

1.

a) False, for $y=|x|$, there is no gradient at global minimizer $x=0$

b) False, since $\text{prox}_f(\bar{x}) = \arg\min_{x \in \mathbb{R}^d} \{f(x) + \frac{1}{2}\|x - \bar{x}\|_2^2\}$

if $f(x) = -\frac{1}{2}\|x - \bar{x}\|_2^2$, then $\forall x \in \mathbb{R}^d$ could be $\text{prox}_f(\bar{x})$

c) False x^3 , at $x=0 \quad \nabla^2 f(x) = 0 \geq 0 \quad \nabla f(x) = 0$, but

d) False $e^{-x} > 0$, and it is a convex, but it doesn't have a local minimizer
it is not a minimizer

e) True. since $\rho(\bar{x} - \bar{x}^*) \in \partial f(\bar{x}^*)$

then, $f(x) \geq f(\bar{x}^*) + g^T(x - \bar{x}^*)$ for all $x \in \mathbb{R}^d$

$$f(x) \geq f(\bar{x}^*) + \rho(\bar{x} - \bar{x}^*)^T(x - \bar{x}^*)$$

$$f(x) - f(\bar{x}^*) \geq \rho(x \cdot \bar{x} - x \cdot \bar{x}^* - \bar{x} \cdot \bar{x}^* + \bar{x}^* \cdot \bar{x})$$

We want:

$$f(x) + \frac{\rho}{2}(x - \bar{x})^T(x - \bar{x}) \geq f(\bar{x}^*) + \frac{\rho}{2}(\bar{x} - \bar{x}^*)^T(\bar{x} - \bar{x}^*)$$

to make life easier, denote \bar{x} as b \bar{x}^* as c x as a

We want:

$$f(a) + \frac{\rho}{2}(a - b)^T(a - b) \geq f(c) + \frac{\rho}{2}(c - b)^T(c - b)$$

then, we need $f(a) - f(c) \geq \frac{\rho}{2}(c^2 - bc + b^2) - \frac{\rho}{2}(a^2 - ab + b^2) = \rho\left(\frac{1}{2}c^2 - b(c+a) - \frac{1}{2}a^2\right) \quad (1)$

since we have $f(a) - f(c) \geq \rho(b - c)(a - c)$

$$\text{Since } ab - ac - bc + c^2 - \frac{1}{2}c^2 + bc - ab + \frac{1}{2}a^2 = \frac{1}{2}a^2 - ac + \frac{c^2}{2} \geq 0$$

so (1) holds, so \bar{x}^* is the global minimizer

$$Q_2 \text{ let } g(x) = \frac{1}{2a} \|x\|_2^2 = \frac{1}{2a} x^T x$$

Since f has L -Lipschitz gradient, $\Rightarrow \|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|, \forall x, y$

a) $\nabla g(x) = \frac{1}{a} x \text{ so } \|\nabla g(x) - \nabla g(y)\| = \frac{1}{a}\|x-y\|$

So $g(x)$ has $\frac{1}{a}$ -Lipschitz gradient

let $f(x) + g(x) = (f+g)(x)$, we have $\nabla(f+g)(x) = \nabla f(x) + \nabla g(x) \leq L\|x\| + \frac{1}{a}\|x\|$

so $f(x) + \frac{1}{2a}\|x\|_2^2$ has $(\frac{1}{a} + L)$ -Lipschitz continuous gradient.

b) $f(x) + \frac{1}{2a}\|x\|_2^2 = (f+g)(x) \quad \checkmark 9.16 PB. lemma(ii)$

$(f+g)(x)$ is $(\frac{1}{a} - L)$ -strongly convex $\Leftrightarrow (f+g)(y) \geq (f+g)(x) + \nabla(f+g)(x)^T(y-x) + (\frac{1}{2a} - \frac{L}{2})\|y-x\|_2^2$
if $a < \frac{1}{L}$ for $\forall x, y \in \mathbb{R}^d$.

So we need to prove:

$$f(y) + \frac{1}{2a} y^T y \geq f(x) + \frac{1}{2a} x^T x + \left(\nabla f(x) + \frac{1}{a} x\right)^T(y-x) + \left(\frac{1}{2a} - \frac{L}{2}\right)\|y-x\|^2 \dots \dots \quad (1)$$

since f has L -Lipschitz gradient, $\Rightarrow \|f(y) - f(x) - \nabla f(x)^T(y-x)\| \leq \frac{L}{2}\|y-x\|^2$

so we can have $f(y) - f(x) - \nabla f(x)^T(y-x) \geq -\frac{L}{2}\|y-x\|^2$ terms in (1)

now we need to prove $\frac{1}{2a} y^T y \geq \frac{1}{2a} x^T x + \frac{1}{a} (y-x)^T x + \frac{1}{2a} (y-x)^T (y-x)$

multiply $2a$ on both sides we need $y^T y \geq x^T x + 2x^T y - 2x^T x + y^T y - 2x^T y + x^T x$

~~$y^T y \geq x^T x + 2x^T y - 2x^T x + y^T y - 2x^T y + x^T x$~~ so left hand side = right hand side.

so (1) holds for $\forall x, y \in \mathbb{R}^d$: Hence $f(x) + \frac{1}{2a}\|x\|_2^2$ is $(\frac{1}{2} - L)$ -strongly convex if $a < 1/L$

c) in HW3. we proved strongly convex function f has a unique minimizer

$$\text{prox}_{af}(\bar{x}) = \arg \min_x \left\{ f(x) + \frac{1}{2a} \|x - \bar{x}\|_2^2 \right\} = \arg \min_x \left\{ f(x) + \frac{1}{2a} x^T x - \frac{1}{a} x^T \bar{x} + \frac{1}{2a} \bar{x}^T \bar{x} \right\}$$

$\frac{1}{2a} \bar{x}^T \bar{x}$ is a constant,

for $-\frac{1}{a} x^T \bar{x}$: since $-\frac{1}{a} y^T \bar{x} = -\frac{1}{a} x^T \bar{x} + (-\frac{1}{a} \bar{x})^T (y-x) + \frac{M}{2} \|y-x\|^2$, so $M=0$

so $\frac{1}{2a} \bar{x}^T \bar{x}$ is a 0-strongly convex,

so $f(x) + \frac{1}{2a} x^T x - \frac{1}{a} x^T \bar{x} + \frac{1}{2a} \bar{x}^T \bar{x}$ is a $(\frac{1}{a} - L)$ -strongly convex when $a < 1/L$

then, $\text{prox}_{af}(\bar{x})$ is a singleton for every $a < 1/L$ and $\bar{x} \in \mathbb{R}^d$

d) define a new operator: inexact proximal operator: $\text{Prox}_{\lambda h}^{\epsilon}(y)$

$$x \in \text{Prox}_{\lambda h}^{\epsilon}(y) = \left\{ z \in \mathbb{R}^N : \frac{1}{2\lambda} \|z - y\|^2 + h(z) \leq \epsilon + \min_x \frac{1}{2\lambda} \|x - y\|^2 + h(x) \right\}$$

ϵ is error, we can set it to a small constant, everytime we just need to find an $x \in \text{Prox}_{\lambda f}^{\epsilon}(y)$ if we can't, just make ϵ_k bigger.

algorithm: $f = f(x)$ in the problem $h(x) = \frac{1}{2} \|x\|_2^2$

input m , error ϵ_k ($k=1 \dots m$), $t_0=0$, $t_1=1$, stepsize $\lambda < \frac{1}{\sum}$

output X_m .

initialize $x_0 \in \mathbb{R}^d$ and $x_1 = z_1 = v_0$

AIPG method.

for $k=1, 2, \dots, m$ do

$$y_k = x_k + \frac{t_{k-1}}{t_k} (z_k - x_k) + \frac{t_{k-1}-1}{t_k} (x_k - x_{k-1})$$

compute z_{k+1} that $z_{k+1} \in \text{Prox}_{\lambda h}^{\epsilon_k}(y_k - \lambda \nabla f(y_k))$

compute v_{k+1} that $v_{k+1} \in \text{Prox}_{\lambda h}^{\epsilon_k}(x_k - \lambda \nabla f(x_k))$

$$t_{k+1} = \sqrt{4t_k^2 + 1} + 1$$

$$x_{k+1} = \begin{cases} z_{k+1} & \text{if } f(z_{k+1}) \leq f(v_{k+1}) \\ v_{k+1} & \text{otherwise} \end{cases}$$

from Bi.G, De.w

"Inexact Proximal Gradient Methods for Non-convex and Non-smooth optimization"

convergence rate: $O(\frac{1}{T^2})$

$\text{Prox}_{\lambda h}^{\epsilon}(x)$ is an approximation of $\text{Prox}_{\lambda h}(x)$

Q3

a) if x^* is a global minimizer of f A

$$0 \in A_i^T(Ax - b) + \begin{cases} \{r\} & \text{if } x_i > 0 \\ [-r, r] & \text{if } x_i = 0 \\ \{-r\} & \text{if } x_i < 0 \end{cases} \dots B$$

$$\partial \frac{1}{2} \|Ax - b\|_2^2 = A^T(Ax - b)$$

$$\partial \|x\|_1 = \begin{cases} 1 & \text{if } x_i > 0 \\ [-1, 1] & \text{if } x_i = 0 \\ -1 & \text{if } x_i < 0 \end{cases}$$

proof of $A \Rightarrow B$

since x^* is a global minimizer, then $f(y) \geq f(x^*)$ for $\forall y \in \mathbb{R}^d$

so $f(y) \geq f(x^*) + 0^T(y - x^*)$ for $\forall y \in \mathbb{R}^d$

$$\text{so } 0 \in \partial f(x^*) \Rightarrow 0 \in A_i^T(Ax^* - b) + \begin{cases} \{r\} & \text{if } x_i > 0 \\ [-r, r] & \text{if } x_i = 0 \\ \{-r\} & \text{if } x_i < 0 \end{cases} \quad (\text{10.b. lemma}) \quad (\text{sum rule})$$

proof of $B \Rightarrow A$

$$\text{if } 0 \in A_i^T(Ax^* - b) + \begin{cases} \{r\} & \text{if } x_i > 0 \\ [-r, r] & \text{if } x_i = 0 \\ \{-r\} & \text{if } x_i < 0 \end{cases}$$

then, $f(y) \geq f(x^*) + 0^T(y - x^*) \Rightarrow f(y) \geq f(x^*)$ for $\forall y \in \mathbb{R}^d$.

so x^* is a global minimizer.

Q4

a) in Hw3 Q2a we have $f(x) \geq f(\bar{x}) - \frac{1}{2L} \|\nabla f(\bar{x})\|_2^2$
 for all $f(x) \leq f(\bar{x})$ and $f(x)$ is a $L > 0$ -strongly convex

to make the statement in the same form, let's replace x with \bar{x}

then, we need to prove, $\frac{1}{2} \|\nabla f(\bar{x})\|_2^2 \geq L(f(\bar{x}) - \min_x f(x))$

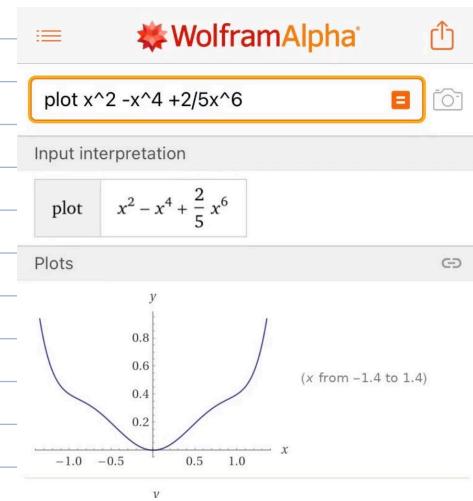
since $\min_{x'} f(x') \leq f(\bar{x})$ so $x' \in X$ that $f(x) \leq f(\bar{x})$

so $f(x') \geq f(\bar{x}) - \frac{1}{2L} \|\nabla f(\bar{x})\|_2^2 \Rightarrow \frac{1}{2} \|\nabla f(\bar{x})\|_2^2 \geq L(f(\bar{x}) - \min_x f(x'))$

Hence, (a) holds for every L -strongly convex f

$$b) x^2 - x^4 + \frac{2}{5}x^6$$

$$\text{let } L = \frac{1}{100}$$



c). from class, we have $f(x_{k+1}) \leq f(x_k) - \frac{1}{2L} \|\nabla f(x_k)\|_2^2$

from (a), we have $\frac{1}{2} \|\nabla f(x)\|_2^2 \geq L(f(x) - \min_x f(x))$

so $f(x_{k+1}) \leq f(x_k) - \frac{L}{L} (f(x_k) - f(x^*))$ $x^* = \underset{x}{\operatorname{argmin}} f(x)$

so $f(x_{k+1}) - f(x^*) \leq f(x_k) - f(x^*) - \frac{L}{L} (f(x_k) - f(x^*)) = (1 - \frac{L}{L}) (f(x_k) - f(x^*))$

since $f(x_{k+1}) - f(x^*) \leq (1 - \frac{L}{L}) (f(x_k) - f(x^*))$

so $f(x_{k+1}) - \min_x f(x) \leq (1 - \frac{L}{L})^{k+1} (f(x_0) - \min_x f(x))$

$f(x_k) - \min_x f(x) \leq (1 - \frac{L}{L})^k (f(x_0) - \min_x f(x))$

(4d.) since $f(x_{k+1}) \leq f(x_k) - \frac{1}{2L} \|\nabla f(x_k)\|_2^2$
 and we have $\|\nabla f(x)\|_2 \geq M(f(x) - f(x^*))$ $x^* = \arg \min_x f(x)$
 so $f(x_{k+1}) \leq f(x_k) - \frac{M^2}{2L} (f(x_k)^2 - 2f(x_k)f(x^*) + f(x^*)^2)$

then, we can have

$$f(x_k) \leq f(x_{k-1}) - \frac{M^2}{2L} (f(x_{k-1}) - f(x^*))^2$$

$$\text{so } f(x_k) - f(x^*) \leq f(x_{k-1}) - f(x^*) - \frac{M^2}{2L} (f(x_{k-1}) - f(x^*))^2$$

$$\text{let } f(x_k) - f(x^*) = A_k \text{ so } A_k \leq A_{k-1} - \frac{M^2}{2L} (A_{k-1})^2$$

$$\text{since } f(x_k) - f(x^*) \leq f(x_{k-1}) - f(x^*)$$

$$\text{so } A_k \leq A_{k-1} - \frac{M^2}{2L} A_k \cdot A_{k-1}$$

so we can have

$$\frac{1}{A_k} \leq \frac{1}{A_{k-1}} - \frac{M^2}{2L}$$

$$\text{so } \frac{1}{A_k} \geq \frac{1}{A_{k-1}} + \frac{M^2}{2L} \geq \frac{1}{A_{k-2}} + \frac{M^2}{2L} + \frac{M^2}{2L} \dots \geq \frac{1}{A_0} + \frac{kM^2}{2L} \geq \frac{kM^2}{2L}$$

$$\text{so } A_k \leq \frac{2L}{kM^2}$$

$$\text{Hence, } f(x_k) - f(x^*) \leq \frac{2L}{kM^2}$$

when $k \rightarrow \infty$ $f(x_k) = \min f(x)$ is true

5a) Since f is a convex
then $f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{L}{2} \|y-x\|^2 \quad \forall x, y$

$$\text{So } f(y) \leq f(x) + \frac{L}{2} ((y-x)^T(y-x) + \frac{2}{L} (y-x)^T \nabla f(x) + \frac{1}{L^2} \|\nabla f(x)\|^2)$$

$$\text{So } f(y) \leq f(x) - \frac{1}{2L} \|\nabla f(x)\|_2^2 - \frac{L}{2} \cdot \frac{1}{L^2} \|\nabla f(x)\|^2$$

$$\text{So } \frac{1}{2L} \|\nabla f(y_k)\|_2^2 \leq f(y_k) - f(x^*)$$

$$\text{and } f(y_k) - f(x^*) \leq \frac{1}{k^2} \cdot 2L \cdot \|x_0 - x^*\|_2^2$$

$$\text{So } \|\nabla f(y_k)\|_2^2 \leq \frac{4L^2}{k^2} \|x_0 - x^*\|_2^2$$

$$\text{So } \|\nabla f(y_k)\|_2 \leq \frac{2L}{k} \|x_0 - x^*\|_2$$

5b) from (a) we have $f(y) \leq f(x) - \frac{1}{2L} \|\nabla f(x)\|_2^2$

$$\text{So } f(\bar{x}_{k+1}) \leq f(\bar{x}_k) - \frac{1}{2L} \|\nabla f(\bar{x}_k)\|_2^2$$

$$f(\bar{x}_k) \leq f(\bar{x}_{k-1}) - \frac{1}{2L} \|\nabla f(\bar{x}_{k-1})\|_2^2$$

$$f(\bar{x}_2) \leq f(\bar{x}_1) - \frac{1}{2L} \|\nabla f(\bar{x}_1)\|_2^2$$

$$f(\bar{x}_1) \leq f(\bar{x}_0) - \frac{1}{2L} \|\nabla f(\bar{x}_0)\|_2^2$$

$$\text{So } \frac{1}{2L} \sum_{i=0}^{k-1} \|\nabla f(\bar{x}_i)\|_2^2 \leq f(\bar{x}_0) - f(\bar{x}_k) = f(y_k) - f(x^*) \leq f(\bar{y}_k) - f(x^*) \leq \frac{2L \|\bar{x}_0 - x^*\|^2}{k^2}$$

$$\text{So } \frac{1}{2L} k \cdot \min_{0 \leq i \leq k-1} \|\nabla f(\bar{x}_i)\|_2^2 \leq \frac{2L \|\bar{x}_0 - x^*\|^2}{k^2}$$

$$\Rightarrow \min_{0 \leq i \leq k-1} \|\nabla f(\bar{x}_i)\|_2^2 \leq \frac{4L^2 \|\bar{x}_0 - x^*\|^2}{k^3} C$$

$$\text{So } \exists \bar{x}_i \text{ such that } \|\nabla f(\bar{x}_i)\|_2 \leq \frac{C}{k^{\frac{3}{2}}} \quad C = 2L \|\bar{x}_0 - x^*\|$$