1.

1. By smoothness,

$$f(x^*) \le f(x) + \langle \nabla f(x), x^* - x \rangle + \frac{L}{2}\| x^* - x \|_2^2$$

$$f(x) - f^* \ge \langle \nabla f(x), x - x^* \rangle - \frac{L}{2}\| x - x^* \|_2^2 = \langle \nabla f(x), u \rangle - \frac{L}{2}\|u\|_2^2 := h(u).$$

$h(u)$ is concave. $\nabla h(u) = 0$ $\nabla f(x) - Lu^* = 0.$ $u^* = \frac{1}{L}\nabla f(x)$

$$\Rightarrow \sup_u h(u) = \frac{1}{L}\|\nabla f(x)\|_2^2 - \frac{L}{2}\cdot\frac{1}{L^2}\|\nabla f(x)\|_2^2 = \frac{1}{2L}\|\nabla f(x)\|_2^2.$$

2. $h_x(z) = f(z) - \langle \nabla f(x), z \rangle$.   How do we understand this function?

is cvx: $h_x(z) \ge h_x(y) + \langle \nabla h_x(y), z - y \rangle$.

$$\nabla h_x(y) = \nabla f(y) - \nabla f(x).$$

$$h_x(z) - h_x(y) = f(z) - f(y) - \langle \nabla f(x), z - y \rangle$$

Only to prove $f(z) - f(y) \ge \langle \nabla f(y) - \nabla f(x), z - y \rangle + \langle \nabla f(x), z - y \rangle$

$$= \langle \nabla f(y), z - y \rangle. \quad \square.$$

is L-smooth: $\|\nabla^2 h_x(y)\| = \|\nabla^2 f(y)\|_2 \underset{\uparrow}{\le} L$

L-smooth.

x, y are symmetric so same res works for $h_y(z)$.

Apply part I to $h_x(z)$, and set $z = y$,

$$f(y) - \langle \nabla f(x), y \rangle - h_x^* \ge \frac{1}{2L}\|\nabla h_x(y)\|_2^2 = \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2$$

Apply part I to $h_y(z)$, and set $z = x$.

$$f(x) - \langle \nabla f(y), x \rangle - h_y^* \ge \frac{1}{2L}\|\nabla h_y(x)\|_2^2 = \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2$$

Consider h_x^*: By convexity,

$$f(z) \ge f(x) + \langle \nabla f(x), z-x \rangle.$$

$$h_x(z) = f(z) - \langle \nabla f(x), z \rangle \ge f(x) - \langle \nabla f(x), x \rangle.$$

Found that equality holds if $z = x$ $\Rightarrow h_x^* = f(x) - \langle \nabla f(x), x \rangle$.

Similarly, $h_y^* = f(y) - \langle \nabla f(y), y \rangle$. Plug back,

$$\begin{cases} f(y) - f(x) - \langle \nabla f(x), y-x \rangle \ge \frac{1}{2L} \| \nabla f(y) - \nabla f(x) \|_2^2 & ① \\ f(x) - f(y) - \langle \nabla f(y), x-y \rangle \ge \frac{1}{2L} \| \nabla f(y) - \nabla f(x) \|_2^2 & ② \end{cases}$$

$① + ② \Rightarrow \langle \nabla f(x) - \nabla f(y), x-y \rangle \ge \frac{1}{L} \| \nabla f(x) - \nabla f(y) \|_2^2$.

2. 1. Note that $\forall j \neq i$, $\nabla_j f(x + \delta e_i) - \nabla_j f(x) = 0$.

Hence condition implies $\forall i$, $\| \nabla f(x + \delta e_i) - \nabla f(x) \| \le L_i |\delta| = L_i \| (x + \delta e_i) - x \|$,

which is a $\underset{(\text{I D K})}{\underline{\text{detailed}}}$ version of smoothness (defined on different directions)

Apply L_i-smoothness,

$$\forall i, \ f(x + \delta e_i) - f(x) \le \langle \nabla f(x), \delta e_i \rangle + \frac{L_i}{2} \| \delta e_i \|_2^2$$
$$= \delta \nabla_i f(x) + \frac{L_i}{2} |\delta|^2$$

2. $X_{k+1} - X_k = -\underbrace{\alpha_{i_k} \nabla_{i_k} f(X_k)}_{\delta} e_{i_k}$ Apply res in part I,

$$\Rightarrow f(X_{k+1}) - f(X_k) \le -\alpha_{i_k} (\nabla_{i_k} f(X_k))^2 + \frac{L_{i_k}}{2} (\alpha_{i_k} \nabla_{i_k} f(X_k))^2$$

$$= (\nabla_{i_k} f(x_k))^2 \left(\underbrace{\frac{L_{i_k}}{2}\alpha_{i_k}^2 - \alpha_{i_k}}\right)$$

We wish RHS to be smaller. Since x_k is fixed,

we can only adjust $\square$ part.

Set $\alpha_{i_k} = \frac{-1}{-2\cdot\frac{L_{i_k}}{2}} = \frac{1}{L_{i_k}}$ and we got it.

This is our choice of α_i. (α_{i_k}) $\Rightarrow f(x_{k+1}) - f(x_k) \le -\frac{1}{2L_{i_k}}(\nabla_{i_k} f(x_k))^2$

Then

$$\mathbb{E}[f(x_{k+1}) - f(x_k)] \le \mathbb{E}_{\substack{i_k \\ \downarrow \\ \text{uni}}}\left[-\frac{1}{2L_{i_k}}(\nabla_{i_k} f(x_k))^2\right]$$

$$= \frac{1}{d}\sum_{i=1}^{d} -\frac{1}{2L_i}(\nabla_i f(x_k))^2$$

Set $L = \max_i L_i$

$$\le -\frac{1}{2Ld}\sum_{i=1}^{d}(\nabla_i f(x_k))^2 = -\frac{1}{2Ld}\|\nabla f(x_k)\|_2^2$$

$\beta = \frac{1}{Ld}$, the largest β I can get.

2.3. follow part 2, for fixed k, $\|\nabla f(x_k)\|_2^2 \le -\frac{2}{\beta}\mathbb{E}_{i_k}[f(x_{k+1}) - f(x_k)]$

$$\min_{0\le k\le K} \mathbb{E}[\|\nabla f(x_k)\|_2^2] \le -\frac{2}{\beta}\min_{0\le k\le K}\mathbb{E}[f(x_{k+1}) - f(x_k)]$$

$$= -\frac{2}{\beta}\cdot\frac{1}{K+1}\sum_{i=0}^{K} f(x_{i+1}) - f(x_i)$$

$$\le -\frac{2(f(x_{K+1}) - f(x_0))}{\beta(K+1)} = \frac{2(f(x_0) - f(x_{K+1}))}{\beta(K+1)}$$

$\ge f^*$

$$\le \frac{2(f(x_0)-f^*)}{\beta(k+1)}$$

3.

1. $$\nabla\varphi(y) = \frac{1}{2}\cdot 2\|y-x_0\|\cdot\frac{y-x_0}{\|y-x_0\|} = y-x_0.$$

$$D_\varphi(x,y) = \frac{1}{2}\|x-x_0\|^2 - \frac{1}{2}\|y-x_0\|^2 - \langle y-x_0, (x-x_0)-(y-x_0)\rangle$$

$$= \frac{1}{2}\|x-x_0\|^2 + \frac{1}{2}\|y-x_0\|^2 - \langle y-x_0, x-x_0\rangle$$

$$= \frac{1}{2}\|(x-x_0)-(y-y_0)\|_2^2 = \frac{1}{2}\|x-y\|^2.$$

2. $\phi(x) = \psi(x) + \langle z, x\rangle.$ $\quad \nabla\phi(y) = \nabla\psi(y) + z.$

$$D_\phi(x,y) = \phi(x)-\phi(y) - \langle\nabla\varphi(y)+z, x-y\rangle$$

$$= \psi(x)-\psi(y) + \underline{\langle z, x-y\rangle} - \langle\varphi(y), x-y\rangle \underline{- \langle z, x-y\rangle}$$

$$= D_\psi(x,y).$$

3. $$\text{RHS} = \underline{\psi(z)} - \psi(y) - \langle\nabla\psi(y), z-y\rangle + \underline{\langle\psi(z), x-z\rangle} - \langle\psi(y), x-z\rangle$$

$$+ \psi(x) - \underline{\psi(z)} - \underline{\langle\nabla\psi(z), x-z\rangle}$$

$$= \varphi(x) - \psi(y) - \langle\psi(y), x-y\rangle. = \text{LHS}.$$

4. $z, \bar{x}$ fixed.

$h(x) = \langle z, x\rangle + D_\varphi(x, \bar{x})$ $\qquad \nabla h(y) = z + \nabla_y D_\varphi(y, \bar{x})$

$$= z + \nabla\varphi(y) - \nabla\varphi(\bar{x})$$

Follow hint, $\langle\nabla h(y), x-y\rangle = \langle z + \nabla\varphi(y) - \nabla\varphi(\bar{x}), x-y\rangle \ge 0.$

$$\langle z, x\rangle \geq \langle z, y\rangle + \langle \nabla\varphi(\bar{x}) - \nabla\varphi(y), x-y\rangle$$

$$h(x) = \langle z, x\rangle + \varphi(x) - \varphi(\bar{x}) - \langle \nabla\varphi(\bar{x}), x-\bar{x}\rangle$$

$$\geq \langle z, y\rangle + \langle \nabla\varphi(\bar{x}) - \nabla\varphi(y), x-y\rangle + \varphi(x) - \varphi(\bar{x}) - \langle \nabla\varphi(\bar{x}), x-\bar{x}\rangle$$

$$= \langle z, y\rangle \underbrace{- \langle \nabla\varphi(\bar{x}), \bar{x}-y\rangle} \underbrace{- \langle \nabla\varphi(y), x-y\rangle + \big(\varphi(x) - \varphi(y)\big)} - \underbrace{\big(\varphi(\bar{x}) - \varphi(y)\big)}$$

$$= \langle z, y\rangle + D_\varphi(x, y) + D_\varphi(y, \bar{x})$$

$\square$.

4.

1. Follow HW I, Q2

$$\nabla h_z(x) = z + \nabla\left(\tfrac{1}{2}\|x\|_p^2\right)$$

$$h_z(y) - h_z(x) - \langle \nabla h_z(x), y-x\rangle = \langle z, y-x\rangle + \tfrac{1}{2}\|y\|_p^2 - \tfrac{1}{2}\|x\|_p^2 - \langle z + \nabla(\tfrac{1}{2}\|x\|_p^2), y-x\rangle$$

$$= \tfrac{1}{2}\|y\|_p^2 - \tfrac{1}{2}\|x\|_p^2 - \langle \nabla(\tfrac{1}{2}\|x\|_p^2), y-x\rangle$$

$$\geq \tfrac{1}{2}\|y\|_p^2 - \tfrac{1}{2}\|x\|_p^2 - \|x\|_p\|y-x\|_p \qquad \|x\|_p\left(\tfrac{1}{2}\|x\|_p + \|y-x\|_p\right)$$

$$\|y\|_p = \|x+y-x\|_p$$

(p-norm) $= \tfrac{1}{2}\|y-x\|^2 + \tfrac{1}{2}\|x\|^2 + \|x\|\|y-x\| - \tfrac{1}{2}\|x\|^2 - \|x\|\|y-x\| = \tfrac{1}{2}\|y-x\|^2 \geq 0.$

$\Rightarrow h_z$ is cvx.

Set $\nabla h_z(x^*) = 0$. $\quad \nabla(\tfrac{1}{2}\|x^*\|_p^2) + z = 0$.

By Fenchel's conjugate theory, $\quad x^* = -\nabla(\tfrac{1}{2}\|z\|_q^2)$.

*

$$\min_x h_z(x) = h_z(x^*) = \langle z, -\nabla(\tfrac{1}{2}\|z\|_q^2)\rangle + \tfrac{1}{2}\|-\nabla(\tfrac{1}{2}\|z\|_q^2)\|_p^2$$

$$= \langle \underbrace{z, -\nabla \tfrac{1}{2}\|z\|_q^2}_{\text{same direction}} \rangle + \tfrac{1}{2}\|z\|_q^2$$

$$= -\|z\|_q \underbrace{\|\nabla(\tfrac{1}{2}\|z\|_q^2)\|_p}_{\|z\|_q} + \tfrac{1}{2}\|z\|_q^2 = -\tfrac{1}{2}\|z\|_q^2 .$$

2. Let $s = u - x_k$.

$h(s) = \langle \nabla f(x_k), s \rangle + \frac{L}{2}\|s\|_p^2$.

Apply part 1, $\Rightarrow h(s)_{min} = -\frac{1}{2L}\|\nabla f(x_k)\|_q^2$

L-smooth,

$$f(x_{k+1}) \le f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_p^2$$

$$\text{RHS min} = -\frac{1}{2L}\|\nabla f(x_k)\|_q^2 + f(x_k). \Rightarrow f(x_{k+1}) - f(x_k) \le -\frac{1}{2L}\|\nabla f(x_k)\|_q^2 .$$

$$\sum_{i=0}^{K} f(x_{i+1}) - f(x_i) = f(x_{K+1}) - f(x_0) \le -\frac{1}{2L}\sum_{i=0}^{K}\|\nabla f(x_k)\|_q^2 \le -\frac{K+1}{2L}\min_{i\in[K]}\|\nabla f(x_i)\|_q^2$$

$$\min_{i\in[K]}\|\nabla f(x_i)\|_q^2 \le \frac{2L(f(x_0) - f(x_{K+1}))}{K+1} \le \frac{2L(f(x_0) - f^*)}{K+1}.$$

$$\min_{i\in[K]}\|\nabla f(x_i)\|_q \le \sqrt{\frac{2L(f(x_0) - f^*)}{K+1}}.$$

Similar res.

3. $L_2 = L$ because $p=2, q=2$ is just a special case.

$$\sqrt{\frac{2L(f(x_0) - f^*)}{K+1}}.$$