1.

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

b) False, since $prox_f(\bar{x}) = \arg\min\{f(x)+\frac{1}{2}\|x-\bar{x}\|_2^2\}$ $x\in R^d$

if $f(x) = -\frac{1}{2}\|x-\bar{x}\|_2^2$, then $Prox_f(\bar{x})=0$, $\forall x\in R^d$ could be $prox_f(\bar{x})$

c) False x^3, at $x=0$ $\nabla^2 f(x) = 0 \succeq 0$ $\nabla f(x)=0$, but it is not a minimizer

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

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

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

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:

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

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

Since we have $f(x)-f(\bar{x}^*) \geq \rho(b-c)(a-c)$

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

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

a) $\nabla g(x) = \frac{1}{a}x$ 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 $\nabla(f+g)(x) = \nabla f(x) + \nabla g(x) \le L\|x-y\| + \frac{1}{a}\|x-y\|$

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

b) $f(x) + \frac{1}{2a}\|x\|_2^2 = (f+g)(x)$ ↓9.16 P6. lemma(ii)

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

So we need to prove:

$$f(y) + \frac{1}{2a}y^Ty \ge f(x) + \frac{1}{2a}x^Tx + (\nabla f(x) + \frac{1}{a}x)^T(y-x) + (\frac{1}{2a} - \frac{L}{2})\|y-x\|^2 \quad \cdots\cdots \quad (1)$$

Since f has L-lipschitz gradient, $\Rightarrow \|f(y) - f(x) - \nabla f(x)^T(y-x)\| \le \frac{L}{2}\|y-x\|_2^2$

So we can have $f(y) - f(x) - \nabla f(x)^T(y-x) \ge -\frac{L}{2}\|y-x\|_2^2$ terms in (1)

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

multiply $2a$ on both sides we need $y^Ty \ge x^Tx + 2x^Ty - 2x^Tx + y^Ty - 2x^Ty + x^Tx$

$\cancel{y^Ty} \ge \cancel{x^Tx} + \cancel{2x^Ty} - \cancel{2x^Tx} + \cancel{y^Ty} - \cancel{2x^Ty} + \cancel{x^Tx}$ so left hand side = right hand side.

so (1) holds for $\forall x,y \in R^d$.

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

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

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

for $-\frac{1}{a}x^T\bar{x}$: since $-\frac{1}{a}y^T\bar{y} = -\frac{1}{a}x^T\bar{x} + (-\frac{1}{a}\bar{x})^T(y-x) + \frac{M}{2}\|y-x\|_2^2$, so $M=0$

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

so $f(x) + \frac{1}{2a}x^Tx - \frac{1}{a}x^T\bar{x} + \frac{1}{2a}\bar{x}^T\bar{x}$ is a $(\frac{1}{a}-L)$-Strongly Convex when $a<\frac{1}{L}$

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

d) define a new operater: inexact proximal operator: $\mathrm{Prox}^{\varepsilon}_{\gamma h(y)}$

$$x \in \mathrm{Prox}^{\varepsilon}_{\alpha h}(y) = \left\{ z \in \mathbb{R}^N : \frac{1}{2\alpha}\|z-y\|^2 + h(z) \le \varepsilon + \min_x \frac{1}{2\gamma}\|x-y\|^2 + h(x) \right\}$$

ε is error, we can set it to a small constant, everytime we just need to find an $X \in \mathrm{Prox}^{\varepsilon}_{\alpha f}(y)$ 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 $(k=1 \cdots m)$, $t_0 = 0$, $t_1 = 1$, stepsize $\alpha < \frac{1}{L}$

output X_{m+1}.

AIPG method.

initialize $X_0 \in R^d$ and $X_1 = Z_1 = X_0$

for $k = 1, 2 \cdots m$ 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 $Z_{k+1} \in \mathrm{Prox}^{\varepsilon_k}_{\alpha h}(y_k - \alpha \nabla f(y_k))$

compute V_{k+1} that $V_{k+1} \in \mathrm{Prox}^{\varepsilon_k}_{\alpha h}(X_k - \alpha \nabla f(X_k))$

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

$$X_{k+1} = \begin{cases} Z_{k+1} & \text{if } f(Z_{k+1}) \le f(V_{k+1}) \\ V_{k+1} & \text{otherwise} \end{cases}$$

from Bin.G, De.w

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

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

Q3

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

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

proof of $A \Rightarrow B$

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

$$f(y) \geq f(x^*) + 0^T(y - x^*) \quad \text{for } \forall y$$

$$\text{so } 0 \in \partial f(x^*) \Rightarrow \quad 0 \in A_i^T(Ax^* - b) + \begin{cases} \{r\} & \text{if } x_i > 0 \\ [-r, r] & \text{if } x_i = 0 \\ \{-r\} & \text{if } x_i < 0 \end{cases} \quad \text{(10.6. lemma sum Rule)}$$

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 $f(x) \geq f(\bar{x}) - \frac{1}{2\mu}\|\nabla f(\bar{x})\|_2^2$

for all $f(x) \leq f(\bar{x})$ and $f(x)$ is a $\mu > 0$-strongly convex

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

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

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

So $f(x') \geq f(\bar{x}) - \frac{1}{2\mu}\|\nabla f(\bar{x})\|_2^2 \Rightarrow \frac{1}{2}\|\nabla f(\bar{x})\|_2^2 \geq \mu(f(\bar{x}) - \min_{x'} f(x'))$

Hence, (1) holds for every μ-strongly convex f

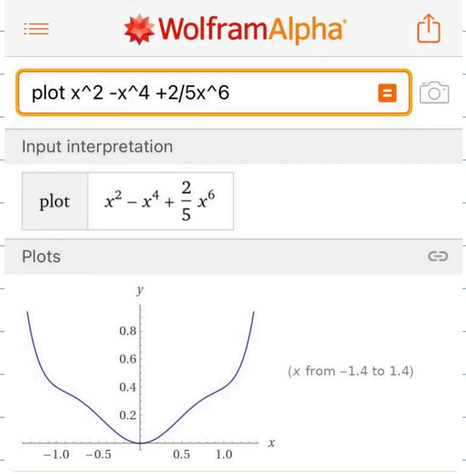
b) $x^2 - x^4 + \frac{2}{5}x^6$ let $\mu = \frac{1}{100}$

```
clc;
clear;
a = 0.1
b = 0.01



flag = 0;
for i=-5:0.01:5
    plot(i, 1/2*(12/5*(i^5) - 4*(i^3) +2*i)^2 - 1/100*(2/5*i^6 - i^4 +i*i) ,"*b")
    hold on
    if ( 1/2*(12/5*(i^5) - 4*(i^3) +2*i)^2 - 1/100*(2/5*i^6 - i^4 +i*i) <0 )
        flag = 1
    end
end
flag
```

```
a =
    0.1000
b =
    0.0100
flag =
    0
>>
```



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 $\frac{1}{2}\|\nabla f(x)\|_2^2 \geq \mu(f(x) - \min_{x'} f(x'))$

So $f(x_{k+1}) \leq f(x_k) - \frac{\mu}{L}(f(x_k) - f(x^*))$ $\quad x^* : \operatorname{argmin}_{x^*} f(x)$

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

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

So $f(x_{k+1}) - \min_{x'} f(x') \leq (1 - \frac{\mu}{L})^{k+1}(f(x_0) - \min_{x'} f(x'))$

$\Downarrow$

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

4d.) since $f(x_{k+1}) \le f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|_2^2$ $x^* = \underset{x}{\operatorname{argmin}} f(x)$

and we have $\|\nabla f(x)\|_2 \ge \mu (f(x) - f(x^*))$

so $f(x_{k+1}) \le f(x_k) - \frac{\mu^2}{2L}\left(f^2(x_k) - 2f(x_k)f(x^*) + f^2(x^*)\right)$

then, we can have

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

so $f(x_k) - f(x^*) \le f(x_{k-1}) - f(x^*) - \frac{\mu^2}{2L}(f(x_{k-1}) - f(x^*))^2$

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

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

so $A_k \le A_{k-1} - \frac{\mu^2}{2L} A_k \cdot A_{k-1}$

so we can have $\frac{1}{A_{k-1}} \le \frac{1}{A_k} - \frac{\mu^2}{2L}$

so $\frac{1}{A_k} \ge \frac{1}{A_{k-1}} + \frac{\mu^2}{2L} \ge \frac{1}{A_{k-2}} + \frac{\mu^2}{2L} + \frac{\mu^2}{2L} \dots \ge \frac{1}{A_0} + \frac{k\cdot\mu^2}{2L} \ge \frac{k\cdot\mu^2}{2L}$

so $A_k \le \frac{2L}{k\cdot\mu^2}$

Hence, $f(x_k) - f(x^*) \le \frac{2L}{k\cdot\mu^2}$

when $k \to \infty$ $f(x_k) = \min f(x)$

5a) Since f is a convex

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

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

$- \frac{L}{2} \cdot \frac{1}{L^2} \nabla f(x)^T \nabla f(x)$

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

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

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

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

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

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

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

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

$\vdots$

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

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

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

$= f(y_k) - f(\bar{x}_k)$

$\leq f(y_k) - f(x^*) \leq \frac{2L \|\bar{x}_0 - x^*\|^2}{k^2}$

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

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

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