Optimization Methods - A2

Naitik Kariwal (2023201044)

April 8, 2025

1 Trid Function

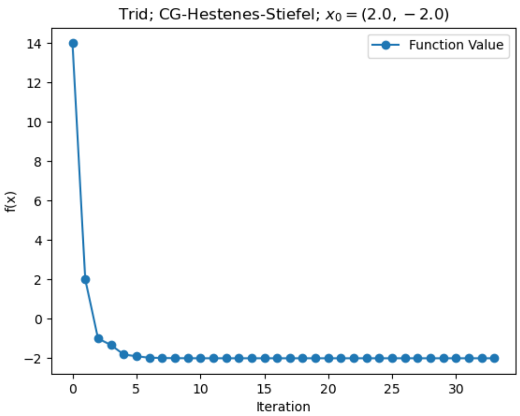
The Trid function is defined as:

$$f(\mathbf{x}) = \sum_{i=1}^{n}(x_i - 1)^2 - \sum_{i=2}^{n} x_i x_{i-1} \tag{1}$$

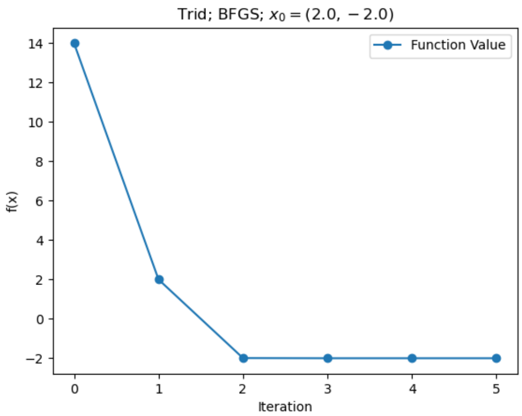
1.1 Plots

Below are the plots illustrating the optimization process:

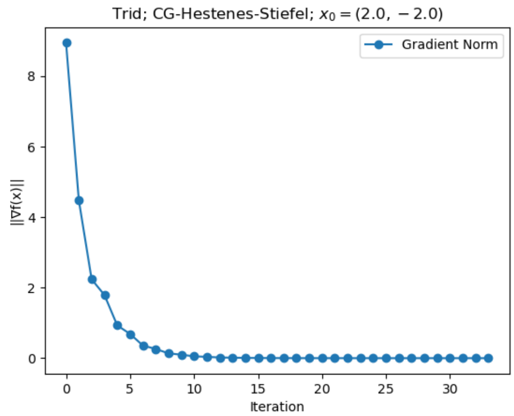
1. Plot of function values $f(x)$ vs. iterations.
2. Plot of gradient norm $\|\nabla f(x)\|$ vs. iterations.
3. Contour plot with the optimization path.



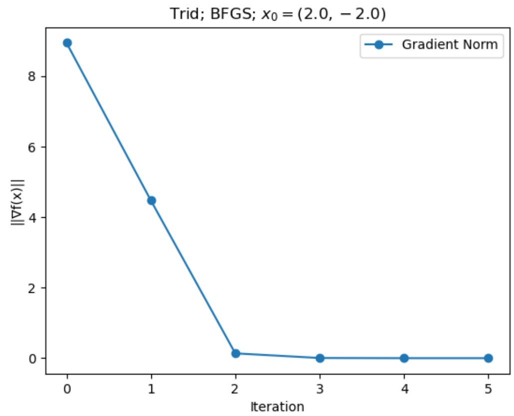
(a) $f(x)$ vs. iterations (Backtracking Armijo)



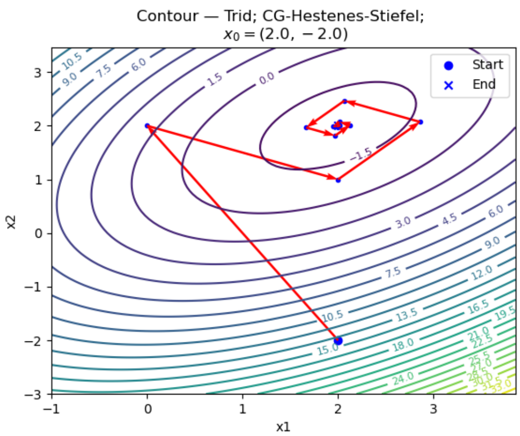
(b) $f(x)$ vs. iterations (Pure Newton's)



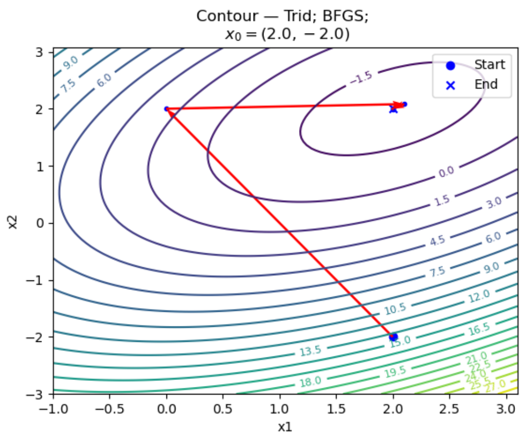
(a) $\|\nabla f(x)\|$ vs. iterations (Backtracking Armijo)



(b) $\|\nabla f(x)\|$ vs. iterations (Pure Newton's)



(a) Contour plot (Backtracking Armijo)



(b) Contour plot (Pure Newton's)

1.2 Results

Table 1: Trid Function

Test	Conjugate Gradient: HS Quasi-Newton: SR1	PR DFP	FR BFGS
0	[2. 2.] [2. 2.]	[2. 2.] [2. 2.]	[2. 2.] [2. 2.]
1	[2. 2.] [2. 2.]	[2. 2.] [2. 2.]	[2. 2.] [2. 2.]

Comments: All the algorithms led to convergence to the unique minimum.

2 Three-Hump Camel Function

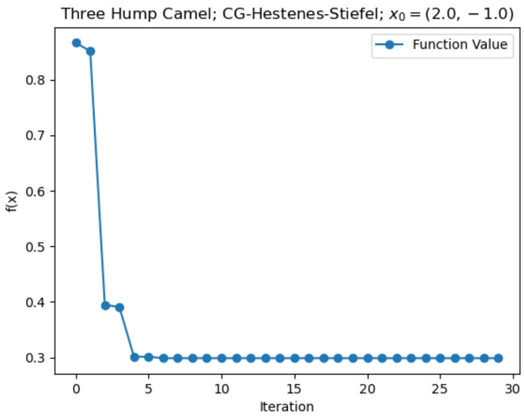
The function is defined as:

$$f(x_1, x_2) = 2x_1^2 - 1.05x_1^4 + \frac{x_1^6}{6} + x_1x_2 + x_2^2 \tag{2}$$

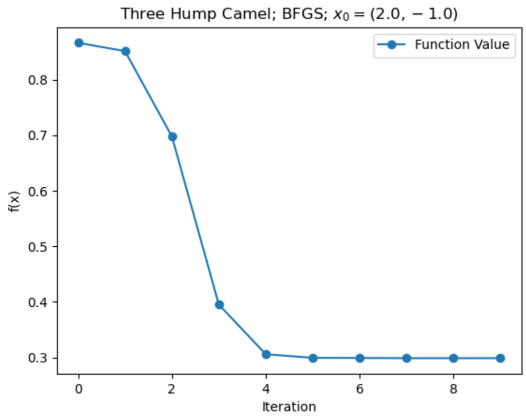
2.1 Plots

Below are the plots illustrating the optimization process:

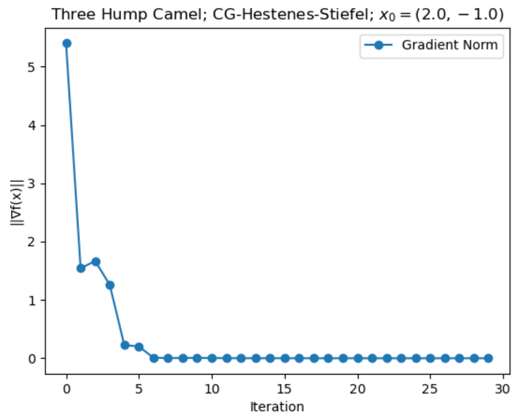
1. Plot of function values $f(x)$ vs. iterations.
2. Plot of gradient norm $\|\nabla f(x)\|$ vs. iterations.
3. Contour plot with the optimization path.



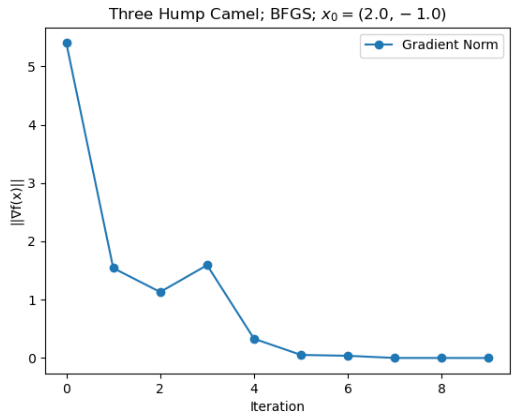
(a) $f(x)$ vs. iterations (Backtracking Armijo)



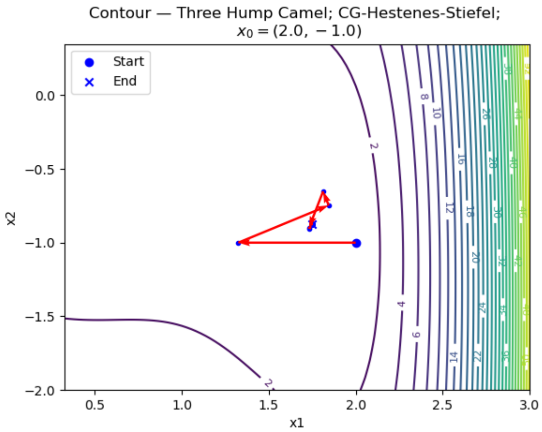
(b) $f(x)$ vs. iterations (Pure Newton's)



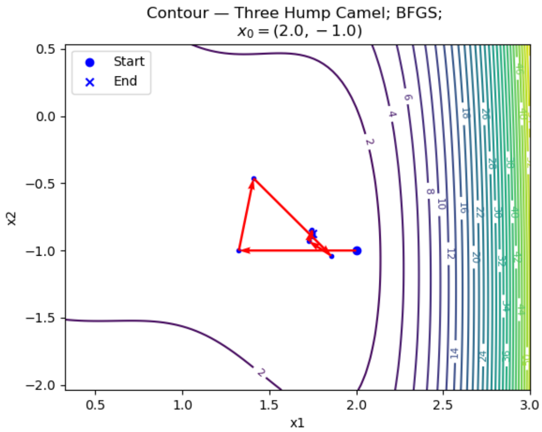
(a) $\|\nabla f(x)\|$ vs. iterations (Backtracking Armijo)



(b) $\|\nabla f(x)\|$ vs. iterations (Pure Newton's)



(a) Contour plot (Backtracking Armijo)



(b) Contour plot (Pure Newton's)

2.2 Results

Table 2: Three-Hump Camel Function

Test	Conjugate Gradient: HS Quasi-Newton: SR1	PR DFP	FR BFGS
2	[-1.748, 0.874] [-1.748, 0.874]	[-1.748, 0.874] [-1.748, 0.874]	[-0., 0.] [-1.748, 0.874]
3	[1.748, -0.874] [1.748, -0.874]	[1.748, -0.874] [1.748, -0.874]	[0., -0.] [1.748, -0.874]
4	[-0., 0.] [0., 0.]	[1.748, -0.874] [1.748, -0.874]	[0., -0.] [1.748, -0.874]
5	[0., -0.] [-0., -0.]	[-1.748, 0.874] [-1.748, 0.874]	[-0., 0.] [-1.748, 0.874]

Comments: All the algorithms led to convergence to the either of the three minimizers.

3 Styblinski-Tang Function

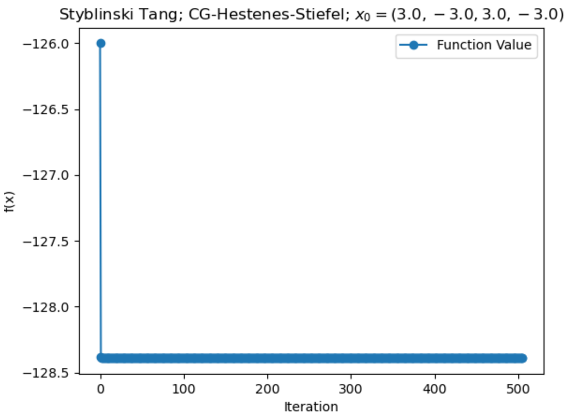
The function is defined as:

$$f(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{n}(x_i^4 - 16x_i^2 + 5x_i) \tag{3}$$

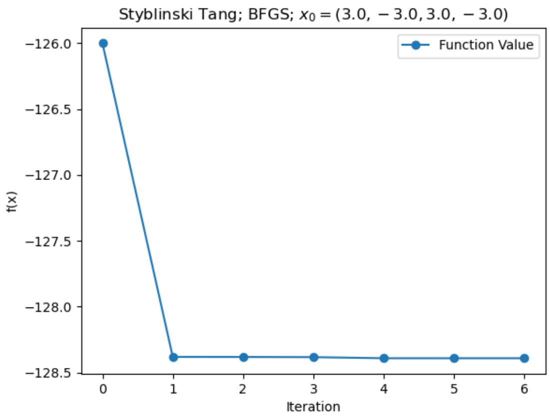
3.1 Plots

Below are the plots illustrating the optimization process:

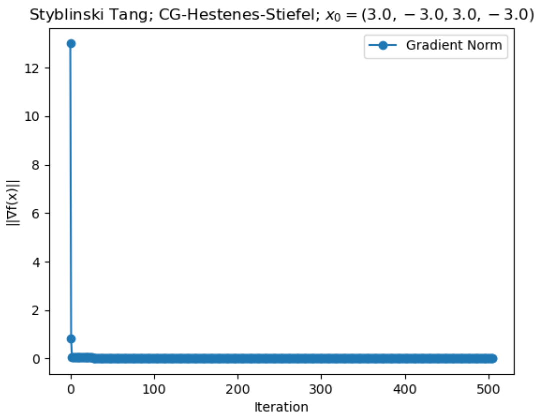
1. Plot of function values $f(x)$ vs. iterations.
2. Plot of gradient norm $\|\nabla f(x)\|$ vs. iterations.
3. Contour plot with the optimization path.



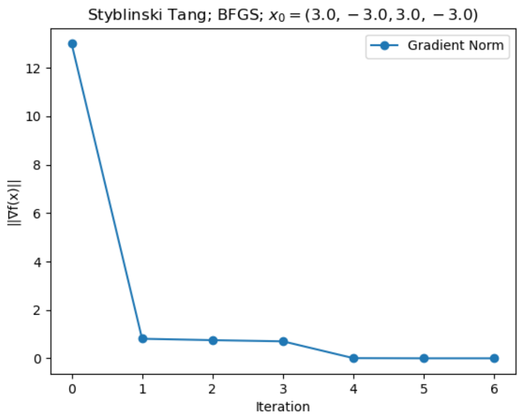
(a) $f(x)$ vs. iterations (Backtracking Armijo)



(b) $f(x)$ vs. iterations (Pure Newton's)



(a) $\|\nabla f(x)\|$ vs. iterations (Backtracking Armijo)



(b) $\|\nabla f(x)\|$ vs. iterations (Pure Newton's)

3.2 Results

Comments: All the methods led to convergence to the one of the possible minimizers.

Table 3: Styblinski-Tang Function

Test	Conjugate Gradient: HS Quasi-Newton: SR1	PR DFP	FR BFGS
10	[-2.904 -2.904 -2.904 -2.904] [-2.904 -2.904 -2.904 -2.904]	[-2.904 -2.904 -2.904 -2.904] [-2.904 -2.904 -2.904 -2.904]	[-2.904 -2.904 -2.904 -2.904] [-2.904 -2.904 -2.904 -2.904]
11	[-2.904 -2.904 -2.904 -2.904] [2.747 2.747 2.747 2.747]	[-2.904 -2.904 -2.904 -2.904] [2.747 2.747 2.747 2.747]	[2.747 2.747 2.747 2.747] [2.747 2.747 2.747 2.747]
12	[-2.904 -2.904 -2.904 -2.904] [-2.904 -2.904 -2.904 -2.904]	[-2.904 -2.904 -2.904 -2.904] [-2.904 -2.904 -2.904 -2.904]	[-2.904 -2.904 -2.904 -2.904] [-2.904 -2.904 -2.904 -2.904]
13	[2.747 -2.904 2.747 -2.904] [2.747 -2.904 2.747 -2.904]	[2.747 -2.904 2.747 -2.904] [2.747 -2.904 2.747 -2.904]	[2.747 -2.904 2.747 -2.904] [2.747 -2.904 2.747 -2.904]

4 Rosenbrock Function

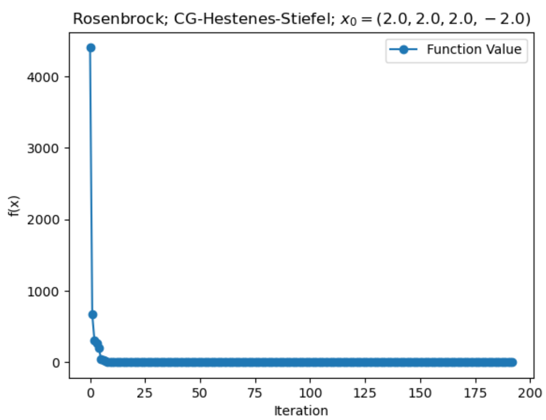
The Rosenbrock function is given by:

$$f(\mathbf{x}) = \sum_{i=1}^{n-1} \left(100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\right) \tag{4}$$

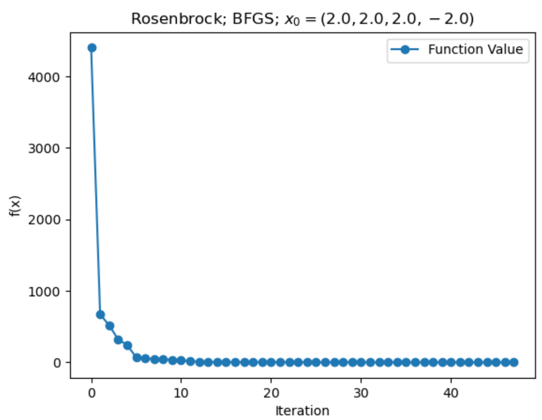
4.1 Plots

Below are the plots illustrating the optimization process:

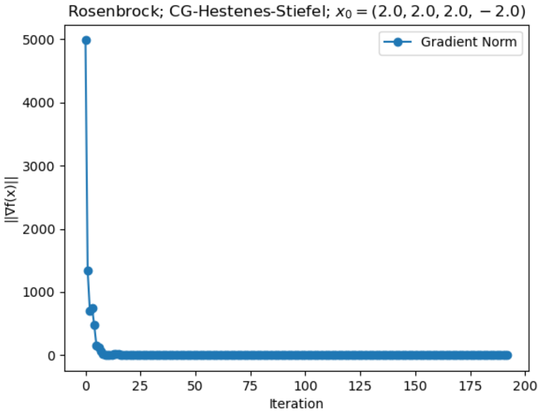
1. Plot of function values $f(x)$ vs. iterations.
2. Plot of gradient norm $\|\nabla f(x)\|$ vs. iterations.
3. Contour plot with the optimization path.



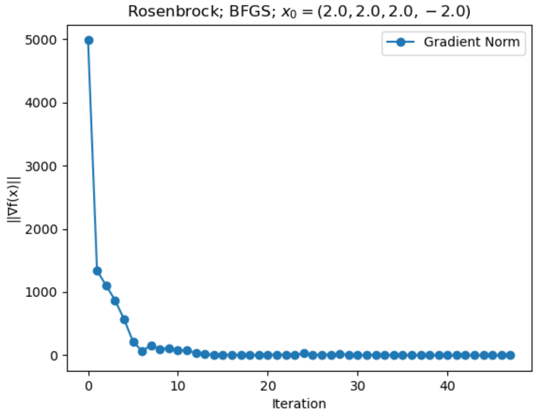
(a) $f(x)$ vs. iterations (Backtracking Armijo)



(b) $f(x)$ vs. iterations (Pure Newton's)



(a) $\|\nabla f(x)\|$ vs. iterations (Backtracking Armijo)



(b) $\|\nabla f(x)\|$ vs. iterations (Pure Newton's)

4.2 Results

Table 4: Rosenbrock Function

Test	Conjugate Gradient: HS Quasi-Newton: SR1	PR DFP	FR BFGS
6	[1. 1. 1. 1.] [0.442 0.161 0.022 0.003]	[1. 1. 1. 1.] [1. 1. 1. 1.]	[1. 1. 1. 1.] [1. 1. 1. 1.]
7	[1. 1. 1. 1.] [-0.636 0.417 0.182 0.033]	[1. 1. 1. 1.] [-0.629 0.408 0.175 0.03]	[1. 1. 1. 1.] [1. 1. 1. 1.]
8	[1. 1. 1. 1.] [0.726 0.489 0.202 0.027]	[1. 1. 1. 1.] [0.973 0.945 0.892 0.794]	[-0.776 0.613 0.382 0.146] [1. 1. 1. 1.]
9	[1. 1. 1. 1.] [-0.183 0.048 0.046 0.011]	[1. 1. 1. 1.] [0.785 0.608 0.365 0.123]	[1. 1. 1. 1.] [1. 1. 1. 1.]

Comments: All the methods led to convergence to the one of the many minimizers.

5 Root of Square Function

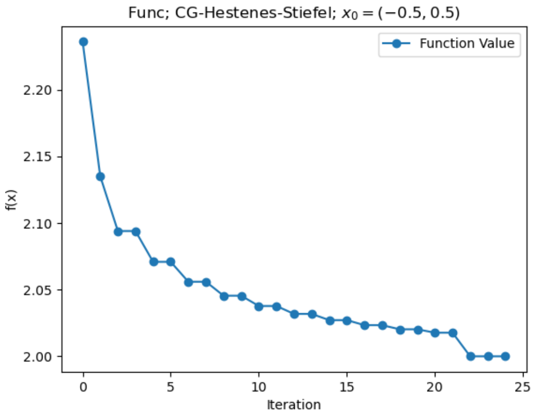
The function is defined as:

$$f(\mathbf{x}) = \sqrt{x_1^2 + 1} + \sqrt{x_2^2 + 1} \tag{5}$$

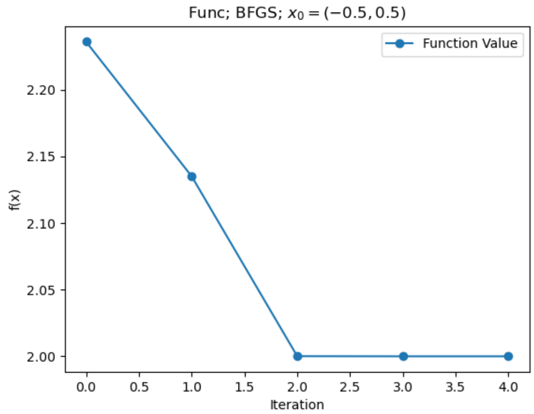
5.1 Plots

Below are the plots illustrating the optimization process:

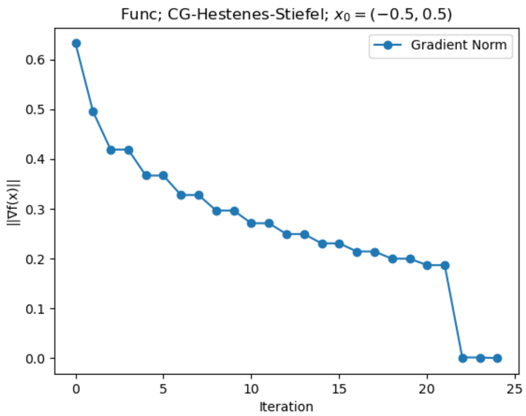
1. Plot of function values $f(x)$ vs. iterations.
2. Plot of gradient norm $\|\nabla f(x)\|$ vs. iterations.
3. Contour plot with the optimization path.



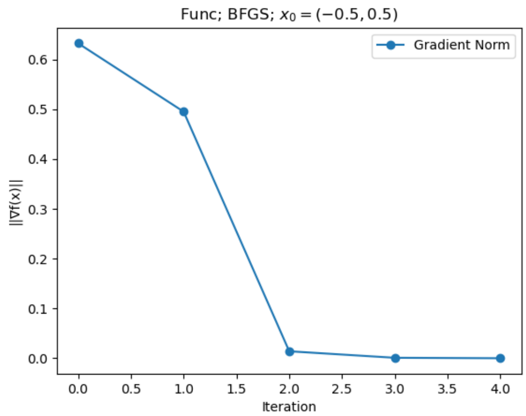
(a) $f(x)$ vs. iterations (Backtracking Armijo)



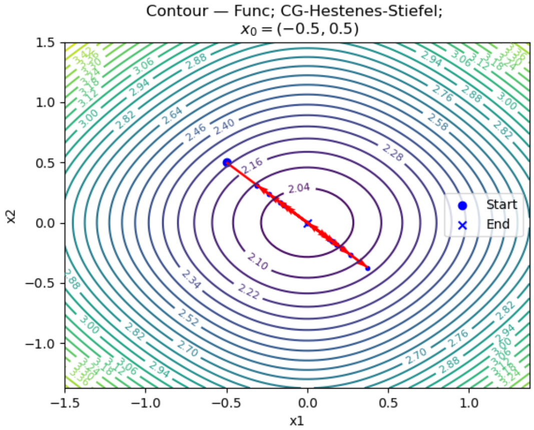
(b) $f(x)$ vs. iterations (Pure Newton's)



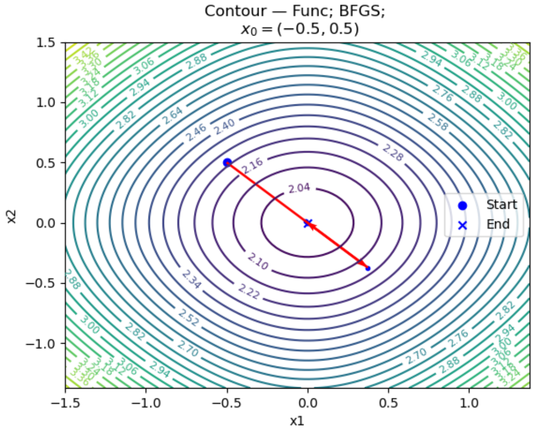
(a) $\|\nabla f(x)\|$ vs. iterations (Backtracking Armijo)



(b) $\|\nabla f(x)\|$ vs. iterations (Pure Newton's)



(a) Contour plot (Backtracking Armijo)



(b) Contour plot (Pure Newton's)

5.2 Results

Table 5: Function 1

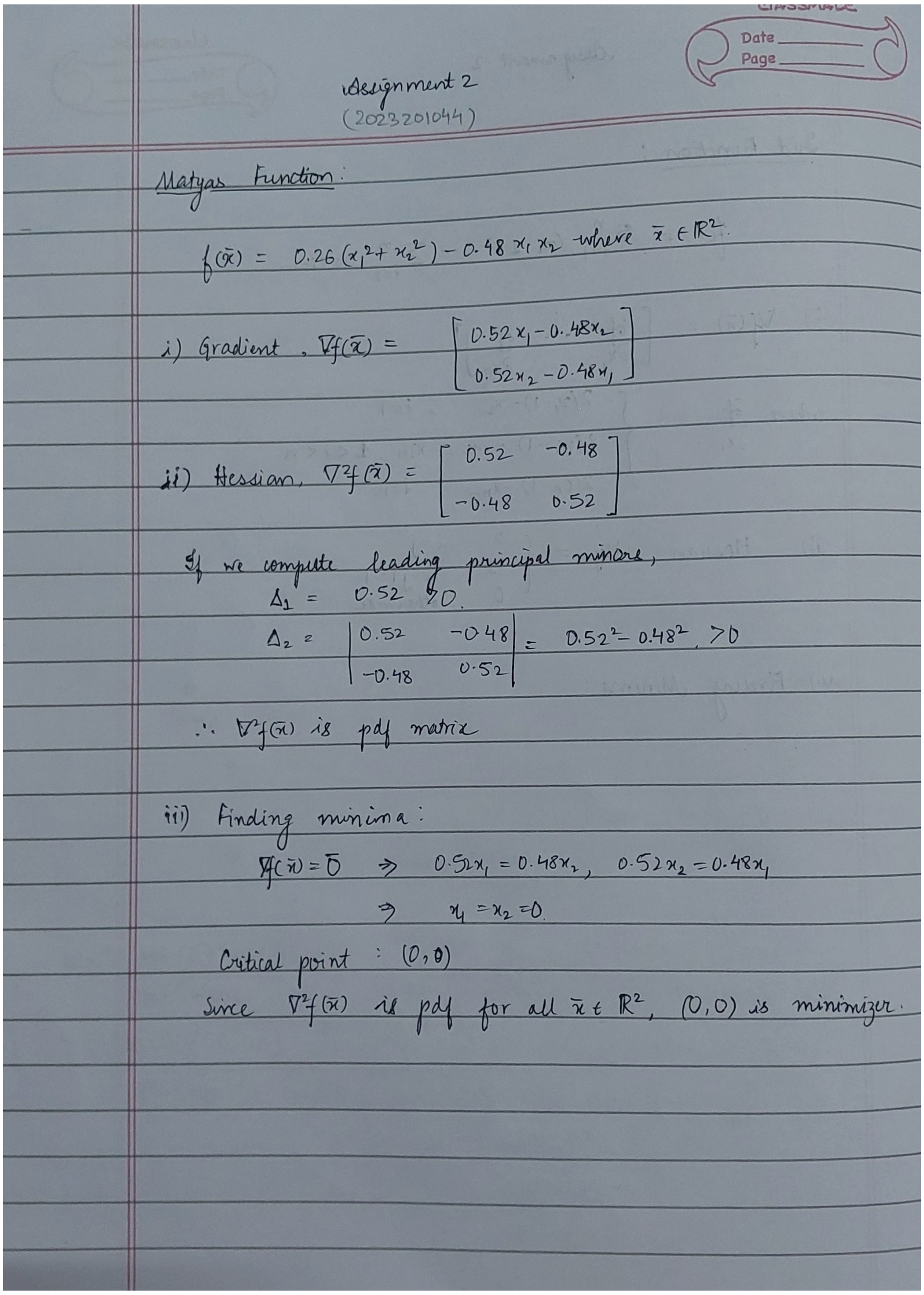
Test	Conjugate Gradient: HS Quasi-Newton: SR1	PR DFP	FR BFGS
14	[-0. -0.] [0. 0.]	[-0. -0.] [0. 0.]	[0. 0.] [0. 0.]
15	[0. -0.] [-0. 0.]	[0. -0.] [-0. 0.]	[0. -0.] [-0. 0.]
16	[0. -0.] [0. -0.]	[0. -0.] [0. -0.]	[-0. 0.] [-0. -0.]

Comments: All the methods led to convergence to the unique minimum.

6 Matyas Function

The function is defined as:

$$f(\mathbf{x}) = 0.26(x_1^2 + x_2^2) - 0.48x_1x_2 \tag{6}$$



Date
Page

Assignment 2
(2023201044)

Matyas Function:

$f(\bar{x}) = 0.26(x_1^2 + x_2^2) - 0.48\, x_1 x_2$ where $\bar{x} \in \mathbb{R}^2$.

i) Gradient, $\nabla f(\bar{x}) = \begin{bmatrix} 0.52x_1 - 0.48x_2 \\ 0.52x_2 - 0.48x_1 \end{bmatrix}$

ii) Hessian, $\nabla^2 f(\bar{x}) = \begin{bmatrix} 0.52 & -0.48 \\ -0.48 & 0.52 \end{bmatrix}$

If we compute leading principal minors,

$\Delta_1 = 0.52 > 0.$

$\Delta_2 = \begin{vmatrix} 0.52 & -0.48 \\ -0.48 & 0.52 \end{vmatrix} = 0.52^2 - 0.48^2 > 0$

$\therefore \nabla^2 f(\bar{x})$ is pdf matrix

iii) Finding minima:

$\nabla f(\bar{x}) = \bar{0} \Rightarrow 0.52x_1 = 0.48x_2,\ 0.52x_2 = 0.48x_1$

$\Rightarrow x_1 = x_2 = 0.$

Critical point : $(0,0)$

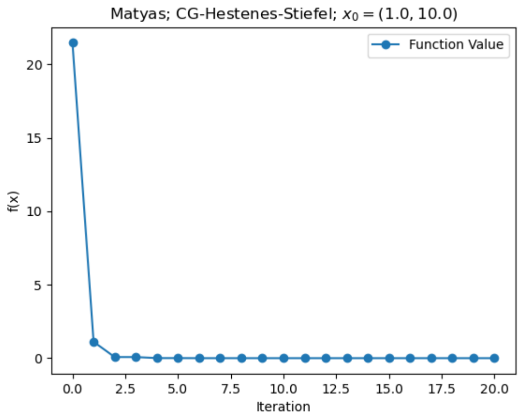
Since $\nabla^2 f(\bar{x})$ is pdf for all $\bar{x} \in \mathbb{R}^2$, $(0,0)$ is minimizer.

Figure 14: Derivation of gradient, Hessian and minimizer.

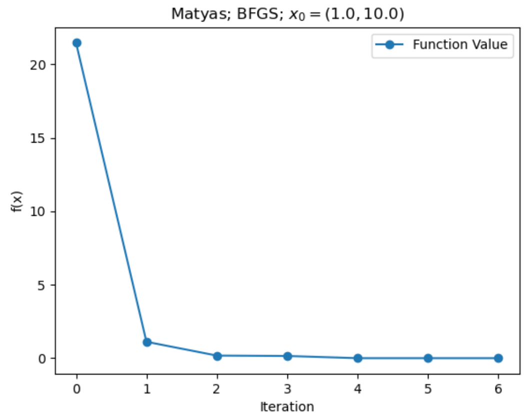
6.1 Plots

Below are the plots illustrating the optimization process:

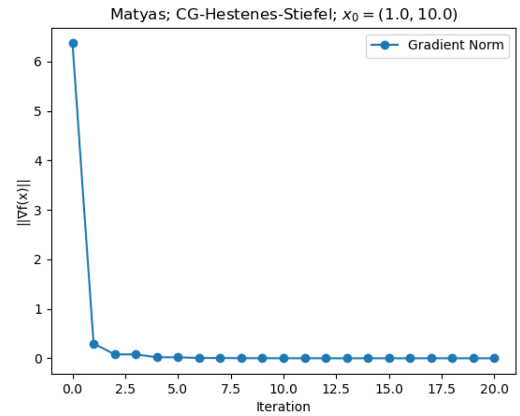
1. Plot of function values $f(x)$ vs. iterations.
2. Plot of gradient norm $\|\nabla f(x)\|$ vs. iterations.
3. Contour plot with the optimization path.



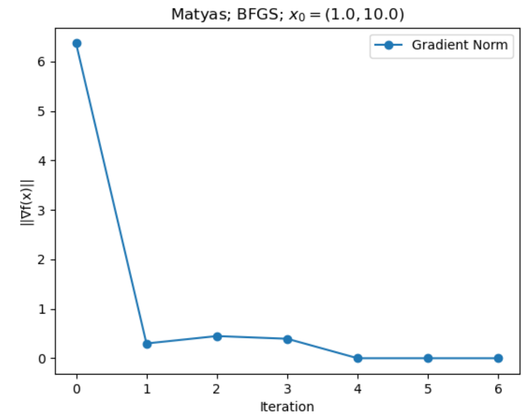
(a) $f(x)$ vs. iterations (Backtracking Armijo)



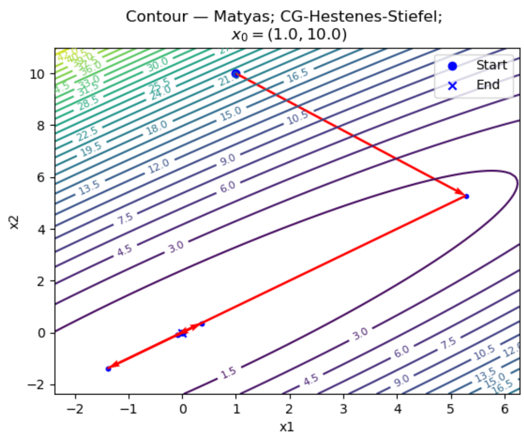
(b) $f(x)$ vs. iterations (Pure Newton's)



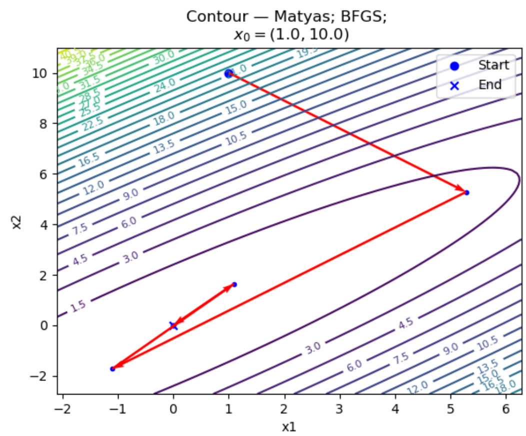
(a) $\|\nabla f(x)\|$ vs. iterations (Backtracking Armijo)



(b) $\|\nabla f(x)\|$ vs. iterations (Pure Newton's)



(a) Contour plot (Backtracking Armijo)



(b) Contour plot (Pure Newton's)

6.2 Results

Comments: All the methods led to convergence to the unique minimum.

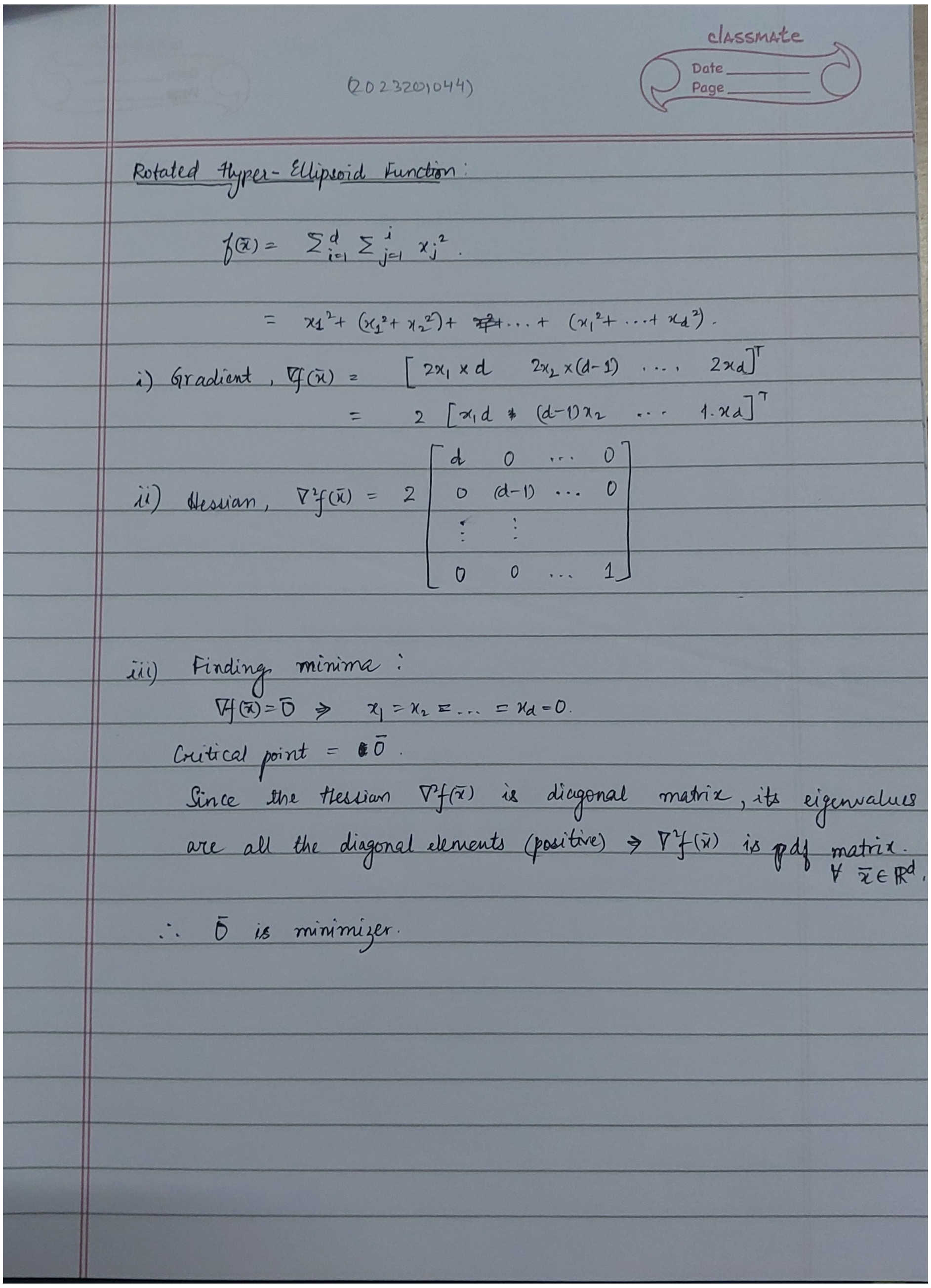
Table 6: Matyas Function

Test	Conjugate Gradient: HS Quasi-Newton: SR1	PR DFP	FR BFGS
17	[0. 0.] [0. 0.]	[0. 0.] [0. 0.]	[0. 0.] [0. 0.]
18	[0. 0.] [0. 0.]	[0. 0.] [0. 0.]	[-0. -0.] [-0. -0.]

7 Rotated Hyper-Ellipsoid Function

The function is defined as:

$$f(\mathbf{x}) = \sum_{i=1}^{d} \sum_{j=1}^{i} x_j^2 \tag{7}$$



(2023201044)

Rotated Hyper-Ellipsoid Function:

$$f(\bar{x}) = \sum_{i=1}^{d} \sum_{j=1}^{i} x_j^2 .$$

$$= x_1^2 + (x_1^2 + x_2^2) + \dots + (x_1^2 + \dots + x_d^2).$$

i) Gradient, $\nabla f(\bar{x}) = \begin{bmatrix} 2x_1 \times d & 2x_2 \times (d-1) & \dots & 2x_d \end{bmatrix}^T$

$= 2 \begin{bmatrix} x_1 d & (d-1)x_2 & \dots & 1 \cdot x_d \end{bmatrix}^T$

ii) Hessian, $\nabla^2 f(\bar{x}) = 2 \begin{bmatrix} d & 0 & \dots & 0 \\ 0 & (d-1) & \dots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \dots & 1 \end{bmatrix}$

iii) Finding minima:

$\nabla f(\bar{x}) = \bar{0} \Rightarrow x_1 = x_2 = \dots = x_d = 0.$

Critical point $= \bar{0}$.

Since the Hessian $\nabla^2 f(\bar{x})$ is diagonal matrix, its eigenvalues are all the diagonal elements (positive) $\Rightarrow \nabla^2 f(\bar{x})$ is pdf matrix. $\forall \bar{x} \in \mathbb{R}^d$.

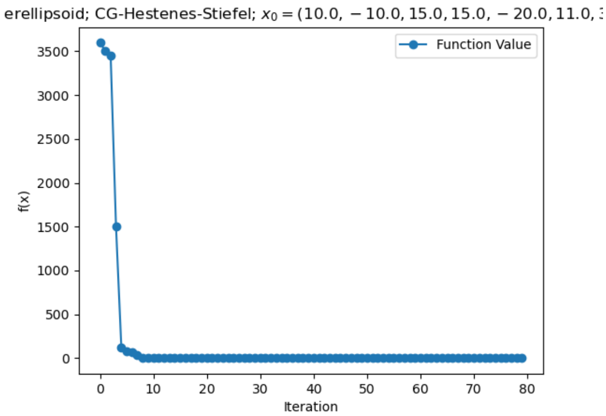
$\therefore$ $\bar{0}$ is minimizer.

Figure 18: Derivation of gradient, Hessian and minimizer.

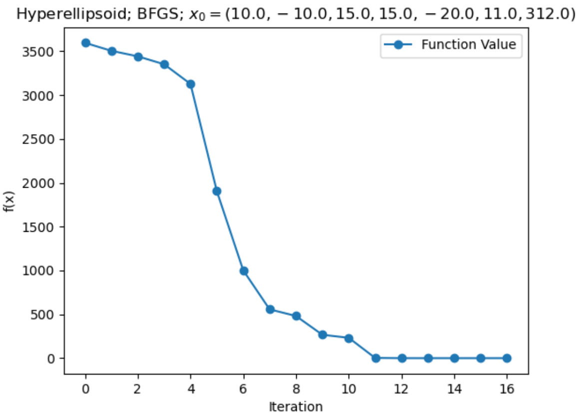
7.1 Plots

Below are the plots illustrating the optimization process:

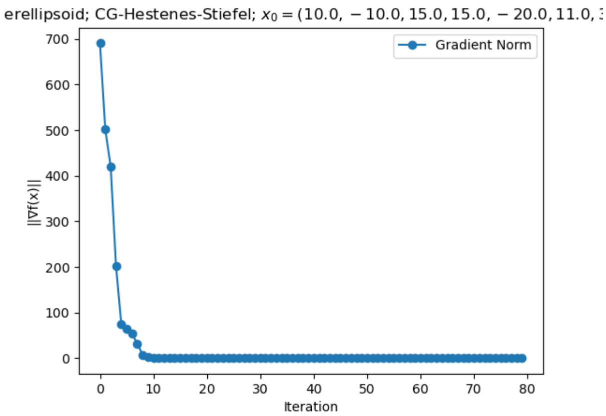
1. Plot of function values $f(x)$ vs. iterations.
2. Plot of gradient norm $\|\nabla f(x)\|$ vs. iterations.
3. Contour plot with the optimization path.



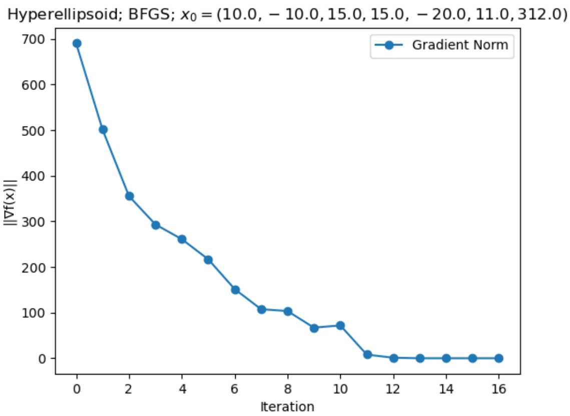
(a) $f(x)$ vs. iterations (Backtracking Armijo)



(b) $f(x)$ vs. iterations (Pure Newton's)



(a) $\|\nabla f(x)\|$ vs. iterations (Backtracking Armijo)



(b) $\|\nabla f(x)\|$ vs. iterations (Pure Newton's)

7.2 Results

Table 7: Rotated Hyper-Ellipsoid Function

Test	Conjugate Gradient: HS Quasi-Newton: SR1	PR DFP	FR BFGS
19	[0. 0. 0.] [0. 0. 0.]	[0. 0. 0.] [0. -0. -0.]	[-0. 0. -0.] [0. 0. 0.]
20	[0. -0. 0. -0. 0. 0. 0.] [0. 0. 0. -0. -0. -0. 0.]	[0. 0. -0. 0. -0. -0. 0.] [-0. -0. -0. 0. 0. 0. 0.]	[0. -0. -0. 0. -0. -0. -0.] [0. 0. -0. -0. -0. 0. -0.]

Comments: All the methods led to convergence to the unique minimum.