Constrained Optimization

January 7th, 2025

Constrained Optimization

To fix the ideas and get intuition for how to use the KKT theorem, consider the following one-dimensional optimization problem:

$$\max_{0 \leq x \leq 5} x^3 - 5x^2 + x$$

This is a simple optimization problem that you can do without using the KKT theorem explicitly (i.e. use first order condition, find local maximum, and check the endpoints). However, what I want to demonstrate is that this process is exactly equivalent to what the KKT procedure is doing, though in this case, it is like taking a sledgehammer to crack a nut.

The objective function is $f(x) = x^3 - 5x^2 + x$. This function is not quasi-concave so the KKT conditions are <u>not sufficient</u>. The constraints are $g_1(x) = x - 5 \leq 0$ and $g_2(x) = -x \leq 0$.

Now, what does the KKT theorem state? **IF** the following three conditions hold:

1. f, $g_1, \ldots, g_K$ are continuously differentiable in x
2. $\mathcal{X}(\theta)$ is non-empty (i.e. the constraint inequalities can be satisfied)
3. For a solution x^*, the vectors in $\{\nabla g_j(x^*, \theta) | g_j(x^*, 0) = 0\}$ are linearly independent

THEN there exist $\lambda_1, \ldots, \lambda_K \geq 0$ such that:

$$\nabla f(x^*, \theta) = \sum_{j=1}^{K} \lambda_j \nabla g_j(x^*, \theta)$$

$$\lambda_j g_j(x^*, \theta) = 0$$

Now, #1 and #2 obviously hold. What about #3? Clearly, $\nabla g_1 = 1$ and $\nabla g_2 = -1$ are not linearly independent. However, at most only one of the constraints will ever bind. If $g_1(x^*) = 0$ then $x^* = 5$ and so $g_2(x^*) < 0$. If $g_2(x^*) = 0$, then $x^* = 0$ and so $g_1(x^*) < 0$. Similarly, if $x^* \in (0, 5)$, then $g_1(x^*) < 0$ and $g_2(x^*) < 0$. Thus, for any solution x^*, $\{\nabla g_j(x^*, \theta) | g_j(x^*, 0) = 0\}$

will be linearly independent (the set will either be empty or have one element!). Therefore, all three conditions are satisfied.

Given this, the theorem states that for a solution x^*, there exist $\lambda_1, \lambda_2 \geq 0$ such that:

$$3x^2 - 10x + 1 = \lambda_1 - \lambda_2$$
$$\lambda_1 \cdot (x^* - 5) = 0$$
$$\lambda_2 \cdot (-x^*) = 0$$

There are three cases to consider, depending on which constraints bind. Notice that when a constraint binds, it's equivalent to "checking the endpoint".

Case #1: $5 - x \leq 0$ **binds**
$\Longrightarrow x = 5 \Longrightarrow -x$ is strictly less than 0. Therefore, $\lambda_2 = 0$. Since $x = 5$ and $\lambda_2 = 0 \Longrightarrow \lambda_1 = 26$.

Case #2: $-x \leq 0$ **binds**
$\Longrightarrow x = 0 \Longrightarrow 5-x$ is strictly less than 0. Therefore, $\lambda_1 = 0$. Since $x = 0$ and $\lambda_1 = 0, \Longrightarrow \lambda_2 = -1$.

Case #3: Neither constraint binds
$\Longrightarrow \lambda_1 = \lambda_2 = 0$. However, then the first equation reduces to $3x^2 - 10x + 1 = \lambda_1 - \lambda_2 = 0$, which is just the standard first order condition! To understand why, recognize that if neither constraint binds, x is in the interior. *If* the maximum were in the interior, then the usual first order condition should hold. Solving the quadratic yields $x =\approx 0.103$ or $x =\approx 3.23$

We now have <u>four</u> solutions to the KKT conditions. This is because the KKT conditions are necessary but not sufficient. We know that the solution will solve the KKT conditions. However, any $(x, \lambda_1, \lambda_2)$ that satisfies the KKT conditions is not necessarily a solution to the optimization problem. All that we need to do now is check which of these four candidates yields the maximal value. If you do this, you will find that the optimal solution is $x = 5$.

Envelope Theorem Assuming Differentiability

$$x^*(\theta) = \arg\max_{x \in \mathcal{X}(\theta)} f(x, \theta)$$

$$V(\theta) = \max_{x \in \mathcal{X}(\theta)} f(x, \theta)$$

To build intuition for the envelope theorem, we will assume that everything is differentiable. Moreover, we will assume a given choice $x = (x_1, \ldots, x_n)$ is a vector in $\mathbb{R}^n$ and a parameter $\theta = (\theta_1, \ldots, \theta_m)$ is a vector in $\mathbb{R}^m$. This is just to say is that a parameter may include multiple items (e.g. prices of goods and total budget) and choices might be over a bundle of items (e.g. apples and bananas).
The purpose of the envelope theorem is to understand how $V(\theta)$ changes when we vary a *single component* of the parameter vector.

0.1 Choice Set Independent of Parameter Set

The first case we will look at is the one we examined in class. When the choice set $\mathcal{X}(\theta)$ is independent of θ it means there is a fixed choice set $\mathcal{X}$. We want to compute $\frac{\partial V(\theta)}{\partial \theta_j}$ for each j.

By definition, $V(\theta) = f(x^*(\theta), \theta)$. The chain rule yields:

$$\frac{\partial V(\theta)}{\partial \theta_j} = \sum_{i=1}^{n} \frac{\partial f(x^*(\theta), \theta)}{\partial x_i} \cdot \frac{x_i^*(\theta)}{\partial \theta_j} + \frac{\partial f(x^*(\theta), \theta)}{\partial \theta_j}$$

Think about $\frac{\partial f(x_i^*(\theta), \theta)}{\partial x_i}$. By definition, $x^*(\theta)$ maximizes $f(x, \theta)$, which means it satisfies the first-order condition. Therefore, $\frac{\partial f(x_i^*(\theta), \theta)}{\partial x_i} = 0$. Thus:

$$\frac{\partial V(\theta)}{\partial \theta_j} = \frac{\partial f(x^*(\theta), \theta)}{\partial \theta_j}$$

Example 1. *Let $f(x, \theta) = -\frac{1}{2}(x - \theta)^2 + x$. We can solve for the optimal x by using a first-order condition since f is concave. This yields $x^* = \theta + 1$. If you wanted to, you could plug this back into f and compute the value function directly: $V(\theta) = f(x^*(\theta), \theta) = -\frac{1}{2}(1 + \theta - \theta)^2 + \theta + 1 = 1 + \theta$. Taking the derivative yields $\frac{\partial V}{\partial \theta} = 1$.*

Using the envelope theorem, $\frac{\partial V}{\partial \theta} = \frac{\partial f(x^*(\theta), \theta)}{\partial \theta} = x^*(\theta) - \theta = 1$.

0.2 Choice Set As Constraints

Suppose the choice set is dependent on the parameter set and can be described by a series of constraints. In other words, there are functions $g_1, \ldots, g_K$ such that $\mathcal{X}(\theta) = \{x | g_1(x, \theta), \ldots, g_K(x, \theta) \leq 0\}$. This is the case in standard consumer optimization. The optimization problem is:

$$\max_{x \in \mathbb{R}^n} f(x, \theta)$$

$$\text{subject to } g_j(x, \theta) \leq 0 \text{ for } j = 1, \ldots, K$$

We want to compute $\frac{\partial V(\theta)}{\partial \theta_j}$. Using the chain rule, we still have the following expression:

$$\frac{\partial V(\theta)}{\partial \theta_j} = \sum_{i=1}^{n} \frac{\partial f(x^*(\theta), \theta)}{\partial x_i} \cdot \frac{x_i^*(\theta)}{\partial \theta_j} + \frac{\partial f(x^*(\theta), \theta)}{\partial \theta_j}$$

However, we can not simply replace $\frac{\partial f(x_i^*(\theta), \theta)}{\partial x_i}$ with 0 because x^* does not satisfy the traditional first-order condition. Luckily, we can make use of the KKT conditions.

At the solution x^*, we know that there are multipliers $\lambda_1, \ldots, \lambda_K \geq 0$ such that:

$$\frac{\partial f(x^*(\theta), \theta)}{\partial x_i} = \sum_{t=1}^{K} \lambda_k \frac{\partial g_t(x^*(\theta), \theta)}{\partial x_i}$$

Substituting this expression into the one for $\frac{\partial V(\theta)}{\partial \theta_j}$, we have:

$$\begin{aligned}
\frac{\partial V(\theta)}{\partial \theta_j} &= \sum_{i=1}^{n} \Big(\sum_{t=1}^{K} \lambda_t \frac{\partial g_t(x^*(\theta), \theta)}{\partial x_i} \Big) \cdot \frac{x_i^*(\theta)}{\partial \theta_j} + \frac{\partial f(x^*(\theta), \theta)}{\partial \theta_j} \\
&= \sum_{t=1}^{K} \lambda_t \Big(\sum_{i=1}^{n} \frac{\partial g_t(x^*(\theta), \theta)}{\partial x_i} \frac{\partial x_i^*(\theta)}{\theta_j} \Big) + \frac{\partial f(x^*(\theta), \theta)}{\partial \theta_j}
\end{aligned}$$

This still looks like a complete mess, but we can clean it up a bit. First recognize that the constraints that dont bind have multipliers equal to 0, which means they drop out. We can effectively ignore them. Thus, without loss of generality, assume that all the constraints bind. When a constraint g_t binds, it means that $g_t(x^*(\theta), \theta) = 0$. Differentiating $g_t(x^*(\theta), \theta) = 0$ with respect to θ_j yields:

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\partial g_t(x^*(\theta), \theta)}{\partial x_i} \frac{\partial x_i^*(\theta)}{\theta_j} + \frac{\partial g_t(x^*(\theta), \theta)}{\theta_j} &= 0 \\
\implies -\frac{\partial g_t(x^*(\theta), \theta)}{\theta_j} &= \sum_{i=1}^{n} \frac{\partial g_t(x^*(\theta), \theta)}{\partial x_i} \frac{\partial x_i^*(\theta)}{\theta_j}
\end{aligned}$$

We can now substitute this into our expression for $\frac{\partial V(\theta)}{\partial \theta_j}$:

$$\frac{\partial V(\theta)}{\partial \theta_j} = \frac{\partial f(x^*(\theta), \theta)}{\partial \theta_j} - \sum_{t=1}^{K} \lambda_t \frac{\partial g_t(x^*(\theta), \theta)}{\theta_j}$$

Recognize the similarity between this and the envelope theorem when the choice set is independent of the parameter (e.g. all the λ's are 0). Also, the right-hand side expression should be familiar. Replace f with the utility function and the θ_j with p_j (so this becomes a standard consumer maximization problem), and the right-hand side is exactly the derivative of the lagrangian with respect to the price of good j.