

Nonlinear Optimization - Homework 1

Christian Segercrantz 481056

October 21, 2021

2.1

$$f(x) = x_1^3 - x_1 + x_2^3 - x_2 \quad (1)$$

(a)

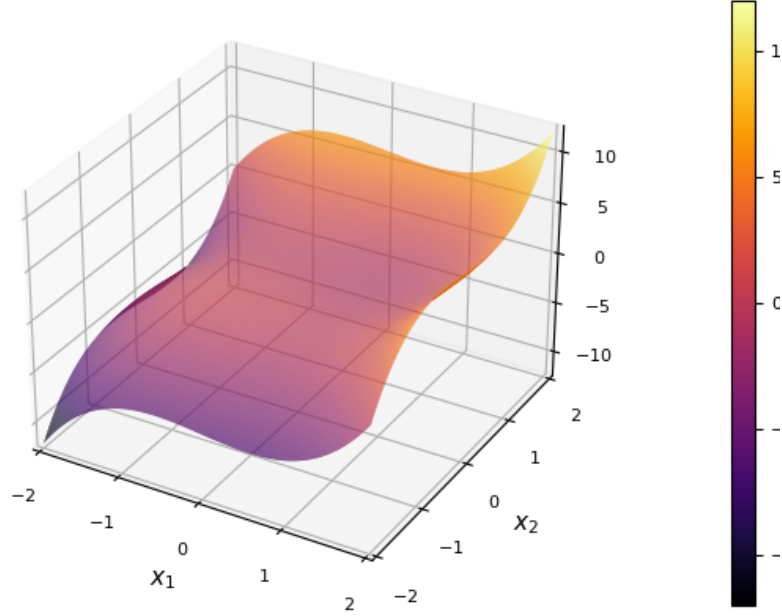


Figure 1: The surface plot of equation (1).

By examining Figure 1, I would conclude that the function is non-convex as we can clearly see areas that do not follow the definition of a convex function $f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$.

(b)

In order for us to find the critical point, we will solve the first-order derivative of the function.

$$\nabla f(x_1, x_2) = \begin{bmatrix} 3x_1^2 - 1 \\ 3x_2^2 - 1 \end{bmatrix}. \quad (2)$$

The first-order necessary condition is $\nabla f(\bar{x}) = 0$ for our unconstrained problem. We can solve the equations to get the critical points:

$$2x^2 - 1 = 0 \iff x = \pm\sqrt{1/3}. \quad (3)$$

This gives us the four points $(\sqrt{1/3}, \sqrt{1/3})$, $(-\sqrt{1/3}, \sqrt{1/3})$, $(\sqrt{1/3}, -\sqrt{1/3})$, and $(-\sqrt{1/3}, -\sqrt{1/3})$.

(c)

We calculate the hessian from the gradient and receive

$$Hf(x) = \begin{bmatrix} 6x_1 & 0 \\ 0 & 6x_2 \end{bmatrix} \quad (4)$$

In order for us to know how the function behaves, we need to examine the semi-definiteness of the Hessian. A positive semi-definite matrix has all of its eigenvalues non-negative and vice versa for a negative semi-definite matrix for all x . Additionally, a convex(/concave) functions hessian, if twice differentiable, is positive(/negative) semi-definite. The two eigen values of the function is $\lambda_1 = 6x_1$ and $\lambda_2 = 6x_2$. We can see that only for values where both x_1 and x_2 are non-negative(/-positive) is the hessian positive(/negative) semi-definite. In other words, the function is not convex(concave).

Using Julia we can solve the eigen values in order to get the curvature around the points. The point $(\sqrt{1/3}, \sqrt{1/3})$ has all positive eigen values, $(-\sqrt{1/3}, -\sqrt{1/3})$ has all negative eigen values and the other two one positive and one negative. From this we can conclude that $(\sqrt{1/3}, \sqrt{1/3})$ is a local minima, $(-\sqrt{1/3}, -\sqrt{1/3})$ a local maxima and the two others saddle points.

2.2

$$f(x_1, x_2) = 2x_1^2 - x_1x_2 + x_2^2 - 3x_1 + e^{2x_1+x_2} \quad (5)$$

(a)

By corollary 6 of lecture 4, we know that the first-order necessary condition for unconstrained problems is that "Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable at \bar{x} . If \bar{x} is a local minimum, then $\nabla f(\bar{x}) = 0$ ". The equation to be satisfied is thus $\nabla f(\bar{x}) = 0$. We calculate ∇f :

$$\nabla f = \begin{bmatrix} 4x_1 - x_2 - 3 + 2e^{2x_1+x_2} \\ -x_1 + 2x_2 + e^{2x_1+x_2} \end{bmatrix} \quad (6)$$

For this to be a sufficient condition for optimality the function $f(x_1, x_2)$ has to be convex. Then, based on theorem 8 from lecture 4 can we have sufficient conditions. Since we know the following things:

1. Polynomials are convex,
2. The linear combination of convex functions are convex,

we can conclude that the function $f(x_1, x_2)$ is convex. We thus have the necessary and sufficient condition for optimality.

(b)

If $\bar{x} = (0, 0)$ is to be a optimal point, must it satisfy $\nabla f(0, 0) = \mathbf{0}$.

$$\nabla f = \begin{bmatrix} 4x_1 - x_2 - 3 + 2e^{2x_1+x_2} \\ -x_1 + 2x_2 + e^{2x_1+x_2} \end{bmatrix} \quad (7)$$

$$= \begin{bmatrix} 0 - 0 - 3 + 2e^0 \\ -0 + 0 + e^0 \end{bmatrix} \quad (8)$$

$$= \begin{bmatrix} -1 \\ 1 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (9)$$

We can see that the point $(0, 0)$ does not satisfy the condition.

The direction d that makes the function decrease must satisfy the condition $\nabla f(\bar{x})^\top d < 0$.

(c)

To find the minimum for $f(x)$ in the direction $d = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ we must calculate the step size $\bar{\lambda} = \operatorname{argmin}_{\lambda} d^T \nabla f(x + \lambda d)$.

$$\theta'(\lambda) = \begin{bmatrix} 1 & 0 \end{bmatrix} \nabla f \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \quad (10)$$

$$= \begin{bmatrix} 1 & 0 \end{bmatrix} \nabla f \left(\begin{bmatrix} x_1 + \lambda \\ x_2 \end{bmatrix} \right) \quad (11)$$

$$= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 4(x_1 + \lambda) - x_2 - 3 + 2e^{2(x_1 + \lambda) + x_2} \\ -(x_1 + \lambda) + 2x_2 + e^{2(x_1 + \lambda) + x_2} \end{bmatrix} \quad (12)$$

$$= 4\lambda - x_2 + 4x_1 - 3 + 2e^{2\lambda + 2x_1 + x_2} \quad (13)$$

$$(14)$$

We substitute (x_1, x_2) the point $(0, 0)$ now and get

$$\theta'(\lambda) = 4\lambda + 2e^{2\lambda}. \quad (15)$$

In order for this to be optimal we set the point zero and solve for λ :

$$0 = 4\lambda + 2e^{2\lambda} \quad (16)$$

$$\implies \lambda = \frac{1}{4}(3 - 2W(e^{3/2})) \approx 0.118, \quad (17)$$

where W is the log product function.

We can now calculate the new point now by noticing that $\lambda d = \begin{bmatrix} \lambda \\ 0 \end{bmatrix}$. Since the starting point is $(0, 0)$ we get the new point by simply setting $f(\lambda, 0) \approx 0, 94$.

2.3

The code can be seen in the attached jupyter notebook. The iterations of all three methods can be seen from Table 1. Since they can easily be seen from the table, the answers have not been split into a-d parts.

In the case of the Newton method, we are not only employing first-order derivative information but also second order derivative information. This leads to a more complex computations but need for less iterations. For the conjugate gradient, which we can see lies between the gradient and newton methods in iteration amount, we are not directly using the Hessian but still using more information than in the gradient method.

For the Conjugate gradient method,

Table 1: The iterations for both functions f and g of exercise 2.3.

Function	Gradient Descent	Newton's method	Conjugate Gradient method
$f(x_1, x_2)$	11	1	8
$g(x_1, x_2)$	18	4	6