HW1 (a) : $f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \| x_2 - x_1 \|_2^2$

equivalent

$* \; g(x) = f(x) - \frac{\mu}{2} \| x \|^2$ is convex

from first-order condition for convexity of $g(x)$:

$g(x)$ is convex if and only if: $g(x_2) \geq g(x_1) + \nabla g(x_1)^T (x_2 - x_1)$

$$\equiv f(x_2) - \frac{\mu}{2} \| x_2 \|^2 \geq f(x_1) - \frac{\mu}{2} \| x_1 \|^2 + \nabla \left(f(x_1) - \frac{\mu}{2} \| x_1 \|^2 \right)^T (x_2 - x_1)$$

$$\Rightarrow f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \| x_2 - x_1 \|^2$$

State ①

from second-order condition for convexity

$\nabla^2 f(x) \succeq 0 \qquad g(x) = f(x) - \frac{\mu}{2} \| x \|^2$

$\nabla^2 \left(f(x) - \frac{\mu}{2} \| x \|^2 \right) \succeq 0 \qquad \nabla^2 f(x) - \mu I_d \succeq 0 \Rightarrow \nabla^2 f(x) \succeq \mu I_d$

State ② : It is monotone gradient condition for convexity of $g(x)$; $g(x)$ is convex if and only if :

$$\left(\nabla g(x) - \nabla g(y) \right)^T (x - y) \geq 0 \quad \forall x, y$$

$$\left(\nabla \left[f(x_1) - \frac{\mu}{2} \| x_1 \|^2 \right] - \nabla \left[f(x_2) - \frac{\mu}{2} \| x_2 \|^2 \right] \right)^T (x_1 - x_2) \geq 0$$

$$\left(\nabla f(x_1) - \nabla f(x_2) - \mu \| x_1 \|^2 + \mu \| x_2 \|^2 \right)^T (x_1 - x_2) \geq 0$$

$$\left(\nabla f(x_2) - \nabla f(x_1) \right)^T (x_2 - x_1) \geq \mu \| x_2 - x_1 \|_2^2$$

Hw 1 (a)

part a)

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \| x_2 - x_1 \|_2^2$$

$$f(x^*) \geq f(x) + \nabla f(x) (x^* - x) + \frac{\mu}{2} \| x^* - x \|_2^2$$

$$f(x) - f^* \leq \nabla f(x) (x - x^*) - \frac{\mu}{2} \| x^* - x \|_2^2$$

$$\leq -\frac{1}{2} \left[-2 \nabla f(x) (x - x^*) + \mu (x^* - x)^T (x^* - x) + \frac{1}{\mu} \| \nabla f \|_2^2 \right] + \frac{1}{2\mu} \| \nabla f \|_2^2$$

$$= \underbrace{-\frac{1}{2} \left\| \sqrt{\mu} (x - x^*) + \frac{1}{\sqrt{\mu}} \nabla f(x) \right\|_2^2}_{\text{always positive}} + \frac{1}{2\mu} \| \nabla f \|_2^2$$

$$\leq \frac{1}{2\mu} \| \nabla f \|_2^2$$

Hw 1 (a)

part b) $(\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \geq \mu \| x_2 - x_1 \|_2^2$

using Cachy - schwertz :

$$\mu \| x_2 - x_1 \|_2^2 \leq (\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \leq \| \nabla f(x_2) - \nabla f(x_1) \| \, \| x_2 - x_1 \|$$

$$\Rightarrow \frac{1}{\mu} \| \nabla f(x_2) - \nabla f(x_1) \| \geq \| x_2 - x_1 \|$$

HW 1(a)
part C)

we consider $\varphi_x(z) = f(z) - \nabla f(x)^T z$ that $\varphi_x(z)$ is strongly convex with the same μ since:

From equivalent 2 $\longrightarrow (\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \geq \mu \| x_2 - x_1 \|_2^2$

$$(\nabla \varphi_x(z_1) - \nabla \varphi_x(z_2))^T (z_1 - z_2) = (\nabla f(z_1) - \nabla f(z_2))(z_1 - z_2) \geq \mu \| z_1 - z_2 \|$$

by using a) $\longrightarrow$ $f(x) - p^* \leq \frac{1}{2\mu} \| \nabla f(x) \|_2^2 \quad \forall x$

by applying a) to the $\varphi_x(z)$ that is strongly convex:

* with $z^* = x_1$

$$\left(f(x_2) - \nabla f(x_1)^T x_2\right) - \left(f(x_1) - \nabla f(x_1)^T x_1\right) = \varphi_x(x_2) - \varphi_x(x_1)$$

$$\leq \frac{1}{2\mu} \| \nabla \varphi_x(x_2) \|^2 = \frac{1}{2\mu} \| \nabla f(x_2) - \nabla f(x_1) \|_2^2$$

$$f(x_2) - f(x_1) + \nabla f(x_1)^T (x_1 - x_2) \leq \frac{1}{2\mu} \| \nabla f(x_2) - \nabla f(x_1) \|_2^2$$

$$* \; f(x_2) \leq f(x_1) + \nabla f(x_1)(x_2 - x_1) + \frac{1}{2\mu} \| \nabla f(x_2) - \nabla f(x_1) \|_2^2$$

interchanging x_2 & x_1 in *

$$** \; f(x_1) \leq f(x_2) + \nabla f(x_2)(x_1 - x_2) + \frac{1}{2\mu} \| \nabla f(x_1) - \nabla f(x_2) \|_2^2$$

Sum * & ** $\quad \cancel{f(x_1)} + \cancel{f(x_2)} \leq \cancel{f(x_1)} + \cancel{f(x_2)} - (x_2 - x_1)\left[\nabla f(x_1) + \nabla f(x_2)\right]$
$$+ \frac{1}{2\mu} \| \nabla f(x_1) - \nabla f(x_2) \|^2$$

$$\left[\nabla f(x_2) - \nabla f(x_1)\right]^T (x_2 - x_1) \leq \frac{1}{\mu} \| \nabla f(x_1) - \nabla f(x_2) \|^2 \quad \forall x_1, x_2$$

HW 1 (a)

Part d) $f(x) + r(x)$ is strongly convex for any convex f and strongly convex r.

$$h(x) = f(x) + r(x)$$

$$f \text{ is convex} \xrightarrow[\text{condition}]{\text{firs order}} f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1)$$

$$r \text{ is strongly convex} \longrightarrow r(x_2) \geq r(x_1) + \nabla r(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \| x_2 - x_1 \|_2^2$$

Sum $\Rightarrow$

$$f(x_2) + r(x_2) \geq f(x_1) + r(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \nabla r(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \| x_2 - x_1 \|_2^2$$

$$h(x_2) \geq h(x_1) + (x_2 - x_1)\left(\nabla f(x_1) + \nabla r(x_1)\right)^T + \frac{\mu}{2} \| x_2 - x_1 \|_2^2$$

$$h(x_2) \geq h(x_1) + \nabla h(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \| x_2 - x_1 \|_2^2$$

So $h(x)$ is strongly convex. $\square$

4

H.7.b)

Part a) $f(x_2) \leq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{L}{2} \|x_2 - x_1\|_2^2 \quad \forall x_1, x_2$

① strong convexity $\|\nabla f(x_2) - \nabla f(x_1)\|_2 \geq \mu \|x_2 - x_1\|_2$

L-Smooth : $\|\nabla f(x_2) - \nabla f(x_1)\|_2 \leq L \|x_2 - x_1\|_2$

$\Rightarrow \mu \|x_2 - x_1\| \leq \|\nabla f(x_2) - \nabla f(x_1)\|_2 \leq L \|x_2 - x_1\|_2$

$\mu \leq L$

$f(x_2) \geq f(x_1) + \nabla f^T(x_1)(x_2 - x_1) + \frac{\mu}{2} \|x_2 - x_1\|_2^2$

$L \geq \mu$

$f(x_2) \leq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{L}{2} \|x_2 - x_1\|_2^2$

② $g(x) = \frac{L}{2} x^T x - f(x)$ convex $\forall x$

first order condition:

$f(x_2) \geq f(x_1) + \nabla f^T(x_1)(x_2 - x_1)$

$g(x_2) \geq g(x_1) + \nabla g(x_1)(x_2 - x_1)$

$\frac{L}{2} x_2^T x_2 - f(x_2) \geq \frac{L}{2} x_1^T x_1 - f(x_1) - \nabla f(x_1)(x_2 - x_1)$

$\Rightarrow f(x_2) \leq f(x_1) + \nabla f(x_1)(x_2 - x_1) + \frac{L}{2} \|x_2 - x_1\|_2^2$

H1 - b)

b) $f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{1}{2L}\|\nabla f(x_2) - \nabla f(x_1)\|_2^2 \quad \forall x_1, x_2$

Define: $y = x_2 - \frac{1}{L}\left(\nabla f(x_2) - \nabla f(x_1)\right)$ $\Rightarrow x_2 - x_1 \geq y - x_1$

$y - x_1 = x_2 - x_1 - \frac{1}{L}\left(\nabla f(x_2) - \nabla f(x_1)\right)$

$f(x_2) - f(x_1) = f(y) - f(x_1) - [f(y) - f(x_2)] \geq \nabla f(x_1)^T (y - x_1) - [\nabla f(x_2)^T (y - x_2)$
$+ \frac{L}{2}\|y - x_2\|_2^2]$

$-f(y) + f(x_2) \geq -\nabla f(x_2)^T (y - x_2) - \left(\frac{L}{2}\|y - x_2\|_2^2\right)$ (a)

$f(y) - f(x_1) \geq \nabla f^T(x_1)(y - x_1)$ (b)

Sum (a),(b) $\Rightarrow f(x_2) - f(x_1) \geq -\left[\nabla f(x_2^T)(y - x_2)\right] - \left[\frac{L}{2}\|y - x_2\|_2^2\right) + \nabla f(x_1)(y - x_1)$

$f(x_2) - f(x_1) \geq \nabla f^T(x_2) x_2 - \nabla f^T(x_1) x_2 - [\nabla f(x_2)(y - x_2)] - \frac{L}{2}\|y - x_2\|_2^2$

$+ \nabla f^T(x_1)(y - x_1) = \nabla f^T(x_1)(x_2 - x_1) + \nabla f(x_1)(y - x_2) - \nabla f^T(x_2)(y - x_2)$

$+ \nabla f^T(x_1)(x_2 - x_1) + \left[\nabla f^T(x_1) - \nabla f^T(x_2)\right](y - x_2) - \frac{L}{2}\|y - x_2\|_2^2$

$\Rightarrow f(x_2) - f(x_1) \geq \nabla f(x_1)(x_2 - x_1) + \frac{1}{L}\|\nabla f(x_1) - \nabla f(x_2)\|_2^2$
$- \frac{L}{2} \times \frac{1}{L^2}\|\nabla f(x_1) - \nabla f(x_2)\|^2$

$\Rightarrow f(x_2) - f(x_1) \geq \nabla f^T(x_2 - x_1) + \frac{1}{2L}\|\nabla f(x_1) - \nabla f(x_2)\|_2^2$

H1-b)

part c) $(\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \geq \frac{1}{L} \|\nabla f(x_2) - \nabla f(x_1)\|_2^2 \quad \forall x_1, x_2$

from part(b) $\rightarrow$ $f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{1}{2L} \|x_2 - x_1\|_2^2$

$$f(x_1) \geq f(x_2) + \nabla f(x_2)^T (x_1 - x_2) + \frac{1}{2L} \|x_1 - x_2\|_2^2$$

Sum: $\cancel{f(x_2) + f(x_1)} \geq \cancel{f(x_1) + f(x_2)} + \nabla f(x_1)^T (x_2 - x_1) + \nabla f(x_2)^T (x_1 - x_2)$

$$+ \frac{1}{2L} \|x_2 - x_1\|^2 + \frac{1}{2L} \|x_1 - x_2\|_2^2$$

$$[\nabla f(x_2) - \nabla f(x_1)] (x_2 - x_1) \geq \frac{1}{2L} \|x_2 - x_1\|^2 + \frac{1}{2L} \|x_1 - x_2\|_2^2$$

$$[\nabla f(x_2) - \nabla f(x_1)] (x_2 - x_1) \geq \frac{1}{L} \|x_2 - x_1\|^2 \geq \frac{1}{L} \|\nabla f(x_2) - \nabla f(x_1)\|_2^2$$

$$\|\nabla f(x_2) - \nabla f(x_1)\|_2^2 \leq L^2 \|x_2 - x_1\|_2^2$$

$$\|x_2 - x_1\|_2^2 \geq \frac{\|\nabla f(x_2) - \nabla f(x_1)\|_2^2}{L^2}$$

HW1 (C): use Lagrange dual function, λ is the dual variable. so, the Lagrangian function can be shown as:

$$= \mathcal{L}(x,\lambda) = \frac{1}{N}\sum_{i\in N} f_i(x_i) + \lambda(b - Ax) = \frac{1}{N}\sum f_i(x_i) + \lambda(b_1 - a_1x_1 - a_2x_2 \cdots - a_Nx_N)$$

→ Now, we have unconstrained problem & can use descent methods:

$$\nabla^2 \mathcal{L} = \frac{1}{N}\begin{bmatrix} \frac{\partial^2 f_1}{\partial x_1^2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{\partial^2 f_2}{\partial x_2^2} & 0 & \cdots & 0 \\ 0 & & & & \vdots \\ \vdots & & & & \frac{\partial^2 f_N}{\partial x_N^2} \\ 0 & \cdots & & & \end{bmatrix} \quad \nabla \mathcal{L} = \frac{1}{N}\begin{bmatrix} \frac{\partial f_1}{\partial x_1} - \lambda a_1 \\ \frac{\partial f_2}{\partial x_2} - \lambda a_2 \\ \vdots \\ \frac{\partial f_N}{\partial x_N} - \lambda a_N \end{bmatrix} \leftarrow p=1$$

part (a) The Gradient Descent method will be used, because N is small.

part (b) Finding a good coordinate for GD is usually very hard in high-dimension! Therefore, GD is not useful for $N = 10^9$. In this case we use Newton. since Hessian is diagonal, finding the inverse of it, is easy!

part (c) $\boxed{\text{for the case } p=1,\ N=10^9}$ → yes we can use Newton because the hessian is a diagonal matrix and finding the inverse of a diagonal matrix is easy

$\boxed{\text{for the case } 1<p\ll N,\ N=10^9}$ → Again we use Newton, because Hessian is diagonal & in below form:

$$\nabla^2 = \frac{1}{N}\begin{bmatrix} \frac{\partial^2 f_1}{\partial x_1^2} & 0 & \cdots & \\ 0 & \ddots & & \\ \vdots & & & \frac{\partial^2 f_N}{\partial x_N^2} \end{bmatrix} - \begin{bmatrix} \lambda a_{11} + \cdots + \lambda_p a_{p1} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & \lambda a_{1N} + \cdots + \lambda_p a_{pN} \end{bmatrix}$$

$\lambda = [\lambda_1, \ldots, \lambda_p]$

↳ dual variable

so, because Hessian is diagonal, finding its inverse is easy & we use Newton.

part

(d)

(d) → for part (a) $P=1, \ N=1000$

$$L(x,\lambda)=\frac{1}{N}\sum_{i\in N} f_i(x_i)+r(x)+\lambda(b-Ax)$$

$$\nabla=\frac{1}{N}\begin{bmatrix}\frac{\partial f_1}{\partial x_1}+\frac{\partial r}{\partial x_1}-\lambda a_1\\ \frac{\partial f_2}{\partial x_2}+\frac{\partial r}{\partial x_2}-\lambda a_2\\ \vdots\\ \frac{\partial f_N}{\partial x_N}+\frac{\partial r}{\partial x_N}-\lambda a_N\end{bmatrix}$$

$\Rightarrow$ so, we can compute ∇, and we can use GD again.

(d) → for part (b)

$N=1e^9, \ P=1$

$$\nabla^2=\frac{1}{N}\begin{bmatrix}\frac{\partial^2 f_1}{\partial x_1^2} & 0 & 0 & \cdots & 0\\ 0 & \frac{\partial^2 f_2}{\partial x_2^2} & & & 0\\ \vdots & & \ddots & & \\ 0 & 0 & & & \frac{\partial^2 f_N}{\partial x_N^2}\end{bmatrix}+\begin{bmatrix}\frac{\partial^2 r}{\partial x_1^2} & \frac{\partial^2 r}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 r}{\partial x_1\partial x_N}\\ \frac{\partial^2 r}{\partial x_2\partial x_1} & \frac{\partial^2 r}{\partial x_2^2} & & \vdots\\ \vdots & & \ddots & \\ & & \cdots & \frac{\partial^2 r}{\partial x_N^2}\end{bmatrix}$$

Because ∇^2 is a Hermitian and positive-definite matrix, so we can use cholesky decomposition / decompose it into a lower triangular matrix & its conjugate transpose. so we can find the inverse of it, easily $\Rightarrow$ we use Newton.

(d) → for part c:

Yes, we use Newton.

As described in previous section we can use cholesky decomposition.

Then we have low rank matrix & diagonal matrix, which is easy to find its inverse.

In $\boxed{P=1}$ the closed form of inverse matrix can be found & in $K\,P \ll 1$, still it has a good structure, because it is low rank, so the inverse can be found.

H1.d

prove: $(\nabla f(x) - \nabla f(y))^T (x-y) \geq \frac{\mu L}{\mu+L} \|x-y\|_2^2 + \frac{1}{\mu+L} \|\nabla f(x) - \nabla f(y)\|_2^2$

f : L smooth & μ strongly convex

g : $f(x) - \frac{\mu}{2}\|x\|^2$

$\mu \leq \nabla^2 g - \nabla^2 f \leq L$

$0 \leq \nabla^2 g - \nabla^2 f - \mu \leq L-\mu \quad \Rightarrow g \quad L-\mu$ smooth

$(\nabla g_y - \nabla g_x)^T (y-x) \geq \frac{1}{L-\mu} \|\nabla g_y - \nabla g_x\|^2$

$(\nabla f_y - \nabla f_x - \mu(y-x))^T (y-x) \geq \frac{1}{L-\mu} \|\nabla f_y - \nabla f_x - \mu(y-x)\|^2$

$(L+\mu)(\nabla f_y - \nabla f_x)^T (y-x) \geq \|\nabla f_y - \nabla f_x\|^2 + \underbrace{(\mu^2 + \mu(L-\mu))}_{\mu L}\|y-x\|^2$

84