Problem 1.

(a.) To show: $f(z) \geq f(\bar{x}) + g_c(x)^T(z-x) + \frac{1}{2L}\|g_c(x)\|^2$

Where $\bar{x} = \Pi_c(x - \frac{1}{L}\nabla f(x))$ and $g_c(x) = L(x-\bar{x})$

$\Rightarrow$ Consider $f(z) - f(\bar{x}) = (f(z) - f(x)) - (f(\bar{x}) - f(x))$

We have $f(z) - f(x) \geq \nabla f(x)^T(z-x)$ by convexity, and

$f(\bar{x}) - f(x) \leq \nabla f(x)^T(\bar{x}-x) + \frac{L}{2}\|\bar{x}-x\|^2$ by smoothness.

Thus, we can get $f(z) - f(\bar{x}) \geq (\nabla f(x)^T(z-x)) - (\nabla f(x)^T(\bar{x}-x) + \frac{L}{2}\|\bar{x}-x\|^2)$

$= \nabla f(x)^T(z-\bar{x}) - \frac{L}{2}\|\bar{x}-x\|^2 \cdots ①$

Note that $(\bar{x} - x - \frac{1}{L}\nabla f(x))^T(\bar{x}-z) \leq 0$ for any z by projection theorem.

Rearrange the terms into $(g_c(x) - \nabla f(x))^T(z-\bar{x}) \leq 0$

Then we get $\nabla f(x)^T(z-\bar{x}) \geq g_c(x)^T(z-\bar{x}) \cdots ②$

Plug ② into ①, we get
$$
\begin{aligned}
f(z) - f(\bar{x}) &\geq \nabla f(x)^T(z-\bar{x}) - \frac{L}{2}\|\bar{x}-x\|^2 \\
&\geq g_c(x)^T(z-\bar{x}) - \frac{L}{2}\|\bar{x}-x\|^2 \\
&= g_c(x)^T(z-x+x-\bar{x}) - \frac{L}{2}\|x-\bar{x}\|^2 \\
&= g_c(x)^T(z-x) + L\|x-\bar{x}\|^2 - \frac{L}{2}\|x-\bar{x}\|^2 \\
&= g_c(x)^T(z-x) + \frac{1}{2L}\|g_c(x)\|^2
\end{aligned}
$$

Therefore, $f(z) \geq f(\bar{x}) + g_c(x)^T(z-x) + \frac{1}{2L}\|g_c(x)\|^2$ #

(b.) Take $x = x_t$, $z = x_t$ in (a.). then $\bar{x} = x_{t+1}$

$\Rightarrow f(x_t) \geq f(x_{t+1}) + g_c(x_t)^T(x_t - x_t) + \frac{1}{2L}\|g_c(x_t)\|^2$

Rearrange into $f(x_{t+1}) - f(x_t) \leq -\frac{1}{2L}\|g_c(x_t)\|^2$ #

(c.) Take $x = x_t$, $z = x^*$ in (a.), then $\bar{x} = x_{t+1}$: $f(x^*) \geq f(x_{t+1}) + g_c(x_t)^T(x^* - x_t) + \frac{1}{2L}\|g_c(x_t)\|^2 \cdots ③$

Take $x = x^*$, $z = x_t$ in (a.), then $\bar{x} = x^*$ and $g_c(x) = 0$: $f(x_t) \geq f(x^*) \cdots ④$

$③ + ④ \Rightarrow f(x_t) \geq f(x_{t+1}) + g_c(x_t)^T(x^* - x_t) + \frac{1}{2L}\|g_c(x_t)\|^2$

$\geq f(x_{t+1}) + g_c(x_t)^T(x^* - x_t)$ since $\|g_c(x_t)\|^2 \geq 0$

Rearrange into $f(x_{t+1}) - f(x_t) \leq g_c(x_t)^T(x_t - x^*) \underset{\text{taking } |\cdot|}{\leq} |g_c(x_t)^T(x_t - x^*)| \underset{\text{Cauchy-Schwarz}}{\leq} \|g_c(x_t)\|\|x_t - x^*\|$

Then we get $\|g_c(x_t)\| \geq \dfrac{f(x_{t+1}) - f(x_t)}{\|x_t - x^*\|}$ #

(d.) Define $\Delta_t := f(x_t) - f(x^*)$

$\Delta_{t+1} - \Delta_t = f(x_{t+1}) - f(x_t) \underset{(b.)}{\leq} -\frac{1}{2L}\|g_c(x_t)\|^2 \underset{(c)}{\leq} -\frac{1}{2L}\frac{1}{\|x_t - x^*\|^2}\|f(x_{t+1}) - f(x_t)\|^2$

Note that $\|x_t - x^*\| \le \|x_0 - x^*\|$ and $f(x_t) \ge f(x^*)$ for $t \ge 0$

So, we have $\Delta_{t+1} - \Delta_t \le -\frac{1}{2L}\frac{1}{\|x_t - x^*\|^2}\|f(x_{t+1}) - f(x_t)\|^2$

$$\le -\frac{1}{2L}\frac{1}{\|x_0 - x^*\|^2}\|f(x_{t+1}) - f(x_t)\|^2 \quad \text{by } \|x_0 - x^*\| \le \|x_t - x^*\|$$

$$\le -\frac{1}{2L}\frac{1}{\|x_0 - x^*\|^2}\|f(x_{t+1}) - f(x^*)\|^2 \quad \text{by } f(x^*) \le f(x_t)$$

$$= \frac{-\Delta_{t+1}^2}{2L\|x_0 - x^*\|^2} \#$$

$*$ To prove $\|x_t - x^*\| \le \|x_0 - x^*\|$

$$\|x_{k+1} - x^*\|^2 = \|\Pi_C(x_k - \tfrac{1}{L}\nabla f(x_k)) - \Pi_C(x^* - \tfrac{1}{L}\nabla f(x^*))\|^2$$

(Since Π_C is non-expansive) $\le \|(x_k - \frac{1}{L}\nabla f(x_k)) - (x^* - \frac{1}{L}\nabla f(x^*))\|^2$

$$= \|(x_k - x^*) - \tfrac{1}{L}(\nabla f(x_k) - \nabla f(x^*))\|^2$$

$$= \|x_k - x^*\|^2 + \tfrac{1}{L^2}\|\nabla f(x_k) - \nabla f(x^*)\|^2 - \tfrac{2}{L}\langle x_k - x^*, \nabla f(x_k) - \nabla f(x^*)\rangle$$

L-smooth

$(0 \le \|\nabla f(x_k) - \nabla f(x^*)\|^2 \le L\langle x_k - x^*, \nabla f(x_k) - \nabla f(x^*)\rangle)$ $\le \|x_k - x^*\|^2 + \frac{1}{L}\langle x_k - x^*, \nabla f(x_k) - \nabla f(x^*)\rangle - \frac{2}{L}\langle x_k - x^*, \nabla f(x_k) - \nabla f(x^*)\rangle$

$\Downarrow$

$0 \le \langle x_k - x^*, \nabla f(x_k) - \nabla f(x^*)\rangle$ $\quad = \|x_k - x^*\|^2 - \frac{1}{L}\langle x_k - x^*, \nabla f(x_k) - \nabla f(x^*)\rangle$

$\hookrightarrow \le \|x_k - x^*\|^2$

So, we know $\|x_t - x^*\| \le \|x_{t-1} - x^*\| \le \cdots \le \|x_0 - x^*\| \#$

(e) To show: $\Delta_t \le \frac{3L\|x_0 - x^*\|^2 + f(x_0) - f(x^*)}{t+1} = \frac{1}{t+1}(3L\|x_0 - x^*\|^2 + \Delta_0)$

When $t=0$, $\Delta_0 \le \Delta_0 + 3L\|x_0 - x^*\|^2$ holds

For $t=1$, by (a), we have $f(x^*) \le f(x_1) + L(x_0 - x_1)^T(x^* - x_0) + \frac{L}{2}\|x_0 - x_1\|^2$

$$f(x_1) - f(x^*) \le L(x_1 - x_0)^T(x^* - x_0) - \tfrac{L}{2}\|x_0 - x_1\|^2$$

$$= L(x_1 - x_0)^T(x^* - x_0) - \tfrac{L}{2}\|x_0 - x^* + x^* - x_1\|^2$$

$$= L(x_1 - x_0)^T(x^* - x_0) - \tfrac{L}{2}\|x_0 - x^*\|^2 - \tfrac{L}{2}\|x^* - x_1\|^2 - L(x_0 - x^*)^T(x^* - x_1)$$

$$= L\|x^* - x_0\|^2 - \tfrac{L}{2}\|x_0 - x^*\|^2 - \tfrac{L}{2}\|x^* - x_1\|^2$$

$$= \tfrac{L}{2}\|x^* - x_0\|^2 - \tfrac{L}{2}\|x^* - x_1\|^2$$

$$= \tfrac{3}{2}L\|x^* - x_0\|^2 - L\|x^* - x_0\|^2 - \tfrac{L}{2}\|x^* - x_1\|^2$$

(L-smooth) $= \frac{3}{2}L\|x^* - x_0\|^2 - 2(f(x^*) - f(x_0) - \nabla f(x_0)^T(x^* - x_0)) - (f(x^*) - f(x_1) - \nabla f(x_1)(x^* - x_1))$

$$= \tfrac{3}{2}L\|x^* - x_0\|^2 + 2(f(x_0) - f(x^*)) + 2\nabla f(x_0)^T(x^* - x_0) + (f(x_1) - f(x^*) + \nabla f(x_1)(x^* - x_1))$$

$$= \frac{3L\|x_0 - x^*\|^2 + \Delta_0}{2} + \tfrac{1}{2}(3f(x_0) + 2f(x_1) - 5f(x^*)) + 2\nabla f(x_0)^T(x^* - x_0) + \nabla f(x_1)^T(x^* - x_1)$$

$$\le \frac{3L\|x_0 - x^*\|^2 + \Delta_0}{2} \#$$

Suppose it holds up to $t=T$,

then for $t=T+1$:

$$\text{we have } \Delta_{T+1} \overset{(d.)}{\leq} \Delta_T - \frac{-\Delta_{T+1}}{2L\|x_0-x^*\|^2} \overset{\text{assumption}}{\leq} \frac{3L\|x_0-x^*\|^2+\Delta_0}{T+1} - \frac{-\Delta_{T+1}^2}{2L\|x_0-x^*\|^2}$$

Suppose $\Delta_{T+1} > \frac{1}{T+2}(3L\|x_0-x^*\|^2+\Delta_0)$

$$\text{then } \Delta_{T+1}-\Delta_T > \frac{1}{T+2}(3L\|x_0-x^*\|^2+\Delta_0) - \frac{1}{T+1}(3L\|x_0-x^*\|^2+\Delta_0)$$

$$= \frac{-1}{(T+2)(T+1)}(3L\|x_0-x^*\|^2+\Delta_0)$$

Note that $\frac{T+1}{T+2}\times\frac{3}{2}\geq 1$ for all $T\geq 1$

$$\Delta_{T+1}-\Delta_T > \frac{-1}{(T+2)(T+1)}\cdot\frac{T+1}{T+2}\cdot\frac{3}{2}(3L\|x_0-x^*\|^2+\Delta_0)$$

$$\geq -\frac{1}{(T+2)^2}\cdot\frac{3}{2}\cdot(3L\|x_0-x^*\|^2+\Delta_0)\cdot\frac{L\|x_0-x^*\|^2+\Delta_0}{L\|x_0-x^*\|^2}$$

$$= \frac{-(3L\|x_0-x^*\|+\Delta_0)^2}{(T+2)^2\cdot 2L\|x_0-x^*\|^2} > \frac{-\Delta_{T+1}^2}{2L\|x_0-x^*\|^2}, \text{ which contradicts to (d.)}$$

By mathematical induction, $\Delta_t \leq \frac{3L\|x_0-x^*\|^2+\Delta_0}{t+1}$ for $t\geq 0$ #

Problem 2. $\Pi_C(x) = \underset{z \in C}{\operatorname{argmin}} \|x - z\|$

(a.) If $x_c = \Pi_C(x)$, then x_c is the global minimizer of $f(z) = \|x - z\|^2$, $z \in C$

Then $\nabla f(z) = -2(x - z)$, by FONC-C, we know $\nabla f(x_c)^T (z - x_c) \geq 0 \quad \forall z \in C$.

So we have $\nabla f(x_c)(z - x_c) = -2(x - x_c)^T (z - x_c) \geq 0 \quad \forall z \in C$

$\Rightarrow (x - x_c)^T (z - x_c) \leq 0 \quad \forall z \in C$

If $(x - x_c)^T (z - x_c) \leq 0 \ \forall z \in C$

which means the angle $\angle x x_c z$ is not acute angle for all $z \in C$

then we can construct a hyperplane to separate $x x_c$ and $x_c z$,

and we can make the hyperplane to pass x_c and be perpendicular to $\overrightarrow{x x_c}$.

So that, x_c would be the closest point on the hyperplane with x

and since $x_c \in C$, so x_c is also the closest point in C with x.

Thus, x_c is the projection of x on C.

(b.) From (a.), we have $\langle x - x_c, z - x_c \rangle \leq 0 \quad \forall z \in C$

Since $z_c \in C$, so we can replace z with z_c

$\Rightarrow \quad \langle x - x_c, z_c - x_c \rangle \leq 0 \quad \cdots ①$

Similarly, we have $\langle z - z_c, x_c - z_c \rangle \leq 0 \Leftrightarrow \langle z_c - z, z_c - x_c \rangle \leq 0 \quad \cdots ②$

$① + ② : \langle z_c - x_c - (z - x), z_c - x_c \rangle \leq 0$

$\langle z_c - x_c, z_c - x_c \rangle \leq \langle z - x, z_c - x_c \rangle \leq \|z - x\| \|z_c - x_c\|$

Thus, we have

$\|z_c - x_c\|^2 \leq \|z - x\| \|z_c - x_c\|$

$\|z_c - x_c\| \leq \|z - x\|$ #

Problem 3. $D_\phi(y\|x) = \phi(y) - \phi(x) - \nabla\phi(x)^T(y-x)$

(a.) $\nabla_y D_\phi(y\|x) = \nabla_y \phi(y) - \nabla_y \phi(x) - \nabla_y(\nabla\phi(x)^T(y-x))$

$= \nabla\phi(y) - \nabla\phi(x)$ #

(b.) $D_{\phi_1+\lambda\phi_2}(y\|x) = (\phi_1+\lambda\phi_2)(y) - (\phi_1+\lambda\phi_2)(x) - \nabla(\phi_1+\lambda\phi_2)(x)^T(y-x)$

$= \phi_1(y) + \lambda\phi_2(y) - \phi_1(x) - \lambda\phi_2(x) - \nabla\phi_1(x)^T(y-x) - \lambda\nabla\phi_2(x)^T(y-x)$

$= D_{\phi_1}(y\|x) + \lambda D_{\phi_2}(y\|x)$

(c.) $D_\phi(z\|x) = \phi(z) - \phi(x) - \nabla\phi(x)^T(z-x)$

$= \phi(z) - \phi(\bar{x}) + \phi(\bar{x}) - \phi(x) - \nabla\phi(x)^T(z-\bar{x}+\bar{x}-x)$

$= D_\phi(z\|\bar{x}) + \phi(\bar{x}) - \phi(x) - \nabla\phi(x)^T(\bar{x}-x)$

Since $\bar{x} = \underset{x' \in C}{\operatorname{argmin}}\, D_\phi(x'\|x)$

by FONC-C, we have $\langle \nabla_{\bar{x}} D_\phi(\bar{x}\|x), z-\bar{x} \rangle \geq 0 \quad \forall z \in C$

From (a.), we have $\langle \nabla\phi(\bar{x}) - \nabla\phi(x), z-\bar{x} \rangle \geq 0 \quad \forall z \in C$

$D_\phi(z\|\bar{x}) + D_\phi(\bar{x}\|x) = \phi(z) - \phi(\bar{x}) + \phi(\bar{x}) - \phi(x) - \nabla\phi(\bar{x})^T(z-\bar{x}) - \nabla\phi(x)^T(\bar{x}-x)$

$= \phi(z) - \phi(x) - \nabla\phi(x)^T(z-x) + \nabla\phi(x)^T(z-x) - \nabla\phi(\bar{x})^T(z-\bar{x}) - \nabla\phi(x)^T(\bar{x}-x)$

$= D_\phi(z\|x) + \nabla\phi(x)^T(z-\bar{x}) - \nabla\phi(\bar{x})^T(z-\bar{x})$

$= D_\phi(z\|x) - \underbrace{\langle \nabla\phi(\bar{x}) - \nabla\phi(x), z-\bar{x} \rangle}_{\geq 0}$

$\leq D_\phi(z\|x)$ #

Problem 4.

Take step size = $\frac{2}{k+2}$, k =0, 1, 2, 3, …, 19999

Total iterations = 20000

Initial x_0 : all elements are 1/p

Gradient Calculation: using PyTorch

```
        obj = objective(x_, a)
        obj.backward()
        grad = x_.grad
```