CS1.404: Assignment 1

Himanshu Singh

March 11, 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

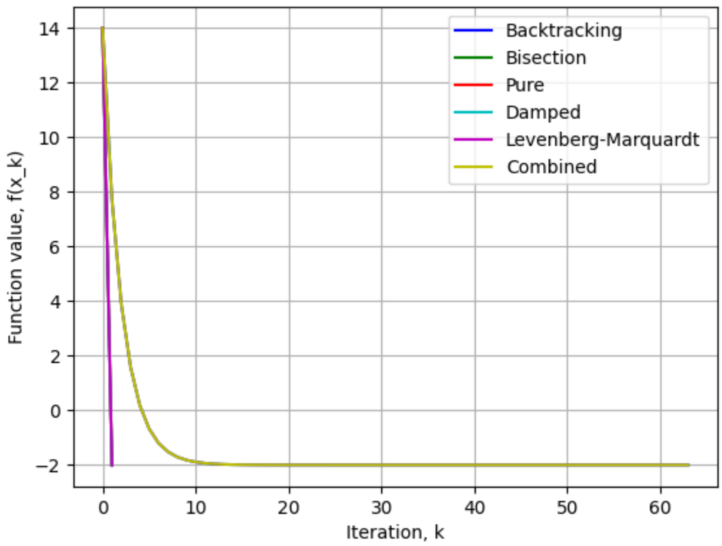

Figure 1: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (-2, -2)$

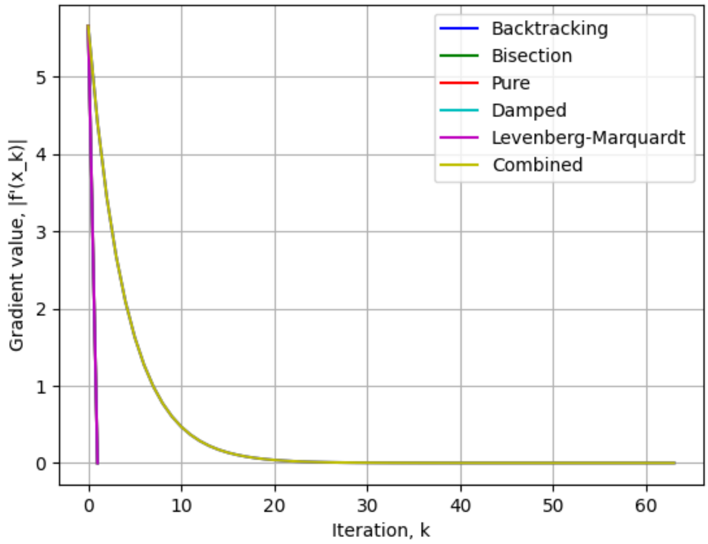

Figure 2: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (-2, -2)$

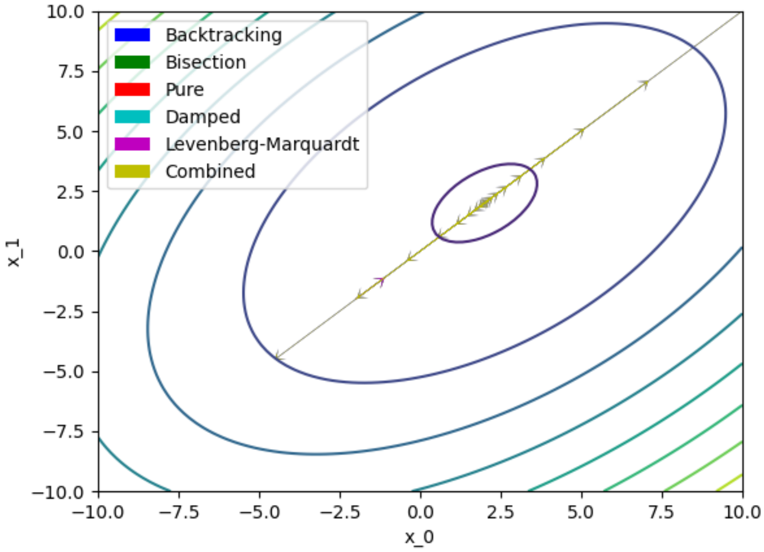

Figure 3: Contour plot, with direction of updates for $\mathbf{x}_0 = (-2, -2)$

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.

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

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

Solving (4) and (5), we get

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

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

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

2.3 Convergence of Algorithms

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

2.4 Plots

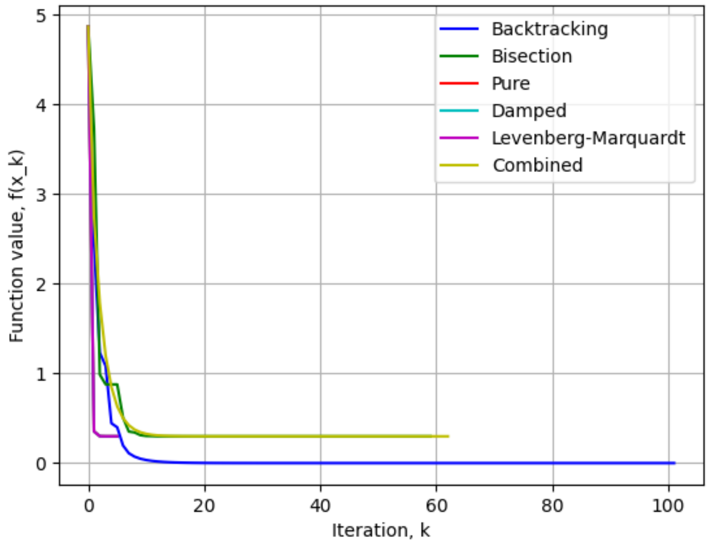

Figure 4: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (-2, -1)$

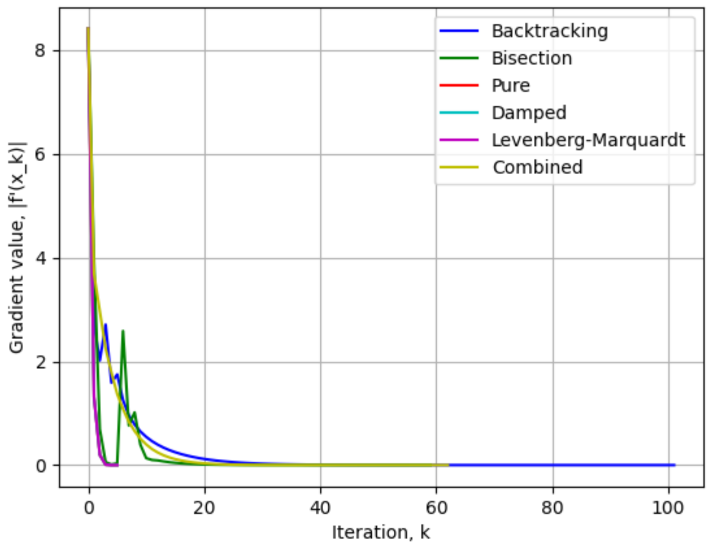

Figure 5: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (-2, -1)$

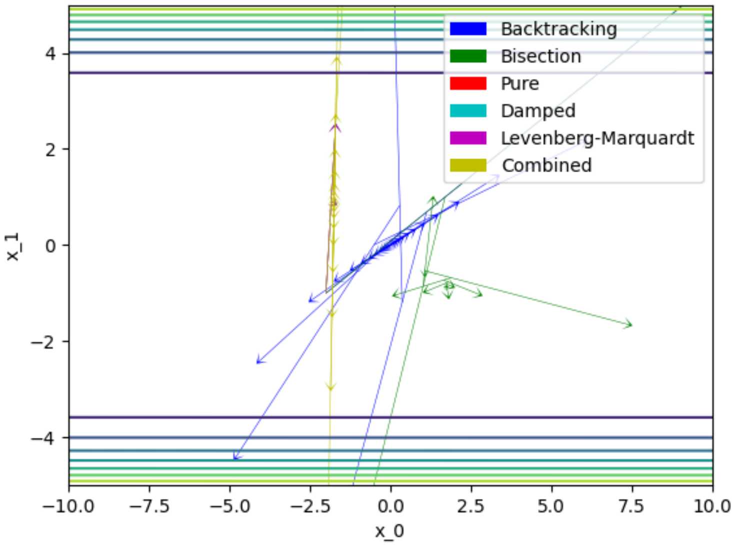

Figure 6: Contour plot, with direction of updates for $\mathbf{x}_0 = (-2, -1)$

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.

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

$$\implies 2x_i^3 - 16x_i + \frac{5}{2} = 0 \qquad \forall i \in [1, d]$$

$$
\begin{aligned}
&\implies x_i \approx -2.9035, 0.15673, 2.7468 && \forall i \in [1, d] && (9)
\end{aligned}
$$

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

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

1. Pure and Damped Newton's Method converged, for all except the first test case. Notably, the Pure Netwon's Method returned $\mathbf{x} = (0.157, 0.157, 0.157, 0.157)$ on failure, which as we saw earlier satisfies the FONC.
2. Steepest Descent with Backtracking, Steepest Descent with Bisection search, Levenberg-Marquardt Modification and Combined Damped Netwon's Method converged to one of the local minimas, for all the test cases.

3.4 Plots

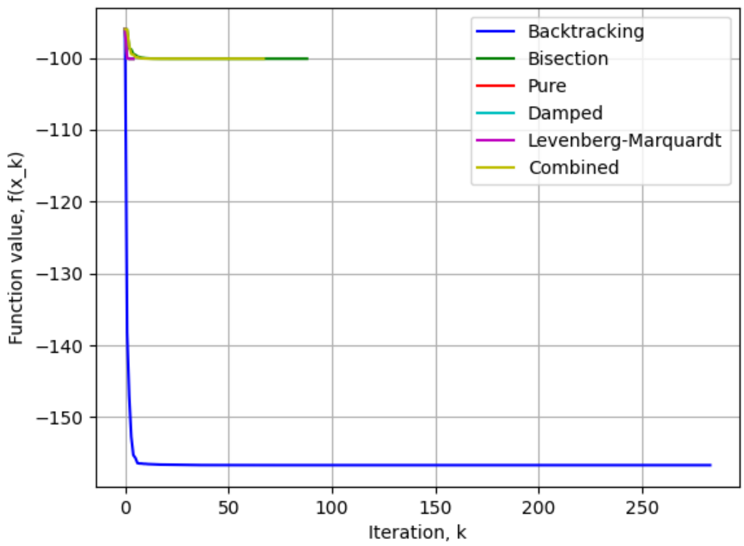

Figure 7: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (3, 3, 3, 3)$

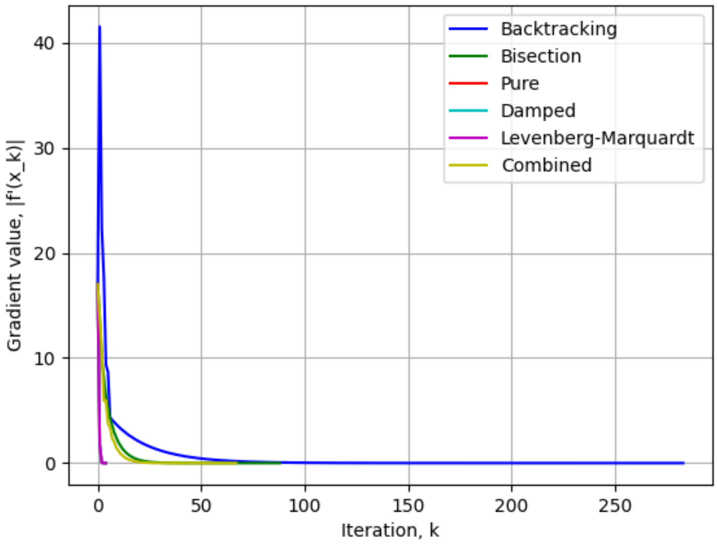

Figure 8: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (3, 3, 3, 3)$

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. Steepest Descent with Bisection search converged, for all except the first test case.
2. Pure Newton's Method, Damped Newton's Method and Levenberg-Marquardt Modification converged, for all except the third test case.

3. Combined Damped Netwon's Method converged only for the first and fourth test case.
4. Steepest Descent with Backtracking converged for all the test cases.

4.3 Plots

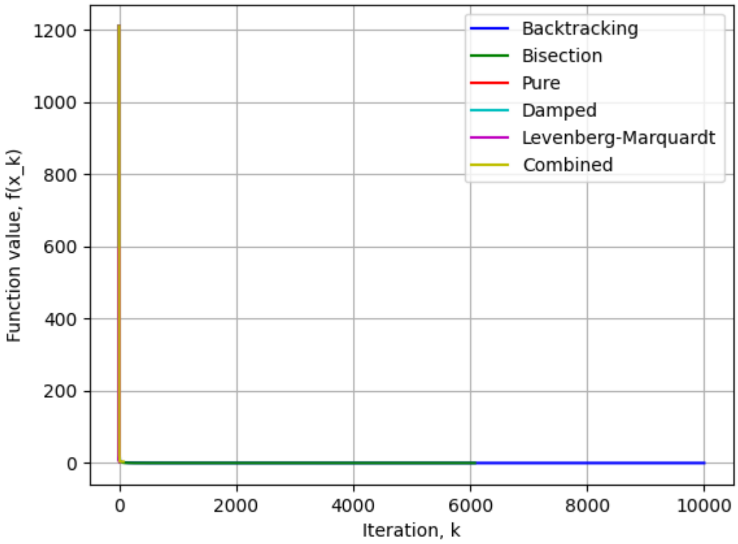

Figure 9: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (-2, 2, 2, 2)$

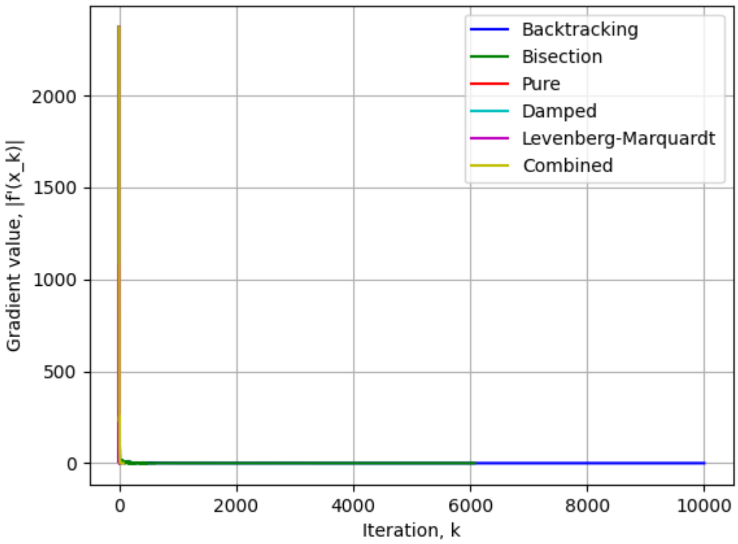

Figure 10: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (-2, 2, 2, 2)$

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.

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

$$\begin{aligned} &\mathbf{y}^T \nabla^2 f(\mathbf{x}^*) \mathbf{y} > 0 && \forall \mathbf{y} \in \mathbb{R}^2 \\ &\implies \frac{y_i^2}{\sqrt{(1+x_i^2)^3}} > 0 && \forall y_i \in \mathbb{R}, i \in [1,2] \\ &\implies (1+x_i^2)^3 > 0 && \forall i \in [1,2] \\ &\implies x_i^2 > -1 && \forall i \in [1,2] \end{aligned} \tag{13}$$

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

1. Pure Newton's Method and Levenberg-Marquardt Modification converged, only for the second test case.

2. Steepest Descent with Backtracking, Steepest Descent with Bisection search, Damped Newton's Method and Combined Damped Netwon's Method converged to the local minima $\mathbf{x}^* = (0,0)$, for all the test cases.

5.4 Plots

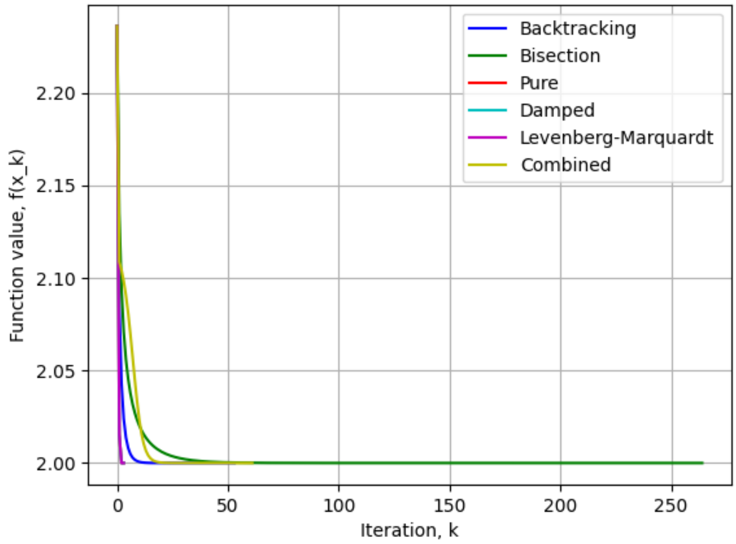

Figure 11: $f(\mathbf{x}_k)$ vs k, for $\mathbf{x}_0 = (-0.5, 0.5)$

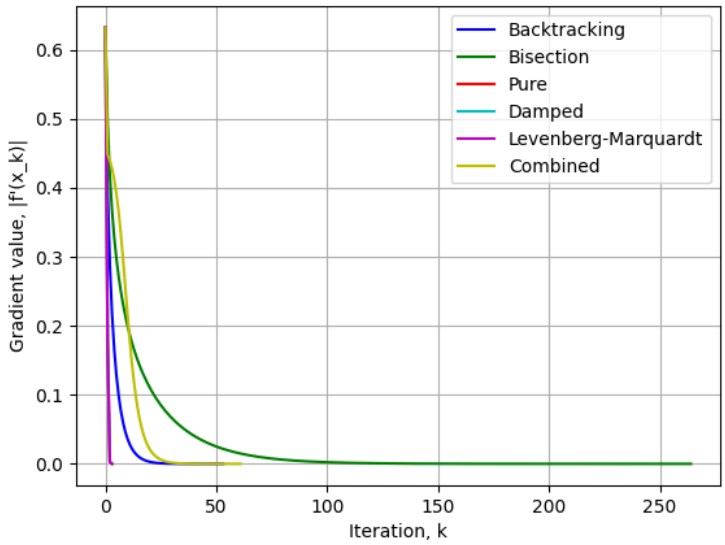

Figure 12: $|\nabla f(\mathbf{x}_k)|$ vs k, for $\mathbf{x}_0 = (-0.5, 0.5)$

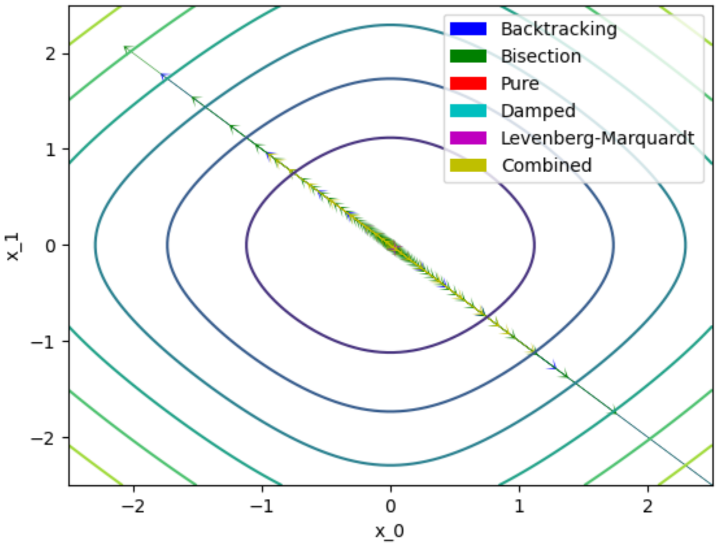

Figure 13: Contour plot, with direction of updates for $\mathbf{x}_0 = (-0.5, 0.5)$

Appendix: Output for all test cases

Test Case	Function	Initial Point	Backtracking	Bisection	Pure	Damped	Levenberg-Marquardt	Combined
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)	(-1.748, 0.874)	(-1.748, 0.874)	(-1.748, 0.874)	(-1.748, 0.874)	(0, 0)
3		(2, -1)	(1.748, -0.874)	(1.748, -0.874)	(1.748, -0.874)	(1.748, -0.874)	(1.748, -0.874)	(0, 0)
4		(-2, -1)	(0, 0)	(1.748, -0.874)	(-1.748, 0.874)	(-1.748, 0.874)	(-1.748, 0.874)	(-1.748, 0.874)
5		(2, 1)	(0, 0)	(-1.748, 0.874)	(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)	(-0.776, 0.613, 0.382, 0.146)	(1, 1, 1, 1)	(1, 1, 1, 1)	(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)	(1, 1, 1, 1)	(1, 1, 1, 1)	(-0.776, 0.613, 0.382, 0.146)
8		(-2, 2, 2, 2)	(1, 1, 1, 1)	(1, 1, 1, 1)	(-0.776, 0.613, 0.382, 0.146)	(-0.776, 0.613, 0.382, 0.146)	(-0.776, 0.613, 0.382, 0.146)	(-0.776, 0.613, 0.382, 0.146)
9		(3, 3, 3, 3)	(1, 1, 1, 1)	(1, 1, 1, 1)	(1, 1, 1, 1)	(1, 1, 1, 1)	(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)	(0.157, 0.157, 0.157, 0.157)	(0, 0, 0, 0)	(-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)	(2.747, 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)	(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)	(-8.72e+115, -8.72e+115)	(0, 0)	(-8.72e+115, -8.72+115)	(0, 0)

Test Case	Function	Initial Point	Backtracking	Bisection	Pure	Damped	Levenberg-Marquardt	Combined
15		(-0.5, 0.5)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
16		(-3.5, 0.5)	(0, 0)	(0, 0)	(4.9e+14, 0)	(0, 0)	(4.9e+14, 0)	(0, 0)