CSE257 - Search and Optimization - Assignment 1

Saqib Azim (A59010162)
UC San Diego
sazim@ucsd.edu

October 24, 2021

1 Question 1

1.1 $f_1(x_1, x_2) = (x_1 - x_2)^2$

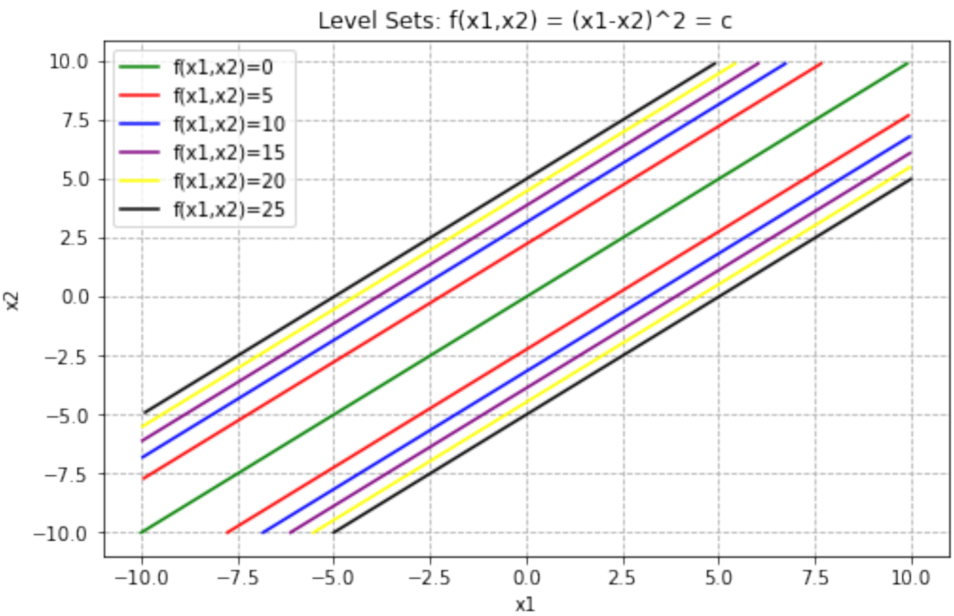

Figure 1: Level Set Plot for $f_1(x_1, x_2) = (x_1 - x_2)^2$

1.2 $f_2(x_1, x_2) = (x_1^2 - 3x_2^2)$

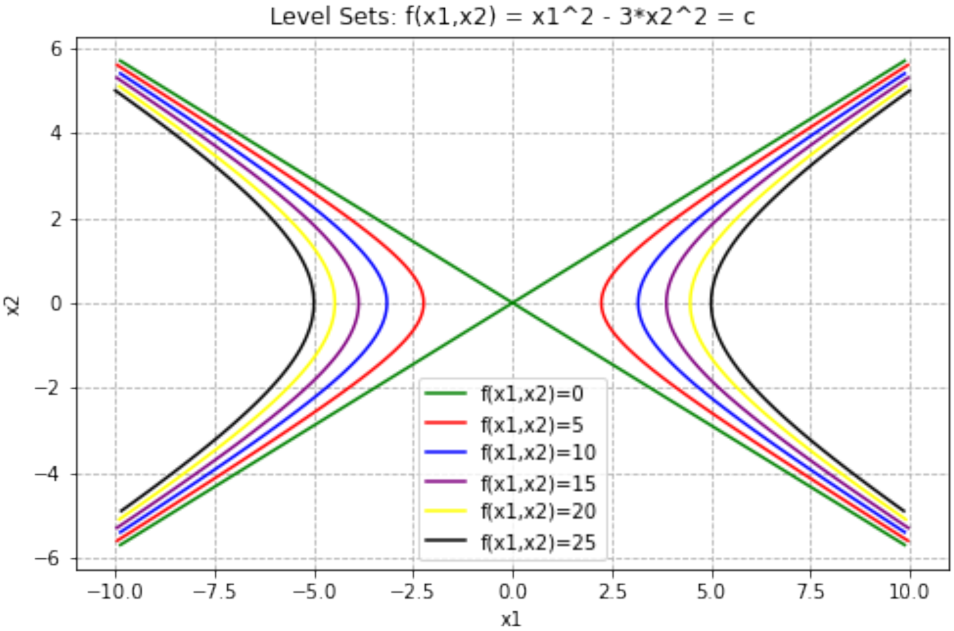

Figure 2: Level Set Plot for $f_2(x_1, x_2) = (x_1^2 - 3x_2^2)$

1.3 $f_3(x_1, x_2, x_3) = (x_1^2 + 5x_2^2 + 3x_3^2)$

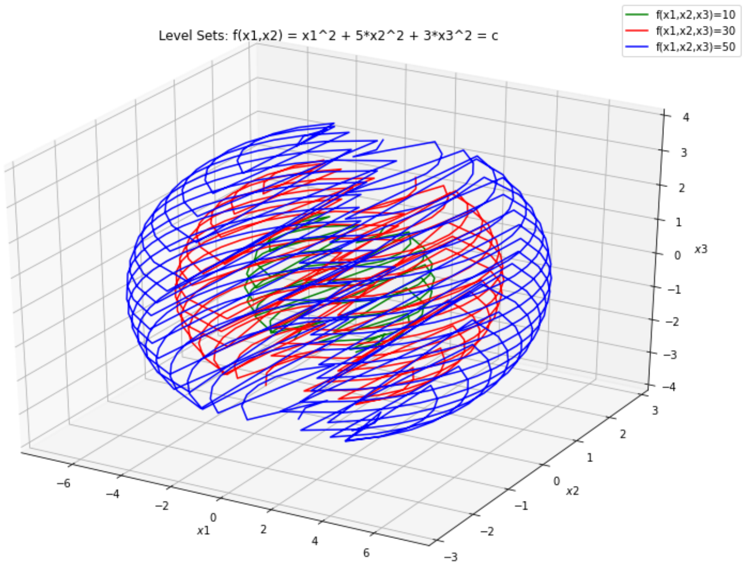

Figure 3: Level Set Plot for $f_3(x_1, x_2, x_3) = (x_1^2 + 5x_2^2 + 3x_3^2)$

2 Question 2

2.1 Gradient Descent with Fixed Step Size $\alpha = 0.3$

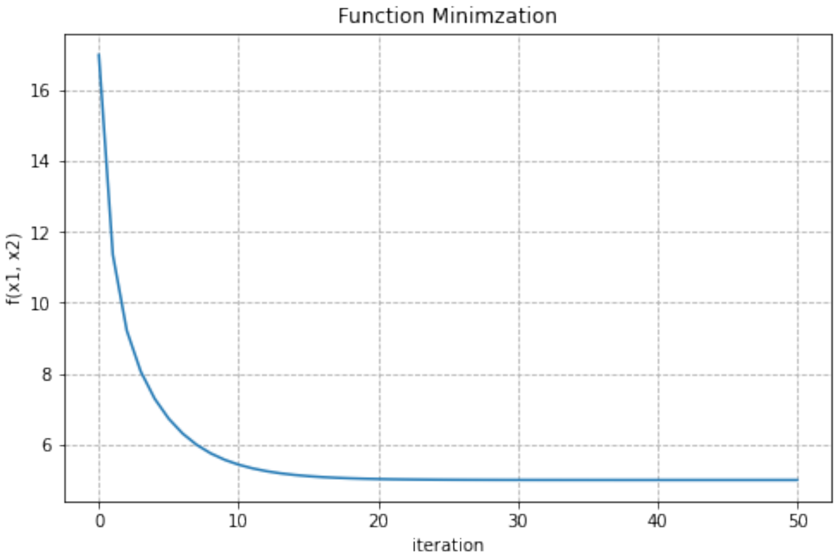

Figure 4: Plot of $f(x_1, x_2)$ with iterations

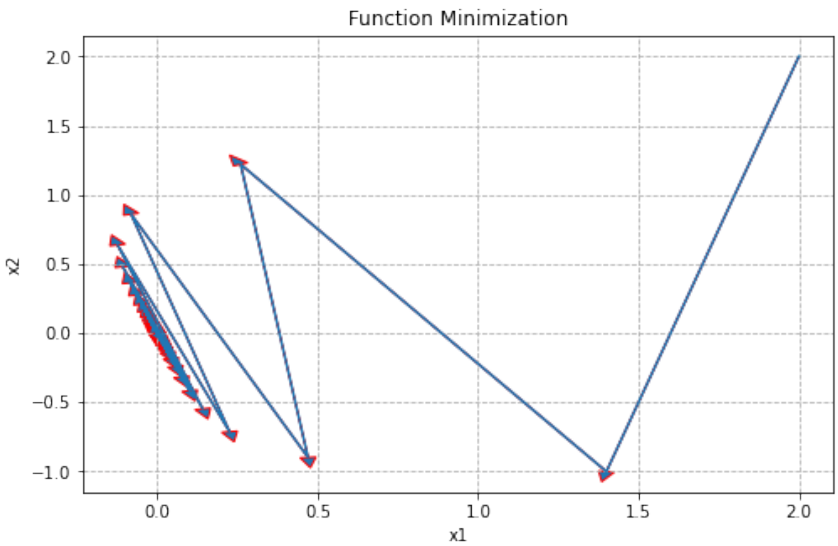

Figure 5: Plot of sequence of points and gradient descent direction

2.2 Gradient Descent with Optimal Step Size

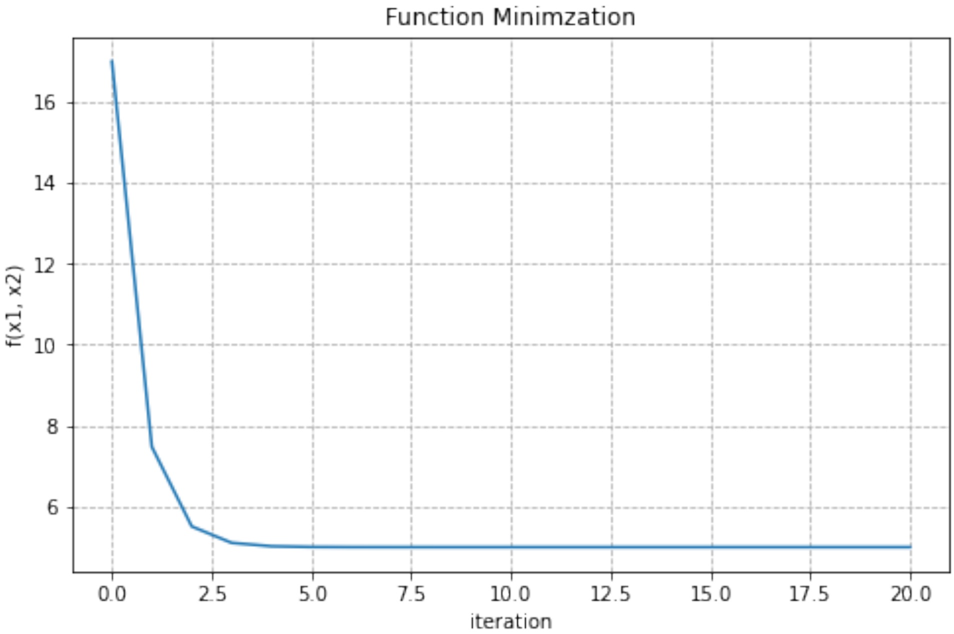

Figure 6: Plot of $f(x_1, x_2)$ with iterations

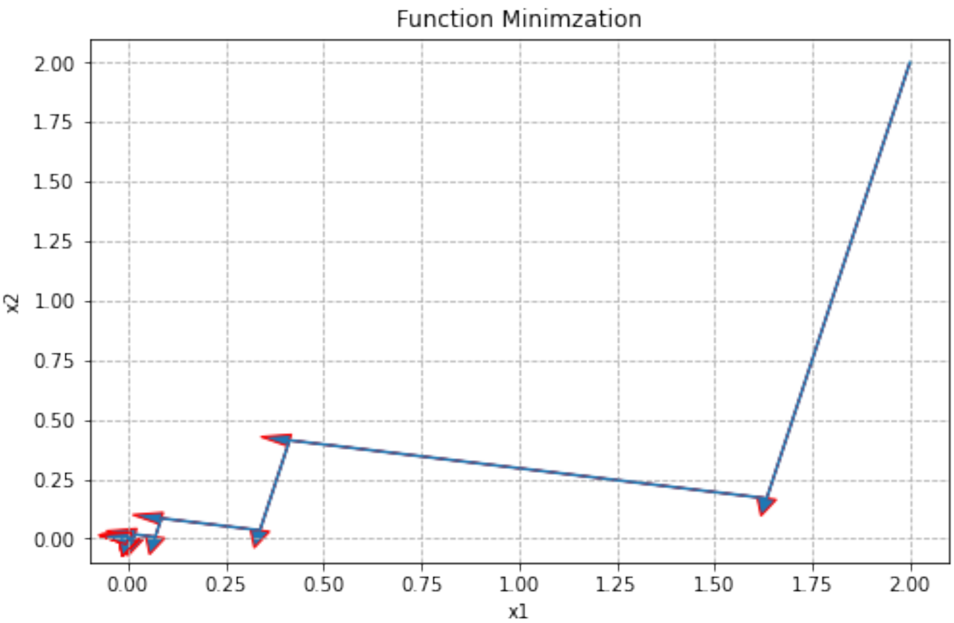

Figure 7: Plot of sequence of points and gradient descent direction

2.3 Newton Descent with Step Size $\alpha = 1$

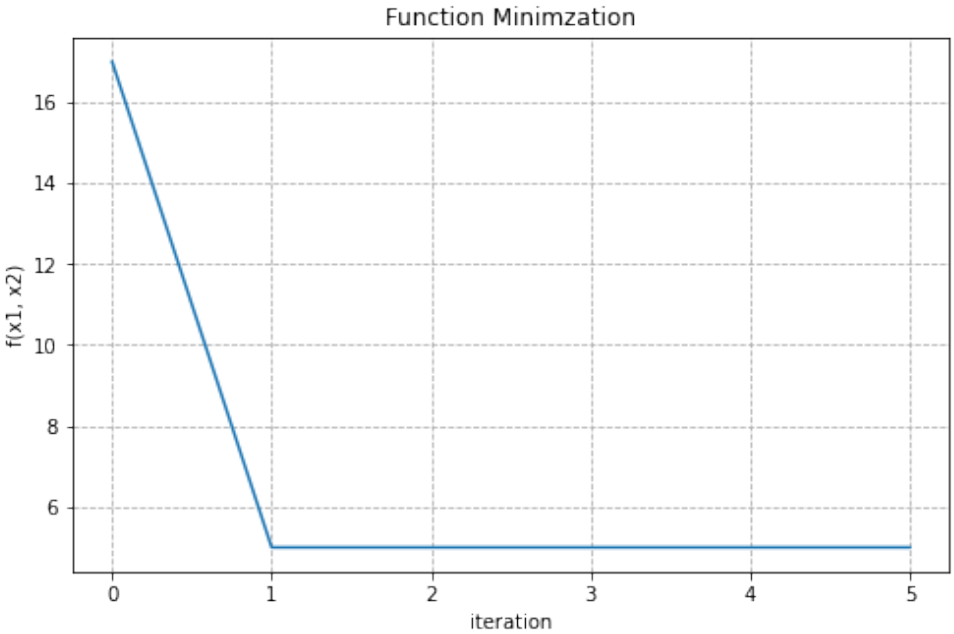

Figure 8: Plot of $f(x_1, x_2)$ with iterations

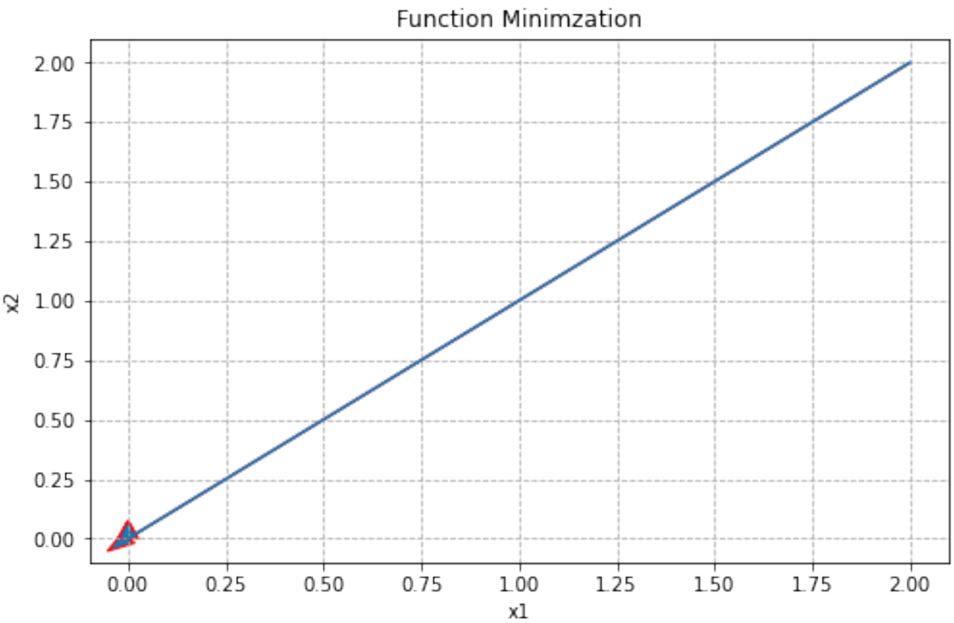

Figure 9: Plot of sequence of points and gradient descent direction

3 Question 4

3.1 Gradient Descent to optimize $L(A, B)$

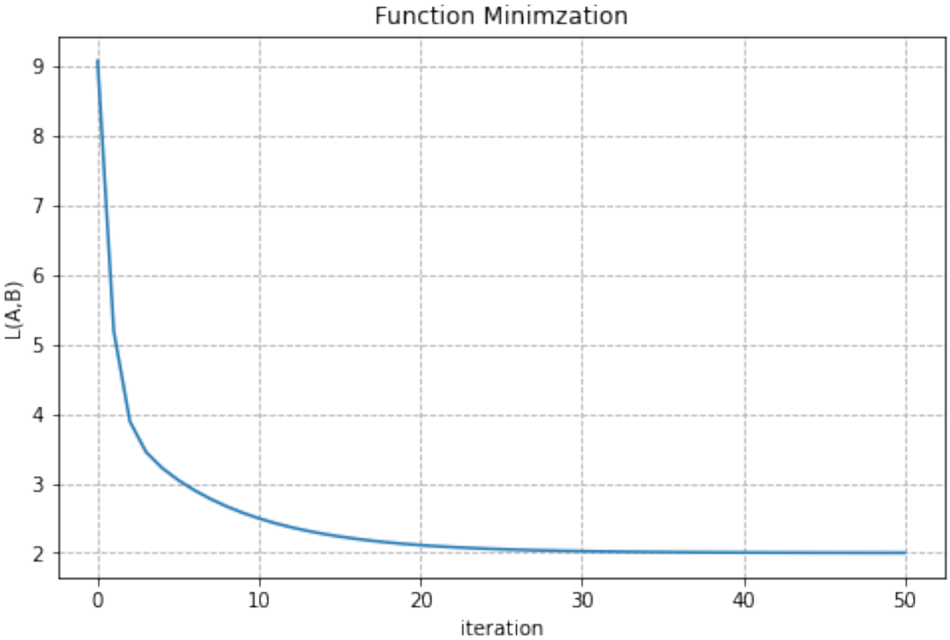

Figure 10: Plot of Loss value $L(A, B)$ with iterations

```
Minimum L(A,B) = 2.00145 after 50 iterations of gradient descent with alpha = 0.01
 A =
 [[-0.04640704 -0.01967163  0.02325636]
 [-0.38630694 -0.08384057 -0.80668798]
 [-0.36242432  0.64452212 -0.06844823]]
 B =
 [[ 0.01599    ]
 [-0.02592165]
 [ 0.27878116]]
```

Figure 11: Loss value after 50 iterations. Optimal A and B

(4) Model $f_{A,B}(x) = Ax + B \quad x \in \mathbb{R}^{3\times1} \quad f(x) \in \mathbb{R}^{3\times1} \quad A \in \mathbb{R}^{3\times3} \quad B \in \mathbb{R}^{3\times1}$

$$L(A,B) = \sum_{i=1}^{3} \| f_{A,B}(x_i) - y_i \|_2^2$$

$$= \sum_{i=1}^{3} \| Ax_i + B - y_i \|_2^2$$

$$= \sum_{i=1}^{3} (Ax_i + B - y_i)^T (Ax_i + B - y_i)$$

$$= \sum_{i=1}^{3} (x_i^T A^T + B^T - y_i^T)(Ax_i + B - y_i)$$

$$= \sum_{i=1}^{3} \left(x_i^T A^T A x_i + 2x_i^T A^T B - 2x_i^T A^T y_i - 2B^T y_i + B^T B + y_i^T y_i \right)$$

$$L(A,B) = \sum_{i=1}^{3} \left(x_i^T A^T A x_i + 2B^T A x_i - 2y_i^T A x_i - 2y_i^T B + B^T B + y_i^T y_i \right)$$

$$\frac{\delta L}{\delta A} = 2 \sum_{i=1}^{3} \left(A x_i x_i^T + B x_i^T - y_i x_i^T \right) \in \mathbb{R}^{3\times3}$$

$$\frac{\delta L}{\delta B} = 2 \sum_{i=1}^{3} (Ax_i - y_i + B) \in \mathbb{R}^{3\times1}$$

let $A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \qquad B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

$$\frac{\delta L}{\delta B} = 2 \sum_{i=1}^{3} \begin{bmatrix} a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} - y_{i1} + b_1 \\ a_4 x_{i1} + a_5 x_{i2} + a_6 x_{i3} - y_{i2} + b_2 \\ a_7 x_{i1} + a_8 x_{i2} + a_9 x_{i3} - y_{i3} + b_3 \end{bmatrix}$$

$$\frac{\delta L}{\delta A} = 2\sum_{i=1}^{3}\left\{ \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix}\begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{bmatrix}\begin{bmatrix} x_{i1} & x_{i2} & x_{i3} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}\begin{bmatrix} x_{i1} & x_{i2} & x_{i3} \end{bmatrix} - \begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix}\begin{bmatrix} x_{i1} & x_{i2} & x_{i3} \end{bmatrix} \right\}$$

$$\frac{\delta L}{\delta A} = 2\sum_{i=1}^{3}\left\{ \begin{bmatrix} a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} \\ a_4 x_{i1} + a_5 x_{i2} + a_6 x_{i3} \\ a_7 x_{i1} + a_8 x_{i2} + a_9 x_{i3} \end{bmatrix}\begin{bmatrix} x_{i1} & x_{i2} & x_{i3} \end{bmatrix} + \begin{bmatrix} b_1 x_{i1} & b_1 x_{i2} & b_1 x_{i3} \\ b_2 x_{i1} & b_2 x_{i2} & b_2 x_{i3} \\ b_3 x_{i1} & b_3 x_{i2} & b_3 x_{i3} \end{bmatrix} - \begin{bmatrix} y_{i1} x_{i1} & y_{i1} x_{i2} & y_{i1} x_{i3} \\ y_{i2} x_{i1} & y_{i2} x_{i2} & y_{i2} x_{i3} \\ y_{i3} x_{i1} & y_{i3} x_{i2} & y_{i3} x_{i3} \end{bmatrix} \right.$$

$$\nabla L(A,B) = \begin{bmatrix} \frac{\delta L}{\delta a_1} & \frac{\delta L}{\delta a_2} & \frac{\delta L}{\delta a_3} & \frac{\delta L}{\delta a_4} & \frac{\delta L}{\delta a_5} & \frac{\delta L}{\delta a_6} & \frac{\delta L}{\delta a_7} & \frac{\delta L}{\delta a_8} & \frac{\delta L}{\delta a_9} & \frac{\delta L}{\delta b_1} & \frac{\delta L}{\delta b_2} & \frac{\delta L}{\delta b_3} \end{bmatrix}^T$$

$$\frac{\delta L}{\delta a_1} = 2\sum_{i=1}^{3} (a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3}) x_{i1} + b_1 x_{i1} - y_{i1} x_{i1}$$

$$\frac{\delta L}{\delta a_2} = 2\sum_{i=1}^{3} (a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3}) x_{i2} + b_1 x_{i2} - y_{i1} x_{i2}$$

$$\frac{\delta L}{\delta a_3} = 2\sum_{i=1}^{3} (a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3}) x_{i3} + b_1 x_{i3} - y_{i1} x_{i3}$$

$$\frac{\delta L}{\delta a_4} = 2\sum_{i=1}^{3} (a_4 x_{i1} + a_5 x_{i2} + a_6 x_{i3}) x_{i1} + b_2 x_{i1} - y_{i2} x_{i1}$$

$$\frac{\delta L}{\delta a_5} = 2\sum_{i=1}^{3} (a_4 x_{i1} + a_5 x_{i2} + a_6 x_{i3}) x_{i2} + b_2 x_{i2} - y_{i2} x_{i2}$$

$$\frac{\delta L}{\delta a_6} = 2\sum_{i=1}^{3} (a_4 x_{i1} + a_5 x_{i2} + a_6 x_{i3}) x_{i3} + b_2 x_{i3} - y_{i2} x_{i3}$$

$$\frac{\delta L}{\delta a_7} = 2\sum_{i=1}^{3} (a_7 x_{i1} + a_8 x_{i2} + a_9 x_{i3}) x_{i1} + b_3 x_{i1} - y_{i3} x_{i1}$$

$$\frac{\delta L}{\delta a_8} = 2\sum_{i=1}^{3} (a_7 x_{i1} + a_8 x_{i2} + a_9 x_{i3}) x_{i2} + b_3 x_{i2} - y_{i3} x_{i2}$$

$$\frac{\delta L}{\delta a_9} = 2\sum_{i=1}^{3} (a_7 x_{i1} + a_8 x_{i2} + a_9 x_{i3}) x_{i3} + b_3 x_{i3} - y_{i3} x_{i3}$$

$$\frac{\delta L}{\delta b_1} = 2\sum_{i=1}^{3} a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} - y_{i1} + b_1$$

$$\frac{\delta L}{\delta b_2} = 2\sum_{i=1}^{3} a_4 x_{i1} + a_5 x_{i2} + a_6 x_{i3} - y_{i2} + b_2$$

$$\frac{\delta L}{\delta b_3} = 2\sum_{i=1}^{3} a_7 x_{i1} + a_8 x_{i2} + a_9 x_{i3} - y_{i3} + b_3.$$

$$\nabla^2 L(A,B) = 2\sum_{i=1}^{3} \begin{bmatrix}
x_{i1}^2 & x_{i1}x_{i2} & x_{i1}x_{i3} & 0 & 0 & 0 & 0 & 0 & 0 & x_{i1} & 0 & 0 \\
x_{i1}x_{i2} & x_{i2}^2 & x_{i2}x_{i3} & 0 & 0 & 0 & 0 & 0 & 0 & x_{i2} & 0 & 0 \\
x_{i1}x_{i3} & x_{i2}x_{i3} & x_{i3}^2 & 0 & 0 & 0 & 0 & 0 & 0 & x_{i3} & 0 & 0 \\
0 & 0 & 0 & x_{i1}^2 & x_{i1}x_{i2} & x_{i1}x_{i3} & 0 & 0 & 0 & 0 & x_{i1} & 0 \\
0 & 0 & 0 & x_{i1}x_{i2} & x_{i2}^2 & x_{i2}x_{i3} & 0 & 0 & 0 & 0 & x_{i2} & 0 \\
0 & 0 & 0 & x_{i1}x_{i3} & x_{i2}x_{i3} & x_{i3}^2 & 0 & 0 & 0 & 0 & x_{i3} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & x_{i1}^2 & x_{i1}x_{i2} & x_{i1}x_{i3} & 0 & 0 & x_{i1} \\
0 & 0 & 0 & 0 & 0 & 0 & x_{i1}x_{i2} & x_{i2}^2 & x_{i2}x_{i3} & 0 & 0 & x_{i2} \\
0 & 0 & 0 & 0 & 0 & 0 & x_{i1}x_{i3} & x_{i2}x_{i3} & x_{i3}^2 & 0 & 0 & x_{i3} \\
x_{i1} & x_{i2} & x_{i3} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & x_{i1} & x_{i2} & x_{i3} & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & x_{i1} & x_{i2} & x_{i3} & 0 & 0 & 1
\end{bmatrix}$$

The Hessian $\nabla^2 L(A,B)$ is a 12×12 matrix corresponding to the 12 parameters in A and B combined.

We computed the Hessian Matrix using python code.

$$
\nabla^2 L(A,B) = 2 \times \begin{bmatrix}
3 & 2 & 7 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
2 & 12 & 10 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 0 & 0 \\
7 & 10 & 19 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 \\
0 & 0 & 0 & 3 & 2 & 7 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 2 & 12 & 10 & 0 & 0 & 0 & 0 & 6 & 0 \\
0 & 0 & 0 & 7 & 10 & 19 & 0 & 0 & 0 & 0 & 5 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 3 & 2 & 7 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 2 & 12 & 10 & 0 & 0 & 6 \\
0 & 0 & 0 & 0 & 0 & 0 & 7 & 10 & 19 & 0 & 0 & 5 \\
1 & 6 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 6 & 5 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 6 & 5 & 0 & 0 & 1
\end{bmatrix}_{12 \times 12}
$$

$\boxed{4.3}$

$$\det(\nabla^2 L(A,B)) = 1.1 \times 10^{-39} \simeq 0$$ (computed using numpy python).

Therefore, the inverse of the Hessian Matrix does not exist

In Newton descent method, there is an assumption that $\nabla^2 L(A,B) > 0$ has to be positive definite and Newton direction ~~$p = -(\nabla^2 f(x))$~~

$$p = -(\nabla^2 L(A,B))^{-1} \nabla L(A,B)$$

Since the inverse does not exist, Newton directions cannot be used for this minimization.

④ ① Prove that any local minimum of $L(A, B)$ is also a global minimum of $L(A, B)$

Proof: In Question ③ of this assignment, we proved that if a function is convex, then any of its local minima is also a global minima.

Therefore, if we can prove that $L(A, B)$ is a convex function, then we can use the result from Question ③ to show that any local minimum of $L(A, B)$ is also a global minimum.

① To prove : $L(A, B) = \sum_{i=1}^{3} \|f_{A,B}(x_i) - y_i\|_2^2$ is a convex function

② Already Proven : for a convex function, any local minima is a global minima

Proof ① : $L(A, B)$ can be viewed as a composition of multiple functions summation, L2-norm, square.

We will first prove that norms are convex.

Let $\|.\|$ be a norm on a vector space V. Then for all $x, y \in V$ and $\lambda \in [0, 1]$

$\|\lambda x + (1-\lambda) y\| \leq \|\lambda x\| + \|(1-\lambda) y\|$ [∵ using triangle inequality]

$\leq |\lambda| \|x\| + |1-\lambda| \|y\|$ [∵ Norms are homogeneous]

$\leq \lambda \|x\| + (1-\lambda) \|y\|$ [∵ $\lambda \geq 0$, $(1-\lambda) \geq 0$]

Thus,
$$\|\lambda x + (1-\lambda) y\| \leq \lambda \|x\| + (1-\lambda)\|y\|$$

The above condition ~~hold~~ holds for any $x, y \in V$ (vector space) and $\lambda \in [0,1]$. which proves that norms are convex function.

Now, consider $g(x) = x^2$, $x \in V$ vector space.

Let $x_1, x_2 \in V$. then, for any $\lambda \in [0,1]$

$$g(\lambda x_1 + (1-\lambda) x_2) = [\lambda x_1 + (1-\lambda) x_2]^2$$
$$= \lambda^2 x_1^2 + (1-\lambda)^2 x_2^2 + 2\lambda(1-\lambda) x_1 x_2$$

$$[\lambda x_1 + (1-\lambda) x_2]^2 = \lambda^2 x_1^2 + (1-\lambda)^2 x_2^2 + 2\lambda(1-\lambda) x_1 x_2$$

$$[\lambda x_1 + (1-\lambda) x_2]^2 - (\lambda x_1^2 + (1-\lambda) x_2^2) = \lambda^2 x_1^2 + (1-\lambda)^2 x_2^2 + 2\lambda(1-\lambda) x_1 x_2 - (\lambda x_1^2 + (1-\lambda) x_2^2)$$

subtracting $(\lambda x_1^2 + (1-\lambda) x_2^2)$ from both LHS & RHS.

$$[\lambda x_1 + (1-\lambda) x_2]^2 - (\lambda x_1^2 + (1-\lambda) x_2^2) = (\lambda^2 - \lambda) x_1^2 + 2 x_1 x_2 \lambda(1-\lambda) + (\lambda^2 - \lambda) x_2^2$$
$$= \lambda(1-\lambda)[-x_1^2 + 2x_1 x_2 - x_2^2]$$
$$= -\lambda(1-\lambda)(x_1 - x_2)^2 \leq 0$$

Since $\lambda \geq 0$ and $(1-\lambda) \geq 0$.

$$[\lambda x_1 + (1-\lambda) x_2]^2 \leq \lambda x_1^2 + (1-\lambda) x_2^2 \Rightarrow g(x) = x^2 \text{ is a convex function.}$$

Now we want to prove that for any $x, y \in \Omega$, if f and g are convex functions, and

g is non-decreasing in the range of f, then $g \circ f$ is also a convex function.

Proof:

$$g \circ f(\lambda x + (1-\lambda)y) = g[f(\lambda x + (1-\lambda)y)]$$

$$g[f(\lambda x + (1-\lambda)y)]$$

Since f is convex $\Rightarrow f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$

Since g is non-decreasing function in the range of f.

$$g[f(\lambda x + (1-\lambda)y)] \leq g[\lambda f(x) + (1-\lambda)f(y)]$$

$$\leq \lambda g(f(x)) + (1-\lambda)g(f(y)) \quad [\because g \text{ is convex}]$$

$$\leq \lambda g \circ f(x) + (1-\lambda) g \circ f(y) \quad \forall \lambda \in [0,1]$$

Now, let g be the function $g(x) = x^2$ and $f(x) = ||x||$

Both f and g are convex and g is non-decreasing in the range of f.

Thus, $g \circ f(x) = ||x||^2$ is a convex function for any $x \in V$. where V is a vector space.

finally the sum of convex functions is again convex.

~~Let f and g be convex f~~

Let h_1 and h_2 be convex functions.

$$(h_1+h_2)(\lambda x+(1-\lambda)y) = h_1(\lambda x+(1-\lambda)y) + h_2(\lambda x+(1-\lambda)y)$$

$$\leq \lambda h_1(x) + (1-\lambda)h_1(y) + \lambda h_2(x) + (1-\lambda)h_2(y)$$

$$\leq \lambda(h_1+h_2)(x) + (1-\lambda)(h_1+h_2)y.$$

Thus, (h_1+h_2) is a convex function.

Now,

$$L(A,B) = \sum_{i=1}^{3} \| f_{A,B}(x_i) - y_i \|_2^2$$

$$L(A,B) = h_1 + h_2 + h_3 \quad \text{where } h_i = \| f_{A,B}(x_i) - y_i \|_2^2$$

Now, $h_i = g(f(x_i))$

where $g(x) = x^2 \longrightarrow$ convex

$f(x) = \|x\| \longrightarrow$ convex

$g\circ f(x) = g(f(x)) \longrightarrow$ convex

$h_i \longrightarrow$ convex.

$L(A,B) = h_1 + h_2 + h_3 \longrightarrow$ convex.

Thus, $L(A,B)$ is a convex function.

and it follows from proof in question 3 that any local minimum of $L(A,B)$ is also a global minimum.

③ Prove that if a function is convex, then any of its local minimum is also a global minimum.

Proof: Let f be a convex function and Ω be a convex set.

Let x^* be a local minima of f which implies

$x^* \in \Omega$ and $\exists\ \varepsilon > 0$ such that $f(x^*) \leq f(x)$

$$\forall \quad x \in \underbrace{B(x^*, \varepsilon)}_{\text{neighborhood of } x^*.}$$

Now, Assume for the sake of contradiction that

$\exists\ x_0 \in \Omega$ with $f(x_0) < f(x^*)$ ——— ①

Since Ω is a convex set and x^* and x_0 belong to Ω., they should satisfy the convexity property.

$$(\lambda x^* + (1-\lambda) x_0) \in \Omega \qquad \forall\ \lambda \in [0,1]$$

In addition, since f is convex function in set Ω

Convexity of $f \Rightarrow$ $f(\lambda x^* + (1-\lambda) x_0) \leq \lambda f(x^*) + (1-\lambda) f(x_0)$

Using $f(x_0) < f(x^*)$ from ①

$$f(\lambda x^* + (1-\lambda) x_0) < \lambda f(x^*) + (1-\lambda) f(x^*)$$

$$\Rightarrow \quad f(\lambda x^* + (1-\lambda) x_0) < f(x^*) \qquad \text{——— ②}$$

As $\lambda \to 1 \Rightarrow (\lambda x^* + (1-\lambda) x_0) \to x^* \Rightarrow \lim_{\lambda \to 1} (\lambda x^* + (1-\lambda) x_0) \in B(x^*, \varepsilon)$ for some $\varepsilon > 0$.

Eqn. 2 is valid for all $\lambda \in [0,1]$

$$\lim_{\lambda \to 1} f\left[\lambda x^* + (1-\lambda) x_0\right] < \lim_{\lambda \to 1} f(x^*)$$

$$f(x^*) < f(x^*)$$

Thus, as $\lambda \to 1$, eqn. 2 contradicts the local optimality of $f(x)$ at $x = x^*$.

which means, our assumption that there exists an $x_0 \in \Omega$ with $f(x_0) < f(x^*)$ cannot be true for a convex function $f(x)$ with local minima at x^*.

$\Rightarrow$ x^* is the global minima of f and there cannot exist an $x_0 \in \Omega$ for which $f(x_0) < f(x^*)$

$\Rightarrow$ Any local minima for f is also a global minima if f is a convex function

Proved.

⑤ Prove that gradient descent with exact line search takes always orthogonal steps.

Proof:

Gradient descent + Exact line search.

iteration :	1	2	3	. . . . . .	$k-1$	k	$k+1$
direction :	p_1	p_2	p_3	. . . .	p_{k-1}	p_k	p_{k+1}
step size :	α_1	α_2	α_3	. . . .	α_{k-1}	α_k	α_{k+1}.

Let p_k be the gradient descent (GD) direction in the kth iteration. $p_k \in R^n$, $x_k \in R^n$ and $\alpha_k \in R$, $\alpha_k \geq 0$

$f(x) : R^n \to R$ is a continuously differentiable and lower-bounded function.

In gradient descent, at each step, direction $p_k = -\nabla f(x_k)$

In exact line search, at each step, after computing the direction p_k, we find the optimal α_k that minimizes $f(x_k)$ along that direction.

$$\Rightarrow \quad \underbrace{\alpha_k = \underset{\alpha \geq 0}{\operatorname{argmin}} \; f(x_k + \alpha p_k)} = \underset{\alpha \geq 0}{\operatorname{argmin}} \; f(x_k + \alpha(-\nabla f(x_k)))$$

Using the 1st order necessary condition for local minima to compute α_k.

$$\frac{\delta}{\delta \alpha} f(x_k + \alpha p_k) \Big|_{\alpha = \alpha_k} = 0$$

$$\Rightarrow \quad \nabla f(x_k + \alpha p_k)^T p_k \Big|_{\alpha = \alpha_k} = 0$$

$$\Rightarrow \nabla f(x_k + \alpha_k p_k)^T p_k = 0 \quad \text{——— (1)}$$

Gradient update step: $x_{k+1} = x_k + \alpha_k p_k$. ——— (2)

We need to prove that successive directions (say p_k and p_{k+1}) in gradient descent with exact line search are always orthogonal.

$\Rightarrow \boxed{p_{k+1}^T p_k = 0} \longrightarrow$ to prove. $\forall$ iteration steps.

$$\begin{aligned} LHS = p_{k+1}^T p_k &= (-\nabla f(x_{k+1}))^T p_k \\ &= -\nabla f(x_{k+1})^T p_k \\ &= -\nabla f(x_k + \alpha_k p_k)^T p_k \quad [\because \text{using } (2)] \\ &= 0 \quad [\because \text{using } (1)] \end{aligned}$$

Proved.

6 follow the proof sketches in slides and prove the 1st order and 2nd order necessary conditions for local minima

Proof for 1st order local minima – necessary conditions.

→ If x^* is a local minimizer and f is C^1 (continuously differentiable) in an open neighborhood of x^* then $\nabla f(x^*) = 0$.

Let $f: R^n \to R$ and f is continuously differentiable. C^1.

Let $x^* \in R^n$ is a local minima of f but $\nabla f(x^*) \neq \vec{0}$

we will prove by contradiction that if $\nabla f(x^*) \neq 0$ then x^* cannot be a local minima

Consider the direction $p = -\nabla f(x^*) \in R^n$.

Then $\nabla f(x^*)^T p = -\|\nabla f(x^*)\|_2^2 < 0$ —— 1

Now, consider the neighborhood of x^* i.e, $B(x^*, \varepsilon)$ where $\varepsilon > 0$ such that $\forall x \in B(x^*, \varepsilon)$, $f(x^*) \leq f(x)$ as x^* is a local minima of f.

for a point in the neighborhood of x^* and in the direction p from x^*, we can write the Taylor series expansion around x^*.

$$\underbrace{f(x^* + \lambda p) = f(x^*) + \nabla f(x^*)^T . \lambda p}_{\text{1st order truncated Taylor series approximation.}} + \underbrace{o(\|\lambda p\|_2^2)}_{\text{higher order terms.}}$$

~~for some λ, where $\lambda \in [0,1]$ $\lambda \in (0,1)$~~

for any $\lambda \in (0, \varepsilon)$ for some $\varepsilon > 0$ s.t $(x^* + \lambda p) \in B(x^*, \varepsilon)$

$$f(x^* + \lambda p) = f(x^*) + \lambda \nabla f(x^*)^T . p$$

Using direction $p = -\nabla f(x^*)$

$$\Rightarrow f(x^* + \lambda p) = f(x^*) - \lambda \nabla f(x^*)^T \nabla . f(x^*)$$

$$\Rightarrow f(x^* + \lambda p) = f(x^*) - \underbrace{\lambda \| \nabla f(x^*) \|_2^2}_{>0} \quad [\because \text{from } 1]$$

$$\Rightarrow f(x^* + \lambda p) - f(x^*) = -\lambda \| \nabla f(x^*) \|_2^2 < 0 \quad \forall \lambda \in (0,1)$$

$$\Rightarrow f(x^* + \lambda p) < f(x^*)$$

That is, there exists a point $(x^* + \lambda p)$ in the neighborhood of x^* ~~such that~~ $B(x^*, \varepsilon)$ for some $\varepsilon > 0$ such that

$$f(x^* + \lambda p) < f(x^*)$$

which implies that x^* cannot be a local minima for $f(x)$ if

$$\nabla f(x^*) \neq \vec{0}$$

$\Rightarrow$ for $x^* \in R^n$ to be a local minima of f, $\nabla f(x^*) = 0$

Proof of 2nd order necessary condition for local minima.

→ If x^* is a local minimizer and f is C^2 in an open neighborhood of x^* then,

$$\nabla f(x^*) = 0 \quad \text{and} \quad \underbrace{\nabla^2 f(x^*)}_{\text{positive semi-definite.}} \geq 0$$

If x^* is a local minimizer then $\nabla f(x^*) = 0$ [already proved in 1st part]

↳ we will assume this to be proven and true from hence forth.

Let $\nabla^2 f(x^*) < 0$. Then we will prove by contradiction that x^* cannot be a local minimizer of f.

$\nabla^2 f(x^*) < 0 \Rightarrow$ for any vector $p \in R^n$ and $p \neq 0$

$$p^T \nabla^2 f(x^*) p < 0 \quad \text{———} \quad (1)$$

Now, consider the neighborhood of x^*, $B(x^*, \varepsilon)$ for some $\varepsilon > 0$ such that for all $x \in B(x^*, \varepsilon)$, $f(x^*) \leq f(x)$ as we have assumed that x^* is a local minima of f.

for any point in the neighborhood of x^* and in the direction $p \in R^n$ from x^*, we can write Taylor series expansion centered around x^*.

$$f(x^* + \lambda p) = f(x^*) + \nabla f(x^*)^T \lambda p + \frac{1}{2} \lambda p^T \nabla^2 f(x^*) \lambda p. + \text{higher order terms.}$$

Using 2nd order truncated Taylor series approximation for f.

for any λ, where $\lambda \in (0, \varepsilon)$ for some $\varepsilon > 0$ such that $(x^* + \lambda p) \in B(x^*, \varepsilon)$

Using 1st order necessary condition for local minima

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

$$\Rightarrow f(x^* + \lambda p) = f(x^*) + 0 + \frac{1}{2}\lambda^2 p^T \nabla^2 f(x^*) p$$

Since $p^T \nabla^2 f(x^*) p < 0 \quad \forall p \in R^n$ [Assumption]

$$\frac{1}{2}\lambda^2 p^T \nabla^2 f(x^*) p < 0 \quad \forall p \in R^n \ \forall \lambda \in (0, \epsilon) \text{ for some } \epsilon > 0$$

Thus we have

$$f(x^* + \lambda p) - f(x^*) < 0$$

$$\Rightarrow f(x^* + \lambda p) < f(x^*)$$

That is $\exists$ a point $(x^* + \lambda p)$ in the neighborhood of x^* $B(x^*, \epsilon)$

such that ~~The above~~

$$f(x^* + \lambda p) < f(x^*)$$

$$\Downarrow$$

x^* cannot be a local minima for fem if ~~$\nabla f(x^*) \neq 0$~~

$\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) < 0$

$\Rightarrow$ for $x^* \in R^n$ to be a local minima, $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \geq 0$