Question-1 AI21BTECH11011 Gunjit Mittal

Question

Consider the following unconstrained optimization problem.

$$\min \quad \frac{1}{3}x_1^6 - 2.1x_1^4 + 4x_2^4 + 4(x_1^2 - x_2^2) + x_1x_2$$

$$\text{Sub. to} \;\; x \in \mathbb{R}^2$$

Using a programming language of your choice, obtain the minima of the optimization problem using the Gradient Descent method (with a fixed step size of 0.1) and Newton's method. Use the initial point as $x^0 = [1, -1]^T$.

Answer

Gradient Descent

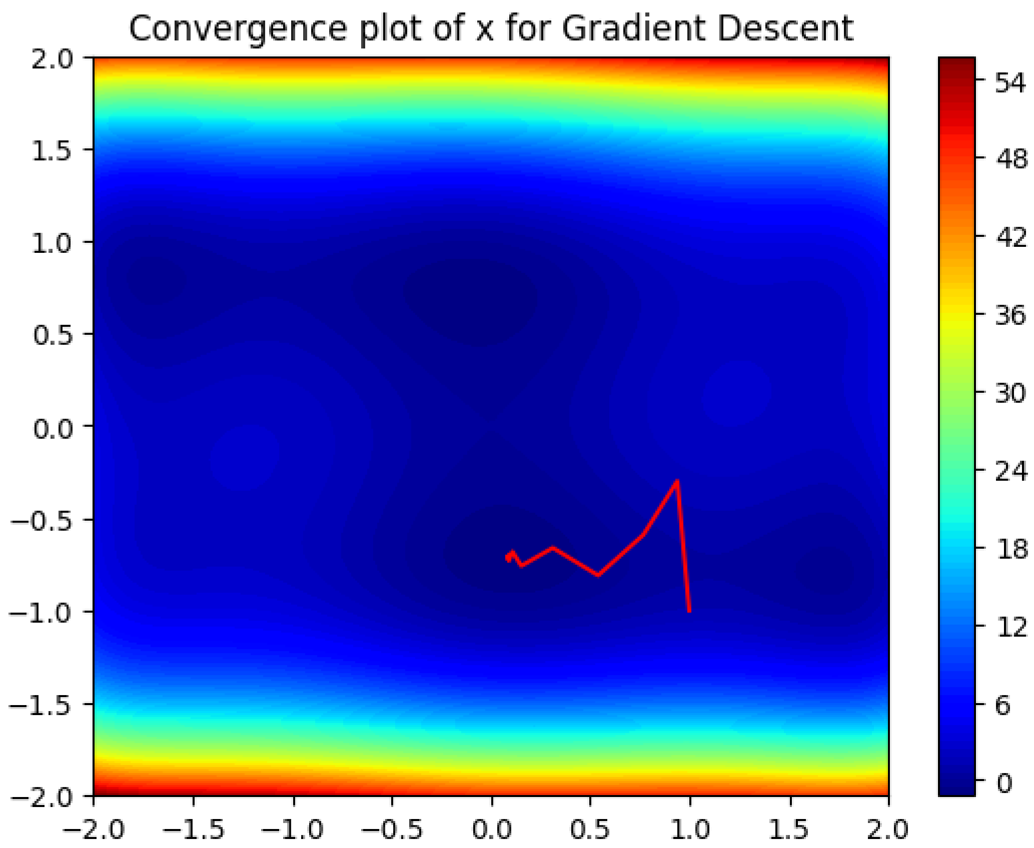
Given $f(x) = \frac{1}{3}x_1^6 - 2.1x_1^4 + 4x_2^4 + 4(x_1^2 - x_2^2) + x_1x_2$. For gradient descent, we use the direction as the negative of the gradient and since the step size is fixed $(\alpha = 0.1)$ we can directly use it.

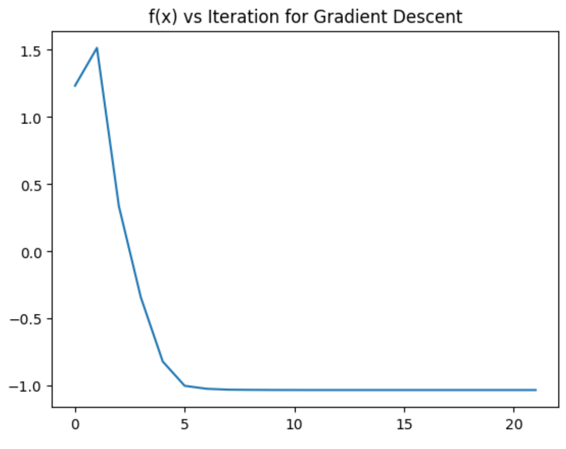
The gradient for this function is $\nabla f(x) = \begin{bmatrix} 2x_1^5 - 8.4x_1^3 + 8x_1 + x_2 \\ 16x_2^3) - 8x_2 + x_1 \end{bmatrix}$

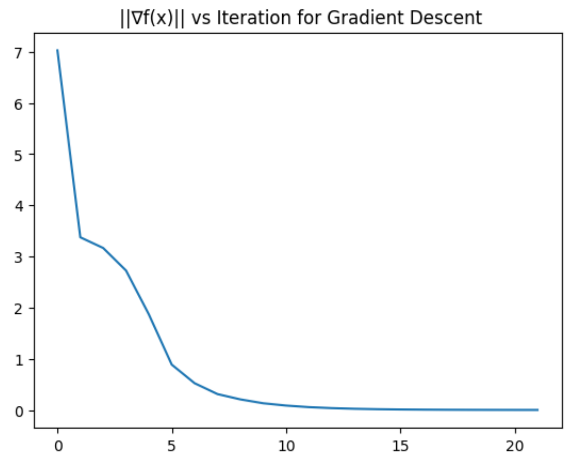
Using the initial conditions provided in the question and the following update equation:

$$x^{k+1} = x^k - \alpha \nabla f(x)$$

Solving through code, we get the following plots:







The minima is $x = [0.089836 - 0.712701]$ with the minimum value being -1.031628, which can be seen from the plots as well. The tolerance(ϵ) used for this code is 0.001. The code converges in 21 iterations.

Newton's method

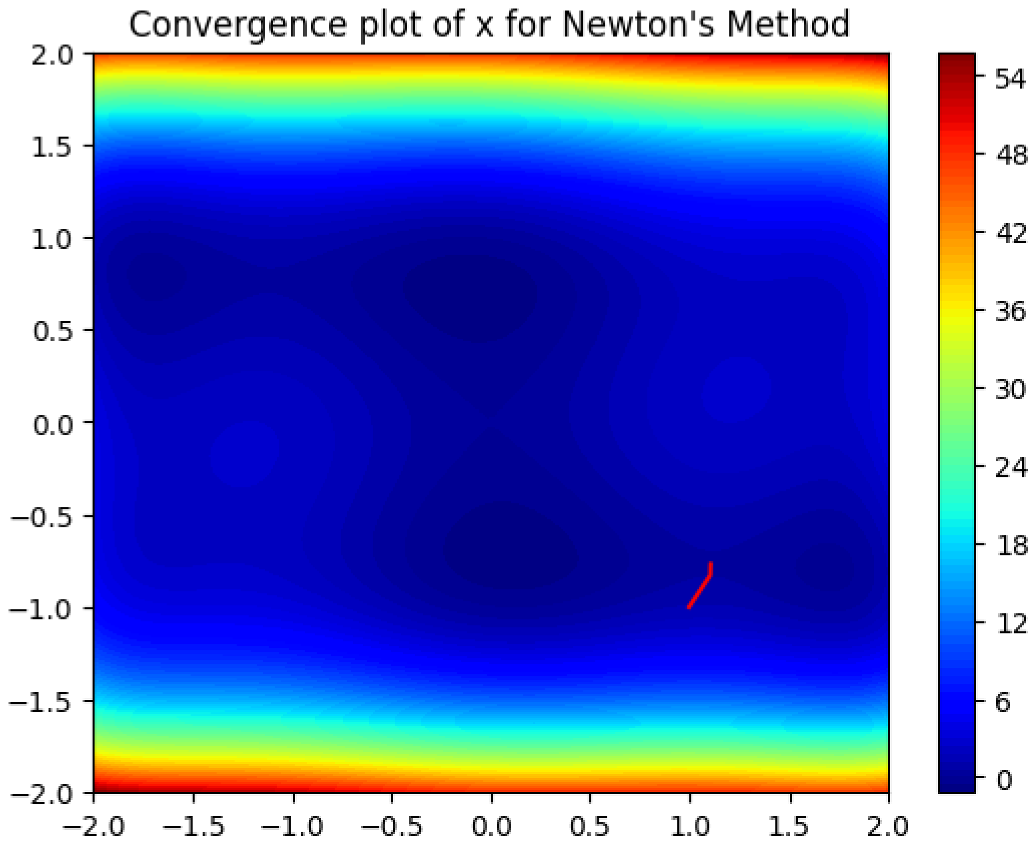
Other than $f(x)$ and $\nabla f(x)$ we also need $\nabla^2 f(x)$ for Newton's method.

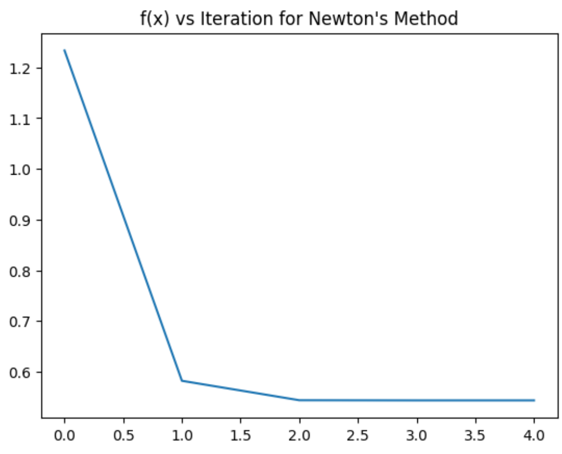
Calculating it, we get $\nabla^2 f(x) = \begin{bmatrix} 10x_1^4 - 25.2x_1^2 + 8 & 1 \\ 1 & 48x_2^2 - 8 \end{bmatrix}$

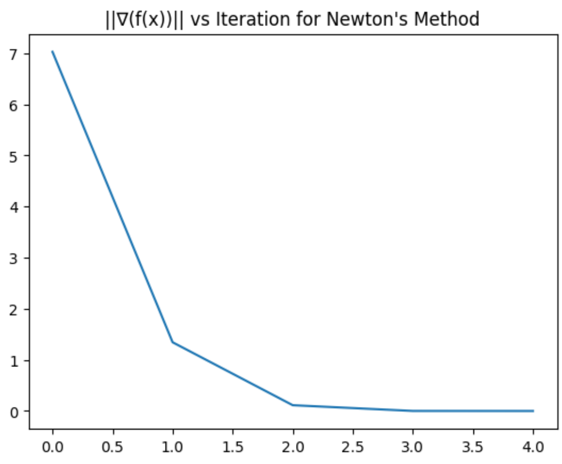
Using the initial conditions provided in the question and the following update equation:

$$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$$

Solving through code, we get the following plots:

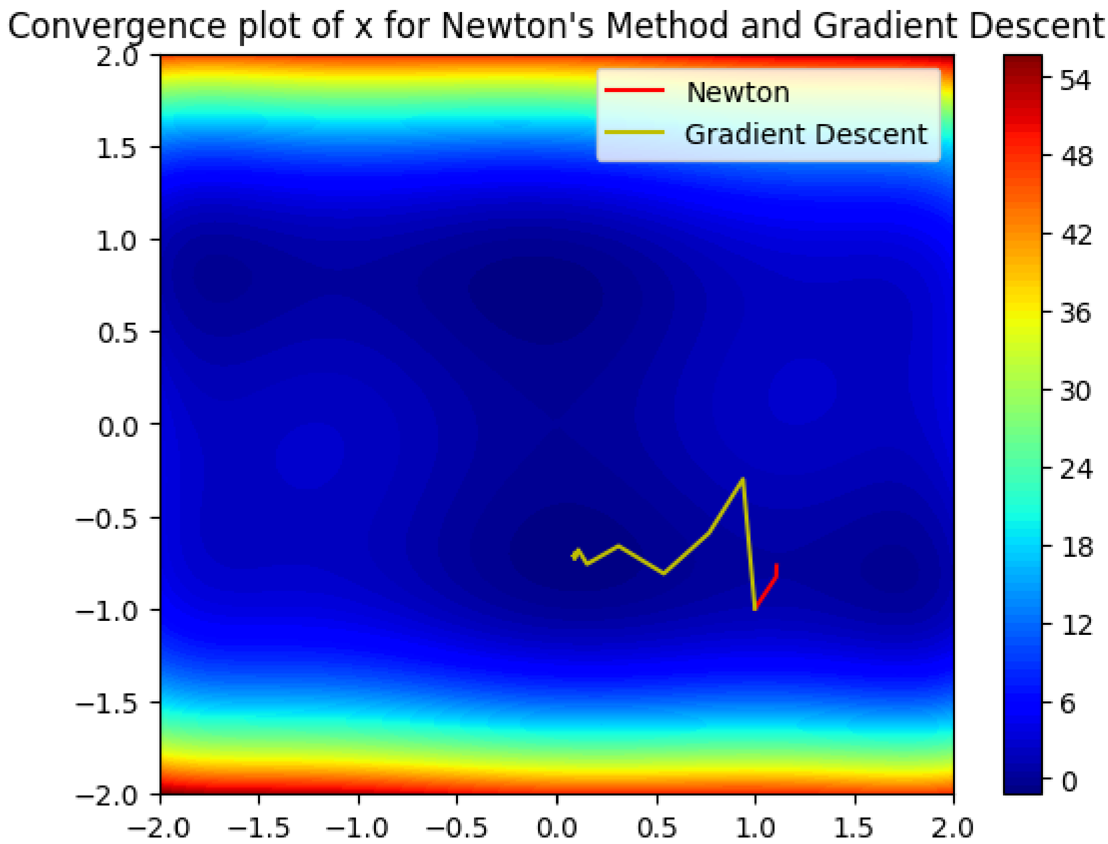






The minima is $x = [1.109205, -0.768268]$ with the minimum value being 0.543718, which can be seen from the plots as well. The tolerance(ϵ) used for this code is $1e^{-6}$. The code converges in 4 iterations.

Gradient descent vs Newton's method



Newton's method converged in 4 iterations compared to 21 for the Gradient Descent, which is significantly faster. Though the minimum value achieved by Gradient descent

From the combined convergence plot for Newton's method and Gradient descent, we can see that both converge to different points despite starting from the same starting point. Newton's method converged to a saddle point and failed to converge further. But since the step size for Gradient descent was large enough, it could escape the saddle point and was able to converge further.