Part 2

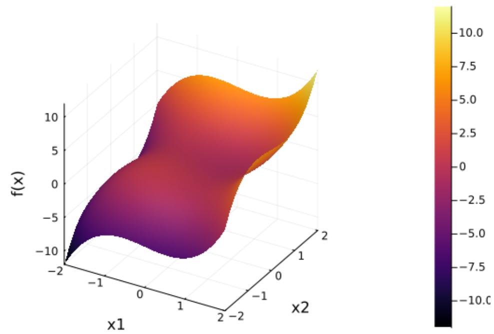

Figure 1. As we can see the function is not convex.

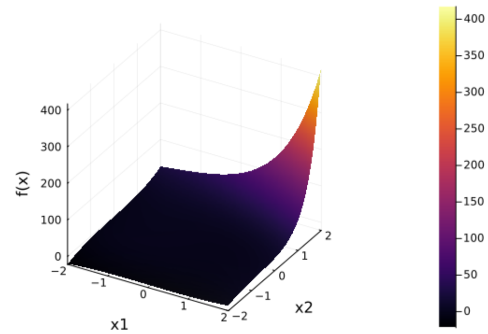

Figure 2. As we can see the function is not convex.

I Exercise 1

Let be $f : \mathbb{R}^2 \to \mathbb{R}$, given by $f(\boldsymbol{x}) = x_1^3 - x_1 + x_2^3 - x_2$. So $f \in C^\infty(\mathbb{R}^2)$ since is polynomial.

1 Point a

To determine whether the function is convex or not, we need to examine the curvature. If the function is convex, the curvature will be non-negative everywhere. If it's not convex, there will be regions with negative curvature. Based on the plot(figure 2), you can observe the curvature of the surface. If the surface is always above its tangent planes, it's convex; otherwise, it's not convex. In this case, it appears that the function is not convex because it has regions where the surface dips below its tangent planes. As a proof, we compute the hessian:

$$\nabla f = \begin{bmatrix} 3x_1^2 - 1 \\ 3x_2^2 - 1 \end{bmatrix}, \quad \mathbb{H}_f = \begin{bmatrix} 6x_1 & 0 \\ 0 & 6x_2 \end{bmatrix}.$$

So the function cannot be convex since x_1 and x_2 are eigenvalues of its hessian, so the sign is variable.

2 Point b

By definition, since f is continuously differentiable maximum and minimum are critical points(the reverse is not true). So we look for $\boldsymbol{x} \in \mathbb{R}^2$ such that $\nabla f(\boldsymbol{x}) = 0$. So:

$$\boldsymbol{x} = \left(\pm\frac{1}{\sqrt{3}}, \pm\frac{1}{\sqrt{3}} \right),$$

are the four critical points.

3 Point c

We observe that, for a fixed value of $\boldsymbol{x}$, one eigenvalue is negative while the other is positive. This observation indicates the existence of isotropic vectors.

II Exercise 2

Let be $f(x_1, x_2) = 2x_1^2 - x_1x_2 + x_2^2 - 3x_1 + e^{2x_1+x_2}$, and consider the problem.

1 Point a

Since the function is infinitely differentiable. So due to Fermat's Theorem an optimal point is a stationary point. So if $\boldsymbol{x}$ is an optimum, we should have:

$$0 = \nabla f = \begin{bmatrix} 4x_1 - x_2 - 3 + 2e^{2x_1+x_2} \\ -x_1 + 2x_2 + e^{2x_1+x_2} \end{bmatrix}.$$

Now we compute the hessian:

$$\mathbb{H}_f = \begin{bmatrix} 4 + 4e^{2x_1+x_2} & -1 + 2e^{2x_1+x_2} \\ -1 + 2e^{2x_1+x_2} & 2 + e^{2x_1+x_2} \end{bmatrix}.$$

By the plot that you can see in Figure 2, the function seems not to be concave/convex. So Fermat's condition is not sufficient.

In fact take points $(0, 0)$ and $(1, 1)$, we observe that, $f(0, 0) = 1$ and $f(1, 1) = e^3 - 2$. Let we consider the convex combination of the two points $(0.5, 0.5)$, we have that:

$$f(0.5, 0.5) = e^{\frac{3}{2}} - 1 < \frac{e^3}{2} - \frac{1}{2} = 0.5f(0, 0) + 0.5f(1, 1),$$

showing that f is not concave.

If we take now the points $(0,0)$ and $(-1,-1)$, we observe that $f(-1,-1) = 3 + \mathrm{e}^{-3}$. Instead $f(-.5,-.5) = 2 + \mathrm{e}^{-1.5}$, and so:

$$
\begin{aligned}
f(-0.5,-0.5) =& 2 + \mathrm{e}^{-1.5} > 2 + \mathrm{e}^{-3/2} \\
=& 0.5 f(0,0) + 0.5 f(-1,-1),
\end{aligned}
$$

showing that f is not convex.

2 Point b and c

We note that, if $\bar{\boldsymbol{x}} = (0,0)$, we have:

$$\nabla f(\bar{\boldsymbol{x}}) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

proving that it is not an optimal point.

If $\boldsymbol{d}$ is a descent direction at $\bar{\boldsymbol{x}}$, by definition we should have:

$$0 > \boldsymbol{d}^T \cdot \nabla f(\bar{\boldsymbol{x}}) = \boldsymbol{d}^T \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix} = d_2 - d_1.$$

This implies that for a valid descent direction, we must have $d_2 < d_1$. To achieve this, we can introduce a positive value $\delta > 0$ such that $d_2 = d_1 - \delta$. In order to maintain a normalised direction, we can further express this as:

$$\boldsymbol{d}^T = \frac{1}{\sqrt{d_1^2 + (d_1 - \delta)^2}} \begin{bmatrix} d_1 \\ d_1 - \delta \end{bmatrix}.$$

This formula provides the steepest descent direction at the point $\bar{\boldsymbol{x}}$, for all $d_1 \in \mathbb{R}$ and $\delta > 0$.

However, it's important to note that the steepest descent corresponds to the direction opposite to the gradient, which can be expressed as:

$$\boldsymbol{d}^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

This observation holds in general; a descent direction is opposite to the gradient, and as a result, their scalar product should be negative. Since the gradient points in the direction of the greatest increase, the opposite direction of the gradient is the steepest descent direction.

III Exercise 3

1 Newton's Method

For convex functions, such as those in your case, Newton's method is known to converge rapidly and directly to the global minimum. This is a well-established property of convex optimization. Convexity guarantees that there are no local minima other than the global minimum, making it easier for Newton's method to find the optimal solution quickly. The observed number of iterations being very low (1 for function f and 6 for function g) aligns with the expected behavior of Newton's method for convex functions.

2 Conjugate Gradient Method

Convexity is also beneficial for the Conjugate Gradient method. Convex functions have the property that the sequence of iterates generated by the Conjugate Gradient method converges to the global minimum in a finite number of steps. When dealing with convex functions, the Conjugate Gradient method is expected to converge to the optimum within a number of iterations proportional to the dimensionality of the problem. The fact that the Conjugate Gradient method was able to converge rapidly in your 2D optimization problems (20 iterations for function f and 12 iterations for function g) is consistent with the convexity of the functions.

In summary, the convexity of the functions f and g provides strong theoretical support for the observed behavior of both Newton's method and the Conjugate Gradient method. These methods are well-suited for convex optimization problems and are expected to converge quickly to the global minimum when applied to convex functions.

Table 1. Results for Gradient, Conjugate Gradient, and Newton Methods (Function f), starting point $(7, 3)$

Method	Gradient	Conjugate Gradient	Newton
Optimal Solution	$[1.02 \cdot 10^{-3}, 1.02 \cdot 10^{-3}]$	$[2.63 \cdot 10^{-4}, 3.29 \cdot 10^{-4}]$	$[-1.22 \cdot 10^{-7}, -5.24 \cdot 10^{-8}]$
Dist. from Optimizer	$1.442 \cdot 10^{-3}$	$4.212 \cdot 10^{-4}$	$1.331 \cdot 10^{-7}$
Dist. from Optimal Value	$4.159 \cdot 10^{-8}$	$4.596 \cdot 10^{-9}$	$1.528 \cdot 10^{-15}$
Tot. Iter.	11	20	1
Norm at Optimizer	$5.795 \cdot 10^{-5}$	$4.963 \cdot 10^{-5}$	$4.969 \cdot 10^{-8}$

Table 2. Results for Gradient, Conjugate Gradient, and Newton Methods (Function g), starting point $(7, 3)$

Method	Gradient	Conjugate Gradient	Newton
Optimal Solution	$[-0.3465, -1.820 \cdot 10^{-6}]$	$[-0.3466, 5.551 \cdot 10^{-6}]$	$[-0.3466, 1.819 \cdot 10^{-8}]$
Dist. from Optimizer	$2.521 \cdot 10^{-5}$	$8.246 \cdot 10^{-6}$	$4.758 \cdot 10^{-7}$
Dist. from Optimal Value	$3.303 \cdot 10^{-6}$	2.559	$3.303 \cdot 10^{-6}$
Total Iterations	17	12	6
Norm at Optimizer	$6.668 \cdot 10^{-5}$	$6.607 \cdot 10^{-5}$	$2.687 \cdot 10^{-7}$

Table 3. Results for Function f with starting point $(-4, -2)$

Method	Gradient	Conjugate Gradient	Newton
Optimal Solution	$[-0.6 \cdot 10^{-3}, -0.61 \cdot 10^{-3}]$	$[-0.35 \cdot 10^{-3}, -0.42 \cdot 10^{-3}]$	$[6.99 \cdot 10^{-8}, 3.5 \cdot 10^{-8}]$
Dist. from Optimum	$0.854 \cdot 10^{-3}$	$0.554 \cdot 10^{-3}$	$7.818 \cdot 10^{-8}$
Dist. from Optimal Value	$1.459 \cdot 10^{-8}$	$7.349 \cdot 10^{-9}$	$4.156 \cdot 10^{-16}$
Total Iterations	13	20	1
Norm at Optmizer	$3.440 \cdot 10^{-5}$	$5.475 \cdot 10^{-5}$	$2.490 \cdot 10^{-8}$

Table 4. Results for Function g with starting point $(-4, -2)$

Method	Gradient	Conjugate Gradient	Newton
Optimal Solution	$[-0.347, 1.493 \cdot 10^{-6}]$	$[-0.347, 1.189 \cdot 10^{-8}]$	$[-0.3466, -1.075 \cdot 10^{-8}]$
Dist. from Optimum	$1.461 \cdot 10^{-5}$	$9.032 \cdot 10^{-7}$	$3.673 \cdot 10^{-7}$
Dist. from Optimal Value	$3.303 \cdot 10^{-6}$	2.559	$3.303 \cdot 10^{-6}$
Total Iterations	18	6	4
Norm at Optmizer	$4.004 \cdot 10^{-5}$	$1.270 \cdot 10^{-6}$	$1.649 \cdot 10^{-7}$