf{x}) = \mathbf{u}^* d_{-\mathbf{Au}} P_S[(1,0)] = \mathbf{u}^*(-[\mathbf{Au}]_1, 0) = -(1 - 4/3) = \frac{1}{3}. \tag{12}$$

Problem 7.6

From (7.1), we know that

$$\|\mathbf{M}\|_\dagger = \|\mathbf{UDU}^{-1}\|_\dagger = \|\mathbf{U}^{-1}(\mathbf{UDU}^{-1})\mathbf{U}\|_\dagger = \|\mathbf{D}\|_\dagger. \tag{13}$$

We find $\|\mathbf{D}\|_\dagger$ by

$$\|\mathbf{D}\|_\dagger = \sup_{\|\mathbf{x}\|_\dagger=1} \|\mathbf{D}\mathbf{x}\|_\dagger. \tag{14}$$

Here, $\mathbf{D} = \mathbf{IDI}$ and $\mathbf{u}_i = \mathbf{e}_i$. The optimal $\mathbf{x}$ satisfies $\mathbf{x}_j = 1$ where $|d_j|$ is largest and 0 else where. And we note that $\mathbf{Dx} = d_j\mathbf{e}_j$, thus $\|\mathbf{M}\|_\dagger = \|\mathbf{D}\|_\dagger = \|\mathbf{Dx}\|_\dagger = \max_{1\leqslant i\leqslant n}|d_i|$.

To show contraction, we notice that

$$\|\mathbf{Mx} + \mathbf{v} - (\mathbf{My} + \mathbf{v})\|_\dagger = \|\mathbf{M}(\mathbf{x} - \mathbf{y})\|_\dagger \leqslant \|\mathbf{M}\|_\dagger\|\mathbf{x} - \mathbf{y}\|. \tag{15}$$

For $\mathbf{M}$ with spectral radius strictly less than 1, we have $\|\mathbf{Mx}+\mathbf{v}-(\mathbf{My}+\mathbf{v})\|_\dagger < \|\mathbf{x}-\mathbf{y}\|$.

The fixed point of the map satisfies $\mathbf{x} = \mathbf{Mx} + \mathbf{v}$. And therefore the fixed point satisfies $\mathbf{x} = -(\mathbf{M} - \mathbf{I})^{-1}\mathbf{v}$.

Problem 7.10

From the obtuse angle criterion, we have the in equality

$$\left[\mathbf{x}_n - \frac{\rho}{L}\nabla f(\mathbf{x}_n) - \mathbf{x}_{n+1}\right]^* [\mathbf{x}_n - \mathbf{x}_{n+1}] \leqslant 0. \tag{16}$$

Rearranging the inequality gives

$$\begin{aligned}
&\|\mathbf{x}_n - \mathbf{x}_{n+1}\|^2 - \frac{\rho}{L}df(\mathbf{x}_n)(\mathbf{x}_n - \mathbf{x}_{n+1}) \leqslant 0\\
&\|\mathbf{x}_n - \mathbf{x}_{n+1}\|^2 \leqslant -\frac{\rho}{L}df(\mathbf{x}_n)(\mathbf{x}_{n+1} - \mathbf{x}_n)\\
&df(\mathbf{x}_n)(\mathbf{x}_{n+1} - \mathbf{x}_n) \leqslant -\frac{L}{\rho}\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2.
\end{aligned} \tag{17}$$

Following the steps in solution to **Problem 7.22** (see Equation 26 to 28), we arrive at the inequality based on quadratic upper bound majorization

$$\begin{aligned}
f(\mathbf{x}_{n+1}) &\leqslant f(\mathbf{x}_n) + df(\mathbf{x}_n)(\mathbf{x}_{n+1} - \mathbf{x}_n) + \frac{L}{2}\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2\\
&\leqslant f(\mathbf{x}_n) - \frac{L}{\rho}\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 + \frac{L}{2}\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2\\
&= f(\mathbf{x}_n) - \left[\frac{L}{\rho} - \frac{L}{2}\right]\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2.
\end{aligned} \tag{18}$$

For $f(\mathbf{x})$ continuous, when $f(\mathbf{x})$ is coercive or S is compact, the minimum of $f(\mathbf{x})$ is attained. Since the sequence $f(\mathbf{x}_n)$ is monotonically decreasing, the limit $\lim_{n\to\infty} f(\mathbf{x}_n)$ exists. To show that $\lim_{n\to\infty}\|\mathbf{x}_{n+1} - \mathbf{x}_n\| = 0$, we notice that

$$0 \leqslant \|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 \leqslant -\frac{f(\mathbf{x}_{n+1}) - f(\mathbf{x}_n)}{L/\rho - L/2}. \tag{19}$$

As n approaches ∞, we have $0 \leqslant \|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 \leqslant 0$, which implies $\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 = 0$. Thus, $\lim_{n\to\infty}\|\mathbf{x}_{n+1} - \mathbf{x}_n\| = 0$ holds.

For the no-progress point $\mathbf{y}$, we have the obtuse angle criterion

$$\left[\mathbf{y} - \frac{\rho}{L}\nabla f(\mathbf{y}) - \mathbf{z}\right]^* (\mathbf{y} - \mathbf{z}) \leqslant 0, \tag{20}$$

which entails

$$\frac{\rho}{L} df(\mathbf{y})(\mathbf{y} - \mathbf{z}) \geqslant \mathbf{y}^*(\mathbf{y} - \mathbf{z}) - \mathbf{z}^*(\mathbf{y} - \mathbf{z}) = \|\mathbf{y} - \mathbf{z}\|^2 \geqslant 0. \tag{21}$$

Thus, the condition $\frac{\rho}{L} df(\mathbf{y})(\mathbf{y} - \mathbf{z}) \geqslant 0$ is a necessary condition for a fixed point, which is a necessary condition for optimality.

When $f(\mathbf{x})$ is convex, the fixed point implies global optimality.

Problem 7.15

When $\mathbf{x}$ occurs on the interior of C

When $\mathbf{x}$ occurs on the interior of C, then $\text{dist}(\mathbf{x}, C) = 0$, and $F_\rho(\mathbf{y})$ and $f(\mathbf{y})$ coincide in a neighborhood of $\mathbf{x}$. Since C is closed and convex, and $\mathbf{x}$ minimizes $f(\mathbf{y})$, it also minimizes $F_\rho(\mathbf{y})$. When $\mathbf{x}$ minimizes $F_\rho(\mathbf{y}) = f(\mathbf{y})$, it clearly minimizes $f(\mathbf{y})$. So the converse is also true.

When $\mathbf{x}$ occurs outside C

If $\mathbf{x}$ minimizes $F_\rho(\mathbf{y})$ but occurs outside C, then the directional derivative in the direction $\mathbf{v}$ for $F_\rho(\mathbf{x})$ is

$$d_{\mathbf{v}} F_\rho(\mathbf{x}) = d_{\mathbf{v}} f(\mathbf{x}) + \rho \frac{[\mathbf{x} - P_C(\mathbf{x})]^* \mathbf{v}}{\|\mathbf{x} - P_C(\mathbf{x})\|}. \tag{22}$$

Setting $\mathbf{v} = -\frac{\mathbf{x} - P_C(\mathbf{x})}{\|\mathbf{x} - P_C(\mathbf{x})\|}$ gives $d_{\mathbf{v}} F_\rho(\mathbf{x}) = d_{\mathbf{v}} f(\mathbf{x}) - \rho$. Since $\|\mathbf{v}\| = 1$, by proposition 3.2.3 we have $|d_{\mathbf{v}} f(\mathbf{x})| \leqslant L$, which entails $d_{\mathbf{v}} F_\rho(\mathbf{x}) = d_{\mathbf{v}} f(\mathbf{x}) - \rho \leqslant L - \rho < 0$ for $\rho > L$. By proposition 3.3.1, this contradicts the assumption that $\mathbf{x}$ minimizes $F_\rho(\mathbf{y})$. Therefore, if $\mathbf{x}$ minimizes $F_\rho(\mathbf{y})$, it must occurs in C. And we can apply the first and third conclusion to prove necessity and sufficiency.

When $\mathbf{x}$ occurs on the boundary of C

If $\mathbf{x}$ minimizes $F_\rho(\mathbf{y})$ and is a boundary point of C, then by (3.5)

$$d_{\mathbf{v}} F_\rho(\mathbf{x}) = d_{\mathbf{v}} f(\mathbf{x}) + \rho \|\mathbf{v} - P_{T_C(\mathbf{x})}(\mathbf{v})\|. \tag{23}$$

For tangent vectors $\mathbf{v} \in P_{T_C(\mathbf{x})}$, $f(\mathbf{y})$ and $F_\rho(\mathbf{y})$ have the same directional derivatives. For $f(\mathbf{y})$ convex, if $\mathbf{x}$ minimizes $f(\mathbf{y})$, then $d_{\mathbf{v}} f(\mathbf{x}) \geqslant 0$ for all tangent directions $\mathbf{v}$, which entails $d_{\mathbf{v}} F_\rho(\mathbf{x}) \geqslant 0$ for all $\mathbf{v}$ as well, justifying the optimality of $\mathbf{x}$ for $F_\rho(\mathbf{y})$. The converse is clearly also true.

Problem 7.16

Let $S = \{(\mathbf{u}, \mathbf{v}) : \mathbf{v} = f(\mathbf{u})\}$ be the graph of $f(\mathbf{x})$. Since $f(\mathbf{x})$ is continuous, for every sequence $\mathbf{x}_n$ with $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$, we have $\lim_{n\to\infty} f(\mathbf{x}_n) = f(\mathbf{x})$. Let $\mathbf{y}_n = f(\mathbf{x}_n)$ and $\mathbf{y} = f(\mathbf{x})$, then $\lim_{n\to\infty} f(\mathbf{x}_n) = \lim_{n\to\infty} \mathbf{y}_n = \mathbf{y}$. Therefore, every sequence of points, $(\mathbf{x}_n, \mathbf{y}_n)$, converges to $(\mathbf{x}, \mathbf{y}) \in S$. And therefore, S is a closed set. Thus, S being a closed set is a necessary condition for $f(\mathbf{x})$ to be continuous. The converse, however, is not true as shown by the counterexample in the problem.

Problem 7.19

We prove $S = T_{r-1} \circ \cdots \circ T_0$ is paracontractive through induction. The base case, $S_1 = T_0$, is paracontractive with fixed point set F_0 by problem definition. Assume $S_k = T_{k-1} \circ \cdots \circ T_0$ is paracontractive with fixed point set $\cap_{i=0}^{k-1} F_i$. Let $\mathbf{y} \in F_k$, since T_k is paracontractive, we have

$$\begin{aligned} \|T_k[T_{k-1} \circ \cdots \circ T_0(\mathbf{x})] - \mathbf{y}\|_\dagger &< \|T_{k-1} \circ \cdots \circ T_0(\mathbf{x}) - \mathbf{y}\|_\dagger \\ &< \|T_{k-2} \circ \cdots \circ T_0(\mathbf{x}) - \mathbf{y}\|_\dagger \\ &\cdots \\ &< \|T_0(\mathbf{x}) - \mathbf{y}\|_\dagger < \|\mathbf{x} - \mathbf{y}\|_\dagger \end{aligned} \tag{24}$$

Therefore, the map $S_{k+1} = T_k \circ T_{k-1} \circ \cdots \circ T_0$ is also paracontractive. Let $\mathbf{z}_{m+1} = S_{k+1}(\mathbf{z}_m)$, we then have the inequality $\|\mathbf{z}_{m+1} - \mathbf{y}\|_\dagger \leqslant \|\mathbf{z}_m - \mathbf{y}\|_\dagger$ which entails a lower bound $\|\mathbf{z}_m - \mathbf{y}\|_\dagger = d \geqslant 0$ and the cluster point $\mathbf{z}_\infty$ at which the lower bound is attained, suggesting

$$\|T_k \circ S_k(\mathbf{x}_\infty) - \mathbf{y}\|_\dagger \leqslant \|\mathbf{x}_\infty - \mathbf{y}\|_\dagger. \tag{25}$$

Therefore, $\mathbf{x}_\infty$ is both a stationary point in $\cap_{i=0}^{k-1} F_i$ (otherwise, $S_k(\mathbf{x}_\infty)$ will not be stationary making $T_k \circ S_k(\mathbf{x}_\infty)$ not stationary) and also a stationary point in F_k. In other words, $\mathbf{x}_\infty \in \cap_{i=0}^{k} F_i$.

In conclusion, $S = T_{r-1} \circ \cdots \circ T_0$ is paracontractive with fixed point set $F = \cap_{i=0}^{r-1} F_i$.

Problem 7.22

We first show that $g(\mathbf{x}) = \frac{L}{2}\mathbf{x}^*\mathbf{x} - f(\mathbf{x})$ is convex. Because the gradient of $f(\mathbf{x})$ is Lipschitz continuous, $f(\mathbf{x})$ is twice differentiable and has the inequality $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leqslant L\|\mathbf{x} - \mathbf{y}\|$. From Cauchy-Schwarz inequality, we have

$$\begin{aligned} [\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})]^*(\mathbf{x} - \mathbf{y}) &\leqslant \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|\|\mathbf{x} - \mathbf{y}\| \\ &\leqslant L\|\mathbf{x} - \mathbf{y}\|^2. \end{aligned} \tag{26}$$

The above inequality entails

$$\begin{aligned} [\nabla g(\mathbf{x}) - \nabla g(\mathbf{y})]^*(\mathbf{x} - \mathbf{y}) &= \{[L\mathbf{x} - \nabla f(\mathbf{x})] - [L\mathbf{y} - \nabla f(\mathbf{y})]\}^*(\mathbf{x} - \mathbf{y}) \\ &= L\|\mathbf{x} - \mathbf{y}\|^2 - [\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})]^*(\mathbf{x} - \mathbf{y}) \geqslant 0. \end{aligned} \tag{27}$$

It follows that $\nabla g(\mathbf{x})$ is monotone, that $g(\mathbf{x})$ is convex by proposition 5.2.2, and that $L\mathbf{I} - d^2 f(\mathbf{x})$ is positive semi-definite.

By the quadratic upper bound principle, we have

$$f(\mathbf{x}) \leqslant f(\mathbf{y}) + \nabla f(\mathbf{y})^*(\mathbf{x}-\mathbf{y}) + \frac{L}{2}\|\mathbf{x}-\mathbf{y}\|^2 = f(\mathbf{y}) + \frac{L}{2}\|\mathbf{x}-\mathbf{y}\|^2, \tag{28}$$

where the equality follows from $\nabla f(\mathbf{y}) = \mathbf{0}$ when $\mathbf{y}$ attains the minimum. Thus, we have the inequality $f(\mathbf{x}) - f(\mathbf{y}) \leqslant \frac{L}{2}\|\mathbf{x}-\mathbf{y}\|^2$.

By the quadratic upper bound principle, we also have

$$f(\mathbf{y}) \leqslant f(\mathbf{z}) \leqslant f(\mathbf{x}) + \nabla f(\mathbf{x})^*(\mathbf{z}-\mathbf{x}) + \frac{L}{2}\|\mathbf{z}-\mathbf{x}\|^2. \tag{29}$$

We minimize the right hand side over $\mathbf{z}$, this yields the optimal point $\hat{\mathbf{z}} = \mathbf{x} - \frac{1}{L}\nabla f(\mathbf{x})$, with optimal value $f(\mathbf{x}) - \frac{2}{L}\|\nabla f(\mathbf{x})\|^2$. This gives us a tighter upper bound for $f(\mathbf{y})$, i.e. $f(\mathbf{y}) \leqslant f(\mathbf{x}) - \frac{2}{L}\|\nabla f(\mathbf{x})\|^2$, which entails $f(\mathbf{x}) - f(\mathbf{y}) \geqslant \frac{2}{L}\|\nabla f(\mathbf{x})\|^2$.

Put together, we have

$$\frac{2}{L}\|\nabla f(\mathbf{x})\|^2 \leqslant f(\mathbf{x}) - f(\mathbf{y}) \leqslant \frac{L}{2}\|\mathbf{x}-\mathbf{y}\|^2. \tag{30}$$

Problem 7.23

The convexity of $h(\mathbf{x}|\mathbf{x}_n) = g(\mathbf{x}|\mathbf{x}_n) - \frac{\mu}{2}\|\mathbf{x}\|^2$ entails

$$\begin{aligned} & g(\mathbf{x}_n|\mathbf{x}_n) - \frac{\mu}{2}\|\mathbf{x}_n\|^2 \\ & \geqslant g(\mathbf{x}_{n+1}|\mathbf{x}_n) - \frac{\mu}{2}\|\mathbf{x}_{n+1}\|^2 + [\partial g(\mathbf{x}_{n+1}|\mathbf{x_n}) - \mu\mathbf{x}_{n+1}]^*(\mathbf{x}_n - \mathbf{x}_{n+1}) \\ & = g(\mathbf{x}_{n+1}|\mathbf{x}_n) - \frac{\mu}{2}\|\mathbf{x}_{n+1}\|^2 - \mu\mathbf{x}_{n+1}^*(\mathbf{x}_n - \mathbf{x}_{n+1}), \end{aligned} \tag{31}$$

where the equality follows from the fact that $\partial g(\mathbf{x}_{n+1}|\mathbf{x_n}) = \mathbf{0}$ at $\mathbf{x}_{n+1}$. Rearranging the inequality gives

$$\begin{aligned} g(\mathbf{x}_n|\mathbf{x}_n) &\geqslant g(\mathbf{x}_{n+1}|\mathbf{x}_n) + \frac{\mu}{2}\|\mathbf{x}_{n+1}\|^2 + \frac{\mu}{2}\|\mathbf{x}_n\|^2 - \mu\mathbf{x}_n^*\mathbf{x}_{n+1} \\ &= g(\mathbf{x}_{n+1}|\mathbf{x}_n) + \frac{\mu}{2}\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2. \end{aligned} \tag{32}$$

From the first inequality, we have

$$g(\mathbf{x}_n|\mathbf{x}_n) - g(\mathbf{x}_{n+1}|\mathbf{x}_n) = f(\mathbf{x}_n) - g(\mathbf{x}_{n+1}|\mathbf{x}_n) \geqslant \frac{\mu}{2}\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2. \tag{33}$$

Since $g(\mathbf{x}_{n+1}|\mathbf{x}_n) \geqslant f(\mathbf{x}_{n+1})$, $f(\mathbf{x}_n) - f(\mathbf{x}_{n+1}) \geqslant f(\mathbf{x}_n) - g(\mathbf{x}_{n+1}|\mathbf{x}_n)$. And so,

$$f(\mathbf{x}_n) - f(\mathbf{x}_{n+1}) \geqslant \frac{\mu}{2}\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2. \tag{34}$$

To show the third inequality, we note that

$$\frac{\mu}{2}\sum_{k=1}^{n}\|\mathbf{x}_k - \mathbf{x}_{k+1}\| \leqslant \sum_{k=1}^{n}[f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})] = f(\mathbf{x}_1) - f(\mathbf{x}_{n+1}). \tag{35}$$

Since the objective $f(\mathbf{x})$ is bounded and the sequence $f(\mathbf{x}_n)$ decreases monotonically, the left hand side of the third inequality is bounded as n approaches ∞. Assume that $\lim_{n\to\infty}\|\mathbf{x}_n - \mathbf{x}_{n+1}\| = \epsilon > 0$, then $\frac{\mu}{2}\sum_{i=1}^{\infty}\|\mathbf{x}_k - \mathbf{x}_{k+1}\| = \infty$, contradicting the fact that left hand side is bounded. Therefore, $\lim_{n\to\infty}\|\mathbf{x}_n - \mathbf{x}_{n+1}\| = 0$.

Problem 7.24

If $\mathbf{x}$ minimizes $g(\mathbf{y}|\mathbf{x})$, then by proposition 3.3.1, $d_{\mathbf{v}}g(\mathbf{x}|\mathbf{x}) \geqslant 0$ for all tangent directions $\mathbf{v}$. Since $d_{\mathbf{v}}(\mathbf{x}|\mathbf{x}) = d_{\mathbf{v}}f(\mathbf{x})$ for all $\mathbf{x}$ and $\mathbf{v}$ due to strong tangency condition, we have $d_{\mathbf{v}}f(\mathbf{x}) \geqslant 0$ for all tangent directions $\mathbf{v}$ as well. This justifies $\mathbf{x}$ as a stationary point of $f(\mathbf{y})$.