2. Unconstrained optimization in multi dimensions

Given a function $f : R^n \to R$, we seek to find a minimum of $f(x)$, i.e., we solve the following unconstrained optimization problem :

find x to minimize $f(x)$.

First Derivative or Gradient.

Let $f : R^n \to R$ be a function with continuous derivatives. The gradient of $f(x)$ is defined as the column vector containing the first order partial derivatives of $f(x)$:

$$g(x) = \nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1} \; \frac{\partial f(x)}{\partial x_2} \; \ldots\ldots\ldots\ldots\ldots \; \frac{\partial f(x)}{\partial x_n} \right]^T . \qquad \ldots\ldots\ldots.(2)$$

Second Derivative or Hessian.

The Hessian $\nabla^2 f(x)$ of $f(x)$, is the matrix defined by the second order partial derivatives of $f(x)$, as :

$$G = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \ldots\ldots\ldots\ldots\ldots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial^2 x_2} & \ldots\ldots\ldots\ldots\ldots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ . & . & \ldots\ldots\ldots\ldots\ldots\ldots & . \\ . & . & \ldots\ldots\ldots\ldots\ldots\ldots & . \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \ldots\ldots\ldots\ldots\ldots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \qquad \ldots\ldots\ldots.(2)$$

which is symmetric.

Example :

Compute the gradient and the Hessian of the function $f(x_1, x_2) = x_1^2 - 3x_1x_2 + x_2^2$ at the point $x = (x_1, x_2)^T = (1, 1)^T$. Then :

$$g(x) = \nabla f(x) = \begin{bmatrix} 2x_1 - 3x_2 \\ -3x_1 + 2x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$G(x) = \nabla^2 f(x) = \begin{bmatrix} 2 & -3 \\ -3 & 2 \end{bmatrix}$$

Taylor's Series in n Dimensions.

The Taylor series expansion of $f(x)$ about some $x_k \in R^n$ is :

$$f(x) = f(x_k) + g_k^T (x - x_k) + \frac{1}{2}(x - x_k)^T G_k (x - x_k) + \ldots\ldots$$

where $g_k \in R^n$ and $G_k \in R^{n*n}$.

In multiple dimensions, the conditions are simply the multivariate extensions of the one dimension conditions.

The following theorem can be proved for functions of multivariables.

Theorem (A First-order Necessary conditions)

Let $f : R^n \to R$, have continuous first order partial derivatives. If x^* is a local minimum of $f(x)$, then $f^{/}(x^*) = 0$.

We call a point x^* as a stationary point of $f(x)$, if $f^{/}(x^*) = 0$.

3. Classification of Matrices.

Before we prove a second order sufficient condition for the local minimums, let us first review the positive definiteness of a matrix. We say that a matrix A is symmetric positive definite, if

$$x^T \left[\nabla^2 f(x)\right] x > 0 .$$

For any vector $x \neq 0$.

For the next result we recall that a matrix $\nabla^2 f(x)$ is positive definite if $x^T \left[\nabla^2 f(x)\right] x > 0$ for all $x \neq 0$, and positive semi definite if $x^T \left[\nabla^2 f(x)\right] x \geq 0$ for all $x \neq 0$.

Example : Let

$$G(x) = \nabla^2 f(x) = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

Then for any $x = (x_1 , x_2)^T$.

$$x^T G(x) x = (x_1 , x_2) \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2 - 2x_1 x_2 + x_2^2 = x_1^2 + (x_1 - x_2)^2 \geq 0$$

Thus, G is positive semi definite.

Note :

A matrix $\nabla^2 f(x)$ will be positive definite if all its eigenvalues are positive; that is, all the values of λ that satisfy the determinant equation :

$$\left|\nabla^2 f(x) - \lambda I\right| = 0 \, .$$

should be positive. Similarly, the matrix $\nabla^2 f(x)$ will be negative definite if its eigenvalues are negative.

Another test that can be used to find the positive definiteness of a matrix $\nabla^2 f(x) = G$ of order n involves evaluation of the determinants :

$$G_1 = \left|G_{11}\right|, \quad G_2 = \begin{vmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{vmatrix}, \; G_3 = \begin{vmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{vmatrix}$$

$$G_n = \begin{vmatrix} G_{11} & G_{12} & G_{13} & . & . & . & G_{1n} \\ G_{21} & G_{22} & G_{23} & . & . & . & G_{2n} \\ G_{31} & G_{32} & G_{33} & . & . & . & G_{3n} \\ \\ . \\ . \\ G_{n1} & G_{n2} & G_{n3} & . & . & . & G_{nn} \end{vmatrix}$$

The matrix G will be positive definite if and only if all the values $G_1, G_2, G_3, , \ldots\ldots, G_n$ are positive. The matrix G will be negative definite if and only if the sign of G_j is $(-1)^j$ for $j = 1, 2, 3, \ldots\ldots, n$. If some of the G_j are positive and the remaining G_j are zero, the matrix G will be positive semidefinite.

Example :

Consider the Hessian matrix

$$G(x) = \nabla^2 f(x) = \begin{bmatrix} 6 & -1 \\ 1 & 4 \end{bmatrix}$$

Find all eigenvalues of matrix $G(x)$.

The characteristic equation of matrix $G(x)$ is:

$$G(x)-\lambda I=\nabla^2 f(x)-\lambda I=\begin{bmatrix} 6 & -1 \\ 1 & 4 \end{bmatrix}-\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}=\begin{bmatrix} 6-\lambda & -1 \\ 1 & 4-\lambda \end{bmatrix}$$

The eigenvalues of matrix $G(x)$ is given by:

$$\det(G(x)-\lambda I)=\det(\nabla^2 f(x)-\lambda I)=\begin{vmatrix} 6-\lambda & -1 \\ 1 & 4-\lambda \end{vmatrix}$$
$$=(6-\lambda)(4-\lambda)-(-1)$$
$$=\lambda^2-10\lambda+25$$
$$=(\lambda-5)(\lambda-5) \;=\; 0$$

Therefore, the eigenvalues are $\lambda=5$. As all of eigenvalues are positive, the Hessian is positive definite.

Example :

Consider the functionThe corresponding. $f(x_1,x_2)=x_1^2+x_2^2-3x_1x_2$

Hessian matrix is :

$$G(x)=\nabla^2 f(x)=\begin{bmatrix} 2 & -3 \\ -3 & 2 \end{bmatrix}$$

The determinants of the square sub matrices of $G(x)$ are :

$$|G_{11}|=2, \quad \begin{vmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{vmatrix}=\begin{vmatrix} 2 & -3 \\ -3 & 2 \end{vmatrix}= 4-9= -5$$

the Hessian matrix $G(x)$ may be neither positive nor negative definite. (saddle point).

Theorem (A second-order sufficient conditions)

Let $f:R^n\to R$, have continuous first and second order partial derivatives. If $f^{/}(x^*)=0$ and $\nabla^2 f(x)$ is positive definite, then x^* is a local minimum of the $f(x)$.

Example :

Find the critical point of the function $f(x_1,x_2)=x_1^3+3x_2-x_2^3-3x_1$.

Solution :

Critical points are **:**

$$\frac{\partial f}{\partial x_1} = 3x_1^2 - 3 = 0 \quad \Rightarrow \quad x = \underset{-}{+} 1$$

$$\frac{\partial f}{\partial x_2} = 3 - 3x_2^2 = 0 \quad \Rightarrow \quad x = \underset{-}{+} 1$$

For all four point :

$$(1\,,1)\,, \quad (1\,,-1)\,, \quad (-1,1)\,, \quad (-1\,,-1)$$

The second-order partial derivatives of $f(x)$ are given by :

$$\frac{\partial^2 f}{\partial x_1^2} = 6x_1 \quad , \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0 \quad , \quad \frac{\partial^2 f}{\partial x_2^2} = -6x_2$$

The Hessian matrix of f is given by :

$$G = \begin{bmatrix} 6x_1 & 0 \\ 0 & -6x_2 \end{bmatrix}$$

Evaluating at :

Point X	**Value of** G_1	**Value of** G_2	**Nature of** G	**Nature of X**
$(1\,,1)$	$+6$	-36	Indefinite	Saddle point
$(1\,,-1)$	$+6$	$+36$	Positive definite	minimum
$(-1\,,1)$	-6	$+36$	Negative definite	maximum
$(-1\,,-1)$	-6	-36	Indefinite	Saddle point

Exercises:

1. For which real numbers k is the quadratic $f(x) = kx_1^2 + 6x_1x_2 + kx_2^2$ positive definite ?

2. Apply Sylvester's test to check the positive definiteness of the matrix:

$$A = \begin{bmatrix} 1 & -2 & 3 \\ 4 & 1 & 2 \\ 3 & -1 & 2 \end{bmatrix}.$$

3. Find the minimizers and maximzers of the function $f(x) = \frac{1}{3}x_1^3 + \frac{1}{3}x_2^3 - 16x_1 - 4x_2$