Optimization with Inequality Constraints (Kuhn Tucker Theorem)

September 3, 2020

1 Kuhn Tucker Theorem

In these notes, we will introduce the concept of optimization with inequality constraints. Specifically, we are interested in a problem like the one below :

$$\max f(x) \;\; s.t. \; h(x) = c \;\; and \;\; g(x) \leq b$$

To solve this kind of an optimization problem, one needs the Kuhn-Tucker Theorem. Just like an optimization problem with inequality constraints, one sets up the Lagrangean :

$$\mathcal{L}(x, \lambda, \mu) = f(x) - \mu\,(h(x) - c) - \lambda\,(g(x) - b)$$

and then try and maximize this new unconstrained problem by taking FOC's w.r.t (x^*, λ^*, μ^*) and check a couple of additional conditions which will be formalized in the Kuhn-Tucker Theorem below.

The idea is essentially the same as that you have seen in optimization with equality constraints. The only additional thing we need to be worried about when dealing with inequality constraints is whether a constraint will be binding or not. To deal with this we introduce something called the complementary slackness condition. To see this more formally, the Kuhn-Tucker Theorem is stated below.

Theorem 1.1. *Consider the following maximization problem. Let X be an open set in $\mathbb{R}^n$ and let $f : X \to \mathbb{R}$, $(f(x) = f(x_1, \ldots, x_n))$, $g : X \to \mathbb{R}^k$ and $h : X \to \mathbb{R}^m$ be C^1 functions :*

$$\max_{x \in X} f(x) \;\; s.t. \; g(x) \leq b \;\; and \;\; h(x) = c$$

If x^* is a maximizer of the problem and the Constraint Qualification (CQ) condition holds at x^*, then there exists a unique $(\lambda, \mu) \in \mathbb{R}^k_+ \times \mathbb{R}^m$ s.t. the following two conditions hold :

1. ***First Order Condition (FOC) -***

$$\underbrace{\nabla f(x^*)}_{n\times 1} - \underbrace{g'(x^*)}_{n\times k} \cdot \underbrace{\lambda}_{k\times 1} - \underbrace{h'(x^*)}_{n\times m} \cdot \underbrace{\mu}_{m\times 1} = \underbrace{0}_{n\times 1}$$

   Note that this is a set of n equations , one for each x_i.

2. ***Complementary Slackness (CSC) -***

$$h_l\left(x^*\right) = c_l$$

for each $l \in \{1, \ldots, m\}$

$$\lambda_j \geq 0 \ with \ \lambda_j = 0 \ if \ \ g_j(x^*) < b_j$$

for each $j \in \{1, 2, \ldots, k\}$

Notes :

1. The constraints $g(x) \leq b$ are actually a set of k inequalities i.e.

$$g(x) \leq b$$
$$\implies \begin{pmatrix} g_1(x) \\ \vdots \\ g_k(x) \end{pmatrix} \leq \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}$$

and the constraints $h(x) = c$ are actually a set of m inequalities i.e.

$$h(x) = c$$
$$\implies \begin{pmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$$

2. When there is no constraint, the FOC boils down to $\nabla f\left(x^*\right) = 0$ which is the same as in the case for unconstrained optimization.

3. The FOC is essentially for each $i \in \{1, \ldots, n\}$:

$$\frac{\partial \mathcal{L}}{\partial x_i}\left(x^*, \lambda, \mu\right) = \frac{\partial}{\partial x_i}\left(f\left(x\right) - \lambda^T\left(g(x) - b\right) - \mu^T\left(h(x) - c\right)\right)|_{x^*} = 0$$
$$\frac{\partial f}{\partial x_i}\left(x^*\right) - \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i}\left(x^*\right) - \sum_{l=1}^{m} \mu_l \frac{\partial h_l}{\partial x_i}\left(x^*\right) = 0$$

4. Note that the term $\underbrace{g'\left(x^*\right)}_{n\times k} \cdot \underbrace{\lambda}_{k\times 1}$ is nothing but the expression :

$$\underbrace{g'\left(x^*\right)}_{n\times k} \cdot \underbrace{\lambda}_{k\times 1} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_k}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial x_n} & \cdots & \frac{\partial g_k}{\partial x_n} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{pmatrix}$$

and the term $\underbrace{h'(x^*)}_{n\times m}\cdot\underbrace{\mu}_{m\times 1}$ is given by :

$$\underbrace{h'(x^*)}_{n\times m}\cdot\underbrace{\mu}_{m\times 1}=\begin{pmatrix}\frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_1}\\ \vdots & \ddots & \vdots\\ \frac{\partial h_1}{\partial x_n} & \cdots & \frac{\partial h_m}{\partial x_n}\end{pmatrix}\cdot\begin{pmatrix}\mu_1\\ \vdots\\ \mu_k\end{pmatrix}$$

5. Simply, the Kuhn-Tucker Theorem states that ***if** x^* **is a maximizer and satsifies the CQ, then there exists** λ **and** μ **such that** (x^*,λ,μ) **satisfies FOC+CSC.***

6. ***What is the interpretation of the Complementary Slackness Condition?*** Remember that the multiplier λ_j is the marginal increase in the maximized objective function due to a slight relaxation of the constraint. The CSC simply says that if a constraint is not binding at the optimum, then if we relax the constraint a little, we should not be increasing the value of the objective function i.e. $\lambda_j=0$. In a utility maximizing framework, if the Budget constraint does not bind at the optimum i.e. $p\cdot x^*<m$, then we are leaving some money on the table already. Now if we increase m, then our choice x^* should not change which implies that $U(x^*)$ is unchanged which in turn implies that the multiplier associated with the budget constraint is zero i.e. $\lambda=0$.

7. In practice one writes down what people call the **K-T conditions**:

- $\frac{\partial f}{\partial x_i}(x^*)-\sum_{j=1}^{k}\lambda_j\frac{\partial g_j}{\partial x_i}(x^*)-\sum_{l=1}^{m}\mu_l\frac{\partial h_l}{\partial x_i}(x^*)=0\ \forall i=1,\ldots,n$
- $h_l(x)=c_l\ \forall l=1,\ldots,m$
- $\lambda_j\geq 0, g_j(x)\leq b_j$ and $\lambda_j(g_j(x)-b_j)=0\ \forall j=1,\ldots,k$

Some Additional Comments Regarding Implications of the Kuhn-Tucker Theorem

There are a few important things that you must remember regarding the implications of this theorem.

1. It is **ONLY** a necessary condition and **NOT** a sufficient condition.

2. The Kuhn-Tucker Theorem states that ***if** x^* **is a maximizer and satisfies the CQ, then*** (x^*,λ,μ) ***must solve FOC+CSC***. Hence we can find x^* by solving for all solutions to the K-T conditions. It does ***NOT*** however mean that if some x solves ***FOC+CSC*** then x is a maximizer.

3. The Kuhn-Tucker Theorem ***only works at** x^* **at which the CQ holds***. If the CQ fails at x^*, then there may not exist (x^*,λ,μ) which satisfy the FOC and CSC even if x^* is a maximizer of the problem. So there may exist other feasible points at which the CQ fails which may also be maximizers.

4. So when we are finding the solution to an optimization problem, we should check find the solutions x^* where the KT conditions are satisfied. Call these ***Type 1*** solutions. (Note all these points are not maximizers since the KT conditions are only necessary conditions). We should also find all points at which the ***CQ*** fails. Call these ***Type 2*** solutions. Our solution is then found by comparing all these points i.e ***Type 1 + Type 2***.

1.1 Constraint Qualification

We skipped over what Constraint Qualification meant in the statement of the Kuhn-Tucker Theorem. Let's now go over what it means.

Theorem 1.2. *Let X be an open set in $\mathbb{R}^n$ and let $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}^k$ and $h : X \to \mathbb{R}^m$ be C^1 functions :*

$$\max_{x \in X} f(x) \text{ s.t. } g(x) \leq b \text{ and } h(x) = c$$

For a feasible point $\hat{x} \in X$, the inequality constraint $g_j(x) \leq b_j$ is said to be binding at $\hat{x}$ if $g_j(\hat{x}) = b_j$. We say that the **Constraint Qualification (CQ) holds at** *$\hat{x}$ iff the derivates of all binding constraints :*

$$\{\nabla g_j(\hat{x})\}_{\{j: g_j \text{ binding at } \hat{x}\}} \cup \{\nabla h_l(\hat{x})\}_{l=1}^{m}$$

in $\mathbb{R}^n$ are linearly independent. Otherwise we say that the constraint qualification fails at $\hat{x}$.

What does it mean?

Note that the FOC implies that :

$$\underbrace{\nabla f(x^*)}_{n \times 1} = \underbrace{g'(x^*)}_{n \times k} \cdot \underbrace{\lambda}_{k \times 1} + \underbrace{h'(x^*)}_{n \times m} \cdot \underbrace{\mu}_{m \times 1}$$

First note that by complementary slackness, $\lambda_j = 0$ for those constraints which are not binding. Moreover, since the theorem requires states that a **unique** (λ, μ) exists, then it must be that the gradient vectors of the constraints which bind must be Linearly Independent. This is because we know from Linear Algebra that if a vector y is expressed as a linear combination of some linearly independent vectors, then the representation is unique!

2 Examples of Kuhn Tucker

Now we will go over a couple of examples of applying the KT-Theorem. The first one will take you through a step-by-step introduction of how to work through the different cases when solving such a problem. The second example will revisit a modified version of the example we did last week and do it in a Kuhn-Tucker framework.

Example 2.1. Consider the following problem :

$$\begin{aligned} \max \quad & xy + x^2 \\ s.t \quad & g_1(x,y) = x^2 + y \leq 2 \\ & g_2(x,y) = -y \leq -1 \end{aligned}$$

Solution : We will work through how to do it in a systematic procedure. The first thing to note is that a solution exists by Weierstrass' Theorem since the objective function is contiuous and the constraint set is closed and bounded. Moreover, first note that the Lagrangean of this problem is :

$$\mathcal{L} = xy + y^2 - \lambda_1 \left(x^2 + y - 2\right) - \lambda_2 (-y + 1)$$

The KT conditions are then:

$$\mathcal{L}_x = y + 2x - 2\lambda_1 x = 0 \tag{2.1}$$
$$\mathcal{L}_y = x - \lambda_1 + \lambda_2 = 0 \tag{2.2}$$
$$\lambda_1 \geq 0 \text{ with } \lambda_1 = 0 \text{ if } x^2 + y < 2 \tag{2.3}$$
$$\lambda_2 \geq 0 \text{ with } \lambda_2 = 0 \text{ if } y > 1 \tag{2.4}$$

We will now work through the problem in cases :

Case 1. ***Both constraints are binding.*** This means that $x^2 + y = 2$ and $y = 1$. This implies $x = \pm 1$ and $y = 1$. First consider $(x, y) = (1, 1)$, then from (2.1) and (2.2), we have $\lambda_1 = \frac{3}{2}$ and $\lambda_2 = \frac{1}{2}$. Thus we have our first Type 1 candidate with $(x, y) = (1, 1)$ and $(\lambda_1, \lambda_2) = \left(\frac{3}{2}, \frac{1}{2}\right)$. (Why? This solution candidate solves the FOC and the CSC. Note that we still need to check CQ which we will do later)

Now when $(x, y) = (-1, 1)$, we have from (2.1) and (2.2) that $(\lambda_1, \lambda_2) = \left(\frac{1}{2}, \frac{3}{2}\right)$. Hence we have another Type 1 candidate with $(x, y) = (-1, 1)$ and $(\lambda_1, \lambda_2) = \left(\frac{1}{2}, \frac{3}{2}\right)$

Case 2. ***Constraint 1 is binding and Constraint 2 is not binding.*** This means that $x^2 + y = 2$ and $y > 1$. From the CSC in (2.4), we get that $\lambda_2 = 0$ and hence from (2.2), we get $\lambda_1 = x$. Plugging this in (2.1), we have $y + 2x - 2x^2 = 0$. Moreover from the first constraint we have $y = 2 - x^2$. So combining the previous two equations we have :

$$3x^2 - 2x - 2 = 0$$

The solutions are $x = \frac{1}{3}\left(1 \pm \sqrt{7}\right)$. But since $x = \lambda_1 \geq 0$, we must have $x = \frac{1}{3}\left(1 + \sqrt{7}\right)$. But then that implies $y = 2 - x^2 = \frac{2}{9}\left(5 - \sqrt{7}\right) < 1$ which violates that $y > 1$. Hence there are no solution candidates in this case.

Case 3. ***Constraint 1 is not binding and Constraint 2 is binding.*** This means that $x^2 + y < 2$ and $y = 1$. Again this implies from (2.3) that $\lambda_1 = 0$. Then (2.1) gives $x = -\frac{1}{2}$ and (2.2) gives $\lambda_2 = \frac{1}{2}$. Thus $(x, y) = \left(-\frac{1}{2}, 1\right)$ and $(\lambda_1, \lambda_2) = \left(0, \frac{1}{2}\right)$ is a solution candidate.

Case 4. ***Both constraints are not binding.*** This means that $x^2 + y < 2$ and $y > 1$. Hence from the CSC, we get $\lambda_1 = 0 = \lambda_2$. Then from (2.1) and (2.2), we get that $y = 0$ which contradicts that $y > 1$. Hence there are no solution candidates in this case.

So summing up, we have 3 ***Type-1*** solution candidates. Note that the function evaluated at the first solution is $f(1, 1) = 1$ which is larger than the functional value evaluated at the other 2 points.

So we only need to check the constraint qualification. We need to consider the CQ for the different cases :

1. If we are in *Case* 1, then both constraints are binding and then we have that the gradient of the binding constraints are :

$$\begin{pmatrix} \nabla g_1(x) & \nabla g_2(x) \end{pmatrix} = \begin{pmatrix} 2x & 0 \\ 1 & -1 \end{pmatrix}$$

However when both constraints bind, the only points which satisfy the constraints are (1,1) and (-1,1) both of which are already Type-1 solution candidates. So we don't need to worry about feasible points which don't satisfy CQ in this case. Moreover you can check that both of these points satisfy the CQ i.e. the vectors are linearly independent.

2. In *Case 2,* the first constraint is binding which implies that $\nabla g_1(x) = \begin{pmatrix} 2x & 1 \end{pmatrix}'$ which is linearly independent for all feasible points since it is not the $\mathbf{0}$ vector.
3. In *Case 3,* the second constraint is binding which implies that $\nabla g_2(x) = \begin{pmatrix} 0 & -1 \end{pmatrix}'$ which again is linearly independent for all feasible points
4. In *Case 4,* the CQ holds trivially since no constraint is binding.

So what we found is that there are no feasible points at which the CQ fails i.e. there are no ***Type-2*** candidate solutions. So we only have to choose between our ***Type-1*** candidate solutions. Hence, as discussed above $(x, y) = (1, 1)$ gives the maximum functional value and hence is our solution.

The above is a systematic procedure to solve a Kuhn-Tucker problem. Note that since we had 2 inequality constraints, we had to check 4 cases. In general with n constraints you will have to check 2^n cases which can get unmanageable pretty quickly. So as discussed in the previous recitation, it is better to eliminate some cases with clever arguments. But, let us revisit the problem we did in the previous recitation and see how to do it by rigorously trying to follow the KT Theorem and not by using clever arguments to get rid of some cases.

Example 2.2. Consider the problem :

$$\max_{(x_1,x_2)\in\mathbb{R}^2_+} x_1^{\alpha} x_2^{1-\alpha}$$

$$s.t. p_1 x_1 + p_2 x_2 \leq m$$

where $\alpha \in (0, 1)$, $p_1, p_2 > 0$ and $m \geq 0$ are parameters. **Solve for the maximizers using the Kuhn-Tucker conditions.**

Solution : I will do this step by step. Firstly, note that I assumed that $x_1, x_2 > 0$ unlike in the previous recitation. Let us first take a look at what the CQ entails for this problem. It makes sense to check the CQ first since we only have 1 constraint.

- In this problem we only have 1 constraint, and so the gradient of the constraint is given by $\nabla g(x) = \begin{pmatrix} -p_1 & -p_2 \end{pmatrix} \neq 0$ which implies it is linearly independent (since it is not the $\mathbf{0}$ vector). Thus CQ holds at all feasible points! So we do not need to worry about maximizers at points where CQ is not satisfied i.e. we have no Type 2 points!
- Next we write down the Lagrangean :

$$\mathcal{L}(x_1, x_2, \lambda) = x_1^{\alpha} x_2^{1-\alpha} - \lambda (p_1 x_1 + p_2 x_2 - m)$$

- The K-T conditions are then :

$$\begin{cases} x \in \mathbb{R}^2_{++} \\ \alpha x_1^{\alpha-1} x_2^{1-\alpha} - \lambda p_1 = 0 \\ (1-\alpha) x_1^{\alpha} x_2^{-\alpha} - \lambda p_2 = 0 \\ \lambda \geq 0, m - p_1 x_1 - p_2 x_2 \geq 0, \quad \lambda \left(m - p_1 x_1 - p_2 x_2 \right) = 0 \end{cases}$$

- Now since we have 1 inequality constraint, we need to consider 2 cases, one where the constraint binds and one where it doesn't. But looking at the FOC's and using the fact that $p_1, p_2, x_1, x_2 > 0$, we will be able to eliminate checking the case where the budget constraint does not bind.

- Comparing the FOC's :

$$\alpha x_1^{\alpha-1} x_2^{1-\alpha} - \lambda p_1 = 0$$
$$(1-\alpha) x_1^{\alpha} x_2^{-\alpha} - \lambda p_2 = 0$$

  we see that $\lambda > 0$ which implies by the CSC that we must have $m - p_1 x_1 - p_2 x_2 = 0$

- So now we are back to our standard setting where we have the 2 FOC's and the binding budget constraint. So to find the solution, we solve the FOC's simultaneously which gives us :

$$\frac{p_1 x_1}{p_2 x_2} = \frac{\alpha}{1-\alpha}$$

- Plugging this into the budget constraint, we solve for (x_1^*, x_2^*) which we then use to solve λ^*. So we get :

$$(x_1^*, x_2^*, \lambda^*) = \left(\frac{\alpha m}{p_1}, \frac{(1-\alpha)m}{p_2}, \frac{\alpha^\alpha (1-\alpha)^\alpha}{p_1^\alpha p_2^{1-\alpha}} \right)$$

  as the unique ***Type-1*** solution to the KT condition. Because the CQ holds for all points, we know there does not exist any other x^* that might be a maximizer!

- Hence the unique maximizer in this case is $(x_1^*, x_2^*) = \left(\frac{\alpha m}{p_1}, \frac{(1-\alpha)m}{p_2} \right)$.