CS1.404 (Spring 2024)
Optimization Methods
Assignment - 1

Name - Prashant Kumar
Roll Number - 2022201058
Batch - M.Tech (CSE)

Part 1 & 2. Derive the Jacobians and Hessians for all the functions and calculate the minima except Rosenbrock using the Jacobians and Hessians.

- **Trid Function**

Function-1

Trid function:-

Part-1

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

expanding $f(x)$,

$$f(x) = (x_1-1)^2 + (x_2-1)^2 + \cdots + (x_d-1)^2 - x_1x_2 - x_2x_3 - \cdots - x_{d-1}x_d$$

$$J(x) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_d}\right]^T$$

$$\frac{\partial f}{\partial x_1} = 2(x_1 - 1) - x_2$$

$$\frac{\partial f}{\partial x_2} = 2(x_2 - 1) - (x_1 + x_3)$$

$$\vdots$$

$$\frac{\partial f}{\partial x_d} = 2(x_d - 1) - (x_{d-1})$$

$$J(x) = \begin{bmatrix} 2(x_1-1) - x_2 \\ 2(x_2-1) - (x_1+x_3) \\ 2(x_3-1) - (x_2+x_4) \\ \vdots \\ 2(x_d-1) - x_{d-1} \end{bmatrix}_{d \times 1}$$

$$J(x) = \begin{cases} 2(x_i - 1) - x_{i+1} & ; \ i = 1 \\ 2(x_i - 1) - (x_{i-1} + x_{i+1}) & ; \ i > 1 \ \& \ i < d \\ 2(x_i - 1) - x_{i-1} & ; \ i = d \end{cases}$$

$$\frac{\partial^2 f}{\partial x_1^2} = 2, \quad \frac{\partial^2 f}{\partial x_1 x_2} = -1, \quad \frac{\partial^2 f}{\partial x_1 x_3} = 0$$

$$\frac{\partial^2 f}{\partial x_2 x_1} = -1, \quad \frac{\partial^2 f}{\partial x_2^2} = 2, \quad \frac{\partial^2 f}{\partial x_2 x_3} = -1$$

$$\frac{\partial^2 f}{\partial x_d^2} = 2 \ ; \quad \frac{\partial^2 f}{\partial x_d x_{d-1}} = -1 \ \cdots$$

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_d} \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_d x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}_{d \times d} = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & -1 & 2 \end{bmatrix}$$

$$H = \begin{cases} 2 & ; \ \forall \text{ diagonal } (i == j) \\ -1 & ; \ \text{if } |i - j| = 1 \\ 0 & ; \ \text{otw} \end{cases}$$

Part-2

for $d=2$

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

$$f(x) = (x_1 - 1)^2 + (x_2 - 1)^2 - x_1 x_2$$

$$\frac{\partial f}{\partial x_1} = 2(x_1 - 1) - x_2 = 0$$

$$\frac{\partial f}{\partial x_2} = 2(x_2 - 1) - x_1 = 0$$

$2(x_1 - 1) = x_2$ — (1)

$2(x_2 - 1) = x_1$ — (2)

put x_1 value in eq (1)

$$2(2x_2 - 2 - 1) = x_2$$

$$4x_2 - 6 = x_2$$

$\boxed{x_2 = 2}$ $\boxed{x_1 = 2}$

$$\frac{\partial^2 f}{\partial x_1^2} = 2$$

$$\frac{\partial^2 f}{\partial x_1 x_2} = \frac{\partial^2 f}{\partial x_2 x_1} = -1$$

$$\frac{\partial^2 f}{\partial x_2^2} = 2$$

$$H = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

trace is +ve & determinant is +ve.

Since the hessian is +ve definite bcz all the leading principal minors are +ve, hence [2, 2] is the mima for the function

for general d.

$\nabla(x) = 0$

$$\begin{bmatrix} 2(x_1 - 1) - x_2 \\ 2(x_2 - 1) - x_1 - x_3 \\ \vdots \\ 2(x_d - 1) - x_{d-1} \end{bmatrix} = 0$$

$x_2 = 2x_1 - 2$ — (1)

$$x_3 = 2(x_1 - 1) - x_1$$
$$= 2(2x_1 - 3) - x_1$$
$$= 3x_1 - 6$$ — (2)

~~$x_4 = 2(x_3 - 1) - x_2$~~

~~$= 4(3x_1 - 6 - 1) - x_2$~~

~~$= 8x$~~

x_d.

we get $x^* = \begin{bmatrix} x_1 \\ 2x_1 - 2 \\ 3x_1 - 6 \\ 4x_1 - 12 \\ \vdots \\ i x_i - i(i-1) \\ \vdots \\ d x_d - d(d-1) \end{bmatrix}$

- **Three Hump Camel**

function-2

Three hump Camel :-

Part-1

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

$$J(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}^T$$

$$\frac{\partial f}{\partial x_1} = 4x_1 - 4.20x_1^3 + x_1^5 + x_2$$

$$\frac{\partial f}{\partial x_2} = x_1 + 2x_2$$

$$J(x) = \begin{bmatrix} 4x_1 - 4.20x_1^3 + x_1^5 + x_2 \\ x_1 + 2x_2 \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial x_1^2} = 4 - 12.6x_1^2 + 5x_1^4$$

$$\frac{\partial^2 f}{\partial x_1 x_2} = 1$$

$$\frac{\partial^2 f}{\partial x_2 x_1} = 1$$

$$\frac{\partial^2 f}{\partial x_2^2} = 2$$

$$H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

$$Hf = \begin{bmatrix} 4 - 12.6x_1^2 + 5x_1^4 & 1 \\ 1 & 2 \end{bmatrix}_{2\times 2}$$

Part-2

$$J(x) = 0$$

$$4x_1 - 4.20x_1^3 + x_1^5 + x_2 = 0 \quad \text{— (1)}$$

$$x_1 + 2x_2 = 0 \quad \text{— (2)}$$

$x_2 = -\frac{x_1}{2}$ put in eq-(1)

$$4x_1 - 4.2x_1^3 + x_1^5 - \frac{x_1}{2} = 0$$

$$8x_1 - 8.4x_1^3 + 2x_1^5 - x_1 = 0$$

$$2x_1^5 - 8.4x_1^3 + 7x_1 = 0$$

$$x_1[2x_1^4 - 8.4x_1^2 + 7] = 0$$

$$\boxed{x_1 = 0}$$

$$2x_1^4 - 8.4x_1^2 + 7 = 0$$

let $t = x_1^2$

$$[2t^2 - 8.4t + 7 = 0]$$

after solving quadratic equation.

$t_1 = 3.0539,\ 1.1460$

$x_1^2 = 3.0539 \qquad x_1^2 = 1.1460$

$$\begin{bmatrix} x_1 = -1.7475 \\ +1.7475 \end{bmatrix} \quad \begin{bmatrix} x_1 = -1.0705 \\ +1.0705 \end{bmatrix} \quad \begin{bmatrix} x_1 = 0 \end{bmatrix}$$

(I) stationary points are-

$$\begin{bmatrix} (0,0) \\ (-1.7475, 0.8737) \\ (+1.7475, -0.8737) \\ (-1.0705, 0.5352) \\ (+1.0705, -0.5352) \end{bmatrix}$$

at $(0,0)$. $H(x) = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}$

It is semi +ve definite matrix
minima at $(0,0)$ ✓

at $(-1.7475, 0.8737)$

$H(x) = \begin{bmatrix} 12.1435 & 1 \\ 1 & 2 \end{bmatrix}$

semi +ve definite,
minima exist ✓

at $(+1.7475, -0.8737)$

$H(x) = \begin{bmatrix} 12.1435 & 1 \\ 1 & 2 \end{bmatrix}$

$\det(H(x)) > 0$ & trace $(H) > 0$
it is semi +ve definite.
✓

at $(-1.0705, 0.5352)$

$H(x) = \begin{bmatrix} -3.872 & 1 \\ 1 & 2 \end{bmatrix}$

$\det(H) < 0$ & trace$(H) < 0$
It is not semi +ve def
×

at $(+1.0705, -0.5352)$

$H(x) = \begin{bmatrix} -3.872 & 1 \\ 1 & 2 \end{bmatrix}$

$\det(H) < 0$ & trace$(H) < 0$
not a semi +ve definite
×

- **Styblinski-Tang Function**

Function-3 ③

Styblinski - Tang function

Part-1

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

expanding $f(x)$, we get

$$f(x) = \frac{1}{2}(x_1^4 - 16x_1^2 + 5x_1) + \frac{1}{2}(x_2^4 - 16x_2^2 + 5x_2) + \cdots\cdots$$

$$J(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \frac{\partial f}{\partial x_3} & \cdots\cdots & \frac{\partial f}{\partial x_d} \end{bmatrix}^T$$

$$\frac{\partial f}{\partial x_1} = \frac{1}{2}(4x_1^3 - 32x_1 + 5)$$

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{1}{2}(12x_1^2 - 32)$$

$$\frac{\partial f}{\partial x_2} = \frac{1}{2}(4x_2^3 - 32x_2 + 5)$$

$$\frac{\partial^2 f}{\partial x_1 x_2} = 0$$

$$\vdots$$

$$\frac{\partial f}{\partial x_d} = \frac{1}{2}(4x_d^3 - 32x_d + 5)$$

$$J(x) = \begin{bmatrix} \frac{1}{2}(4x_1^3 - 32x_1 + 5) \\ \frac{1}{2}(4x_2^3 - 32x_2 + 5) \\ \vdots \\ \frac{1}{2}(4x_d^3 - 32x_d + 5) \end{bmatrix}$$

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 f}{\partial x_1 x_3} & \cdots & \frac{\partial^2 f}{\partial x_1 x_d} \\ \vdots & & & & \vdots \\ \frac{\partial^2 f}{\partial x_d x_1} & \frac{\partial^2 f}{\partial x_d x_2} & \frac{\partial^2 f}{\partial x_d x_3} & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

$$H = \begin{bmatrix} \frac{1}{2}(12x_1^2 - 32) & 0 & \cdots & 0 \\ 0 & \frac{1}{2}(12x_2^2 - 32) & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \frac{1}{2}(12x_d^2 - 32) \end{bmatrix}$$

Part-2

To find minimas $\nabla f(x) = 0$

$\frac{1}{2}(4x_1^3 - 32x_1 + 5) = 0$

$\frac{1}{2}(4x_2^3 - 32x_2 + 5) = 0$

$\vdots$

$\frac{1}{2}(4x_d^3 - 32x_d + 5) = 0$

$4x_1^3 - 32x_1 + 5 = 0$ —— ①

after solving the cubic equation, we get

$x_1 = -2.9035$
$x_1 = 0.1567$
$x_1 = 2.7468$

for $d=2$ stationary points in the form of (x_1, x_2) are

$(-2.9035, -2.9035)$
$(0.1567, 0.1567)$
$(2.7468, 2.7468)$

$$H = \begin{bmatrix} 6x_1^2 - 16 & 0 \\ 0 & 6x_2^2 - 16 \end{bmatrix}$$

for $(-2.9035, -2.9035)$, $H = \begin{bmatrix} 34.5818 & 0 \\ 0 & 34.5818 \end{bmatrix}$ H is +ve definite so minima exists. ✓

for $(0.1567, 0.1567)$, $H = \begin{bmatrix} -15.8526 & 0 \\ 0 & -15.8526 \end{bmatrix}$ H is not +ve definite so minima does not exist at this point. ✗

for $(2.7468, 2.7468)$, $H = \begin{bmatrix} 29.2694 & 0 \\ 0 & 29.2694 \end{bmatrix}$ +ve define. So minima exists ✓

- **Rosenbrock Function**

4

Rosenbrock function :-

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

expanding $f(x)$, we get

$$f(x) = \left[100\,(x_2 - x_1^2)^2 + (x_1 - 1)^2\right] + \left[100\,(x_3 - x_2^2)^2 + (x_2 - 1)^2\right] + \dots + \left[100\,(x_d - x_{d-1}^2)^2 + (x_{d-1} - 1)^2\right]$$

$$J(x) = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \dots \quad \frac{\partial f}{\partial x_d}\right]^T$$

$$\frac{\partial f}{\partial x_1} = 200\,(x_2 - x_1^2)(-2x_1) + 2(x_1 - 1)$$

$$\frac{\partial f}{\partial x_2} = 200\,(x_3 - x_2^2)(-2x_2) + 2(x_2 - 1) + 200\,(x_2 - x_1^2)$$

$$\vdots$$

$$\frac{\partial f}{\partial x_d} = 200\,(x_d - x_{d-1}^2)$$

$$J(x) = \begin{bmatrix} -400(x_2 - x_1^2)(x_1) + 2(x_1 - 1) \\ +400(x_2 - x_1^2) - 400\,(x_3 - x_2^2)x_2 + 2(x_2 - 1) \\ \vdots \\ 200\,(x_d - x_{d-1}^2) \end{bmatrix}$$

$$J = \begin{cases} -400x_i\,(x_{i+1} - x_i^2) + 2(x_i - 1); & \text{if } i = 1 \\ 200\,(x_i - x_{i-1}^2); & \text{if } i = d \\ -400x_i(x_{i+1} - x_i^2) + 2(x_i - 1) + 400\,(x_i - x_{i-1}^2); & \text{otw} \end{cases}$$

$$\frac{\partial^2 f}{\partial x_1^2} = -400(x_2 - x_1^2) - 400x_1(-2x_1) + 2$$

$$= -400\left[(x_2 - x_1^2 - 2x_1^2\right] + 2$$

$$= 2 - 400x_2 + 1200x_1^2$$

$$\frac{\partial^2 f}{\partial x_1 x_2} = -400x_1$$

$$\frac{\partial^2 f}{\partial x_1 x_3} = 0$$

$$\frac{\partial^2 f}{\partial x_2 x_1} = -400x_1$$

$$\frac{\partial^2 f}{\partial x_2^2} = -400[x_3 - 3x_2^2] + 2 + 200$$

$$= 202 - 400x_3 + 1200x_2^2$$

$$\frac{\partial^2 f}{\partial x_2 x_3} = -400x_2$$

$$\frac{\partial^2 f}{\partial x_d x_{d-1}} = -400x_{d-1}$$

$$\frac{\partial^2 f}{\partial x_d^2} = 200$$

$$H = \begin{bmatrix} 2 - 400x_2 + 1200x_1^2 & -400x_1 & 0 & \cdots & 0 \\ -400x_1 & 202 - 400x_3 + 1200x_2^2 & -400x_2 & \cdots & 0 \\ \vdots & \vdots & & & \\ 0 & \cdots & 0 & & 200 \end{bmatrix}_{d \times d}$$

- **Root of Square Function**

5

Root of Square function:-

Part-1

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

$$J(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}$$

$$\frac{\partial f}{\partial x_1} = \frac{1}{2\sqrt{1+x_1^2}} * 2x_1$$

$$J = \begin{bmatrix} \frac{x_1}{\sqrt{1+x_1^2}} \\ \frac{x_2}{\sqrt{1+x_2^2}} \end{bmatrix}$$

$$\frac{\partial f}{\partial x_1} = \frac{x_1}{\sqrt{1+x_1^2}}$$

$$\frac{\partial f}{\partial x_2} = \frac{x_2}{\sqrt{1+x_2^2}}$$

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{1}{\sqrt{1+x_1^2}} + x_1 * \left(\frac{-1}{\cancel{2}}\right)(1+x_1^2)^{-\frac{3}{2}} * \cancel{2}x_1$$

$$= \frac{1}{\sqrt{1+x_1^2}} - \frac{x_1^2}{\sqrt{1+x_1^2}\,(1+x_1^2)}$$

$$= \frac{1}{\sqrt{1+x_1^2}}\left(1 - \frac{x_1^2}{1+x_1^2}\right)$$

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{1}{(1+x_1^2)^{3/2}}$$

Similarly:-

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{1}{(1+x_2^2)^{3/2}}$$

$$H = \begin{bmatrix} \frac{1}{(1+x_1^2)^{3/2}} & 0 \\ 0 & \frac{1}{(1+x_2^2)^{3/2}} \end{bmatrix}$$

Part-2

for minima $J(x) = 0$

$$\frac{x_1}{\sqrt{1+x_1^2}} = 0 \qquad \frac{x_2}{\sqrt{1+x_2^2}} = 0$$

$$[x_1 = x_2 = 0]$$

$$H(0,0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$\det(H) > 0$, trace ≥ 0

Semi positive definite.

minima exists at $(0, 0)$

Hence $x^* = [0\ 0]^T$

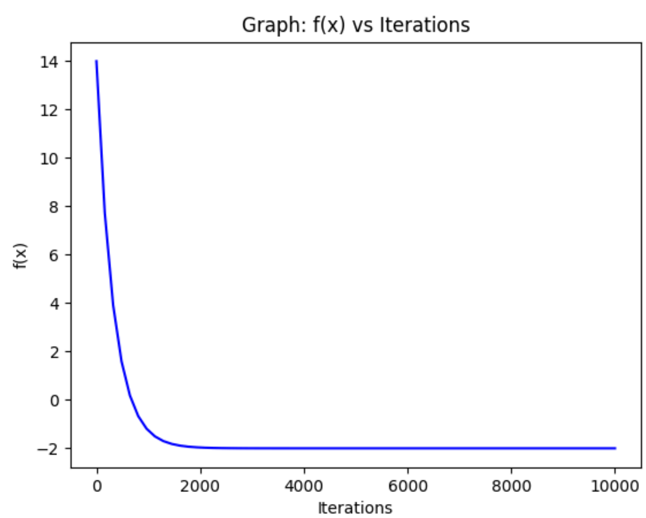
Part 3. State which algorithms failed to converge and under which circumstances.

- Newton's method failed to converge under the Damped Newton method in the case of styblinski_tang_function function at the initial point is [0. 0. 0. 0.].
- Rosenbrock function doesn't converge for initial points, [2,2,2,-2], [2,-2,-2,2], [-2,2,2,2], and [3,3,3,3] for both backtracking and bisection.
- In some iterations, the gradient matrix becomes singular, hence inverse can't be found, therefore we get an error and the function doesn't converge.
  Error in Pure Newton's method
  The test function was func_1 with [3. 3.] as starting point
  Singular matrix
- In some cases, max iterations are exhausted, but still, the function is not convergent.
- **Root of square function -** For TC 14, there was overflow encountered.
  RuntimeWarning: overflow encountered in power return np.diag(1 / ((1 + point**2) ** 1.5))
  The reason could be as stated earlier. Starting from a point that is far away from the minima in Newton's method makes the iterations *divergent*.
  This is the reason for TC14 and TC16, you see such huge values in the Levenberg-Marquard column.

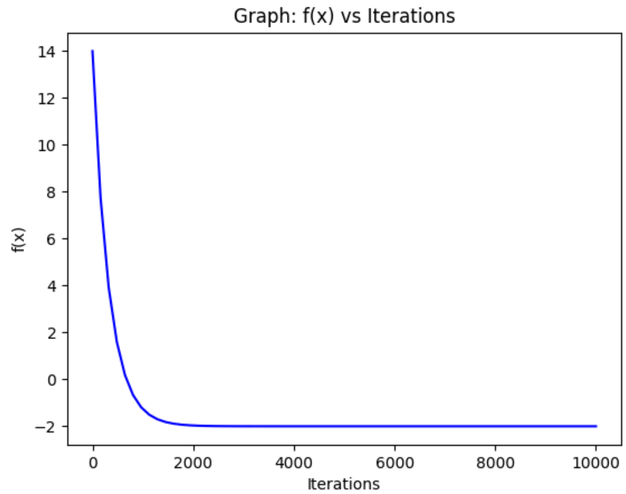
Part 4. Plot f (x) vs iterations and |f ′ (x)| vs iterations

- **f (x) vs iterations**

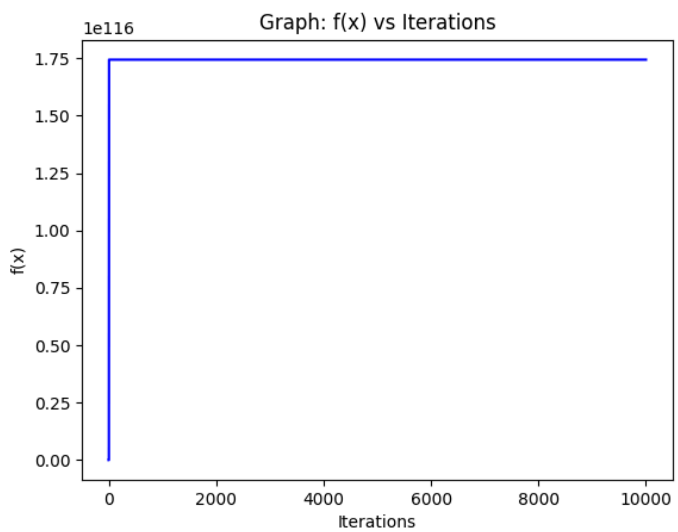
trid_function_[-2. -2.]_Combined_vals



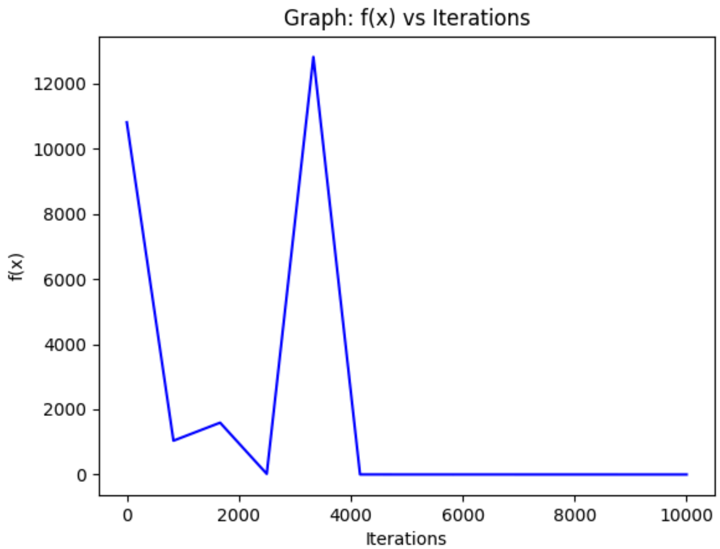
trid_function_[-2. -2.]_Backtracking_vals



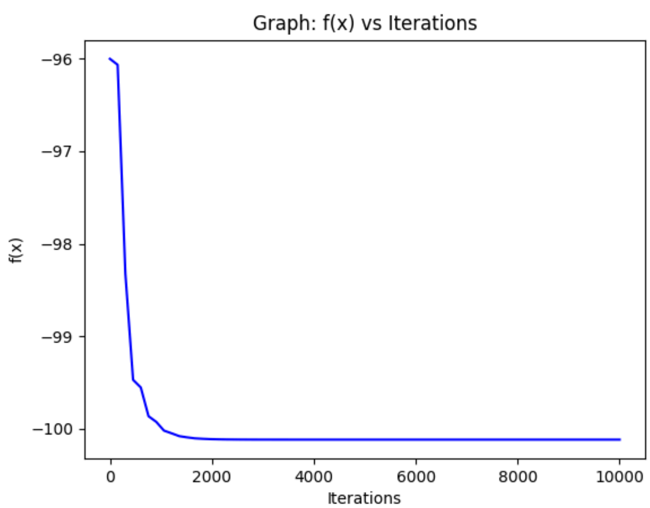
func_1_[3. 3.]_Levenberg-Marquardt_vals At the initial point (3,3) lavenberg marquardt condition the function is not converge.



rosenbrock_function_[3. 3. 3. 3.]_Pure_vals

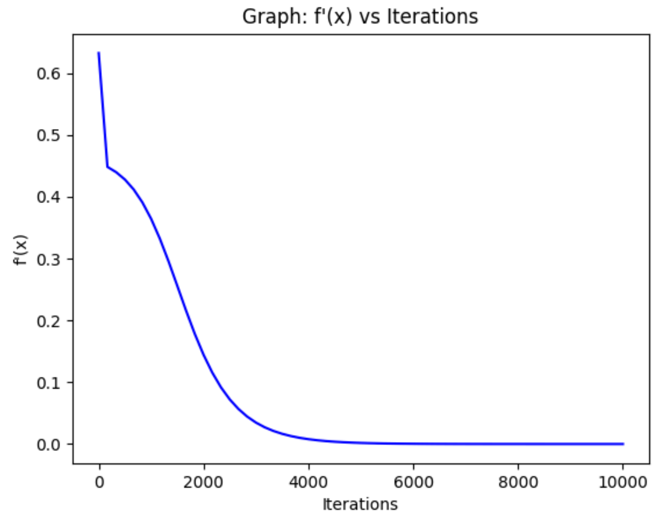


styblinski_tang_function_[3. 3. 3. 3.]_Combined_vals

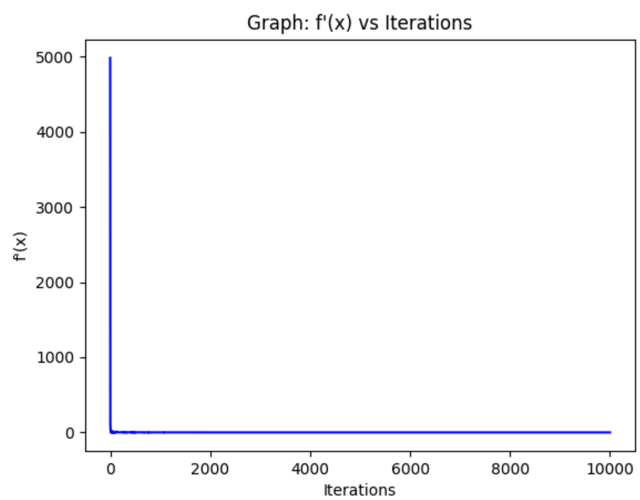


- **|f ′ (x)| vs iterations**

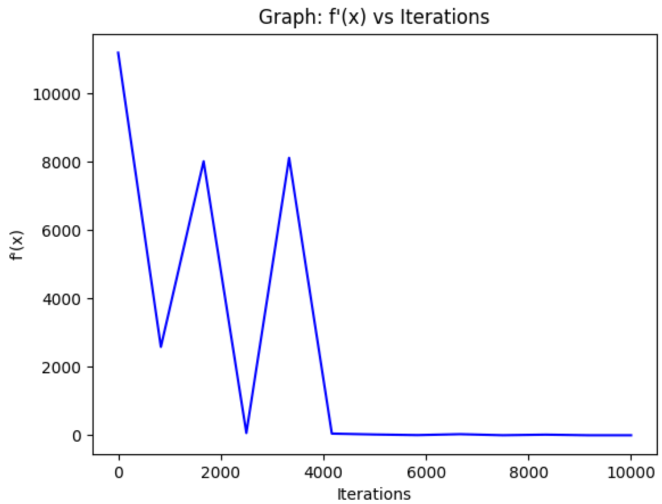
func_1_[-0.5 0.5]_Combined_grad



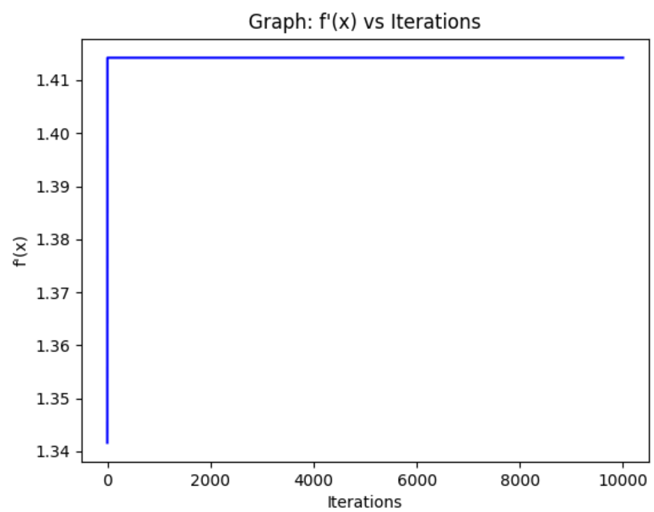
rosenbrock_function_derivative_[2. 2. 2. -2.]_Bisection_grad



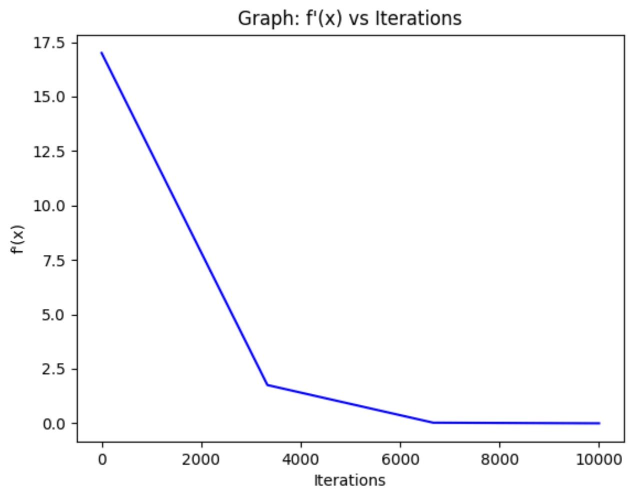
rosenbrock_function_[3. 3. 3. 3.]_Levenberg-Marquardt_grad



func_1_[3. 3.]_Levenberg-Marquardt_grad

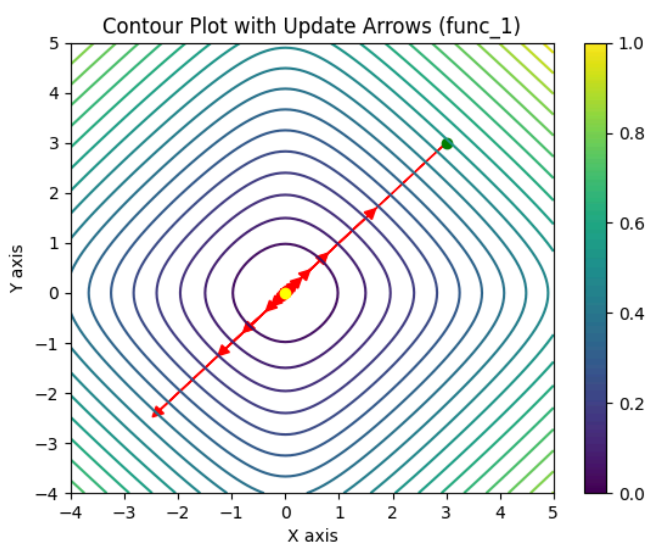


styblinski_tang_function_[3. 3. 3. 3.]_Damped_grad

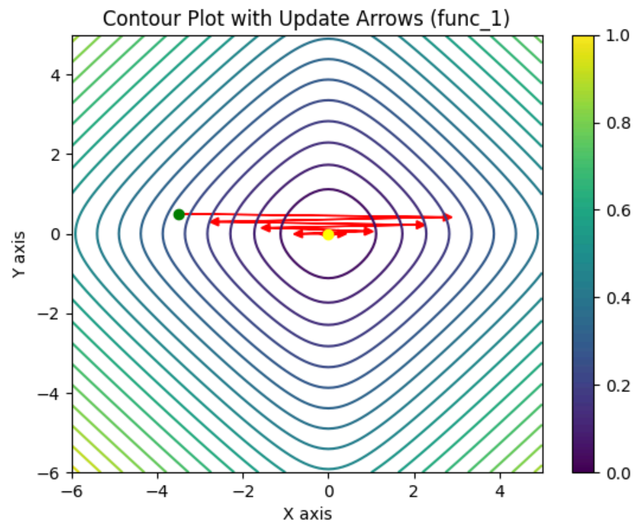


Part 5. Make a contour plot with arrows indicating the direction of updates for all 2-D functions.

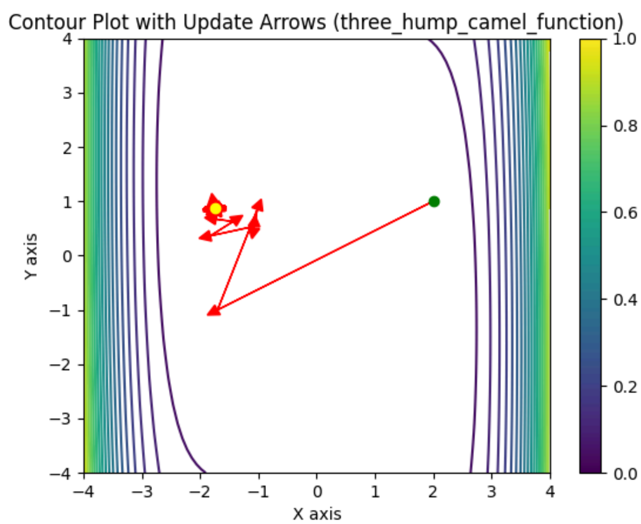
func_1_[3. 3.]_Backtracking_cont



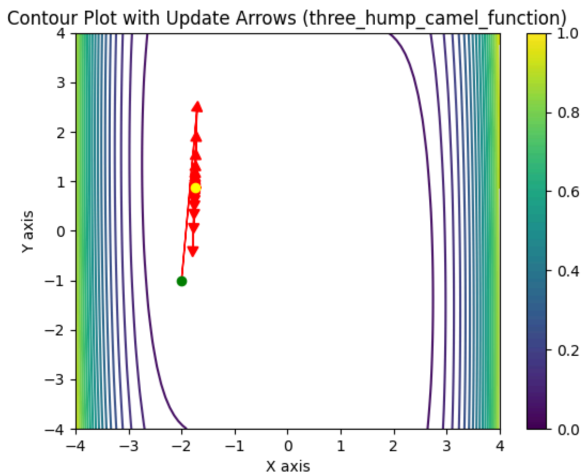
func_1_[-3.5 0.5]_Damped_cont



three_hump_camel_function_[2. 1.]_Bisection_cont



three_hump_camel_function_[-2. -1.]_Combined_cont



trid_function_[-2. -2.]_Combined_cont