Econ 703 TA Note 9

Chia-Min Wei *

August 31, 2025

1 Notation

Let $f : \mathbb{R}^n \to \mathbb{R}$. We define $Df(x) \coloneqq \nabla f(x)^\intercal$. Let $g : \mathbb{R}^n \to \mathbb{R}^k$, where $g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_n(x) \end{bmatrix}$. We define $Dg(x) \coloneqq \begin{bmatrix} \nabla g_1(x)^\intercal \\ \nabla g_2(x)^\intercal \\ \vdots \\ \nabla g_n(x)^\intercal \end{bmatrix}$.

2 Equality Constraints - Lagrange's Method

The Lagrange's method provides a way of *selecting* points that are *possibly* local optimal points when there are only equality constraints. Namely, it provides a *necessary* condition for a point to be a local optimal point.

2.1 First Order Condition

Theorem 2.1 (Lagrange): Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, ..., k$ be C^1 functions. Suppose x^* is a local maximal or minimal point of f on the set

$$\mathcal{D} = \{x : g_i(x) = 0 \text{ for all } 1 \leq i \leq k\}.$$

Write $g = (g_1, ..., g_k)^\intercal : \mathbb{R}^n \to \mathbb{R}^k$. Suppose also that $\text{rank}(Dg(x^*)) = k$. Then there exists $\lambda_1^*, ..., \lambda_k^* \in \mathbb{R}$ such that

$$Df(x^*) + \sum_{i=1}^{k} \lambda_i^* Dg_i(x^*) = 0.$$

Two points should be emphasized.

- Lagrange's Theorem provides a necessary condition for local optimal points, not a sufficient condition. A point satisfying the condition may not be a local optimal point.

*This TA note is based on Prof. John Kennan's math camp lecture taught in 2025 at UW-Madison. All errors are mine.

- If at an optimal point x^*, $Dg(x^*)$ does not have rank k, then the conclusion in Theorem 2.1 does not necessarily hold: there may not exist $\lambda_1, \lambda_2, ..., \lambda_n$ such that

$$Df(x^*) + \sum_{i=1}^{k} \lambda_i^* Dg_i(x^*) = 0.$$

2.2 Intuition

Let us consider the case when there is only one constraint: $g(x_1, ..., x_n) = I - p_1 x_1 - ... - p_n x_n$, and we set the constraint $g(x) = 0$, $x \in \mathbb{R}^n$. Charlie faces the problem:

$$\max U(x) \text{ s.t. } g(x) = 0.$$

(Suppose he is forced to spend up all of his money). At the optimal point $x^* = (x_1^*, ..., x_n^*)$,

$$\frac{MU_1}{p_1} = \frac{MU_2}{p_2} = ... = \frac{MU_n}{p_n} \tag{1}$$

Why? Assume the contrary that at the optimal point x^*,

$$\frac{MU_1}{p_1} > \frac{MU_2}{p_2}.$$

Suppose Charlie sells a small amount ϵ of good 2. By doing so, he obtains $\epsilon \cdot p_2$ units of money. This sale reduces his utility by approximately $MU_2 \cdot \epsilon$. He can then use the money to purchase

$$\frac{\epsilon \cdot p_2}{p_1}$$

units of good 1. The additional amount of good 1 yields an increase in utility of

$$MU_1 \cdot \frac{\epsilon \cdot p_2}{p_1}.$$

Comparing the two changes in utility, Charlie becomes strictly better off since

$$MU_1 \cdot \frac{p_2}{p_1} \cdot \epsilon \;>\; MU_2 \cdot \epsilon.$$

This contradicts the fact that x^* is the optimal point. Therefore, Equation 1 must hold. Note that

$$\begin{aligned} DU(x) &= \begin{bmatrix} MU_1(x) & MU_2(x) & \cdots & MU_n(x) \end{bmatrix} \\ Dg(x) &= \begin{bmatrix} -p_1 & -p_2 & \cdots & -p_n \end{bmatrix} \end{aligned}$$

If we choose $\lambda^* = MU_i/p_i$, then

$$DU(x^*) + \lambda^* Dg(x^*) = 0.$$

This verifies the conclusion of Theorem 2.1.

Suppose Charlie's income increases by ϵ. He can use the additional money to buy ϵ/p_i units of good i. This gives him an additional utility of

$$\frac{MU_1}{p_1}\epsilon = \lambda^* \epsilon.$$

Note that the additional utility he gains is the same regardless of which good he purchases. λ^* is Charlie's marginal utility as his income increases, and is called the **shadow price** of the constraint $g(x)$.

2.3 A Cookbook Procedure

Let an equality constraint problem be written in the following form: $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$ for $1 \le i \le k$.

$$\max f(x) \text{ over } \mathcal{D} = \{x \in \mathbb{R}^n : g_i(x) = 0 \text{ for all } 1 \le i \le k\}.$$

To solve the problem, we follow the procedure:

1. Set up the Lagrangian:
$$L(x; \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x)$$

2. Solve (x, λ) that satisfies the following system of equations:
$$\frac{\partial L}{\partial x_i}(x; \lambda) = 0 \text{ for all } i = 1, ..., n;$$
$$\frac{\partial L}{\partial \lambda_j}(x; \lambda) = 0 \text{ for all } j = 1, ..., k.$$
   The solutions are called the critical points of the Lagrangian.

3. If we are maximizing f, then choose from the critical points that maximizes f. If we are minimizing f, then choose the one that minimizes f.

3 Inequality Constraints - Kuhn and Tucker's Method

Now we turn to a more general case: when some constraints are inequalities. Kuhn and Tucker provided a method to solve the problem:

$$\text{maximize } f \text{ over } \mathcal{D} = \{x \in \mathbb{R}^n : g_j(x) = 0 \text{ for all } 1 \le j \le m_1,\ h_i(x) \ge c_i \text{ for all } 1 \le i \le m_2\},$$

where $f : \mathbb{R}^n \to \mathbb{R}$, each $g_j : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^n \to \mathbb{R}$.

3.1 First Order Condition

Theorem 3.1 (Kuhn and Tucker): Let $f : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^n \to \mathbb{R}$ be C^1 functions, $c_i \in \mathbb{R}$, where $1 \le j \le m_1$ and $1 \le i \le m_2$. Suppose x^* is a local maximal point of f on

$$\mathcal{D} = \{x \in \mathbb{R}^n : g_j(x) = 0 \text{ for all } 1 \le j \le m_1,\ h_i(x) \ge c_i \text{ for all } 1 \le i \le m_2\}.$$

Let $E \subset \{1, ..., l\}$ be the set of inequality constraints that are binding and write $h_E = (h_i)_{i \in E}$. Suppose $D(g, h_E)$ has rank $m_1 + |E|$. Then there exists $(\lambda_1^*, ..., \lambda_{m_1}^*)$ and $(\mu_1^*, ..., \mu_{m_2}^*)$ such that the following conditions are met:

$$\mu_i^* \ge 0 \text{ and } \mu_i^*(h_i(x^*) - c_i) = 0 \text{ for } i = 1, ..., m_2 \tag{KT-1}$$
$$Df(x^*) + \sum_{j=1}^{m_1} \lambda_j^* Dg_j(x^*) + \sum_{i=1}^{m_2} \mu_i^* Dh_i(x^*) = 0. \tag{KT 2}$$

When it is a minimization problem, KT-1 is replaced by

$$\mu_i^* \le 0 \text{ and } \mu_i^*(h_i(x^*) - c_i) = 0 \text{ for } i = 1, ..., m_2.$$

KT-1 is called the *complementary slackness* condition. It is called complementary because at least one of the following must be true: the constraint is binding or the multiplier is 0.

Some points worth noticing:

1. Similar to Lagrange's Theorem, Kuhn and Tucker's Theorem provides a *necessary* condition for local optimal points, not a *sufficient* condition.
2. For local optimal points that does not satisfy $\operatorname{rank}\{D(g, h_E)(x^*)\} = m_1 + |E|$, the conclusion of the theorem may not hold.
3. It is recommended that one write all inequalities in terms of $h_i(x) \geq c_i$ so that one doesn't have to think repeatedly about the sign of λ_i for each i when solving the question. In that way, for maximization problems, $\lambda_i \geq 0$ for all i and, for minimization problems, $\lambda_i \leq 0$ for all i.

3.2 Intuition

Charlie is maximizing his utility $u(x_1, ..., x_n)$ with the following constraints:

$$g(x) = I - p_1x_1 - ... - p_nx_n = 0,$$
$$h_i(x) = x_i \geq c_i, \quad \forall 1 \leq i \leq n.$$

Let $x^* = (x_1^*, ..., x_n^*)$. There are two possible cases:

(i) $x_i^* > c_i$ for all $1 \leq i \leq n$: Then as we discussed in subsection 2.2, at x^*, it must be

$$\frac{MU_1}{p_1} = \frac{MU_2}{p_2} = \cdots = \frac{MU_n}{p_n}.$$

Choose λ^* to be the common ratio, and $\mu_i^* = 0$ for all $1 \leq i \leq n$. Then we would have

$$Df(x^*) + \lambda^* Dg(x^*) + \sum_{i=1}^{n} \mu_i^* Dh_i(x) = 0.$$

(ii) There is some $x_i = c_i$: Consider the case $x_i^* > c_i$ for all $1 \leq i \leq n-1$ while $x_n^* = c_i$. Then our argument in subsection 2.2 no longer works: it may be the case that

$$\frac{MU_1}{p_1} = \cdots = \frac{MU_{n-1}}{p_{n-1}} > \frac{MU_n}{p_n}.$$

Charlie would like to exchange x_n for other goods, but the requirement that he must keep at least c_n units of good n makes this infeasible. Write

$$\lambda^* = \frac{MU_1}{p_1} = ... = \frac{MU_{n-1}}{p_{n-1}}, \quad \mu_n^* = \lambda^* p_n - MU_n > 0.$$

Note that

$$Dh_n(x^*) = \left[0, \cdots, 0, 1\right].$$

Then we would have

$$Df(x^*) + \lambda^* Dg(x^*) + \sum_{i=1}^{n} \mu_i^* Dh_i(x) = 0.$$

This verifies the conclusion of Theorem 3.1.

Suppose Charlie were allowed to hold only $c_n - \epsilon$ units of good n. He could then sell ϵ units of good n for $\epsilon \cdot p_n$ units of money. With this money, he could purchase $(\epsilon \cdot p_n)/p_1$ units of good 1, yielding an additional utility of $(\epsilon \cdot p_n \cdot MU_1)/p_1 = \lambda^* p_n \epsilon$ of utility. At the same time, selling ϵ units of good n reduces his utility by $\epsilon \cdot MU_n$. His net gain of utility is

$$\lambda^* p_n \epsilon - MU_n \epsilon = \mu_n^* \epsilon.$$

Hence, μ_n^* is the marginal utility of *loosening* the constraint $h_n(x) \geq c_n$, and is called the **shadow price** of the constraint $h_n(x)$. (For a constraint $h_i(x) \geq c_i$, loosening it means to decrease c_i.)

This also explained why $\mu_i \leq 0$ in minimization problems: loosening the constraint allows the minimum value to decrease.

3.3 A Cookbook Procedure

Let an inequality constraint problem be written in the following form: $f : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^n \to \mathbb{R}$ be C^1 functions, $c_i \in \mathbb{R}$, where $1 \leq j \leq m_1$ and $1 \leq i \leq m_2$.

$$\max f(x) \text{ over } \mathcal{D} = \{x \in \mathbb{R}^n \in \mathbb{R}^n : g_j(x) = 0 \text{ for all } 1 \leq j \leq m_1,\ h_i(x) \geq c_i \text{ for all } 1 \leq i \leq m_2\}.$$

We follow the procedure to find the solution.

1. Set up the Lagrangian:

$$L(x; \lambda, \mu) = f(x) + \sum_{j=1}^{m_1} \lambda_j g_j(x) + \sum_{i=1}^{m_2} \mu_i (h_i(x) - c_i).$$

2. Solve the following system of equations and inequalities:

$$\frac{\partial L}{\partial \lambda_j}(x; \lambda, \mu) = 0 \text{ for all } j = 1, ..., m_1 \qquad \text{(the equality constraints)}$$

$$\frac{\partial L}{\partial \mu_i}(x; \lambda, \mu) \geq 0 \text{ for all } i = 1, ..., m_2 \qquad \text{(the inequality constraints)}$$

$$\mu_i \geq 0, \quad \mu_i \frac{\partial L}{\partial \mu_i}(x; \lambda, \mu) = 0 \text{ for all } i = 1, ..., m_2 \qquad \text{(KT-1)}$$

$$\frac{\partial L}{\partial x_i}(x; \lambda, \mu) = 0 \text{ for all } i = 1, ..., n. \qquad \text{(KT-2)}$$

Points (x, λ) that solve the system are called critical points.

3. Choose the critical point that maximizes f.

For minimization problems, change (KT-1) to

$$\mu_i \leq 0, \quad \mu_i \frac{\partial L}{\partial \mu_i}(x; \lambda, \mu) = 0 \text{ for all } i = 1, ..., m_2. \qquad \text{(KT-1)}$$

And of course, choose the critical point that minimizes f in the final step.