CS1.404: Assignment 2

Himanshu Singh

April 10, 2024

1 Trid Function

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

1.1 Jacobian and Hessian

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 - x_2 - 2 \\ 2x_2 - x_1 - x_3 - 2 \\ 2x_3 - x_2 - x_4 - 2 \\ \vdots \\ 2x_{d-1} - x_{d-2} - x_d - 2 \\ 2x_d - x_{d-1} - 2 \end{bmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ 0 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & -1 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

1.2 Calculation of Minima

We know from the SOSC for minima that $\mathbf{x}^*$ is a local minima of the function $f(\mathbf{x})$, if the following conditions hold.

$$\begin{aligned}
&\nabla f(\mathbf{x}^*) = \mathbf{0} \\
&\implies 2x_1 - x_2 = 2 \\
&\implies 2x_i - x_{i-1} - x_{i+1} = 2 && \forall i \in [2, d-1] \\
&\implies 2x_d - x_{d-1} = 2
\end{aligned} \tag{1}$$

$$\begin{aligned}
&\mathbf{y}^T \nabla^2 f(\mathbf{x}^*)\mathbf{y} > 0 && \forall \mathbf{y} \in \mathbb{R}^d \\
&\implies \begin{bmatrix} 2y_1 - y_2 & \dots & 2y_i - y_{i-1} - y_{i+1} & \dots & 2y_d - y_{d-1} \end{bmatrix}^T \mathbf{y} > 0 && \forall \mathbf{y} \in \mathbb{R}^d \\
&\implies 2\left(\sum_{i=1}^{d} y_i^2 - \sum_{i=2}^{d} y_{i-1}y_i\right) > 0 && \forall \mathbf{y} \in \mathbb{R}^d
\end{aligned}$$

$$\implies 2\left(\sum_{i=1}^{d} y_i^2 - \sum_{i=2}^{d} y_{i-1}y_i\right) > 0 \qquad \forall \mathbf{y} \in \mathbb{R}^d$$

$$\implies \left(\sum_{i=2}^{d} y_{i-1}^2 + y_i^2 - 2y_{i-1}y_i\right) + y_1^2 + y_d^2 > 0 \qquad \forall \mathbf{y} \in \mathbb{R}^d$$

$$\implies \left(\sum_{i=2}^{d} (y_{i-1} - y_i)^2\right) + y_1^2 + y_d^2 > 0 \qquad \forall \mathbf{y} \in \mathbb{R}^d \tag{2}$$

Assuming $x_i = i(d+1-i)$, and substituting, we see that the inequalities (1) hold; while the inequality (2) holds regardless of $\mathbf{x}$. Thus,

$$\mathbf{x}^* = (d, 2d-2, 3d-6, \ldots, d) \tag{3}$$

1.3 Convergence of Algorithms

For the given test case, all the algorithms converged to the local minima $\mathbf{x}^* = (2, 2)$.

1.4 Plots

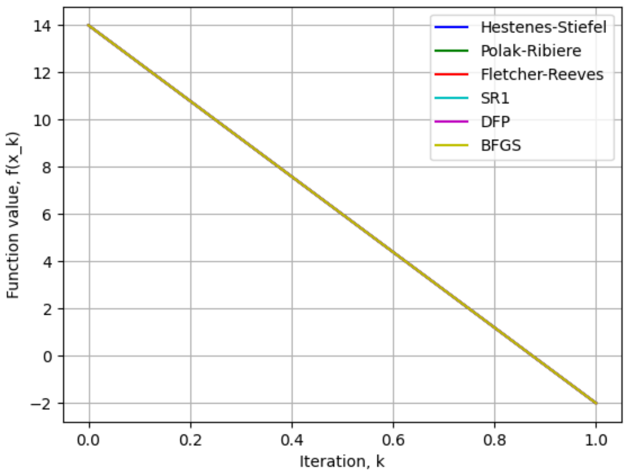

Figure 1: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (-2, -2)$

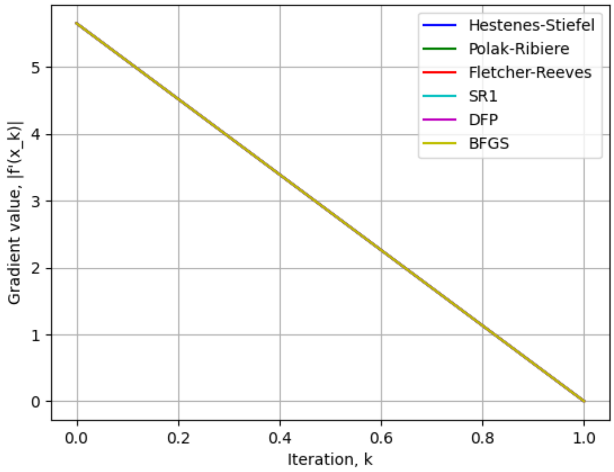

Figure 2: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (-2, -2)$

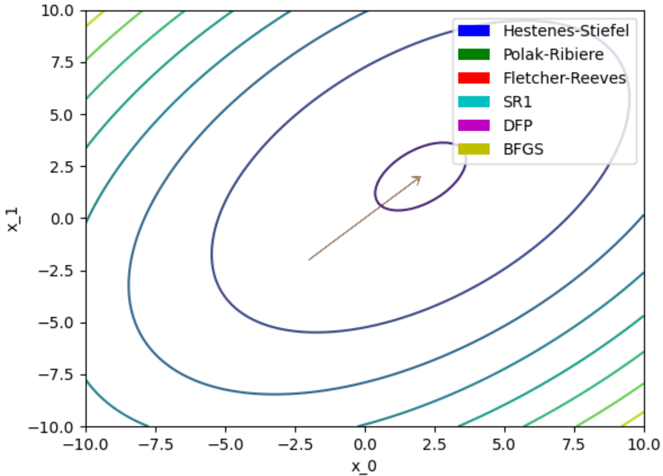

Figure 3: Contour plot, with direction of updates for $\mathbf{x}_0 = (-2, -2)$

2 Three Hump Camel

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

2.1 Jacobian and Hessian

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 4x_1 - 4.2x_1^3 + x_1^5 + x_2 \\ x_1 + 2x_2 \end{bmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 5x_1^4 - 12.6x_1^2 + 4 & 1 \\ 1 & 2 \end{bmatrix}$$

2.2 Calculation of Minima

We know from the SOSC for minima that $\mathbf{x}^*$ is a local minima of the function $f(\mathbf{x})$, if the following conditions hold.

$$
\begin{aligned}
&\nabla f(\mathbf{x}^*) = \mathbf{0} \\
\implies\ & 4x_1 - 4.2x_1^3 + x_1^5 + x_2 = 0 && (4)\\
& x_1 + 2x_2 = 0 && (5)
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{y}^T \nabla^2 f(\mathbf{x}^*) \mathbf{y} > 0 && \forall \mathbf{y} \in \mathbb{R}^2 \\
\implies\ & y_1^2(5x_1^4 - 12.6x_1^2 + 4) + 2y_1y_2 + 2y_2^2 > 0 && \forall y_1, y_2 \in \mathbb{R} \qquad (6)
\end{aligned}
$$

Solving (4) and (5), we get

$$\{\mathbf{x}^*\} = \{(-1.7476, 0.8738), (-1.0705, 0.5353), (0, 0), (1.0705, -0.5353), (1.7476, -0.8738)\} \quad (7)$$

Checking satisfiability of inequality (6) for all possibilities (7), we get

$$\{\mathbf{x}^*\} = \{(-1.7476, 0.8738), (0, 0), (1.7476, -0.8738)\} \quad (8)$$

2.3 Convergence of Algorithms

For the given test cases, all the algorithms converged to one of the local minimas.

2.4 Plots

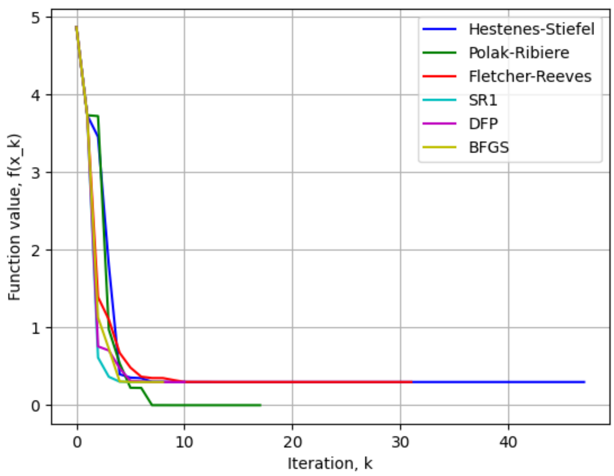

Figure 4: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (-2, -1)$

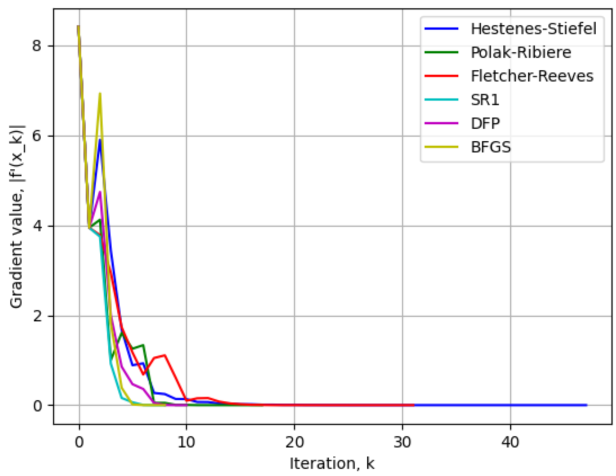

Figure 5: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (-2, -1)$

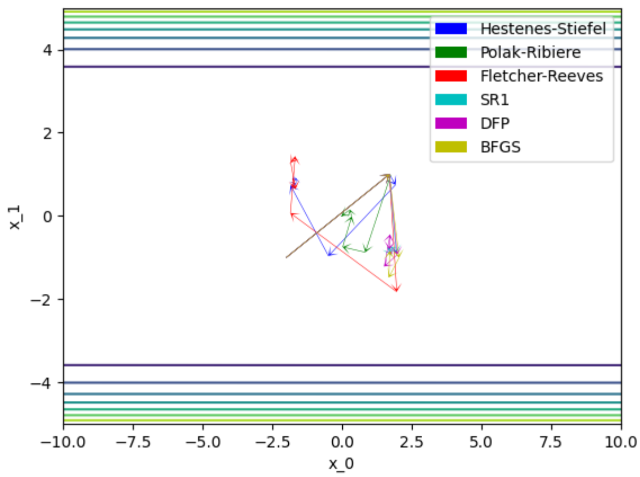

Figure 6: Contour plot, with direction of updates for $\mathbf{x}_0 = (-2, -1)$

3 Styblinski-Tang Function

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

3.1 Jacobian and Hessian

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1^3 - 16x_1 + \frac{5}{2} \\ 2x_2^3 - 16x_2 + \frac{5}{2} \\ \vdots \\ 2x_d^3 - 16x_d + \frac{5}{2} \end{bmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 6x_1^2 - 16 & 0 & \dots & 0 \\ 0 & 6x_2^2 - 16 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 6x_d^2 - 16 \end{bmatrix}$$

3.2 Calculation of Minima

We know from the SOSC for minima that $\mathbf{x}^*$ is a local minima of the function $f(\mathbf{x})$, if the following conditions hold.

$$\begin{aligned} &\nabla f(\mathbf{x}^*) = \mathbf{0} & \\ &\implies 2x_i^3 - 16x_i + \frac{5}{2} = 0 & \forall i \in [1, d] \\ &\implies x_i \approx -2.9035, 0.15673, 2.7468 & \forall i \in [1, d] \end{aligned} \tag{9}$$

$$\begin{aligned} &\mathbf{y}^T \nabla^2 f(\mathbf{x}^*) \mathbf{y} > 0 & \forall \mathbf{y} \in \mathbb{R}^d \\ &\implies y_i^2(6x_i^2 - 16) > 0 & \forall y_i \in \mathbb{R}, i \in [1, d] \\ &\implies x_i^2 > \frac{8}{3} & \forall i \in [1, d] \\ &\implies |x_i| > 1.63 & \forall i \in [1, d] \end{aligned} \tag{10}$$

Combining (9) and (10), we get $x_i = -2.9035, 2.7468$. Thus, the set of local minimas is

$$\{\mathbf{x}^*\} = \{-2.9035, 2.7468\}^d \tag{11}$$

3.3 Convergence of Algorithms

For the given test cases, all the algorithms converged to one of the local minimas.

3.4 Plots

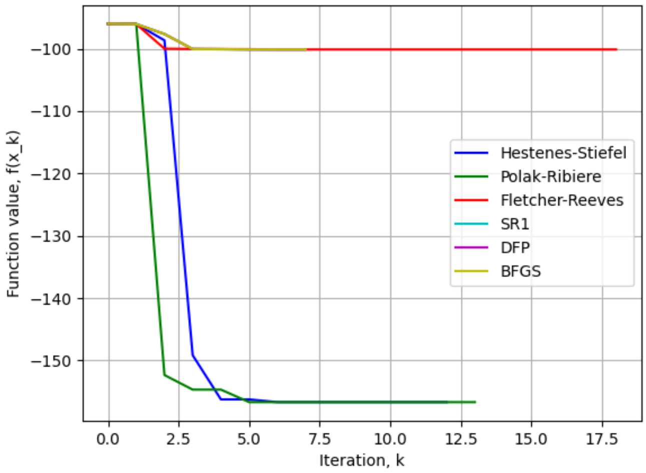

Figure 7: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (3, 3, 3, 3)$

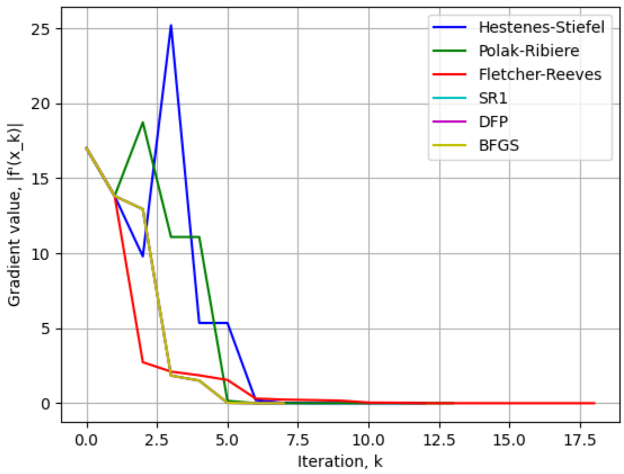

Figure 8: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (3, 3, 3, 3)$

4 Rosenbrock Function

$$f(\mathbf{x}) = \sum_{i=1}^{d-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$$

4.1 Jacobian and Hessian

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 400x_1(x_1^2 - x_2) + 2(x_1 - 1) \\ 400x_2(x_2^2 - x_3) + 2(x_2 - 1) + 200(x_2 - x_1^2) \\ 400x_3(x_3^2 - x_4) + 2(x_3 - 1) + 200(x_3 - x_2^2) \\ \vdots \\ 400x_{d-1}(x_{d-1}^2 - x_k) + 2(x_{d-1} - 1) + 200(x_{d-1} - x_{d-2}^2) \\ 200(x_d - x_{d-1}^2) \end{bmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 & 0 & \dots & 0 \\ -400x_1 & 1200x_2^2 - 400x_3 + 202 & -400x_2 & \dots & 0 \\ 0 & -400x_2 & 1200x_3^2 - 400x_4 + 202 & \dots & 0 \\ 0 & 0 & -400x_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & -400x_{d-1} \\ 0 & 0 & 0 & -400x_{d-1} & 200 \end{bmatrix}$$

4.2 Convergence of Algorithms

1. Quasi-Newton Method with SR1 update failed to converge for all, except the third test case.
2. Conjugate Gradient Method with Fletcher Reeves Formula failed to converge for the first test case.
3. The remaining methods converged to the local minima $\mathbf{x}^* = (0, 0)$, for all the test cases.

4.3 Plots

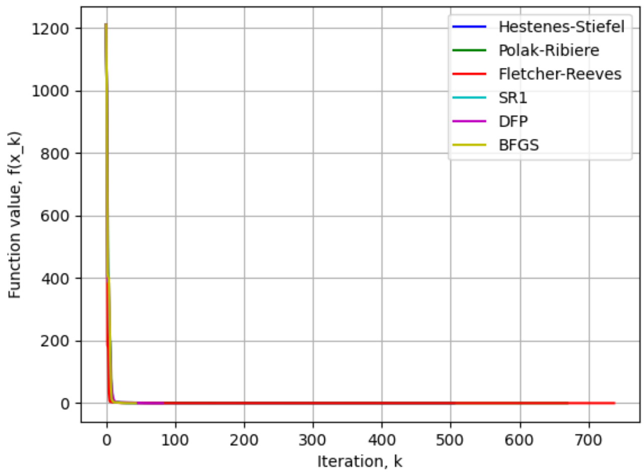

Figure 9: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (-2, 2, 2, 2)$

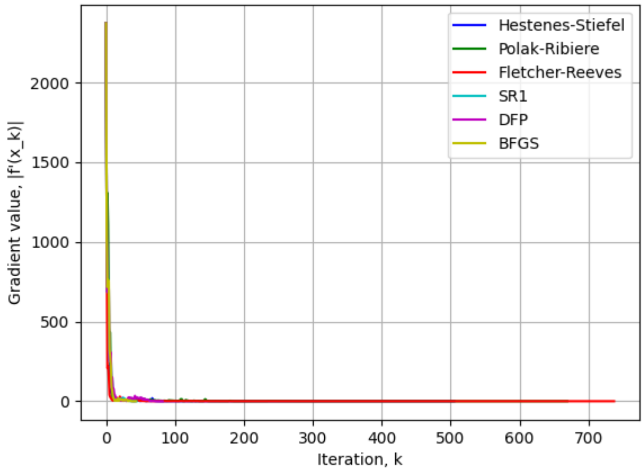

Figure 10: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (-2, 2, 2, 2)$

5 Root of Square Function

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

5.1 Jacobian and Hessian

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \dfrac{x_1}{\sqrt{1 + x_1^2}} \\ \dfrac{x_2}{\sqrt{1 + x_2^2}} \end{bmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \dfrac{1}{\sqrt{(1 + x_1^2)^3}} & 0 \\ 0 & \dfrac{1}{\sqrt{(1 + x_1^2)^3}} \end{bmatrix}$$

5.2 Calculation of Minima

We know from the SOSC for minima that $\mathbf{x}^*$ is a local minima of the function $f(\mathbf{x})$, if the following conditions hold.

$$\begin{aligned} &\nabla f(\mathbf{x}^*) = \mathbf{0} & \\ &\implies \frac{x_i}{\sqrt{1 + x_i^2}} = 0 & \forall i \in [1, 2] \\ &\implies x_i = 0 & \forall i \in [1, 2] \qquad (12) \end{aligned}$$

$$\begin{aligned} &\mathbf{y}^T \nabla^2 f(\mathbf{x}^*) \mathbf{y} > 0 & \forall \mathbf{y} \in \mathbb{R}^2 \\ &\implies \frac{y_i^2}{\sqrt{(1 + x_i^2)^3}} > 0 & \forall y_i \in \mathbb{R}, i \in [1, 2] \end{aligned}$$

$$\implies (1+x_i^2)^3 > 0 \qquad \forall i \in [1,2]$$
$$\implies x_i^2 > -1 \qquad \forall i \in [1,2] \tag{13}$$

Since (13) always hold, from (12) we get $x_i = 0$. Thus, the local minima is

$$\mathbf{x}^* = (0,0) \tag{14}$$

5.3 Convergence of Algorithms

For the given test cases, all the algorithms converged to the local minima $\mathbf{x}^* = (0,0)$.

5.4 Plots

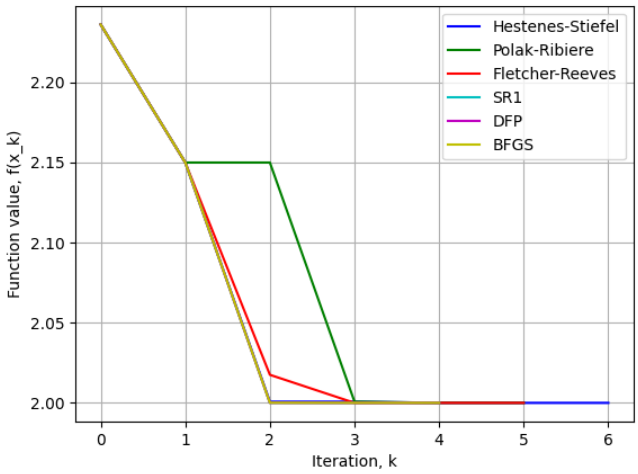

Figure 11: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (-0.5, 0.5)$

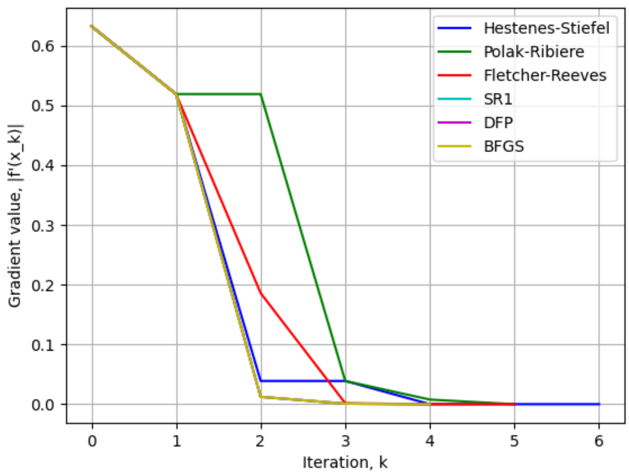

Figure 12: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (-0.5, 0.5)$

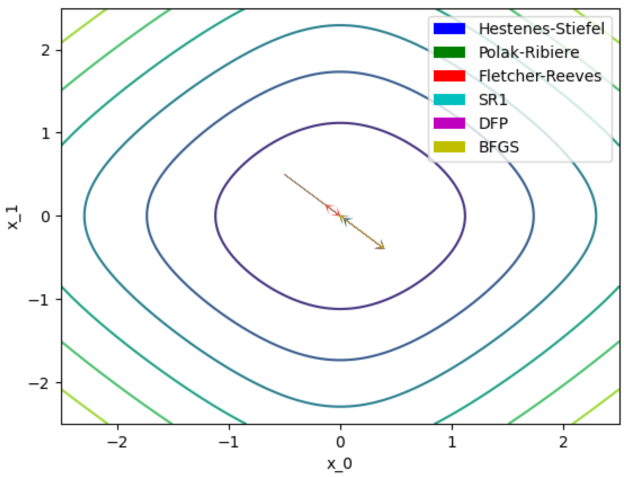

Figure 13: Contour plot, with direction of updates for $\mathbf{x}_0 = (-0.5, 0.5)$

6 Matyas Function

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

6.1 Jacobian and Hessian

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 0.52x_1 - 0.48x_2 \\ 0.52x_2 - 0.48x_1 \end{bmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 0.52 & -0.48 \\ -0.48 & 0.52 \end{bmatrix}$$

6.2 Calculation of Minima

We know from the SOSC for minima that $\mathbf{x}^*$ is a local minima of the function $f(\mathbf{x})$, if the following conditions hold.

$$\begin{aligned} &\nabla f(\mathbf{x}^*) = \mathbf{0} \\ &\implies 0.52x_1 - 0.48x_2 = 0 \qquad (15) \\ &\qquad\quad 0.52x_2 - 0.48x_1 = 0 \qquad (16) \end{aligned}$$

$$\mathbf{y}^T \nabla^2 f(\mathbf{x}^*) \mathbf{y} > 0 \qquad \forall \mathbf{y} \in \mathbb{R}^2 \qquad (17)$$

Solving (15) and (16), we get

$$\mathbf{x}^* = (0, 0) \qquad (18)$$

Finally, to show that (17) holds, we notice that $\nabla^2 f(\mathbf{x}^*)$ is a symmetric matrix. It is thus sufficient to show that its principal minors are positive. Indeed, $D_1(\nabla^2 f(\mathbf{x}^*)) = 0.52, D_2(\nabla^2 f(\mathbf{x}^*)) = 0.04$.

6.3 Convergence of Algorithms

For the given test cases, all the algorithms converged to the local minima $\mathbf{x}^* = (0, 0)$.

6.4 Plots

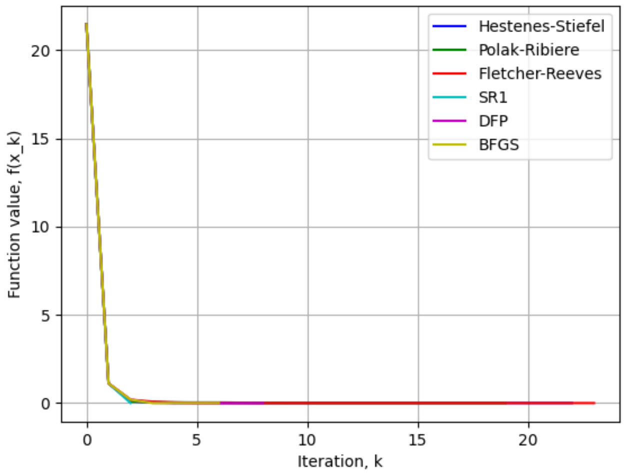

Figure 14: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (1, 10)$

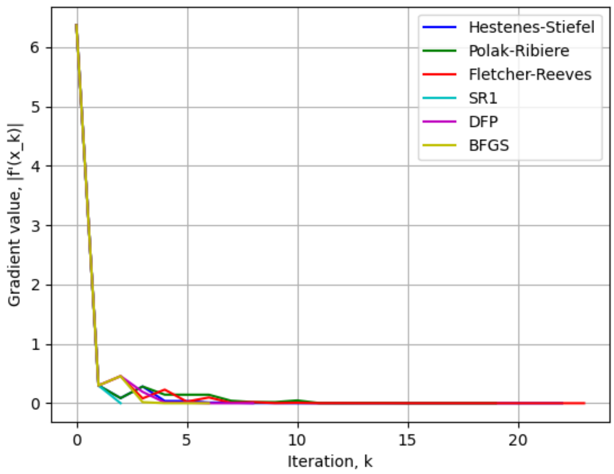

Figure 15: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (1, 10)$

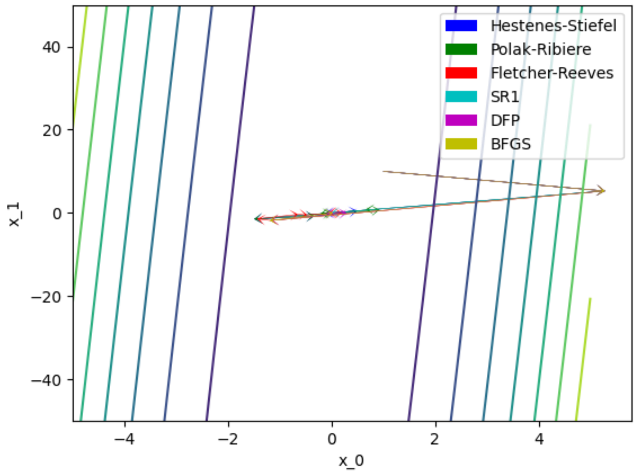

Figure 16: Contour plot, with direction of updates for $\mathbf{x}_0 = (1, 10)$

7 Rotated Hyper-Ellipsoid Function

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

7.1 Jacobian and Hessian

$$\nabla f(\mathbf{x}) = \begin{bmatrix} dx_1^2 \\ (d-1)x_2^2 \\ (d-2)x_3^2 \\ \vdots \\ x_d^2 \end{bmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 2dx_1 & 0 & 0 & \dots & 0 \\ 0 & 2(d-1)x_2 & 0 & \dots & 0 \\ 0 & 0 & 2(d-2)x_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & & 2x_d \end{bmatrix}$$

7.2 Calculation of Minima

We know from the SONC for minima that if $\mathbf{x}^*$ is a local minima of the function $f(\mathbf{x})$, the following conditions hold.

$$\begin{aligned} &\nabla f(\mathbf{x}^*) = \mathbf{0} & \\ &\implies (d-i+1)x_i^2 = 0 & \forall i \in [1, d] \\ &\implies x_i = 0 & \forall i \in [1, d] \end{aligned} \tag{19}$$

$$\mathbf{y}^T \nabla^2 f(\mathbf{x}^*)\mathbf{y} \geq 0 \qquad \forall \mathbf{y} \in \mathbb{R}^2 \tag{20}$$

Since $\nabla^2 f(\mathbf{x}^*)$ is a diagonal matrix with non-negative diagonal entries, it is positive semi-definite, thus establishing (20). To show that (19) is indeed a minima, notice that $f(\mathbf{x})$ is bounded below from $y = 0$, with $f(\mathbf{x}^*)$ precisely achieving this bound. Thus, we get the minima

$$\mathbf{x}^* = \{0\}^d \tag{21}$$

7.3 Convergence of Algorithms

For the given test cases, all the algorithms converged to the local minima $\mathbf{x}^* = \{0\}^d$.

7.4 Plots

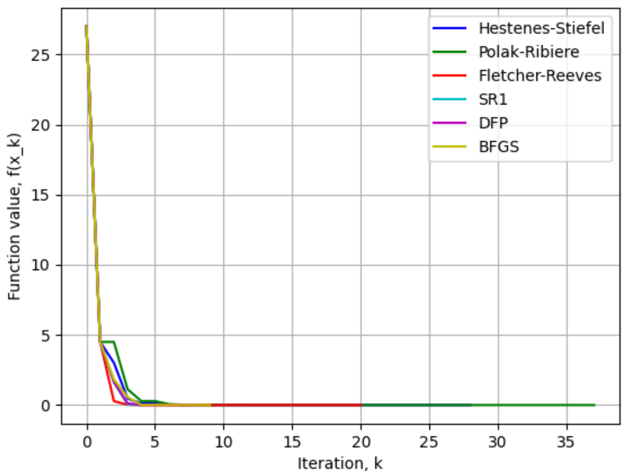

Figure 17: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (-3, 3, 2)$

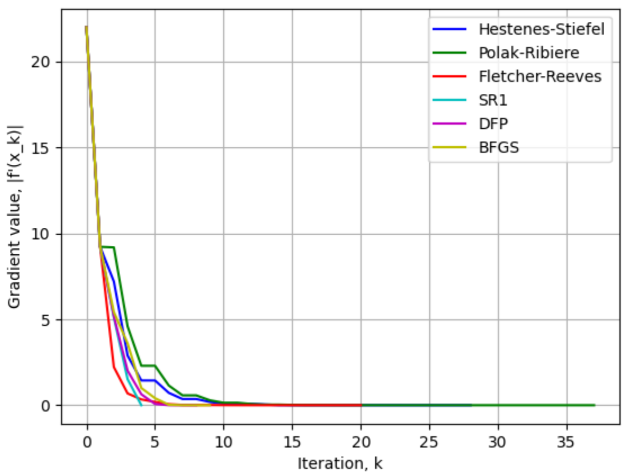

Figure 18: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (-3, 3, 2)$

Appendix: Output for all test cases

Test Case	Function	Initial Point	Conjugate: HS	Conjugate: PR	Conjugate: FR	SR1	DFP	BFGS
1	Trid	(-2, -2)	(2, 2)	(2, 2)	(2, 2)	(2, 2)	(2, 2)	(2, 2)
2	Three Hump Camel	(-2, 1)	(-1.748, 0.874)	(0, 0)	(-1.748, 0.874)	(-1.748, 0.874)	(-1.748, 0.874)	(-1.748, 0.874)
3		(2, -1)	(1.748, -0.874)	(0, 0)	(1.748, -0.874)	(1.748, -0.874)	(1.748, -0.874)	(1.748, -0.874)
4		(-2, -1)	(-1.748, 0.874)	(0, 0)	(-1.748, 0.874)	(1.748, -0.874)	(1.748, -0.874)	(1.748, -0.874)
5		(2, 1)	(1.748, -0.874)	(0, 0)	(1.748, -0.874)	(-1.748, 0.874)	(-1.748, 0.874)	(-1.748, 0.874)
6	Rosenbrock	(2, 2, 2, -2)	(1, 1, 1, 1)	(1, 1, 1, 1)	(-0.776, 0.613, 0.382, 0.146)	(0.212, 0.052, 0.008, 0.002)	(1, 1, 1, 1)	(1, 1, 1, 1)
7		(2, -2, -2, 2)	(1, 1, 1, 1)	(1, 1, 1, 1)	(1, 1, 1, 1)	(-0.635, 0.414, 0.18, 0.033)	(1, 1, 1, 1)	(1, 1, 1, 1)
8		(-2, 2, 2, 2)	(1, 1, 1, 1)	(1, 1, 1, 1)	(1, 1, 1, 1)	(1, 1, 1, 1)	(1, 1, 1, 1)	(1, 1, 1, 1)
9		(3, 3, 3, 3)	(1, 1, 1, 1)	(1, 1, 1, 1)	(1, 1, 1, 1)	(0.953, 0.917, 0.853, 0.736)	(1, 1, 1, 1)	(1, 1, 1, 1)
10	Styblinski-Tang	(0, 0, 0, 0)	(-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		(3, 3, 3, 3)	(-2.904, -2.904, -2.904, -2.904)	(-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)	(2.747, 2.747, 2.747, 2.747)
12		(-3, -3, -3, -3)	(-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		(3, -3, 3, -3)	(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)
14	Root Square	(3, 3)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
15		(-0.5, 0.5)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)

Test Case	Function	Initial Point	Conjugate: HS	Conjugate: PR	Conjugate: FR	SR1	DFP	BFGS
16		(-3.5, 0.5)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
17	Matyas	(2, -2)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
18		(1, 10)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
19	Rotated Hyper Ellipsoid	(-3, 3, 2)	(0, 0, 0)	(0, 0, 0)	(0, 0, 0)	(0, 0, 0)	(0, 0, 0)	(0, 0, 0)
20		(10, -10, 15, 15, -20, 11, 312.0)	(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, 0)