Research note: An Overview of the Traditional Discussion on General Optimization Methodology in the Field of Mathematical Economics

Yoichi SAKURADA Ph.D.

Abstract

Optimization methods used in economics are applied in situations where a certain objective function, such as utility or expenditure, is to be maximized or minimized under constraints such as budget constraints. Typical methods include the methodology of Lagrange multipliers, which is used under equality constraints, and the Karush-Kuhn-Tucker condition, which is used under inequality constraints. Here an overview of each is provided.

Keywords: Lagrange multipliers,
Karush-Kuhn-Tucker condition,
Saddle-point equilibrium

1. Optimization under equality constraints: the methodology of Lagrange multipliers

Here, we consider that the objective function $y = f\,(x_1, x_2, \cdots, x_n)$ is to be maximized under equality condition $g\,(x_1, x_2, \cdots, x_n) = 0$ as below:

$$\begin{aligned}&Max\ y = f\,(x_1, x_2, \cdots, x_n)\\&s.t.^{1)}\\&g\,(x_1, x_2, \cdots, x_n) = 0\end{aligned} \tag{1.1}$$

One of methods for solving optimization problems under equality constraints is the methodology of Lagrange multipliers. The Lagrange multiplier λ to construct the Lagrange function is introduced as follows:

$$L\,(x_1, x_2, \cdots, x_n, \lambda) = f\,(x_1, x_2, \cdots, x_n) - \lambda g\,(x_1, x_2, \cdots, x_n) \tag{1.2}$$

[1] "*s.t.*" is an abbreviation of "*subject to*". It means "subject to the following conditions."

To obtain the solution $x^* = \{x_1^*, x_2^*, \cdots, x_n^*\}$ that maximizes the function f, we set the gradient of the Lagrangian function, i.e., the partial derivative of $L(x_1, x_2, \cdots, x_n, \lambda)$ with respect to the variables $x_1, x_2, \cdots, x_n$, and λ to zero.

$$\begin{cases} \dfrac{\partial L}{\partial \lambda} = 0 \\ \dfrac{\partial L}{\partial x_1} = 0 \\ \cdots\cdots \\ \dfrac{\partial L}{\partial x_n} = 0 \end{cases} \tag{1.3}$$

The first equation is just for deriving the equality constraints, so it does not play any significant role in calculating the optimal solution. The second and subsequent equations are meaningful. That is,

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= \frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} = 0 \\ &\cdots\cdots \\ \frac{\partial L}{\partial x_n} &= \frac{\partial f}{\partial x_n} - \lambda \frac{\partial g}{\partial x_n} = 0 \end{aligned} \tag{1.4}$$

Let's move the second term on the left side of the above equation to the right side and rewrite it in vector form. That is,

$$\begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \cdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix} = \lambda \begin{pmatrix} \dfrac{\partial g}{\partial x_1} \\ \cdots \\ \dfrac{\partial g}{\partial x_n} \end{pmatrix} \tag{1.5}$$

An interesting property can be seen from equation (1.5). It shows that the vector with elements $\partial f / \partial x_i$ $(i = 1, n)$ and the vector with elements $\partial g / \partial x_i$ $(i = 1, n)$ are in a parallel positional relationship via the Lagrange multiplier λ. The vector in equation (1.5), which lists the derivatives of functions f, g, is called the gradient vector, and is expressed as $\nabla f, \nabla g$ using the operator $grad f$, $grad g$, or ∇ (nabla).

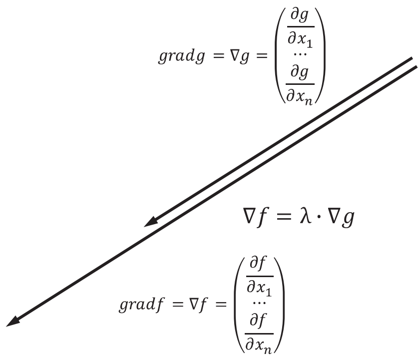

Figure1. 1 Two parallel vectors ∇f, ∇g

Now, consider $f(x) = C$ (C is a constant, $x = (x_1, x_2, \cdots, x_n)$). This can be considered as an isoquant of $f(x)$ projected onto the n-dimensional x-plane. In other words, if we liken $f(x)$ to the height of a mountain drawn against x, then $f(x) = C$ corresponds to the contour line cut off at the value of C. Furthermore, the constraint $g(x) = 0$ is located along the mountainside of the function $f(x)$. Therefore, the problem of maximizing the function $f(x)$ becomes the problem of finding the maximum value among the points on the constraint $g(x) = 0$ that are along the function f.

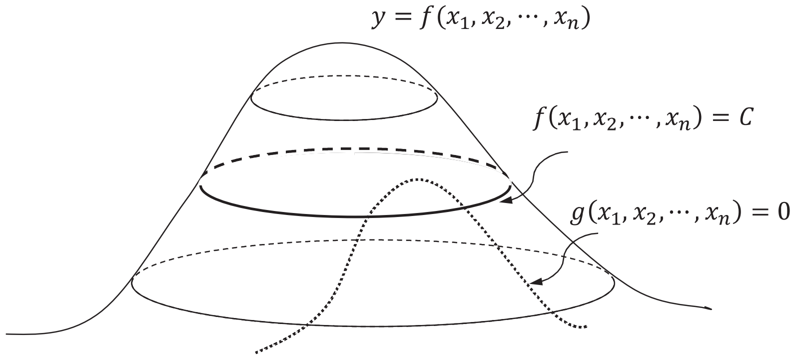

Figure1.2 Contours and constraints of a function f

Take a point $x = (x_1, x_2, \cdots, x_n)$ on this contour line $f(x_1, x_2, \cdots, x_n) = C$, and consider $f(x + \Delta x)$ at a point that is shifted from that point by an infinitesimal amount, $\Delta x = (\Delta x_1, \Delta x_2, \cdots, \Delta x_n)$. This can be calculated by applying the first-order approximation of the Taylor expansion.

$$\begin{aligned} f(x+\Delta x) &\approx f(x)+\frac{\partial f}{\partial x}\Delta x \\ &= f(x_1, x_2, \cdots, x_n)+\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\Delta x_i \\ &= C+\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\Delta x_i \end{aligned} \tag{1.6}$$

In this case, $\Delta x = (\Delta x_1, \Delta x_2, \cdots, \Delta x_n)$ is considered to be the tangent vector of $f(x_1, x_2, \cdots, x_n) = C$ at point x, and by considering Δx to be a sufficiently small quantity, the point $(x_1 + \Delta x_1, x_2 + \Delta x_2, \cdots, x_n + \Delta x_n)$ is also considered to be located on the same curve. Therefore:

$$f(x + \Delta x) = f(x_1 + \Delta x_1, x_2 + \Delta x_2, \cdots, x_n + \Delta x_n) \approx C \tag{1.7}$$

Then, by substituting this into equation (1.6), we get:

$$\begin{gathered} f(x+\Delta x) = C+\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\Delta x_i \\ \therefore \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\Delta x_i = 0 \end{gathered} \tag{1.8}$$

If we express this as a vector dot product, then we get:

$$\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right)(\Delta x_1, \Delta x_n, \cdots, \Delta x_n) = \nabla f \Delta x = 0 \tag{1.9}$$

Equation (1.9) shows that the two vectors are perpendicular. The vector Δx is the tangent vector at point x of the curve $f(x_1, x_2, \cdots, x_n) = C$, and the vector ∇f perpendicular to it is the normal vector of the curve $f(x_1, x_2, \cdots, x_n) = C$. At the same time, according to equation (1.5), since $\nabla f = \lambda \nabla g$, ∇g is parallel to ∇f, and therefore ∇g is also perpendicular to the tangent vector Δx. In other words, at point x where the objective function is maximized, the curve $f(x_1, x_2, \cdots, x_n) = C$ and the curve $g(x_1, x_2, \cdots, x_n) = 0$, which indicates the constraint, are mutually tangent, as shown in the following Figure 1.3.

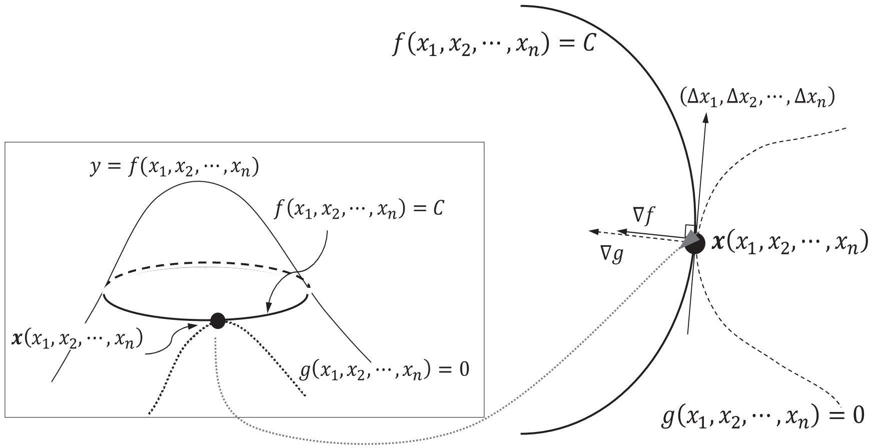

Figure1.3 The case that such two functions as f and g are tangent to each other

Next, let's suppose that $f(x_1, x_2, \cdots, x_n)$ and $g(x_1, x_2, \cdots, x_n)$ do not touch with each other, but intersect. Look at Figure 1.4. The gradient vectors ∇f and ∇g are not parallel but point in different directions. Now, in the area adjacent to the curve $f(x_1, x_2, \cdots, x_n) = C$, for instance, as shown in Figure 1.4, f increases as it moves toward the area to the left of the curve, while f decreases as it moves toward the area to the right. The constraint

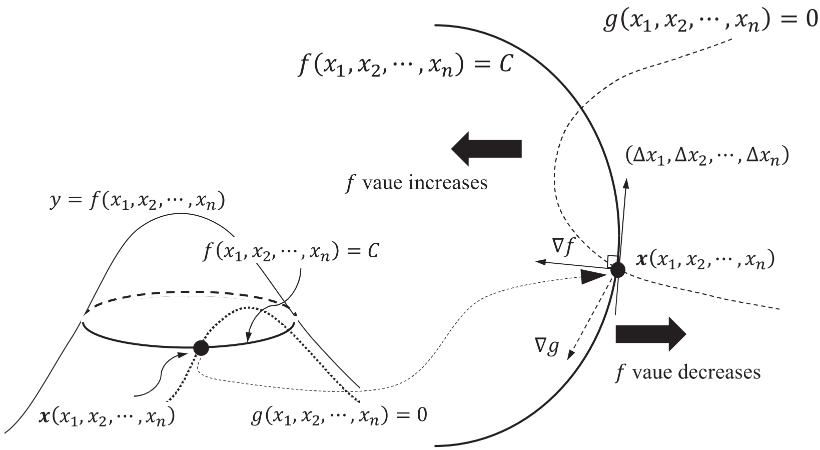

Figure1.4 The case that two functions f and g are not tangent to each other

equation g belongs to both areas, so it is just like lying on a slope halfway up a mountain. Therefore, the point x $(x_1, x_2, \cdots, x_n)$, where $f(x_1, x_2, \cdots, x_n)$ and $g(x_1, x_2, \cdots, x_n)$ intersect, is located halfway down the mountain, so the maximum value can not be found there.

As we have seen above, the Lagrange multiplier method is a technique giving the solution λ^* and x_i^* as the solution that maximizes, or minimizes the function f when the gradient vector ∇L of the Lagrange function $L(f, g, \lambda) = L(x, \lambda)$ is set to zero. Namely:

$$\begin{pmatrix} \dfrac{\partial L}{\partial \lambda^*} \\ \dfrac{\partial L}{\partial x_1^*} \\ \cdots\cdots \\ \dfrac{\partial L}{\partial x_n^*} \end{pmatrix} = \nabla L = 0 \tag{1.10}$$

2. Optimization problems under inequality constraints: Karush-Kuhn-Tucker condition

So far, we have discussed the Lagrange multiplier method, which solves the problem of maximizing, or minimizing the objective function $y = f(x_1, x_2, \cdots, x_n)$ subject to the equality constraint $g(x_1, x_2, \cdots, x_n) = 0$. However, there are cases where there are multiple constraints, or where each constraint is an inequality condition with a specified range. In equation form, the problem can be expressed as follows:

$$\begin{aligned} &Max\ y = f(x_1, x_2, \cdots, x_n) \\ &s.t. \\ &g_i(x_1, x_2, \cdots, x_n) \geq 0 \\ &x_i \geq 0 \\ &\qquad\qquad (i = 1, 2, \cdots, n) \end{aligned} \tag{2.1}$$

The Karush-Kuhn-Tucker condition, or KKT condition for short, gives us a way to solve optimization problems under inequality constraints. When the constraints are inequalities, the discussion becomes a little more complicated. For now, let's consider a one-variable function as a simple case. First, consider the following simple optimization problem.

$$
\begin{aligned}
&Max\ y = f(x)\\
&s.t.\\
&g(x) \geq 0
\end{aligned}
$$

(2.2)

Here, we will construct the Lagrange function $L\ (x, \lambda)$ in the same way as in the Lagrange multiplier method. That is:

$$L\ (x, \lambda) = f(x) - \lambda \cdot g(x)$$

(2.3)

Calculates the gradient of the Lagrange function L (x, λ). That is:

$$\frac{\partial L\ (x,\lambda)}{\partial x} = \frac{\partial f\ (x)}{\partial x} - \lambda \frac{\partial g\ (x)}{\partial x}$$

$$\frac{\partial L\ (x,\lambda)}{\partial \lambda} = -g\ (x)$$

(2.4)

Here, if x^* is the solution that maximizes $y = f(x)$, then there are two possible patterns for x^* in relation to the constraints. One is when x^* is located inside the constraint $g(x) \geq 0$. This type of solution is called an interior solution. It is literally a solution inside the constraints. In this case, x^* is equivalent to the problem of finding the maximum value in a situation where there are no constraints. Therefore, constraints are no longer necessary, which can be expressed by setting $\lambda = 0$ in the first equation in equation (2.4), and the maximum value can be found with only the first-order condition for maximizing the function $y = f(x)$. That is:

【Conditions for maximum value being an interior solution】

$$\frac{\partial L\ (x,\lambda)}{\partial x^*} = \frac{\partial f\ (x)}{\partial x^*} = 0, \quad \lambda = 0$$

(2.5)

On the other hand, there are cases where x^* lies on the edge of the constraints. This is known as a corner solution. In other words, if there were no constraints, it would be possible to find x^* outside the constraints, which would allow the function $y = f(x)$ to be more easily maximized. However, because of the constraints, we are forced to consider the x^*

obtained at the very limit of the range in which the maximization problem can be considered as the solution. Therefore, in this case, the constraints will be in the form of $g(x) = 0$, which is an equality constraint, and so it becomes the same maximization problem as the Lagrange undetermined multiplier method. Namely:

【Conditions when the maximum value is a corner solution】

$$\frac{\partial L(x,\lambda)}{\partial x^*} = \frac{\partial f(x)}{\partial x^*} - \lambda \frac{\partial g(x)}{\partial x^*} = 0$$

$$\frac{\partial L(x,\lambda)}{\partial \lambda} = -g(x^*) = 0, \quad \lambda \neq 0 \tag{2.6}$$

The image of the interior point solution and the the corner solution is shown as below:

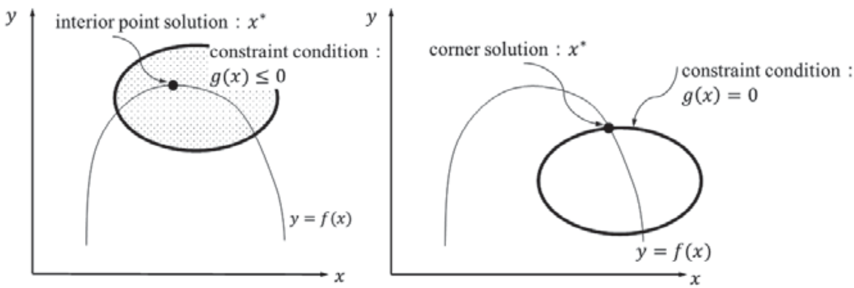

Figure 2.1 Interior and corner solutions

So, what conditions in the maximum problem change between the interior solution and the corner solution? In the case of an interior solution, $\lambda = 0$ is true regardless of the constraint $g(x) \geq 0$. On the other hand, in the case of a corner solution, the constraint $g(x) = 0$ is true regardless of λ. In other words, if either an interior solution or a corner solution is the optimal solution, the possible situations are either $\lambda = 0$ or $g(x) = 0$, and this situation will always occur in one of the two cases. If we were to express this in an equation, we could write it as follows.

$$\lambda \cdot g(x^*) = 0$$

(2.7)

Furthermore, from equation (2.6), $\partial L(x, \lambda)/\partial\lambda = -g(x)$, therefore:

$$\lambda \cdot \frac{\partial L(x^*, \lambda)}{\partial \lambda} = 0$$

(2.8)

This is called the complementarity condition of the KKT conditions. Using this complementarity condition, we can write the maximization problem that encompasses both the interior and extreme solutions as follows. That is, by applying a nonnegative Lagrange multiplier λ.

$$L(x, \lambda) = f(x) - \lambda \cdot g(x)$$

$$\frac{\partial L(x, \lambda)}{\partial x} = 0$$

$$\frac{\partial L(x, \lambda)}{\partial \lambda} \leq 0$$

$$\lambda \cdot \frac{\partial L(x, \lambda)}{\partial \lambda} = 0$$

$$\lambda \geq 0$$

(2.9)

The third equation in the group of equations (2.9) is the constraint equation itself, but because of the form of the Lagrange function, $\partial L(x, \lambda)/\partial\lambda = -g(x)$, and since the constraint $g(x) \geq 0$ has a negative sign, we obtain $\partial L(x, \lambda)/\partial\lambda \leq 0$. Next, let's deepen our understanding of KKT conditions by looking at a more specific problem. This time, we will consider the problem of minimizing an objective function of two variables under three inequality constraints.

$$\begin{aligned} &Min\ y = f(x, y) \\ &s.t. \\ &g_1(x, y) \geq 0,\ g_2(x, y) \geq 0,\ g_3(x, y) \geq 0 \end{aligned}$$

(2.10)

The Lagrangian function is set as follows:

$$L(x,y,\lambda) = f(x,y) - \sum_{i=1}^{3} \lambda_i g_i(x,y) \tag{2.11}$$

The KKT conditions are as follows:

$$\begin{aligned}
&\frac{\partial L(x,y,\lambda)}{\partial x} = \frac{\partial f(x,y)}{\partial x} - \sum_{i=1}^{3} \lambda_i \frac{\partial g_i(x,y)}{\partial x} = 0 \\
&\frac{\partial L(x,y,\lambda)}{\partial y} = \frac{\partial f(x,y)}{\partial y} - \sum_{i=1}^{3} \lambda_i \frac{\partial g_i(x,y)}{\partial y} = 0 \\
&\frac{\partial L(x,y,\lambda)}{\partial \lambda_1} \leq 0, \quad \lambda_1 \cdot \frac{\partial L(x,y,\lambda)}{\partial \lambda_1} = 0 \\
&\frac{\partial L(x,y,\lambda)}{\partial \lambda_2} \leq 0, \quad \lambda_2 \cdot \frac{\partial L(x,y,\lambda)}{\partial \lambda_2} = 0 \\
&\frac{\partial L(x,y,\lambda)}{\partial \lambda_3} \leq 0, \quad \lambda_3 \cdot \frac{\partial L(x,y,\lambda)}{\partial \lambda_3} = 0 \\
&\lambda_1 \geq 0, \quad \lambda_2 \geq 0, \quad \lambda_3 \geq 0
\end{aligned} \tag{2.12}$$

To understand the KKT condition geometrically, let us plot the three constraints and the objective function on the x, y plane.

In Figure 2.2, the curves of the three inequality constraints are shown with dashed lines, and the contour of the objective function $f(x, y) = C$ is drawn as a solid ellipse. The overlapping common area of the inequality constraints $g_i(x, y) \geq 0$ ($i = 1,3$) is shown with hatching, and in this hatched area, a solution (x^*, y^*) that minimizes the objective function $f(x, y)$ is sought. In the figure, the optimal solution is given at the intersection of $g_1(x, y) = 0$ and $g_2(x, y) = 0$, which is the so-called a corner solution for the constraints $g_1(x, y) = 0$ and $g_2(x, y) = 0$. On the other hand, since the point of this optimal solution is within the area of $g_3(x, y) > 0$, it is an interior solution for $g_3(x, y) \geq 0$. From these, the constraints are:

$$g_1(x, y) = 0, \quad \lambda_1 \geq 0$$
$$g_2(x, y) = 0, \quad \lambda_2 \geq 0$$
$$g_3(x, y) \geq 0, \quad \lambda_3 = 0$$

(2.13)

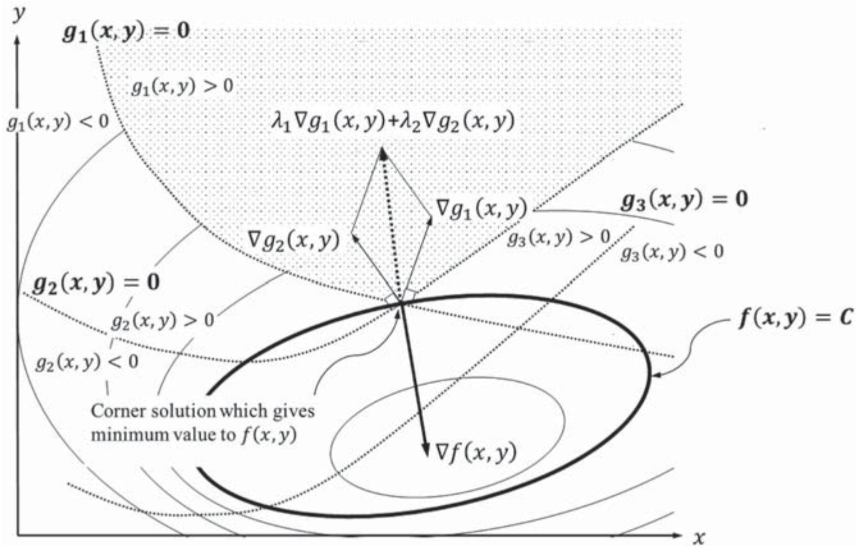

Figure2.2 Geometric interpretation of KKT conditions

Rewriting the KKT condition (2.12), by taking into account of (2.13), it gives:

$$\frac{\partial L(x, y, \lambda)}{\partial x} = \frac{\partial f(x, y)}{\partial x} - \lambda_1 \frac{\partial g_1(x, y)}{\partial x} - \lambda_2 \frac{\partial g_2(x, y)}{\partial x} = 0$$

$$\frac{\partial L(x, y, \lambda)}{\partial y} = \frac{\partial f(x, y)}{\partial y} - \lambda_1 \frac{\partial g_1(x, y)}{\partial y} - \lambda_2 \frac{\partial g_2(x, y)}{\partial y} = 0$$

$$\frac{\partial L(x, y, \lambda)}{\partial \lambda_1} = 0$$

$$\frac{\partial L(x, y, \lambda)}{\partial \lambda_2} = 0$$

$$\frac{\partial L\ (x, y, \lambda)}{\partial \lambda_3} \leq 0$$

$$\lambda_1 \geq 0, \quad \lambda_2 \geq 0, \quad \lambda_3 = 0$$

(2.14)

If we rewrite the right-hand sides of the upper two equations in (2.14) in vector notation, we get:

$$\begin{pmatrix} \dfrac{\partial f\,(x, y)}{\partial x} \\ \dfrac{\partial f\,(x, y)}{\partial y} \end{pmatrix} = \lambda_1 \begin{pmatrix} \dfrac{\partial g_1\,(x, y)}{\partial x} \\ \dfrac{\partial g_1\,(x, y)}{\partial y} \end{pmatrix} + \lambda_2 \begin{pmatrix} \dfrac{\partial g_2\,(x, y)}{\partial x} \\ \dfrac{\partial g_2\,(x, y)}{\partial y} \end{pmatrix}$$

$$\therefore \nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2$$

(2.15)

Equation (2.15) shows that at the optimized point, the gradient of the objective function can be expressed as a linear combination of the gradients of the constraint equations λ_1, λ_2. This shows the mechanical form in which the gradient vector of the objective function, which has been reversed in the Figure 2.2, is balanced by a resultant vector obtained by linearly combining the gradient vector of the constraint equations in a manner similar to the composition of the forces of a parallelogram.

3. Saddle point problems and the Karush-Kuhn-Tucker (KKT) condition

(1) Saddle point and optimization problem

Here, we will deal with the saddle point problem in relation to the KKT conditions. The term “saddle point” is named after the image of a point on a horse’s saddle. For example, the saddle point of the two-variable function $z = f(x, \lambda)$ is the point z^* in Figure 3.1.

At the saddle point $z^* = f(x^*, \lambda^*)$, for $\forall x, \lambda \in \mathbb{R}$, following conditions can be identified:

$$f(x^*, \lambda^*) \leq f(x^*, \lambda)$$
$$f(x, \lambda^*) \leq f(x^*, \lambda^*)$$

(3.1)

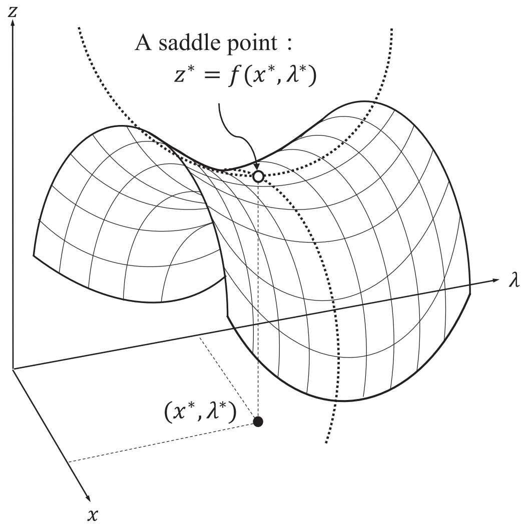

Figure 3.1 A saddle point

When $x = x^*$ is fixed, $f(x^*, \lambda)$ will be at its minimum when $\lambda = \lambda^*$, and on the other hand, when $\lambda = \lambda^*$ is fixed, $f(x, \lambda^*)$ will be at its maximum when $x = x^*$. And as for the Lagrangian function $L\,(x, \lambda)$ that we have been discussing up until now, when (x, λ) is in fact a saddle point (x^*, λ^*), we can say that x^* is the optimal solution that maximizes, or minimizes the objective function $f(x)$, which constitutes the Lagrangian function. In other words:

For $x \in \mathbb{R}^n_+$, consider the following maximization problem as follows:

$$Max\, f(x) \quad s.t. \;\; g_i(x) \geq 0 \quad (i = 1, m) \tag{3.2}$$

In this case, the Lagrange function can be set:

$$L\,(x, \lambda) = f\,(x) + \sum_{i=1}^{m} \lambda_i\, g_i\,(x)$$

(Where $\lambda \in \mathbb{R}^m_+$)

$$(3.3)$$

For (x^*, λ^*), (x^*, λ^*) is a saddle point of the Lagrangian function $L(x, \lambda)$, i.e.,

$$L(x, \lambda^*) \leq L(x^*, \lambda^*) \leq L(x^*, \lambda)$$

$$(3.4)$$

In this case, x^* is the optimal solution that maximizes the objective function $f(x)$. Let us prove this below. Rewrite the inequality $L(x^*, \lambda^*) \leq L(x^*, \lambda)$ on the right side of equation (3.4) as follows:

$$L(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leq L(x^*, \lambda) = f(x^*) + \sum_{i=1}^{m} \lambda_i g_i(x^*)$$

$$\therefore \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leq \sum_{i=1}^{m} \lambda_i g_i(x^*)$$

$$(3.5)$$

In equation (3.5), for any $\lambda_i \geq 0$, $\lambda_i^* g_i(x^*) \leq \lambda_i g_i(x^*)$, but this inequality does not hold when $g_i(x^*) < 0$, considering sufficiently large λ_i, so $g_i(x^*) \geq 0$ must hold. Also, since $\forall \lambda_i^* \geq 0$, therefore:

$$0 \leq \sum_{i=1}^{m} \lambda_i^* g_i(x^*)$$

$$(3.6)$$

On the other hand, if we perform the limit operation on the right-hand side of equation (3.5) by setting $\lambda_i \rightarrow 0$ for all i, we get:

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leq \lim_{\lambda_i \to 0} \sum_{i=1}^{m} \lambda_i g_i(x^*) = 0$$

$$\therefore \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leq 0$$

$$(3.7)$$

From equations (3.6) and (3.7), the following relationship can be derived:

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) = 0$$

(3.8)

Rewrite the inequality $L(x, \lambda^*) \leq L(x^*, \lambda^*)$ on the left side of equation (3.4) as follows:

$$L(x, \lambda^*) = f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) \leq L(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*)$$

(3.9)

Substituting equation (3.8) into the right hand side of equation (3.9), we get:

$$f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) \leq f(x^*)$$

$$\therefore \sum_{i=1}^{m} \lambda_i^* g_i(x) \leq f(x^*) - f(x)$$

(3.10)

From the premise, $\forall \lambda \in \mathbb{R}^m_+$, $g_i(x) \geq 0$ $(i = 1, m)$, so,

$$\sum_{i=1}^{m} \lambda_i^* g_i(x) \geq 0$$

$$\therefore f(x) \leq f(x^*)$$

(3.11)

When (x^*, λ^*) is a saddle point, the objective function $f(x^*)$ reaches its maximum value.

(2) Saddle point problems and KKT conditions (as necessary conditions)

In fact, the KKT condition is a necessary condition for a saddle point problem. However, it is not a sufficient condition. If we illustrate it, the relationship between the two can be expressed as follows:

(Sufficient condition)		(necessary condition)
If a saddle point is given,	then	KKT conditions hold
If the KKT conditions hold,	then	a saddle point is given.

The saddle point conditions as sufficient conditions and the KKT conditions as necessary conditions are listed below.

【Saddle point condition as a sufficient condition】

Consider the following maximization problem for $x \in \mathbb{R}^n_+$:

$$
\begin{aligned}
&Max\, f(x) \\
&s.t. \quad g_i(x) = x_i \geq 0 \quad (i = 1, n)
\end{aligned}
\tag{3.12}
$$

In this case, the Lagrange function is:

$$
L(x, \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x)
\tag{3.13}
$$

(x^*, λ^*) is a saddle point of the Lagrangian function $L(x, \lambda)$, i.e.,

$$
L(x, \lambda^*) \leq L(x^*, \lambda^*) \leq L(x^*, \lambda)
\tag{3.14}
$$

At this time, the following relationship holds at the saddle point (x^*, λ^*).

【KKT conditions as necessary conditions】

$$
\begin{aligned}
&\frac{\partial L(x^*, \lambda^*)}{\partial x_i} \leq 0 \;\; (i = 1, 2, \cdots, n), \quad \sum_{i=1}^{n} x_i^* \frac{\partial L(x^*, \lambda^*)}{\partial x_i} = 0 \\
&\frac{\partial L(x^*, \lambda^*)}{\partial \lambda_j} \geq 0 \;\; (j = 1, 2, \cdots, n), \quad \sum_{j=1}^{n} \lambda_j^* \frac{\partial L(x^*, \lambda^*)}{\partial \lambda_j} = 0
\end{aligned}
\tag{3.15}
$$

This maximization problem is a problem where the constraints are that the variables x_i $(i = 1, \cdots, n)$ are non-negative. Let's look at the proof again. First, from the left side of equation (3.14),

$$L(x, \lambda^*) - L(x^*, \lambda^*) \leq 0 \tag{3.16}$$

Namely:

$$L(x_1, x_2, \cdots, x_i, \cdots, x_n, \lambda^*) - L(x_1^*, x_2^*, \cdots, x_i^*, \cdots, x_n^*, \lambda^*) \leq 0 \tag{3.17}$$

Here, we use the real number h and set $x_i = x_i^* + h$, equation (3.17) becomes:

$$L(x_1, \cdots, x_i^* + h, \cdots, x_n, \lambda^*) - L(x_1^*, \cdots, x_i^*, \cdots, x_n^*, \lambda^*) \leq 0 \tag{3.18}$$

Here, we define the function $F(x^*, \lambda^*, h)$ as follows:

$$F(x^*, \lambda^*, h) = \frac{L(x_1, \cdots, x_i^* + h, \cdots, x_n, \lambda^*) - L(x_1^*, \cdots, x_i^*, \cdots, x_n^*, \lambda^*)}{h} \tag{3.19}$$

In this case, depending on the sign condition of the real number h,

$$\begin{array}{ll} \text{When } h > 0: & F(x^*, \lambda^*, h) \leq 0 \\ \text{When } h < 0: & F(x^*, \lambda^*, h) \geq 0 \end{array} \tag{3.20}$$

When $x_i^* > 0$, for any x_i^*, a real number h such that $x_i^* + h \geq 0$ can be obtained regardless of its sign, so if we perform the limit operation of $h \to 0$ in each of the cases $h > 0$ and $h < 0$, $F(x^*, \lambda^*, h)$, which is between the relational expressions in equation (3.20), will converge to zero. Furthermore, $F(x^*, \lambda^*, h)$ during the limit operation has the meaning of the partial differential coefficient of x_i with respect to the Lagrangian function $L(x, \lambda^*)$, and this means that this converges to zero. That is:

$$\lim_{h \to 0} F(x^*, \lambda^*, h) = \frac{\partial L(x^*, \lambda^*)}{\partial x_i} = 0 \quad \text{(Where } x_i^* > 0\text{)} \tag{3.21}$$

This shows that x_i^* is actually an interior solution to the constraint condition. It can be illustrated as in Figure 3.2 as below. In other words, the fact that the first-order partial differential coefficient of the Lagrangian function at the saddle point x_i^*, which has a positive sign, is zero means that the Lagrangian function has reached its maximum value within the range of the constraint condition $x_i^* > 0$.

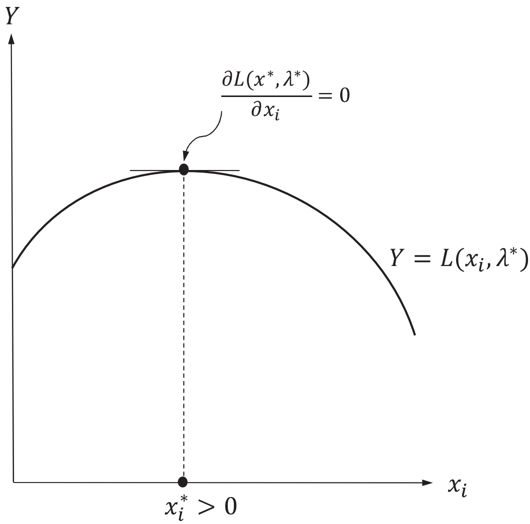

Figure 3.2 A pattern of interior solution

On the other hand, consider the case where the solution is a corner solution, that is, when $L\ (x, \lambda^*)$ is maximized at the point $x_i = 0$ under the non-negativity condition $x_i \geq 0$. When $x_i = 0$, $x_i^* + h \geq 0$ is clearly the case when $h \geq 0$. Therefore, in equation (3.20), if we perform the limit operation of h → 0 using the relationship $F\,(x^*, \lambda^*, h) \leq 0$ when $h > 0$, we will get:

$$\lim_{h \to 0} F\ (x^*, \lambda^*, h) = \frac{\partial L\ (x^*, \lambda^*)}{\partial x_i} \leq 0$$

(Where $x_i^* = 0$)

(3.22)

The solution x_i^* shown in equation (3.22) represents the case of an extreme solution. Let's illustrate this again in Figure 3.3. $x_i^* = 0$ means that the maximum value of the Lagrangian function $L\ (x, \lambda^*)$, which is the objective function, is obtained at the very boundary of the non-negative condition of x_i, which is the constraint. As shown in Figure 3.3, the slope of the function $L\ (x, \lambda^*)$ at the optimal solution $x_i^* = 0$ is not zero, but is negative, which is shown in the inequality condition in equation (3.22).

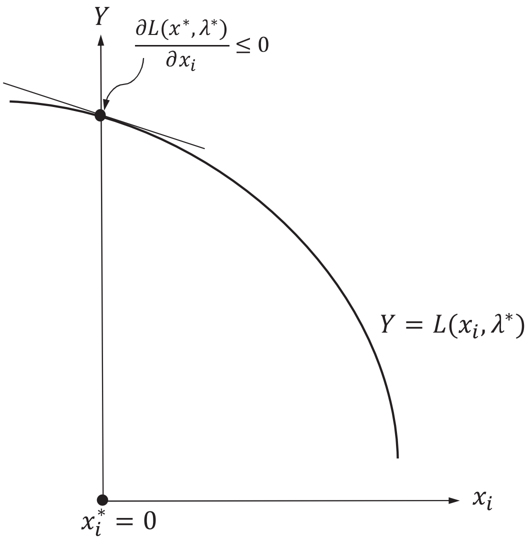

Figure 3.3 A pattern of corner solution

To summarize the above, the KKT condition at the saddle point (x^*, λ^*) is:

$$x_i^* > 0 : \frac{\partial L\,(x^*, \lambda^*)}{\partial x_i} = 0$$
or
$$x_i^* = 0 : \frac{\partial L\,(x^*, \lambda^*)}{\partial x_i} \leq 0$$

(3.23)

If we rewrite this in the form of complementarity conditions, we can get:

$$\frac{\partial L\,(x^*, \lambda^*)}{\partial x_i} \leq 0, \qquad x_i^* \frac{\partial L\,(x^*, \lambda^*)}{\partial x_i} = 0$$

(3.24)

This is true for all $i = 1, \cdots, n$, so:

$$\frac{\partial L\,(x^*, \lambda^*)}{\partial x_i} \leq 0, \qquad \sum_{i=1}^{n} x_i^* \frac{\partial L\,(x^*, \lambda^*)}{\partial x_i} = 0 \qquad (i = 1, 2, \cdots, n)$$

(3.25)

This coincides with equation (3.15). Using the same procedure, we expand the inequality condition on the right side of equation (3.14), $L\,(x, \lambda^*) \leq L\,(x^*, \lambda^*)$, with respect to λ^*, and we get:

$$\lambda_j^* > 0 : \frac{\partial L\,(x^*, \lambda^*)}{\partial \lambda_j} = 0$$

or

$$\lambda_j^* = 0 : \frac{\partial L\,(x^*, \lambda^*)}{\partial \lambda_j} \geq 0$$

(3.26)

The upper equation in (3.26) is the interior solution under the constraints on λ_j^*, and the lower solution corresponds to a corner solution. If illustrated, it can be seen as the following Figure 3.4.

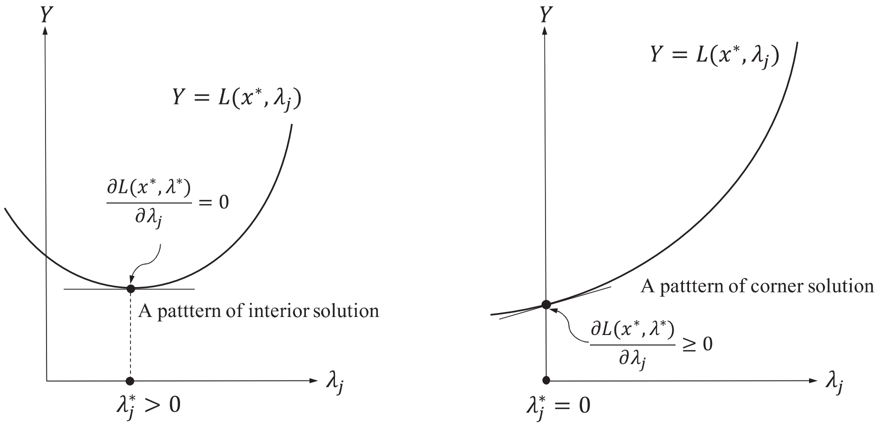

Figure 3.4 Patterns of interior and corner solutions of λ_j^*

If we rewrite this in the form of complementarity conditions, we can get:

$$\frac{\partial L(x^*, \lambda^*)}{\partial \lambda_j} \geq 0, \qquad \lambda_j^* \frac{\partial L(x^*, \lambda^*)}{\partial \lambda_j} = 0$$

(3.27)

This is true for all $j = 1, \cdots, n$, so:

$$\frac{\partial L(x^*, \lambda^*)}{\partial \lambda_j} \geq 0, \qquad \sum_{j=1}^{n} \lambda_j^* \frac{\partial L(x^*, \lambda^*)}{\partial \lambda_j} = 0 \qquad (j = 1, 2, \cdots, n)$$

(3.28)

Now, to reiterate, the KKT conditions are necessary but not sufficient conditions for the conditions that give a saddle point, in other words, the conditions that give an optimal solution. Therefore, just because the KKT conditions are satisfied does not necessarily mean that the optimal solution will be given. However, by adding certain conditions, the KKT conditions become necessary and sufficient conditions for giving an optimal solution. These are the conditions for the convexity and concavity of the objective function. In other words, for the objective function $L(x, \lambda)$,

If the KKT conditions are satisfied and the objective function $L(x, \lambda)$ is concave with respect to x and convex with respect to λ, then the optimal solution x^* ($\equiv$ saddle point) is determined.

Let's prove this. First, the fact that the objective function $L(x, \lambda)$ is a concave function with respect to x can be expressed as follows. That is, after fixing $\lambda = \lambda^*$, for any two points x_i^*, x_i (where $x_i^* \neq x_i$) in $x \in \mathbb{R}^n_+$,

$$\sum_{i=1}^{n} (x_i^* - x_i) \frac{\partial L(x^*, \lambda^*)}{\partial x_i} \leq L(x^*, \lambda^*) - L(x, \lambda^*)$$

(3.29)

We will prove this formula below. First, the condition for the concavity of the objective function $L(x, \lambda)$ is $\forall \alpha \in \mathbb{R}_+$ (where $0 \leq \alpha \leq 1$),

$$L(\alpha x + (1 - \alpha) x^*, \lambda^*) \geq \alpha L(x, \lambda^*) + (1 - \alpha) L(x^*, \lambda^*)$$

$$L(x^* + \alpha(x - x^*), \lambda^*) - L(x^*, \lambda^*) \geq \alpha(L(x, \lambda^*) - L(x^*, \lambda^*))$$

$$\therefore \frac{L(x^* + \alpha(x - x^*), \lambda^*) - L(x^*, \lambda^*)}{\alpha} \geq L(x, \lambda^*) - L(x^*, \lambda^*)$$

(3.30)

Here, let $h = x - x^*$ and use the mean value theorem:

$$\frac{\partial L(x^* + \theta\alpha h, \lambda^*)}{\partial x_i} = \frac{L(x^* + \alpha h, \lambda^*) - L(x^*, \lambda^*)}{\alpha h}$$

$(Where,\ \theta \in (0,1))$

(3.31)

Applying this to the left side of equation (3.30), we get:

$$\begin{aligned}\frac{L(x^* + \alpha(x - x^*), \lambda^*) - L(x^*, \lambda^*)}{\alpha} &= h \cdot \frac{\partial L(x^* + \theta\alpha h, \lambda^*)}{\partial x_i} \\ &= \sum_{i=1}^{n} (x_i - x_i^*) \cdot \frac{\partial L(x^* + \theta\alpha(x - x^*), \lambda^*)}{\partial x_i}\end{aligned}$$

(3.32)

From equations (3.30) and (3.32),

$$\sum_{i=1}^{n} (x_i - x_i^*) \cdot \frac{\partial L(x^* + \theta\alpha(x - x^*), \lambda^*)}{\partial x_i} \geq L(x, \lambda^*) - L(x^*, \lambda^*)$$

(3.33)

Now, when we perform the limit operation of $\alpha \rightarrow 0$,

$$\sum_{i=1}^{n} (x_i - x_i^*) \cdot \frac{\partial L(x^*, \lambda^*)}{\partial x_i} \geq L(x, \lambda^*) - L(x^*, \lambda^*)$$

(3.34)

Furthermore, multiplying both sides of equation (3.34) by -1 and rearranging, we get:

$$\sum_{i=1}^{n} (x_i^* - x_i) \frac{\partial L(x^*, \lambda^*)}{\partial x_i} \leq L(x^*, \lambda^*) - L(x, \lambda^*)$$

(3.35)

Here, we obtain equation (3.29). Up to this point, we have shown the conditions for concaveness with respect to the variable x, but the same can be said for λ. Therefore, the objective function $L(x, \lambda)$ being a convex function with respect to λ is equivalent to the following equation being true.

$$\sum_{i=1}^{n}(\lambda_i^* - \lambda_i)\cdot\frac{\partial L(x^*, \lambda^*)}{\partial \lambda_i} \geq L(x^*, \lambda^*) - L(x^*, \lambda) \tag{3.36}$$

For concave and convex functions, the inequality signs are reversed, as in equations (3.35) and (3.36).

(3) Saddle point problems and KKT conditions (as sufficient conditions)

Based on the above considerations, let us next prove that the KKT conditions and the convexity and concavity conditions of the objective function are sufficient conditions for the saddle point problem. This can be expressed as follows.

【KKT conditions as sufficient conditions, and convexity and concavity conditions】

For $x \in \mathbb{R}^n_+$, the following maximization problem:

$$\begin{aligned} & Max\ f(x) \\ & s.t. \\ & g_i(x) = x_i \geq 0 \quad (i = 1, n) \end{aligned} \tag{3.37}$$

And the Lagrange function constructed from non-negative Lagrange multipliers $\lambda \in \mathbb{R}^n_+$,

$$L(x, \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x) \tag{3.38}$$

For the above function, the KKT condition for the solution (x^*, λ^*) is:

$$\frac{\partial L(x^*, \lambda^*)}{\partial x_i} \leq 0, \qquad \sum_{i=1}^{n} x_i^* \frac{\partial L(x^*, \lambda^*)}{\partial x_i} = 0 \qquad (i = 1, 2, \cdots, n) \tag{3.39}$$

$$\frac{\partial L(x^*, \lambda^*)}{\partial \lambda_j} \geq 0, \qquad \sum_{j=1}^{n} \lambda_j^* \frac{\partial L(x^*, \lambda^*)}{\partial \lambda_j} = 0 \qquad (j = 1, 2, \cdots, n)$$

(3.40)

Furthermore, the concavity condition with respect to x for the Lagrangian function:

$$\sum_{i=1}^{n} (x_i^* - x_i) \frac{\partial L(x^*, \lambda^*)}{\partial x_i} \leq L(x^*, \lambda^*) - L(x, \lambda^*)$$

(3.41)

And the convexity condition for λ:

$$\sum_{i=1}^{n} (\lambda_i^* - \lambda_i) \cdot \frac{\partial L(x^*, \lambda^*)}{\partial \lambda_i} \geq L(x^*, \lambda^*) - L(x^*, \lambda)$$

(3.42)

If the above conditions are met, then the following conditions are met:

【A saddle point condition on (x^*, λ^*) as a necessary condition】

(x^*, λ^*) is a saddle point of the Lagrangian function $L(x, \lambda)$ and satisfies the following relation:

$$L(x, \lambda^*) \leq L(x^*, \lambda^*) \leq L(x^*, \lambda)$$

(3.43)

In other words, the KKT condition and the convexity and concavity conditions of the objective function are necessary and sufficient conditions for the saddle point condition to hold. We will prove this below. The proof will not take many pages. First, from the concavity condition of equation (3.41),

$$L(x, \lambda^*) \leq L(x^*, \lambda^*) - \sum_{i=1}^{n} (x_i^* - x_i) \frac{\partial L(x^*, \lambda^*)}{\partial x_i}$$

(3.44)

From the complementarity condition for x_i in equation (3.39),

$$\sum_{i=1}^{n} x_i^* \frac{\partial L\,(x^*, \lambda^*)}{\partial x_i} = 0$$

(3.45)

Substituting this into equation (3.44), we get

$$L\,(x, \lambda^*) \leq L\,(x^*, \lambda^*) + \sum_{i=1}^{n} x_i \frac{\partial L\,(x^*, \lambda^*)}{\partial x_i}$$

(3.46)

Furthermore, from equation (3.39),

$$\frac{\partial L\,(x^*, \lambda^*)}{\partial x_i} \leq 0 \quad (i = 1, 2, \cdots, n)$$

(3.47)

Therefore, the right hand side of equation (3.46) becomes

$$L\,(x^*, \lambda^*) + \sum_{i=1}^{n} x_i \frac{\partial L\,(x^*, \lambda^*)}{\partial x_i} \leq L\,(x^*, \lambda^*)$$

$$\therefore L\,(x, \lambda^*) \leq L\,(x^*, \lambda^*)$$

(3.48)

Moreover, by the convexity condition in equation (3.42),

$$L\,(x^*, \lambda) \geq L\,(x^*, \lambda^*) - \sum_{i=1}^{n} (\lambda_i^* - \lambda_i) \cdot \frac{\partial L\,(x^*, \lambda^*)}{\partial \lambda_i}$$

(3.49)

Substituting the complementarity condition of (3.40) into (3.49), we get

$$L\,(x^*, \lambda) \geq L\,(x^*, \lambda^*) + \sum_{i=1}^{n} \lambda_i \cdot \frac{\partial L\,(x^*, \lambda^*)}{\partial \lambda_i}$$

(3.50)

From equation (3.40),

$$\frac{\partial L(x^*, \lambda^*)}{\partial \lambda_j} \geq 0 \quad (j = 1, 2, \cdots, n)$$

(3.51)

Therefore, the right hand side of equation (3.50) becomes

$$L(x^*, \lambda^*) + \sum_{i=1}^{n} \lambda_i \cdot \frac{\partial L(x^*, \lambda^*)}{\partial \lambda_i} \geq L(x^*, \lambda^*)$$

$$\therefore L(x^*, \lambda) \geq L(x^*, \lambda^*)$$

(3.52)

From equations (3.48) and (3.52),

$$L(x, \lambda^*) \leq L(x^*, \lambda^*) \leq L(x^*, \lambda)$$

(3.53)

Here we obtain equation (3.43).

Optimization problems under inequality constraints are quite complicated, and there may have been many parts that were difficult to understand compared to the Lagrange multiplier method, which is used to solve optimization problems under equality constraints.

As we have seen, the method using the Karush-Kuhn-Tucker condition, which is an optimization problem under inequality constraints, is a more general method that encompasses the Lagrange multiplier method, which is an optimization problem under equality constraints. In other words, the problem of solving the endpoint solution within the KKT method can be said to be the Lagrange multiplier method.

The method using the Karush-Kuhn-Tucker condition, which is an optimization problem under inequality constraints, requires various case distinctions, so it is much more complicated than the Lagrange multiplier method, which is an optimization problem under equality constraints, and it feels redundant in that the solution is not straightforward.

Bibliography

1) *Andreu Mas-Colell, Michael D. Whinston, Jerry R. Green, "Microeconomic Theory", Oxford University Press, 1995*

2) *Kenneth J. Arrow and Alain C. Enthoven, "Quasi-Concave Programming", Econometrica, Vol. 29, No. 4,, 1961*

3）*Okuno, "An overview of Karush-Kuhn-Tucker Conditions", Opertions research, September, 2015*
4）*Fukukane Nikaido, "Mathematical Methodologies for Modern Economics", Iwanami Press, 1960*
5）*Kamiya and Urai, "Introduction of Mathematics for Economic Studies", Tokyo University Press, 1996*
6）*Toru Maruyama, "Mathematical Economics", Chisen Press, 2002*
7）*Yoichi Sakurada, "Introduction to the Mathematical Economics" Toyo Publishing, 2022*
8）*Yoichi Sakurada, "Introduction to the Numerical Wonderland" Gentoshya Publishing, 2023*