CS 189 HW6 Write-Up

Qingyang Zhao

TOTAL POINTS

146.5 / 190

QUESTION 1

1 Getting Started 10 / 10

✓ + **10 pts Part A, Part B**

+ **5 pts** Missing team
+ **5 pts** Missing signature
+ **0 pts** Incorrect

QUESTION 2

Step Size in Gradient Descent 50 pts

2.1 a 10 / 10

✓ + **10 pts Correct**

+ **8 pts** Mostly Correct
+ **5 pts** Half Correct
+ **2 pts** Trivial Progress
+ **0 pts** Wrong/Missing

2.2 b 10 / 10

✓ + **10 pts Fully Correct**

+ **4 pts** Correct conclusion for general b
+ **4 pts** Correct proof of oscillatory pattern for given b
+ **2 pts** Concluded that descent will not reach the optimum for the given b
+ **0 pts** Incorrect or blank

2.3 c 10 / 10

✓ + **10 pts Correct**

+ **4 pts** Correct rule for general b (travel distance is bounded by 6).
+ **4 pts** Correct proof of failure of GD to reach optimum.
+ **2 pts** Claimed GD does NOT find the optimum for the given b and step size.
+ **0 pts** Incorrect or blank

2.4 d 10 / 10

✓ + **10 pts Correct**

+ **8 pts** Minor error or left out something small
+ **5 pts** Good progress
+ **2 pts** Trivial Progress
+ **0 pts** Incorrect

2.5 e 8 / 10

+ **10 pts** Correct

✓ + **8 pts Almost Correct**

+ **6 pts** Significant Progress
+ **4 pts** Some Progress
+ **0 pts** Incorrect or Blank
+ **2 pts** Trivial Progress

QUESTION 3

Convergence Rate of Gradient Descent 70 pts

3.1 a 8 / 10

+ **10 pts** Correct state evolution

✓ + **8 pts Minor algebraic error / missing last step**

+ **5 pts** Some progress but incorrect final expression
+ **2 pts** Minimal effort
+ **0 pts** Incorrect or blank

Make sure you explicitly write the state evolution in terms of x0!

3.2 b 9.5 / 10

+ **10 pts** States that all eigenvalues of $$(I - \gamma A^TA)$$ must have magnitude strictly less than 1

✓ + **9.5 pts States that all eigenvalues of $$(I - \gamma A^TA)$$ must have magnitude less than or equal to 1, or Did not specify magnitude of eigenvalues**

+ **8 pts** Incorrectly bounds the matrix $$(I - \gamma A^TA)$$ instead of the eigenvalues
+ **7 pts** States that as iterations tends to infinity, the state doesn't reach infinity

+ **5 pts** Recognizes significance of $$(I - \gamma A^TA)$$

+ **0 pts** Incorrect

3.3 c 10 / 10

✓ + **10 pts Getting the bound**

+ **8 pts** substituting into result

+ **5 pts** Rayleigh quotient

+ **2 pts** plug in, simplify equations

+ **0 pts** Incorrect/blank

3.4 d 10 / 10

✓ + **10 pts Correct**

+ **8 pts** almost correct

+ **5 pts** halfway

+ **2 pts** first step

+ **0 pts** Incorrect

3.5 e 10 / 10

✓ + **10 pts Correct**

+ **5 pts** Partial

+ **0 pts** Missing

3.6 f 10 / 10

✓ + **10 pts Correct**

+ **8 pts** Almost Correct

+ **5 pts** Halfway

+ **3 pts** first step

+ **0 pts** Incorrect

3.7 g 0 / 10

+ **10 pts** Correct

+ **8 pts** Almost Correct

+ **5 pts** Halfway Correct

+ **2 pts** Trivial Progress

✓ + **0 pts Wrong/Missing**

QUESTION 4

Sensor Locations 50 pts

4.1 a 10 / 10

✓ + **10 pts Correct**

+ **8 pts** Minor Mistake

+ **5 pts** Partially Correct

+ **0 pts** No Answer

4.2 b 8 / 10

+ **0 pts** Didn't attempt or very very wrong

+ **2 pts** Minimal Progress

+ **5 pts** Right direction and got half-way there

✓ + **8 pts Minor Error(s)**

+ **10 pts** Correct

4.3 c 0 / 10

+ **10 pts** Correct

+ **8 pts** Minor Error

+ **5 pts** Partial

+ **3 pts** Mostly missing

✓ + **0 pts Incorrect/Missing**

4.4 d 3 / 10

+ **10 pts** Correct with reasonable justification

+ **7 pts** Graphs / data but incorrect or missing justification / explanation

✓ + **3 pts Claim without justification or graphs**

+ **0 pts** Incorrect claim / no answer

4.5 f 0 / 10

+ **10 pts** Correct

+ **8 pts** Some Mistake

+ **5 pts** Partial

✓ + **0 pts Wrong/Missing**

QUESTION 5

5 Vegetables 0 / 0

✓ + **0 pts -**

QUESTION 6

6 Write Your Own Question 10 / 10

✓ + **10 pts Correct**

+ **0 pts** No Answer

H06-1

a)

Names	Email Address
Wan	jhun0324@berkeley.edu

Description of Team: Best Group Ever

How did I work?

Comments:

b) I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

Qingyang Zhao

1 Getting Started 10 / 10

- ✓ + **10 pts** **Part A, Part B**
- + **5 pts** Missing team
- + **5 pts** Missing signature
- + **0 pts** Incorrect

HW06-2

a) $f(x)$ is convex iff $\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\lambda x_1 + (1-\lambda) x_2)$.

$f(x) = \|x-b\|_2$

$$f(\lambda x_1 + (1-\lambda) x_2) = \|\lambda x_1 + (1-\lambda) x_2 - \lambda b - (1-\lambda) b\|_2$$
$$= \|\lambda(x_1 - b) + (1-\lambda)(x_2 - b)\|_2$$
$$\leq \|\lambda(x_1 - b)\|_2 + \|(1-\lambda)(x_2 - b)\|_2$$
$$\leq \lambda \|x_1 - b\|_2 + (1-\lambda)\|x_2 - b\|_2$$
$$\leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

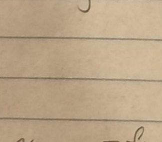
$f(x) = \|x-b\|_2$ is convex.

b) no, cannot find global optimal using this step size.

Reason: t_i is too large, the position will repeat after 8 steps.

0	0
0.55	0.83
1.11	1.66
1.66	2.50
2.22	3.33
2.77	4.16
3.33	4.99
3.88	5.82
4.44	6.66
3.88	5.82
4.44	6.66

Proof: Compute $\nabla f(x) = \frac{\partial f(x)}{\partial x^T} = \frac{1}{2}((x-b)^T(x-b))^{-\frac{1}{2}} \cdot 2 \cdot (x-b)$



$$= \frac{x-b}{\|x-b\|_2}$$

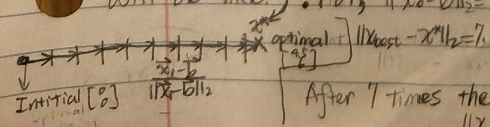
$$x_i = x_{i-1} - \nabla f(x_{i-1}) = x_{i-1} - \frac{x_{i-1} - b}{\|x_{i-1} - b\|_2}$$

Consider, everytime we update x_i, is along the unit vector of $\frac{x_i - b}{\|x_i - b\|_2}$.

We can prove $\frac{x_{i+1} - b}{\|x_{i+1} - b\|_2}$ and $\frac{x_i - b}{\|x_i - b\|_2}$ are in the same direction.

$$\frac{x_{i+1} - b}{\|x_{i+1} - b\|_2} = \frac{x_i - b - \frac{x_i - b}{\|x_i - b\|_2}}{\left\|x_i - b - \frac{x_i - b}{\|x_i - b\|_2}\right\|_2} = (x_i - b)\frac{\left(1 - \frac{1}{\|x_i - b\|_2}\right)}{\left\|(x_i - b)\left(1 - \frac{1}{\|x_i - b\|_2}\right)\right\|_2} = \frac{x_i - b}{\|x_i - b\|_2}$$

which means, when step size is fixed (take 2d case for example) The update will be like ↙. For, $\|x_0 - b\|_2 = 7.5$ with step size $= 1$, the best we can go



$\|x_{best} - x^*\|_2 = 7.5 - \underset{1 \times 7}{\text{stepsize} \times \text{IterateTimes}} = \boxed{0.5 > 0.01}$

After 7 times the distance will remain 0.5. $\|x_i - x^*\|_2,\ i \geq 7$

$\boxed{\text{See Next Page for General, } \vec{b} \neq 0}$

1

2.1 a 10 / 10

✓ + **10 pts** Correct

+ **8 pts** Mostly Correct

+ **5 pts** Half Correct

+ **2 pts** Trivial Progress

+ **0 pts** Wrong/Missing

HW06-2

a) $f(x)$ is convex iff $\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\lambda x_1 + (1-\lambda) x_2)$.

$f(x) = \|x - b\|_2$

$$f(\lambda x_1 + (1-\lambda) x_2) = \| \lambda x_1 + (1-\lambda) x_2 - \lambda b - (1-\lambda) b \|_2$$

$$= \| \lambda (x_1 - b) + (1-\lambda)(x_2 - b) \|_2$$

$$\leq \| \lambda (x_1 - b) \|_2 + \| (1-\lambda)(x_2 - b) \|_2$$

$$\leq \lambda \| x_1 - b \|_2 + (1-\lambda) \| x_2 - b \|_2$$

$$\leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

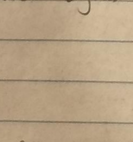
$f(x) = \|x - b\|_2$ is convex.

b) no, cannot find global optimal using this step size.

Reason: t_i is too large, the position will repeat after 8 steps.

0	0
0.55	0.83
1.11	1.66
1.66	2.50
2.22	3.33
2.77	4.16
3.33	4.99
3.88	5.82
4.44	6.66
3.88	5.82
4.44	6.66

Proof: Compute $\nabla f(x) = \frac{\partial f(x)}{\partial x^T} = \frac{1}{2}((x-b)^T(x-b))^{-\frac{1}{2}} \cdot 2 \cdot (x-b)$



$$= \frac{x-b}{\|x-b\|_2}$$

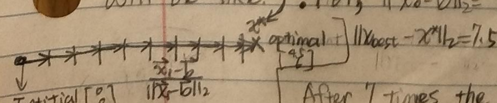
$$x_i = x_{i-1} - \nabla f(x_{i-1}) = x_{i-1} - \frac{x_{i-1} - b}{\|x_{i-1} - b\|_2}$$

Consider, everytime we update x_i, is along the unit vector of $\frac{x_i - b}{\|x_i - b\|_2}$.

We can prove $\frac{x_{i+1} - b}{\|x_{i+1} - b\|_2}$ and $\frac{x_i - b}{\|x_i - b\|_2}$ are in the same direction.

$$\frac{x_{i+1} - b}{\|x_{i+1} - b\|_2} = \frac{x_i - b - \frac{x_i - b}{\|x_i - b\|_2}}{\|x_i - b - \frac{x_i - b}{\|x_i - b\|_2}\|_2} = (x_i - b)\frac{(1 - \frac{1}{\|x_i - b\|_2})}{\|(x_i - b)(1 - \frac{1}{\|x_i - b\|_2})\|_2} = \frac{x_i - b}{\|x_i - b\|_2}$$

which means, when step size is fixed (take 2d case for example) The update will be like: . For, $\|x_0 - b\|_2 = 7.5$ with step size $= 1$, the best we can go



optimal $\begin{bmatrix} 5 \\ 7.5 \end{bmatrix}$ $\|x_{best} - x^*\|_2 = 7.5 - \text{stepsize} \times \text{IterateTimes} = \boxed{0.5 > 0.01}$

1×7

Initial $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\frac{x_i - b}{\|x_i - b\|_2}$

After 7 times the distance will remain 0.5.

$\|x_t - x^*\|_2, \; t \geq 7$

See Next Page for General, $\vec{b} \neq 0$

1

HW02

b) for $\vec{b} \neq 0$, say ε is the optimal solution error we may tolerate (in this case $\varepsilon = 0.01$, $\|x_i - x^*\|_2 \leq \varepsilon$)

if $\|x_0 - b\|_2$ within interval $[\text{integer} - \varepsilon, \text{integer} + \varepsilon]$, where integer $\in N$)

It can get within 0.01 of optimal solution.

2.2 b **10 / 10**

- ✓ **+ 10 pts** Fully Correct
- **+ 4 pts** Correct conclusion for general b
- **+ 4 pts** Correct proof of oscillatory pattern for given b
- **+ 2 pts** Concluded that descent will not reach the optimum for the given b
- **+ 0 pts** Incorrect or blank

HW 06-2

c) No, we cannot reach optimum within 0.01.

Proof: Just the same as part b), we want to check $\|x_{best} - x^*\|_2 \overset{?}{\leq} 0.01$,

Because of every update is along the same direction, when step size is $(\frac{5}{6})^i$, optimal point is $[4.5\ 6]^T$, initial point is $[0,0]^T$.

The distance we want update is $7.5 = \|x_0 - x^*\|_2$

However, using this step size, the largest distance, we can update is $\sum_{i=0}^{\infty} (\frac{5}{6})^i = \frac{1}{1-\frac{5}{6}} = 6 < 7.5$

$7.5 - 6 > 0.01$, never reach within 0.01.

For general $\vec{b} \neq 0$ case, if the $\|x_0 - \vec{b}\|_2 \leq \sum_{i=0}^{\infty} (\frac{5}{6})^i = 6$ (in this case $\|\vec{b}\|_2 \leq 6$), we can reach optimum within $\|x_i - x^*\|_2 \leq 0.01$

d) Yes, it can find, it will take $\boxed{1004}$ steps to get within 0.01.

For $x_0 = [0\ 0]^T$ $b = [4.5\ 6]^T$ $\|x_0 - b\|_2 = \|x_0 - x^*\|_2 = 7.5$

Because, the update also in same direction, all we need to check is

$$\sum_{i=0}^{k} \frac{1}{i+1} \geq 7.5 - 0.01$$

using python get $k = 1004$.

for general $\vec{b} \neq 0$, say we want within ε of the optimal solution

2.3 C 10 / 10

✓ + 10 pts Correct

+ 4 pts Correct rule for general b (travel distance is bounded by 6).

+ 4 pts Correct proof of failure of GD to reach optimum.

+ 2 pts Claimed GD does NOT find the optimum for the given b and step size.

+ 0 pts Incorrect or blank

HW06-2

c) No, we cannot reach optimum within 0.01.

Proof: Just the same as part b), we want to check $\|x_{best} - x^*\|_2 \overset{?}{\leq} 0.01$,

Because of every update is along the same direction, when step size is $(\frac{5}{6})^i$, optimal point is $[4.5\ 6]^T$, initial point is $[0, 0]^T$.

The distance we want update is $7.5 = \|x_0 - x^*\|_2$

However, using this step size, the largest distance, we can update is $\sum_{i=0}^{\infty} (\frac{5}{6})^i = \frac{1}{1-\frac{1}{6}} = 6 < 7.5$

$7.5 - 6 > 0.01$, never reach within 0.01.

For general $\vec{b} \neq 0$ case, if the $\|x_0 - \vec{b}\|_2 \leq \sum_{i=0}^{\infty} (\frac{5}{6})^i = 6$ (in this case $\|\vec{b}\|_2 \leq 6$), we can reach optimum within $\|x_i - x^*\|_2 \leq 0.01$

d) Yes, it can find, it will take $\boxed{1004}$ steps to get within 0.01.

For $x_0 = [0\ 0]^T$ $b = [4.5\ 6]^T$ $\|x_0 - b\|_2 = \|x_0 - x^*\|_2 = 7.5$

Because, the update also in same direction, all we need to check is

$$\sum_{i=0}^{k} \frac{1}{i+1} \geq 7.5 - 0.01$$

using python get $k = 1004$.

for general $\vec{b} \neq 0$, say we want within ε of the optimal solution

2.4 d 10 / 10

✓ + **10 pts** Correct

\+ **8 pts** Minor error or left out something small

\+ **5 pts** Good progress

\+ **2 pts** Trivial Progress

\+ **0 pts** Incorrect

e) fixed size $=1$ or step size $= (\frac{5}{6})^i$ will not suit for both cases.

But step size $= \frac{1}{i+1}$ will always give a optimal solution within 0.01.

Because $\sum_{i=0}^{\infty} \frac{1}{i+1} \to \infty$ the step size of $\frac{1}{i+1}$ series can estimate x (using line search) with any initial points.

However fixed size and $(\frac{5}{6})^i$ has either flaws in cannot reach into ~~interval~~ or, $\sum_{i=1}^{\infty} (\frac{5}{6})^{i}$ is to small.

2.5 e 8 / 10

+ **10 pts** Correct
✓ + **8 pts** **Almost Correct**
+ **6 pts** Significant Progress
+ **4 pts** Some Progress
+ **0 pts** Incorrect or Blank
+ **2 pts** Trivial Progress

HW06-3

a) $f(x) = \frac{1}{2}\|Ax-b\|_2^2 = \frac{1}{2}(Ax-b)^T(Ax-b)$

$\nabla f(x) = A^T(Ax-b)$

$x_{i+1} = x_i - \gamma A^T(Ax_i - b) \qquad \vec{b} = \vec{0}$

$= x_i - \gamma A^T A x_i$

b)

$\lim_{i\to\infty} \|x_i\|_2 = \lim_{i\to\infty} \|x_{i-1} - \gamma A^T A x_{i-1}\|_2$

$A^TA = U\Sigma U^T$

$= \lim_{i\to\infty} \|(I - \gamma A^TA)x_{i-1}\|_2$

$I - \gamma A^TA = U(I-\gamma\Sigma)U^T$

$(I-\gamma A^TA)^i = U(I-\gamma\Sigma)^i U^T$

$= \lim_{i\to\infty} \|(I-\gamma A^TA)^i x_0\|_2$

$x_0' = U^T x_0 \to const.$

$= \lim_{i\to\infty} \|U(I-\gamma\Sigma)^i U^T x_0\|_2 = \lim_{i\to\infty} \|(I-\gamma\Sigma)^i x_0'\|_2$

When viewed as a dynamical system, it is stable $\boxed{\text{when } |1-\gamma\lambda_{max}(A^TA)|, |1-\gamma\lambda_{min}(A^TA)| \text{ both} \le 1}$ Otherwise, $\lim_{i\to\infty} \|x_i\|_2$ will go infinity.

c) $\|\varphi(x) - \varphi(x')\|_2$

$= \|x - \gamma\nabla f(x) - x' + \gamma\nabla f(x')\|_2 = \|x - \gamma A^T(Ax-b) - x' + \gamma A^T(Ax'-b)\|_2$

$A = U\Sigma V^T$

$A^TA = \underset{d\times d}{U\Sigma^2 U^T}$

$= \|(I - \gamma A^TA)(x-x')\|_2$

$= \left\| U \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & 1-\gamma\sigma_2^2 & \\ & & \ddots \\ & & & 1-\gamma\sigma_d^2 \end{bmatrix} U^T(x-x') \right\|_2$

set $y = U^T(x-x')$

$\Sigma' = \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & \ddots & \\ & & 1-\gamma\sigma_d^2 \end{bmatrix}$

$= \left\| \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & \ddots & \\ & & 1-\gamma\sigma_d^2 \end{bmatrix} y \right\|_2 = \|\Sigma' y\|_2$

consider the largest absolute elements in Σ' is either $|1-\gamma\sigma_1^2|$ or $|1-\gamma\sigma_d^2|$

which is $\max\{|1-\gamma\lambda_{max}(A^TA)|, |1-\gamma\lambda_{min}(A^TA)|\}$

See Next Page

3.1 a 8 / 10

+ **10 pts** Correct state evolution

✓ + **8 pts** **Minor algebraic error / missing last step**

+ **5 pts** Some progress but incorrect final expression

+ **2 pts** Minimal effort

+ **0 pts** Incorrect or blank

Make sure you explicitly write the state evolution in terms of x0!

HW06-3

a) $f(x) = \frac{1}{2}\|Ax-b\|_2^2 = \frac{1}{2}(Ax-b)^T(Ax-b)$

$\nabla f(x) = A^T(Ax-b)$

$x_{i+1} = x_i - \gamma A^T(Ax_i - b) \qquad \vec{b} = \vec{0}$

$= x_i - \gamma A^T A x_i$

b)

$\lim_{i\to\infty} \|x_i\|_2 = \lim_{i\to\infty} \|x_{i-1} - \gamma A^T A x_{i-1}\|_2$

$A^T A = U \Sigma U^T$

$= \lim_{i\to\infty} \|(I - \gamma A^T A) x_{i-1}\|_2$

$I - \gamma A^T A = U(I - \gamma\Sigma)U^T$

$(I - \gamma A^T A)^i = U(I-\gamma\Sigma)^i U^T$

$= \lim_{i\to\infty} \|(I - \gamma A^T A)^i x_0\|_2$

$x_0' = U^T x_0 \to const.$

$= \lim_{i\to\infty} \|U(I-\gamma\Sigma)^i U^T x_0\|_2 = \lim_{i\to\infty} \|(I-\gamma\Sigma)^i x_0'\|_2$

When viewed as a dynamical system, it is stable $\boxed{\text{when } |1-\gamma\lambda_{max}(A^TA)|, |1-\gamma\lambda_{min}(A^TA)| \text{ both} \le 1}$ Otherwise, $\lim_{i\to\infty} \|x_i\|_2$ will go infinity.

c) $\|\varphi(x) - \varphi(x')\|_2$

$= \|x - \gamma\nabla f(x) - x' + \gamma\nabla f(x')\|_2 = \|x - \gamma A^T(Ax-b) - x' + \gamma A^T(Ax'-b)\|_2$

$A = U\Sigma V^T$

$\underset{d\times d}{A^TA} = U\Sigma^2 U^T$

$= \|(I - \gamma A^T A)(x - x')\|_2$

$= \left\| U \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & 1-\gamma\sigma_2^2 & \\ & & \ddots \\ & & & 1-\gamma\sigma_d^2 \end{bmatrix} U^T(x-x') \right\|_2$

set $y = U^T(x-x')$

$\Sigma' = \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & \ddots & \\ & & 1-\gamma\sigma_d^2 \end{bmatrix}$

$= \left\| \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & \ddots & \\ & & 1-\gamma\sigma_d^2 \end{bmatrix} y \right\|_2 = \|\Sigma' y\|_2$

consider the largest absolute elements in Σ' is either $|1-\gamma\sigma_1^2|$ or $|1-\gamma\sigma_d^2|$

which is $\max\{|1-\gamma\lambda_{max}(A^TA)|, |1-\gamma\lambda_{min}(A^TA)|\}$

See Next Page

3.2 b **9.5 / 10**

- \+ **10 pts** States that all eigenvalues of $$(I - \gamma A^TA)$$ must have magnitude strictly less than 1

✓ + **9.5 pts** **States that all eigenvalues of $$(I - \gamma A^TA)$$ must have magnitude less than or equal to 1, or Did not specify magnitude of eigenvalues**

- \+ **8 pts** Incorrectly bounds the matrix $$(I - \gamma A^TA)$$ instead of the eigenvalues
- \+ **7 pts** States that as iterations tends to infinity, the state doesn't reach infinity
- \+ **5 pts** Recognizes significance of $$(I - \gamma A^TA)$$
- \+ **0 pts** Incorrect

HW06-3

a) $f(x) = \frac{1}{2}\|Ax-b\|_2^2 = \frac{1}{2}(Ax-b)^T(Ax-b)$

$\nabla f(x) = A^T(Ax-b)$

$x_{i+1} = x_i - \gamma A^T(Ax_i - b)$ $\quad \vec{b} = \vec{0}$

$= x_i - \gamma A^T A x_i$

b)

$\lim_{i\to\infty} \|x_i\|_2 = \lim_{i\to\infty} \|x_{i-1} - \gamma A^T A x_{i-1}\|_2$ $\quad A^TA = U\Sigma U^T$

$= \lim_{i\to\infty} \|(I - \gamma A^TA)x_{i-1}\|_2$ $\quad I - \gamma A^TA = U(I-\gamma\Sigma)U^T$

$(I-\gamma A^TA)^i = U(I-\gamma\Sigma)^i U^T$

$= \lim_{i\to\infty} \|(I-\gamma A^TA)^i x_0\|_2$ $\quad x_0' = U^T x_0 \to \text{const.}$

$= \lim_{i\to\infty} \|U(I-\gamma\Sigma)^i U^T x_0\|_2 = \lim_{i\to\infty} \|(I-\gamma\Sigma)^i x_0'\|_2$

When viewed as a dynamical system, it is stable $\boxed{\text{when } |1-\gamma\lambda_{max}(A^TA)|, |1-\gamma\lambda_{min}(A^TA)| \text{ both} \le 1}$ Otherwise, $\lim_{i\to\infty} \|x_i\|_2$ will go infinity.

c) $\|\varphi(x) - \varphi(x')\|_2$

$= \|x - \gamma\nabla f(x) - x' + \gamma\nabla f(x')\|_2 = \|x - \gamma A^T(Ax-b) - x' + \gamma A^T(Ax'-b)\|_2$ $\quad A = U\Sigma V^T$, $\quad \underset{d\times d}{A^TA = U\Sigma^2 U^T}$

$= \|(I - \gamma A^TA)(x-x')\|_2$

$= \left\| U \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & 1-\gamma\sigma_2^2 & \\ & & \ddots \\ & & & 1-\gamma\sigma_d^2 \end{bmatrix} U^T(x-x') \right\|_2$ $\quad$ set $y = U^T(x-x')$

$\Sigma' = \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & \ddots & \\ & & 1-\gamma\sigma_d^2 \end{bmatrix}$

$= \left\| \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & \ddots & \\ & & 1-\gamma\sigma_d^2 \end{bmatrix} y \right\|_2 = \|\Sigma' y\|_2$

consider the largest absolute elements in Σ' is either $|1-\gamma\sigma_1^2|$ or $|1-\gamma\sigma_d^2|$

which is $\max\{|1-\gamma\lambda_{max}(A^TA)|, |1-\gamma\lambda_{min}(A^TA)|\}$

See Next Page

c)

$$\|\varphi(x)-\varphi(x')\|_2 = \|\Sigma' y\|_2 \qquad \Sigma' = \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & \ddots & \\ & & 1-\gamma\sigma_d^2 \end{bmatrix} \quad y = U^T(x-x')$$

$$\le \left\| \begin{bmatrix} \max\{\ \} & & \\ & \ddots & \\ & & \max\{\ \} \end{bmatrix} y \right\|_2$$

$$= \max\{|1-\gamma\lambda_{max}\{A^TA\}|, |1-\gamma\lambda_{min}\{A^TA\}|\} \|y\|_2$$

$$= \beta \|U^T(x-x')\|_2$$

$$= \beta \|x-x'\|_2$$

$$\|\varphi(x)-\varphi(x')\|_2 \le \beta \|x-x'\|_2$$

Why we are doing this: If β is small, the x will not blow up, converge into small point.

d) $\|x_{k+1}-x^*\|_2 = \|\varphi(x_k) - x^*\|_2$

Quadratic $\nabla f(x^*)=0$

Thus $\varphi(x^*) = x^* - \gamma\nabla f(x^*) = x^*$

① $\boxed{= \|\varphi(x_k) - \varphi(x^*)\|_2}$

$$\le \beta\|x_k - x^*\|_2$$

$$= \beta\|\varphi(x_{k-1}) - \varphi(x^*)\|_2$$

$$\le \beta^2 \|x_{k-1} - x^*\|_2$$

$$\vdots$$

② $\le \beta^{k+1}\|x_0 - x^*\|_2$

3.3 C 10 / 10

✓ + 10 pts Getting the bound

+ 8 pts substituting into result

+ 5 pts Rayleigh quotient

+ 2 pts plug in, simplify equations

+ 0 pts Incorrect/blank

c)

$$\|\varphi(x)-\varphi(x')\|_2 = \|\Sigma' y\|_2 \qquad \Sigma' = \begin{bmatrix} 1-\gamma\sigma_1^2 & & \\ & \ddots & \\ & & 1-\gamma\sigma_d^2 \end{bmatrix} \quad y = U^T(x-x')$$

$$\leq \left\| \begin{bmatrix} \max\{\ \} & & \\ & \ddots & \\ & & \max\{\ \} \end{bmatrix} y \right\|_2$$

$$= \max\{|1-\gamma\lambda_{max}\{A^TA\}|, |1-\gamma\lambda_{min}\{A^TA\}|\} \|y\|_2$$

$$= \beta \|U^T(x-x')\|_2$$

$$= \beta\|x-x'\|_2$$

$$\|\varphi(x)-\varphi(x')\|_2 \leq \beta\|x-x'\|_2$$

Why we are doing this: If β is small, the x will not blow up, converge into small point.

d)

$$\|x_{k+1}-x^*\|_2 = \|\varphi(x_k) - x^*\|_2$$

Quadratic $\nabla f(x^*)=0$

Thus $\varphi(x^*) = x^* - \gamma\nabla f(x^*) = x^*$

① $$\boxed{= \|\varphi(x_k) - \varphi(x^*)\|_2}$$

$$\leq \beta\|x_k - x^*\|_2$$

$$= \beta\|\varphi(x_{k-1}) - \varphi(x^*)\|_2$$

$$\leq \beta^2\|x_{k-1} - x^*\|_2$$

$$\vdots$$

② $$\leq \beta^{k+1}\|x_0 - x^*\|_2$$

3.4 d 10 / 10

- ✓ + **10 pts** Correct
- + **8 pts** almost correct
- + **5 pts** halfway
- + **2 pts** first step
- + **0 pts** Incorrect

HW06 3.

e) $f(x) - f(x^*) = \frac{1}{2}\|Ax-b\|_2^2 - \frac{1}{2}\|Ax^*-b\|_2^2 = \frac{1}{2}\left(x^TA^TAx - 2x^TA^Tb + b^Tb - (x^{*T}A^TAx^* - 2x^{*T}A^Tb + b^Tb)\right)$

$A^TAx^* = A^Tb$

$$= \frac{1}{2}(x^TA^TAx - 2x^TA^TAx^* + 2x^{*}A^TAx^* - x^*A^TAx^*)$$
$$= \frac{1}{2}(x^TA^TA(x-x^*) + (x^*-x)^TA^TAx^*)$$
$$= \frac{1}{2}(x-x^*)^TA^TA(x-x^*)$$
$$= \frac{1}{2}\|A(x-x^*)\|_2^2$$

f) $f(x_k) - f(x^*) = \frac{1}{2}\|A(x_k - x^*)\|_2^2 \qquad A = U\Sigma V^T$

$$= \frac{1}{2}\|U\Sigma V^T(x_k - x^*)\|_2^2$$

Set $y = V^T(x_k - x^*)$.

$$= \frac{1}{2}\|\Sigma V^T(x_k - x^*)\|_2^2$$
$$= \frac{1}{2}\|\Sigma y\|_2^2 \le \frac{1}{2}\lambda_{max}(A^TA)\|y\|_2^2 = \frac{1}{2}\lambda_{max}(A^TA)\|V^T(x_k - x^*)\|_2$$
$$= \frac{\alpha}{2}\|x_k - x^*\|_2$$

$f(x_k) - f(x^*) \le \frac{\alpha}{2}\|x_k - x^*\|_2$

$\to$ Because of part d)

$$f(x_k) - f(x^*) = \frac{1}{2}\|A(x_k - x^*)\|_2^2$$
$$= \frac{1}{2}\|A((I-\gamma A^TA)x_{k-1} - (I-\gamma A^TA)x^* + \cancel{\gamma b} - \cancel{\gamma b})\|_2^2$$
$$= \frac{1}{2}\|A(I-\gamma A^TA)^k(x_0 - x^*)\|_2^2$$
$$= \frac{1}{2}\|U\Sigma V^T(VV^T - \gamma V\Sigma^{2k}V^T)(x_0 - x^*)\|_2^2$$
$$= \frac{1}{2}\|\Sigma(I - \gamma\Sigma^{2k})V^T(x_0 - x^*)\|_2^2$$
$$\le \frac{1}{2}\alpha^2\beta^{2k}\|x_0 - x^*\|_2^2$$

3.5 e 10 / 10

✓ + **10 pts** Correct

+ **5 pts** Partial

+ **0 pts** Missing

HW06 3.

e) $f(x) - f(x^*) = \frac{1}{2}\|Ax-b\|_2^2 - \frac{1}{2}\|Ax^*-b\|_2^2 = \frac{1}{2}\left(x^TA^TAx - 2x^TA^Tb + b^Tb - (x^{*T}A^TAx^* - 2x^{*T}A^Tb + b^Tb)\right)$

$A^TAx^* = A^Tb$ $\quad = \frac{1}{2}(x^TA^TAx - 2x^TA^TAx^* + 2x^{*}A^TAx^* - x^{*}A^TAx^*)$

$= \frac{1}{2}(x^TA^TA(x-x^*) + (x^*-x)^TA^TAx^*)$

$= \frac{1}{2}(x-x^*)^TA^TA(x-x^*)$

$= \frac{1}{2}\|A(x-x^*)\|_2^2$

f) $f(x_k) - f(x^*) = \frac{1}{2}\|A(x_k - x^*)\|_2^2$ $\qquad A = U\Sigma V^T$

$= \frac{1}{2}\|U\Sigma V^T(x_k - x^*)\|_2^2$ $\qquad$ Set $y = V^T(x_k - x^*)$.

$= \frac{1}{2}\|\Sigma V^T(x_k - x^*)\|_2^2$

$= \frac{1}{2}\|\Sigma y\|_2^2 \leq \frac{1}{2}\lambda_{max}(A^TA)\|y\|_2^2 = \frac{1}{2}\lambda_{max}(A^TA)\|V^T(x_k - x^*)\|_2$

$= \frac{\alpha}{2}\|x_k - x^*\|_2$

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

$\longrightarrow$ Because of part d)

$f(x_k) - f(x^*) = \frac{1}{2}\|A(x_k - x^*)\|_2^2$

$= \frac{1}{2}\|A((I-\gamma A^TA)x_{k-1} - (I-\gamma A^TA)x^* + \cancel{\gamma b} - \cancel{\gamma b})\|_2^2$

$= \frac{1}{2}\|A(I-\gamma A^TA)^k(x_0 - x^*)\|_2^2$

$= \frac{1}{2}\|U\Sigma V^T(VV^T - \gamma V\Sigma^{2k}V^T)(x_0 - x^*)\|_2^2$

$= \frac{1}{2}\|\Sigma(I - \gamma\Sigma^{2k})V^T(x_0 - x^*)\|_2^2$

$\leq \frac{1}{2}\alpha^2\beta^{2k}\|x_0 - x^*\|_2^2$

3.6 f **10 / 10**

- ✓ + **10 pts** Correct
- + **8 pts** Almost Correct
- + **5 pts** Halfway
- + **3 pts** first step
- + **0 pts** Incorrect

g) $\beta = \max\left\{ |1-\gamma\lambda_{max}(A^TA)|,\ |1-\gamma\lambda_{min}(A^TA)| \right\}$

3.7 g **0 / 10**

+ **10 pts** Correct
+ **8 pts** Almost Correct
+ **5 pts** Halfway Correct
+ **2 pts** Trivial Progress

✓ + **0 pts** **Wrong/Missing**

HW06-4

a) $L(x_1, y_1) = \sum_{i=1}^{7} \left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)^2$

$$\frac{\partial L(x_1, y_1)}{\partial x_1} = \sum_{i=1}^{7} \frac{-2\left(\sqrt{(a_i-x_1)^2+(b_i-y_1)^2} - d_i\right)(a_i - x_1)}{\sqrt{(a_i-x_1)^2+(b_i-y_1)^2}}$$

$$\frac{\partial L(x_1, y_1)}{\partial y_1} = \sum_{i=1}^{7} \frac{+2\left(\sqrt{(a_i-x_1)^2+(b_i-y_1)^2} - d_i\right)(b_i - y_1)}{\sqrt{(a_i-x_1)^2+(b_i-y_1)^2}}$$

b) $x_0 = \vec{0}$ step size: 0.01 473 steps.

stop criterion: gradient $\leq 10^{-6}$

$x_0 = [100, 0]$ → 466 steps.

c) Average unique points goes up,

proportion of global minimum goes down

d) As variance goes down, the unique point small within $x \in (0, 500)$

the proportion goes below 0.5 after 500.

e) as sensor number goes up, unique points become smaller than before

$\boxed{\text{compare to d}}$

(when variance big, not reach minimum ~~yet~~ within 1000 steps)

proportion becomes smaller than before

4.1 a 10 / 10

✓ + 10 pts Correct

+ 8 pts Minor Mistake

+ 5 pts Partially Correct

+ 0 pts No Answer

HW06-4

a) $L(x_1, y_1) = \sum_{i=1}^{7} \left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)^2$

$$\frac{\partial L(x_1, y_1)}{\partial x_1} = \sum_{i=1}^{7} \frac{-2\left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)(a_i - x_1)}{\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2}}$$

$$\frac{\partial L(x_1, y_1)}{\partial y_1} = \sum_{i=1}^{7} \frac{-2\left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)(b_i - y_1)}{\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2}}$$

b) $x_0 = \vec{0}$ — step size: 0.01 — 473 steps.

stop criterion: gradient $\leq 10^{-6}$

$x_0 = [100, 0]$ → 466 steps.

c) Average unique points goes up,

proportion of global minimum goes down

d) As variance goes down, the unique point small within $x \in (0, 500)$

the proportion goes below 0.5 after 500.

e) as sensor number goes up, unique points become smaller than before

compare to d)

(when variance big, not reach minimum yet within 1000 steps)

proportion becomes smaller than before

4.2 b **8 / 10**

+ **0 pts** Didn't attempt or very very wrong
+ **2 pts** Minimal Progress
+ **5 pts** Right direction and got half-way there
✓ + **8 pts** Minor Error(s)
+ **10 pts** Correct

HW06-4

a) $L(x_1, y_1) = \sum_{i=1}^{7} \left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)^2$

$$\frac{\partial L(x_1, y_1)}{\partial x_1} = \sum_{i=1}^{7} \frac{-2\left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)(a_i - x_1)}{\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2}}$$

$$\frac{\partial L(x_1, y_1)}{\partial y_1} = \sum_{i=1}^{7} \frac{+2\left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)(b_i - y_1)}{\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2}}$$

b) $x_0 = \vec{0}$ — step size: 0.01 — 473 steps.

stop criterion: gradient $\leq 10^{-6}$

$x_0 = [100, 0]$ → 466 steps.

c) Average unique points goes up,

proportion of global minimum goes down

d) As variance goes down, the unique point small within $x \in (0, 500)$

the proportion goes below 0.5 after 500.

e) as sensor number goes up, unique points become smaller than before

compare to d)

(when variance big, not reach minimum yet within 1000 steps)

proportion becomes smaller than before

4.3 C **0 / 10**

- \+ **10 pts** Correct
- \+ **8 pts** Minor Error
- \+ **5 pts** Partial
- \+ **3 pts** Mostly missing
- ✓ + **0 pts** **Incorrect/Missing**

HW06-4

a) $L(x_1, y_1) = \sum_{i=1}^{7} \left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)^2$

$$\frac{\partial L(x_1, y_1)}{\partial x_1} = \sum_{i=1}^{7} \frac{-2\left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)(a_i - x_1)}{\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2}}$$

$$\frac{\partial L(x_1, y_1)}{\partial y_1} = \sum_{i=1}^{7} \frac{+2\left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)(b_i - y_1)}{\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2}}$$

b) $x_0 = \vec{0}$ — step size: 0.01 — 473 steps.

stop criterion: gradient $\leq 10^{-6}$

$x_0 = [100, 0]$ → 466 steps.

c) Average unique points goes up,

proportion of global minimum goes down

d) As variance goes down, the unique point small within $x \in (0, 500)$

The proportion goes below 0.5 after 500.

e) as sensor number goes up, unique points become smaller than before

compare to d)

(when variance big, not reach minimum yet within 1000 steps)

proportion becomes smaller than before

4.4 d 3 / 10

+ **10 pts** Correct with reasonable justification
+ **7 pts** Graphs / data but incorrect or missing justification / explanation
✓ + **3 pts** **Claim without justification or graphs**
+ **0 pts** Incorrect claim / no answer

HW06-4

a) $L(x_1, y_1) = \sum_{i=1}^{7} \left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)^2$

$$\frac{\partial L(x_1, y_1)}{\partial x_1} = \sum_{i=1}^{7} \frac{-2\left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)(a_i - x_1)}{\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2}}$$

$$\frac{\partial L(x_1, y_1)}{\partial y_1} = \sum_{i=1}^{7} \frac{+2\left(\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2} - d_i\right)(b_i - y_1)}{\sqrt{(a_i - x_1)^2 + (b_i - y_1)^2}}$$

b) $x_0 = \vec{0}$ — step size: 0.01 — 473 steps.

Stop criterion: gradient $\leq 10^{-6}$

$x_0 = [100, 0]$ → 466 steps.

c) Average unique points goes up,

proportion of global minimum goes down

d) As variance goes down, the unique point small within $x \in (0, 500)$

the proportion goes below 0.5 after 500.

e) as sensor number goes up, unique points become smaller than before

compare to d)

(when variance big, not reach minimum yet within 1000 steps)

proportion becomes smaller than before

4.5 f **0 / 10**

- + **10 pts** Correct
- + **8 pts** Some Mistake
- + **5 pts** Partial
- ✓ + **0 pts** Wrong/Missing

Question 6 USE Triangle Inequality prove Q2-c)

c) Gradient Descent with this step size cannot reach optimal solution.

We need to check: if exists x_i, s.t. $\|x_i - x^*\|_2 \le 0.01$.

First, $f(x_i) = \|x_i - b\|_2$ $\qquad x_i = x_{i-1} - (\frac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2}$

express $f(x_i)$ in terms of x_{i-1}, we have

$$f(x_i) = \left\| x_{i-1} - b - (\tfrac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2} \right\|_2$$

$$f(x_{i-1}) = \|x_{i-1} - b\|_2$$

According to the Triangle Inequality:

$$\left\| x_{i-1} - b - (\tfrac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2} \right\|_2 + \left\| (\tfrac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2} \right\|_2 \ge \|x_{i-1}-b\|_2$$

$$f(x_i) + \left\| (\tfrac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2} \right\|_2 \ge f(x_{i-1})$$

$$f(x_{i-1}) - f(x_i) \le \left\| (\tfrac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2} \right\|_2 = (\tfrac{5}{6})^{i-1}$$

$\frac{a_0(1-q^n)}{1-q}$

$$\left.\begin{array}{ll} i=1 & f(x_0) - f(x_1) \le (\frac{5}{6})^0 \\ & \vdots \\ i=k & f(x_{k-1}) - f(x_k) \le (\frac{5}{6})^{k-1} \end{array}\right\} \Rightarrow f(x_0) - f(x_k) \le \sum_{i=0}^{k} (\tfrac{5}{6})^{i-1} = \frac{1(1-(\frac{5}{6})^{k-1})}{1-\frac{5}{6}} = 6(1-(\tfrac{5}{6})^{k-1})$$

$$\boxed{f(x_k) \ge f(x_0) - 6[1-(\tfrac{5}{6})^{k-1}]}$$

Back to what we need to check: $\|x_i - x^*\|_2 \overset{?}{\le} 0.01 \iff \|x_i - b\|_2 \le 0.01$

$\iff f(x_i) \overset{?}{\le} 0.01$

We know that $f(x_k) \ge f(x_0) - 6[1-(\frac{5}{6})^{k-1}]$ $\quad$ decrease as k increase, minimum is when $k \to \infty$

$$\min_k f(x_k) = \lim_{k\to\infty} f(x_0) - 6[1-(\tfrac{5}{6})^{k-1}] = \sqrt{4.5^2+6^2} - 6 \approx 1.5$$

$\min_k f(x_k) = 1.5 > 0.01$

Thus $\|x_i - x^*\|_2$ can be never within 0.01, cannot find optimal solution.
The proof is for a general case (either $\vec{b}=\vec{0}$ or $\vec{b}\neq\vec{0}$)

5 Vegetables 0 / 0

✓ + 0 pts -

Question 6 Use Triangle Inequality prove Q2-c)

c) Gradient Descent with this step size cannot reach optimal solution.

We need to check: if exists x_i, s.t. $\|x_i - x^*\|_2 \leq 0.01$.

First, $f(x_i) = \|x_i - b\|_2$ $\qquad x_i = x_{i-1} - (\frac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2}$

express $f(x_i)$ in terms of x_{i-1}, we have

$$f(x_i) = \left\| x_{i-1} - b - (\tfrac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2} \right\|_2$$

$$f(x_{i-1}) = \|x_{i-1} - b\|_2$$

According to the Triangle Inequality:

$$\left\| x_{i-1} - b - (\tfrac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2} \right\|_2 + \left\| (\tfrac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2} \right\|_2 \geq \|x_{i-1} - b\|_2$$

$$f(x_i) + \left\| (\tfrac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2} \right\|_2 \geq f(x_{i-1})$$

$$f(x_{i-1}) - f(x_i) \leq \left\| (\tfrac{5}{6})^{i-1} \frac{x_{i-1}-b}{\|x_{i-1}-b\|_2} \right\|_2 = (\tfrac{5}{6})^{i-1}$$

$\frac{a_0(1-q^n)}{1-q}$

$$\left.\begin{array}{ll} i=1 & f(x_0) - f(x_1) \leq (\frac{5}{6})^0 \\ & \vdots \\ i=k & f(x_{k-1}) - f(x_k) \leq (\frac{5}{6})^{k-1} \end{array}\right\} \Rightarrow f(x_0) - f(x_k) \leq \sum_{i=0}^{k} (\tfrac{5}{6})^{i-1} = \frac{1(1-(\frac{5}{6})^{k-1})}{1-\frac{5}{6}} = 6(1-(\tfrac{5}{6})^{k-1})$$

$$\boxed{f(x_k) \geq f(x_0) - 6[1-(\tfrac{5}{6})^{k-1}]}$$

Back to what we need to check: $\|x_i - x^*\|_2 \overset{?}{\leq} 0.01 \iff \|x_i - b\|_2 \leq 0.01$

$\iff f(x_i) \overset{?}{\leq} 0.01$

decrease as k increase
minimum is when $k \to \infty$

We know that $\quad f(x_k) \geq f(x_0) - 6[1-(\frac{5}{6})^{k-1}]$

$$\min_k f(x_k) = \lim_{k\to\infty} f(x_0) - 6[1-(\tfrac{5}{6})^{k-1}] = \sqrt{4.5^2+6^2} - 6 \approx 1.5$$

$$\min_k f(x_k) = 1.5 > 0.01$$

Thus $\|x_i - x^*\|_2$ can be never within 0.01, cannot find optimal solution.
The proof is for a general case (either $\vec{b} = \vec{0}$ or $\vec{b} \neq \vec{0}$)

6 Write Your Own Question 10 / 10

✓ + 10 pts Correct

+ 0 pts No Answer