Module – 2 Lecture Notes – 4

Optimization of Functions of Multiple Variables subject to Equality Constraints

Introduction
In the previous lecture we learnt the optimization of functions of multiple variables studied for unconstrained optimization. This is done with the aid of the gradient vector and the Hessian matrix. In this lecture we will learn the optimization of functions of multiple variables subjected to equality constraints using the method of constrained variation and the method of Lagrange multipliers.

Constrained optimization
A function of multiple variables, \( f(x) \), is to be optimized subject to one or more equality constraints of many variables. These equality constraints, \( g_j(x) \), may or may not be linear. The problem statement is as follows:

Maximize (or minimize) \( f(X) \), subject to \( g_j(X) = 0, \ j = 1, 2, \ldots, m \)

where

\[
X = \begin{bmatrix}
x_1 \\ x_2 \\ \vdots \\ x_n 
\end{bmatrix}
\]

with the condition that \( m \leq n \); or else if \( m > n \) then the problem becomes an over defined one and there will be no solution. Of the many available methods, the method of constrained variation and the method using Lagrange multipliers are discussed.

Solution by method of Constrained Variation
For the optimization problem defined above, let us consider a specific case with \( n = 2 \) and \( m = 1 \) before we proceed to find the necessary and sufficient conditions for a general problem using Lagrange multipliers. The problem statement is as follows:

Minimize \( f(x_1, x_2) \), subject to \( g(x_1, x_2) = 0 \)

For \( f(x_1, x_2) \) to have a minimum at a point \( X^* = [x_1^*, x_2^*] \), a necessary condition is that the total derivative of \( f(x_1, x_2) \) must be zero at \( [x_1^*, x_2^*] \).

\[
df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 = 0
\]  

(1)

D Nagesh Kumar, IISc, Bangalore  M2L4
Optimization Methods: Optimization using Calculus - Equality constraints

Since \( g(x_1^*, x_2^*) = 0 \) at the minimum point, variations \( dx_1 \) and \( dx_2 \) about the point \( [x_1^*, x_2^*] \) must be admissible variations, i.e. the point lies on the constraint:

\[
g(x_1^* + dx_1, x_2^* + dx_2) = 0
\]

(2)

assuming \( dx_1 \) and \( dx_2 \) are small the Taylor series expansion of this gives us

\[
g(x_1^* + dx_1, x_2^* + dx_2) = g(x_1^*, x_2^*) + \frac{\partial g}{\partial x_1}(x_1^*, x_2^*) dx_1 + \frac{\partial g}{\partial x_2}(x_1^*, x_2^*) dx_2 = 0
\]

(3)

or

\[
dg = \frac{\partial g}{\partial x_1} dx_1 + \frac{\partial g}{\partial x_2} dx_2 = 0 \text{ at } [x_1^*, x_2^*]
\]

(4)

which is the condition that must be satisfied for all admissible variations.

Assuming \( \frac{\partial g}{\partial x_2} \neq 0 \) (4) can be rewritten as

\[
dx_2 = -\frac{\partial g / \partial x_1 - \partial g / \partial x_2 (x_1^*, x_2^*)}{\partial g / \partial x_2} dx_1
\]

(5)

which indicates that once variation along \( x_1 \) \((dx_1)\) is chosen arbitrarily, the variation along \( x_2 \) \((dx_2)\) is decided automatically to satisfy the condition for the admissible variation.

Substituting equation (5) in (1) we have:

\[
\left. df = \left( \frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_2} \frac{\partial g}{\partial x_1} - \frac{\partial f}{\partial x_2} \frac{\partial g}{\partial x_2} \right) dx_1 \right|_{[x_1^*, x_2^*]} = 0
\]

(6)

The equation on the left hand side is called the constrained variation of \( f \). Equation (5) has to be satisfied for all \( dx_i \), hence we have

\[
\left. \left( \frac{\partial f}{\partial x_1} \frac{\partial g}{\partial x_2} - \frac{\partial f}{\partial x_2} \frac{\partial g}{\partial x_1} \right) \right|_{[x_1^*, x_2^*]} = 0
\]

(7)

This gives us the necessary condition to have \([x_1^*, x_2^*]\) as an extreme point (maximum or minimum).

**Solution by method of Lagrange multipliers**

Continuing with the same specific case of the optimization problem with \( n = 2 \) and \( m = 1 \) we define a quantity \( \lambda \), called the Lagrange multiplier as

\[
\lambda = -\frac{\partial f / \partial x_2}{\partial g / \partial x_2} \bigg|_{[x_1^*, x_2^*]}
\]

(8)

Using this in (6)
Optimization Methods: Optimization using Calculus - Equality constraints

$$\left( \frac{\partial f}{\partial x_1} + \lambda \frac{\partial g}{\partial x_1} \right)_{x_1^*, x_2^*} = 0$$ \hspace{1cm} (9)

And (8) written as

$$\left( \frac{\partial f}{\partial x_2} + \lambda \frac{\partial g}{\partial x_2} \right)_{x_1^*, x_2^*} = 0$$ \hspace{1cm} (10)

Also, the constraint equation has to be satisfied at the extreme point

$$g(x_1, x_2)_{x_1^*, x_2^*} = 0$$ \hspace{1cm} (11)

Hence equations (9) to (11) represent the necessary conditions for the point $[x_1^*, x_2^*]$ to be an extreme point.

Note that $\lambda$ could be expressed in terms of $\partial g / \partial x_i$ as well and $\partial g / \partial x_i$ has to be non-zero.

Thus, these necessary conditions require that at least one of the partial derivatives of $g(x_1, x_2)$ be non-zero at an extreme point.

The conditions given by equations (9) to (11) can also be generated by constructing a function $L$, known as the Lagrangian function, as

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda g(x_1, x_2)$$ \hspace{1cm} (12)

Alternatively, treating $L$ as a function of $x_1, x_2$ and $\lambda$, the necessary conditions for its extremum are given by

$$\frac{\partial L}{\partial x_1}(x_1, x_2, \lambda) = \frac{\partial f}{\partial x_1}(x_1, x_2) + \lambda \frac{\partial g}{\partial x_1}(x_1, x_2) = 0$$

$$\frac{\partial L}{\partial x_2}(x_1, x_2, \lambda) = \frac{\partial f}{\partial x_2}(x_1, x_2) + \lambda \frac{\partial g}{\partial x_2}(x_1, x_2) = 0$$ \hspace{1cm} (13)

$$\frac{\partial L}{\partial \lambda}(x_1, x_2, \lambda) = g(x_1, x_2) = 0$$

The necessary and sufficient conditions for a general problem are discussed next.

**Necessary conditions for a general problem**

For a general problem with $n$ variables and $m$ equality constraints the problem is defined as shown earlier

Maximize (or minimize) $f(X)$, subject to $g_j(X) = 0, \ j = 1, 2, \ldots, m$

where

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

D Nagesh Kumar, IISc, Bangalore
In this case the Lagrange function, $L$, will have one Lagrange multiplier $\lambda_j$ for each constraint $g_j(X)$ as

$$L(x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m) = f(X) + \lambda_1 g_1(X) + \lambda_2 g_2(X) + ... + \lambda_m g_m(X)$$  \hspace{1cm} (14)$$

$L$ is now a function of $n + m$ unknowns, $x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m$, and the necessary conditions for the problem defined above are given by

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i}(X) + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i}(X) = 0, \hspace{1cm} i = 1, 2, ..., n$$

$$\frac{\partial L}{\partial \lambda_j} = g_j(X) = 0, \hspace{1cm} j = 1, 2, ..., m$$  \hspace{1cm} (15)$$

which represent $n + m$ equations in terms of the $n + m$ unknowns, $x_i$ and $\lambda_j$. The solution to this set of equations gives us

$$X = \begin{bmatrix} x_1^* \\ \vdots \\ x_n^* \end{bmatrix} \quad \text{and} \quad \lambda^* = \begin{bmatrix} \lambda_1^* \\ \vdots \\ \lambda_m^* \end{bmatrix}$$  \hspace{1cm} (16)$$

The vector $X$ corresponds to the relative constrained minimum of $f(X)$ (subject to the verification of sufficient conditions).

**Sufficient conditions for a general problem**

A sufficient condition for $f(X)$ to have a relative minimum at $X^*$ is that each root of the polynomial in $\epsilon$, defined by the following determinant equation be positive.

$$\begin{vmatrix} L_{11} - \epsilon & L_{12} & \cdots & L_{1n} & g_{11} & g_{21} & \cdots & g_{m1} \\ L_{21} & L_{22} - \epsilon & \cdots & L_{2n} & g_{12} & g_{22} & \cdots & g_{m2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ L_{n1} & L_{n2} & \cdots & L_{nn} - \epsilon & g_{1n} & g_{2n} & \cdots & g_{mn} \\ g_{11} & g_{12} & \cdots & g_{1n} & 0 & \cdots & \cdots & 0 \\ g_{21} & g_{22} & \cdots & g_{2n} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{mn} & 0 & \cdots & \cdots & 0 \end{vmatrix} = 0$$  \hspace{1cm} (17)$$
Optimization Methods: Optimization using Calculus - Equality constraints

where

\[ L_{ij} = \frac{\partial^2 L}{\partial x_i \partial x_j} (X^*, \lambda^*), \quad \text{for } i = 1, 2, ..., n \quad j = 1, 2, ..., m \]  

\[ g_{pq} = \frac{\partial g_p}{\partial x_q} (X^*), \quad \text{where } p = 1, 2, ..., m \text{ and } q = 1, 2, ..., n \]  

(18)

Similarly, a sufficient condition for \( f(X) \) to have a relative maximum at \( X^* \) is that each root of the polynomial in \( \varepsilon \), defined by equation (17) be negative. If equation (17), on solving yields roots, some of which are positive and others negative, then the point \( X^* \) is neither a maximum nor a minimum.

**Example**

Minimize

\[ f(X) = -3x_1^2 - 6x_1x_2 - 5x_2^2 + 7x_1 + 5x_2 \]

Subject to \( x_1 + x_2 = 5 \)

**Solution**

\[ g_1(X) = x_1 + x_2 - 5 = 0 \]

\[ L(x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m) = f(X) + \lambda_1 g_1(X) + \lambda_2 g_2(X) + ... + \lambda_m g_m(X) \] with \( n = 2 \) and \( m = 1 \)

\[ L = -3x_1^2 - 6x_1x_2 - 5x_2^2 + 7x_1 + 5x_2 + \lambda_1 (x_1 + x_2 - 5) \]

\[ \frac{\partial L}{\partial x_1} = -6x_1 - 6x_2 + 7 + \lambda_1 = 0 \]

\[ => x_1 + x_2 = \frac{1}{6} (7 + \lambda_1) \]

\[ => 5 = \frac{1}{6} (7 + \lambda_1) \]

or

\[ \lambda_1 = 23 \]

\[ \frac{\partial L}{\partial x_2} = -6x_1 - 10x_2 + 5 + \lambda_1 = 0 \]

\[ => 3x_1 + 5x_2 = \frac{1}{2} (5 + \lambda_1) \]

\[ => 3(x_1 + x_2) + 2x_2 = \frac{1}{2} (5 + \lambda_1) \]
\[ x_2 = \frac{-1}{2} \]

and,
\[ x_1 = \frac{11}{2} \]

Hence \( X^* = \begin{bmatrix} -\frac{1}{2} & \frac{11}{2} \end{bmatrix}; \lambda^* = [23] \)

\[
\begin{pmatrix}
L_{11} - \varepsilon & L_{12} & g_{11} \\
L_{21} & L_{22} - \varepsilon & g_{21} \\
g_{11} & g_{12} & 0
\end{pmatrix} = 0
\]

\[
L_{11} = \frac{\partial^2 L}{\partial x_1^2} \bigg|_{(X^*, \lambda^*)} = -6
\]

\[
L_{12} = L_{21} = \frac{\partial^2 L}{\partial x_1 \partial x_2} \bigg|_{(X^*, \lambda^*)} = -6
\]

\[
L_{22} = \frac{\partial^2 L}{\partial x_2^2} \bigg|_{(X^*, \lambda^*)} = -10
\]

\[
g_{11} = \frac{\partial g_1}{\partial x_1} \bigg|_{(X^*, \lambda^*)} = 1
\]

\[
g_{12} = g_{21} = \frac{\partial g_1}{\partial x_2} \bigg|_{(X^*, \lambda^*)} = 1
\]

The determinant becomes
\[
\begin{pmatrix}
-6 - \varepsilon & -6 & 1 \\
-6 & -10 - \varepsilon & 1 \\
1 & 1 & 0
\end{pmatrix} = 0
\]

or
\[
(-6 - \varepsilon)[-1] - (-6)[-1] + [1][-6 + 10 + \varepsilon] = 0
\]

\[ \Rightarrow \varepsilon = -2 \]

Since \( \varepsilon \) is negative, \( X^*, \lambda^* \) correspond to a maximum.