Lecture Notes - Static Optimization

MA Math Camp 2021
Columbia University

César Barilla*

This Version : August 18, 2021

Contents

1 General Setup

Definition 1.1. *Let f be a function from X to the poset $(Y, \leq)$, and let $D \subset X$. A* ***maximization problem*** *takes the form*

$$\max_{x \in X} f(x) \ \text{ s.t. } \ x \in D$$

where f is called the ***objective function****, x is called the* ***choice variable****, and D is called the* ***constraint set*** *or* ***feasible set****. A point $x \in X$ is said to be* ***feasible*** *iff $x \in D$.*

The set of ***maximizers****, or* ***maximum points****, of this problem is defined as*

$$\arg\max_{x \in X} \Big\{ f(x) : x \in D \Big\} := \Big\{ x^* \in D : f\left(x^*\right) \geq f(x) \ \forall \ x \in D \Big\}$$

If the set of maximizers is nonempty, then this problem is said to ***have a solution****. In this case, we define the* ***maximum****, or the* ***maximum value****, of this problem as $f(x^*)$, where x^* is an arbitrary maximizer, and denote it as $\max_{x \in X} \big\{ f(x) : x \in D \big\}$.*

*The present lecture notes were largely based on math camp materials from Palaash Bhargava, Paul Koh, and Xuan Li. All errors in this document are mine. If you find a typo or an error, please send me an email at cesar.barilla@columbia.edu.

Notice that although the set of maximizers can be non-singleton when nonempty, the maximum does not depend on the selection of x^* from the set of maximizers. This follows directly from the anti-symmetry property of the partial order on Y.

We can define minimization problem analogously. In fact, we can always transform a minimization problem into a maximization problem by reversing the order $\leq$ on the codomain, and therefore it is without loss to only study maximization problems. In most applications, of course, the codomain of the objective function f is the totally ordered set $(\mathbb{R}, \leq)$. In this case, we can transform a minimization problem of function f to a maximization problem of $-f$.

Proposition 1.2. ***(Variant 1)*** *Let f be a function from X to the poset $(Y, \leq)$, and let $E \subset D \subset X$. Suppose that $\forall\ x \in D$, $\exists\ \hat{x} \in E$ s.t. $f(\hat{x}) \geq f(x)$. Consider the following two problems:*

$$\max_{x \in X} f(x) \ \ s.t.\ x \in D$$

and

$$\max_{x \in X} f(x) \ \ s.t.\ x \in E$$

The maximizers in the two problems have the following relation

$$\arg\max_{x \in X} \left\{f(x) : x \in E\right\} = \left(\arg\max_{x \in X} \left\{f(x) : x \in D\right\}\right) \cap E$$

and if one of the two problems has a solution, then the other also has a solution. Furthermore, when the two problems have a solution, they have the same maximum.

(Variant 2)

Let f be a function from X to the totally ordered set $(Y, \leq)$. Let $D \subset X$, and x_0 be some arbitrary element of D, and define $E := \left\{x \in D : f(x) \geq f(x_0)\right\}$. Then we have

$$\arg\max_{x \in X} \left\{f(x) : x \in E\right\} = \arg\max_{x \in X} \left\{f(x) : x \in D\right\}$$

and the two problems have the same maximum if they have a solution.

Intuitively, Variant 1 of Prop.1.2 says that when we choose $x \in D$ to maximize $f(x)$, we can instead focus only on $E \subset D$ without loss of optimality, if for any alternative $x \in D$ we can find an alternative $\hat{x} \in E$ that is weakly better than x[1].Variant 2 says that if the alternative x_0 is feasible, then we can ignore all alternatives strictly worse than x_0 without loss of optimality[2].

Proposition 1.3. *Let f be a function from X to the poset $(Y, \leq)$, and let $D \subset X$. Let $\{D_\alpha\}_{\alpha \in A}$ be a family of subsets of D s.t.*

$$\bigcup_{\alpha \in A} D_\alpha = D$$

For each $\alpha \in A$, let

$$X^*_\alpha := \arg\max_{x \in X} \left\{f(x) : x \in D_\alpha\right\}$$

[1]The proposition above has an important application in mechanism design. When the principal chooses from the space of all mechanisms to maximize some objective function, the *Revelation Principle* states that any allocation that can be implemented by some mechanism can also be implemented by a direct truthful mechanism. Therefore, the proposition above implies that it is without loss of optimality for the principal to focus only on the space of direct truthful mechanism, which is a much smaller space compared to the space of all mechanisms. In this way, the maximization problem of the principal is greatly simplified.

[2]Notice that Variant 2 requires that the codomain Y is a totally ordered set, and the result does not hold if the order $\leq$ defined on the codomain is not complete.

Suppose that $X_\alpha^* \neq \emptyset$ for any $\alpha \in A$. Then

$$\begin{aligned}&\arg\max_{x\in X}\left\{f(x) : x \in D\right\}\\ &= \arg\max_{x\in X}\left\{f(x) : x \in \bigcup_{\alpha\in A} X_\alpha^*\right\}\end{aligned}$$

Intuitively, this proposition says that when we optimize f over the set D, we can partition D into pieces and optimize in each piece. Then we can collect the maximizers over each piece and compare them.

2 Results on the Set of Maximizers

The first issue about maximization problems is the existence of maximizers. We have seen Weierstrass theorem in Lecture 1, which states that a continuous real-valued function on a compact set must achieve its maximum/minimum. Let's rewrite it as the proposition below.

Proposition 2.1. *Let $f : X \to \mathbb{R}$, $D \subset X$ nonempty, and consider the maximization problem*

$$\max_{x\in X} f(x) \;\; s.t. \;\; x \in D$$

If there exists a metric d defined on the set D s.t. (D, d) is a compact metric space, and the function $f|_D$, i.e. f restricted in D, is continuous w.r.t. the metric d, then

$$\arg\max_{x\in X}\left\{f(x) : x \in D\right\} \neq \emptyset$$

i.e. the maximization problem has a solution.

In the proposition above, we use the usually defined order $\leq$ and the Euclidean distance d_2 for the codomain $\mathbb{R}$. Function f restricted in D is a new function $f|_D : D \to \mathbb{R}$ defined as $f|_D(x) = f(x)$ for any $x \in D$[3].

Weierstrass theorem provides a sufficient condition for a maximization problem to have a solution. However, sometimes we cannot directly apply Weierstrass theorem to argue that a maximization problem has a solution. For example, consider the following maximization problem:

$$\max_{(x_1,x_2)\in\mathbb{R}^2_{++}} \ln x_1 + \ln x_2 \text{ s.t. } x_1 + x_2 = 3$$

Notice that the constraint set D in this problem is

$$D := \left\{(x_1, x_2) \in \mathbb{R}^2_{++} : x_1 + x_2 = 3\right\}$$

which is not compact under the Euclidean distance d_2, since it is not closed in $\left(\mathbb{R}^2, d_2\right)$. Therefore, we cannot directly apply Weierstrass theorem, although the objective function $\ln x_1 + \ln x_2$ is continuous.

However, we can transform this problem to another problem to which Weierstrass theorem applies, using Proposition 1.2. Because $(1, 2) \in D$, and $f(1, 2) = \ln 2$, we can define

$$E := \left\{(x_1, x_2) \in \mathbb{R}^2_{++} : x_1 + x_2 = 3,\; \ln x_1 + \ln x_2 \geq \ln 2\right\}$$

[3] We distinguish between f and $f|_D$ because continuity is not defined without the metric d, while a metric is not necassary on $X \backslash D$.

By Proposition 1.2, the problem

$$\max_{(x_1,x_2)\in\mathbb{R}^2_{++}} \ln x_1 + \ln x_2 \text{ s.t. } (x_1,x_2)\in E$$

has the same set of maximizers as the original problem. It can be shown that E is compact under d_2, and so we can invoke Weierstrass theorem to argue that the maximization problem over E has a solution, and therefore, the original problem over D also has a solution.

The next issue is about the uniqueness of the maximizer, and we have the following result.

Proposition 2.2. *Let X be a set in real vector space $(V,+,\cdot)$, and let $f : X \to \mathbb{R}$. If $D \subset X$ is a convex set in V and $f|_D$ is a strictly quasi-concave function, then $\arg\max_{x\in X}\left\{f(x) : x\in D\right\}$ contains at most one point, i.e. the maximization problem has a unique maximizer if it exists.*

Proof. Suppose $x^*, x^{**} \in \arg\max_{x\in X}\left\{f(x) : x\in D\right\}$ and $x^* \neq x^{**}$. By strict quasi-concavity of f, we have

$$f(x^*) = f(x^{**}) < f\left(\frac{1}{2}x^* + \frac{1}{2}x^{**}\right)$$

Because D is a convex set, we know that $\frac{1}{2}x^* + \frac{1}{2}x^{**} \in D$, and this contradicts the assumption that $x^*, x^{**} \in \arg\max_{x\in X}\left\{f(x) : x\in D\right\}$. □

In the proposition above, if we replace strict quasi-concavity by quasi-concavity, then we don't have this uniqueness result. Instead we have the following result.

Proposition 2.3. *Let X be a set in real vector space $(V,+,\cdot)$, and let $f : X \to \mathbb{R}$. If $D \subset X$ is a convex set in V and $f|_D$ is a quasi-concave function, then $\arg\max_{x\in X}\left\{f(x) : x\in D\right\}$ is a convex set in V.*

Proof. Suppose $x^*, x^{**} \in \arg\max_{x\in X}\left\{f(x) : x\in D\right\}$. Then $\forall\lambda\in[0,1]$, $f(\lambda x^* + (1-\lambda)x^{**}) \geq \min\left\{f(x^*), f(x^{**})\right\}$ because f is quasi-concave. As the value of maximum $f(x^*)$ is unique if exists, we have $f(\lambda x^* + (1-\lambda)x^{**}) = f(x^*)$ and therefore $\lambda x^* + (1-\lambda)x^{**} \in \arg\max_{x\in X}\left\{f(x) : x\in D\right\}$ □

3 Optimization on $\mathbb{R}^n$

From now on, we focus on maximization problems for real-valued functions defined on a subset of $\mathbb{R}^n$.

First, we consider single variable functions for simplicity, and then generalize it to multivariable functions. The next theorem provides the *necessary first order condition* and the *necessary second order condition* for an interior maximizer.

Theorem 3.1. *Let X be a set in $\mathbb{R}$, and $D \subset X$. Let $f : X \to \mathbb{R}$, and consider the problem*

$$\max_{x\in X} f(x) \;\; s.t. \;\; x\in D$$

and let $x^* \in int(D)$ be a maximizer of the problem.

(1) If f is differentiable at x^*, then $f'(x^*) = 0$.

(2) If f is differentiable in an open ball around x^*, and is twice differentiable at x^*, then $f''(x^*) \leq 0$.

In the theorem above, x^* is required to be an interior point of the constraint set D w.r.t. the whole real line $(\mathbb{R}, d_2)$, instead of (X, d_2). Assuming x^* to be an interior point of D implies that x^* is also an interior point of the domain X, and so we are able to talk about the derivative of f at x^*. In (2), we require $f'(x)$ to exist in some open ball around x^*, and so we are able to talk about $f''(x^*)$.

Now let's generalize this result to multivariate functions.

Theorem 3.2. *Let X be a set in $\mathbb{R}^n$, and $D \subset X$. Let $f : X \to \mathbb{R}$, and consider the problem*

$$\max_{x \in X} f(x) \ \ s.t.\ \ x \in D$$

and let $x^* \in int(D)$ be a maximizer of the problem.

(1) If f is differentiable at x^*, then $\nabla f(x^*) = 0$.

(2) If f is differentiable in an open ball around x^*, and is twice differentiable at x^*, then $H_f(x^*)$ is negative semi-definite.

In the theorem above, again x^* is required to be an interior point of the constraint set D w.r.t. the whole Euclidean space $(\mathbb{R}^n, d_2)$, instead of (X, d_2). Assuming x^* to be an interior point of D implies that x^* is also an interior point of the domain X, and so we can talk about total/partial derivatives of f at x^*.

To maximize f, in practice we usually take partials of f and set them equal to 0, and then solve for the maximizers. Setting all partials equal to 0 is called the **(necessary) first order condition (FOC)** of the maximization problem. The theorem above implies that FOC is necessary for interior maximizers at which all partials exist.

There are two things to be careful about when using FOC. First, FOC is not a sufficient condition for an x^* to be a maximizer, i.e. an x^* at which all partials are 0 may or may not be a maximizer. Second, FOC is only necessary for interior maximizers at which all partials exist; if a maximizer x^* is on the boundary of D then FOC may or may not hold at x^*; if some partials do not exist at a maximizer x^*, then it doesn't even make sense to talk about FOC at x^*.

In practice, we solve for all solutions to FOC, and consider them as "type 1" candidates for maximizers. We also collect all points on the boundary of D and all points at which some partials do not exist, and consider them as "type 2" candidates. Then we combine the two types of candidates and examine them carefully. It is possible that the problem does not have a solution at all, in which case no candidate is a maximizer. However, if we know that the problem has a maximizer, possibly by Weierstrass theorem, then we know that it must be among the candidates we have found (Proposition 1.3). Then the maximizers are exactly those candidates that give us the highest value among all candidates.

Negative semi-definite $H_f(x^*)$ is sometimes called the **necessary second order condition (necessary SOC)** of the maximization problem. The theorem above states that necessary SOC is necessary for interior maximizers at which f is twice differentiable, and so it may help us to rule out some solutions to FOC but are not maximizers of the problem.

Negative definite $H_f(x^*)$ is sometimes called the **locally sufficient second order condition (locally sufficient SOC)** of the maximization problem, because when f is C^2 in some open ball around x^*, a negative definite $H_f(x^*)$ is sufficient for x^* to be a *strict local maximizer*, in the sense that $\exists\ \delta > 0$ s.t. $f(x^*) > f(x)$ for any $x \in B_\delta(x^*) \setminus \{x^*\}$. See FMEA Theorem 3.2.1, 2.3.2, and 1.8.1 for a proof. Clearly, a negative definite $H_f(x^*)$ is not sufficient for x^* being a (global) maximizer, since $H_f(x^*)$ only gives us local properties of the function f.

Now let's state several sufficient conditions for x^* being a (global) maximizer.

Theorem 3.3. *Let X be a convex set in $\mathbb{R}^n$, and $D \subset X$. Let $f : X \to \mathbb{R}$ be a concave function, and consider the problem*

$$\max_{x \in X} f(x) \ \ s.t. \ x \in D$$

If f is differentiable at $x^* \in int(X) \cap D$, and $\nabla f(x^*) = 0$, then x^* is a maximizer of the problem.

In the theorem above, the objective function f is assumed to be concave, which is a global property.

Proof. WTS: $f(x^*) \geq f(x)$ for any $x \in D$.

Let's show an even stronger statement: $f(x^*) \geq f(x)$ for any $x \in X$.

Suppose $\exists \ \hat{x} \in X$ s.t. $f(\hat{x}) > f(x^*)$. Clearly, we have $\hat{x} \neq x^*$. By concavity of f, we know that

$$f\left(\lambda \hat{x} + (1-\lambda) x^*\right) \geq \lambda f(\hat{x}) + (1-\lambda) f\left(x^*\right)$$

for any $\lambda \in [0,1]$. Let $z := \hat{x} - x^*$, and for any $\lambda \in (0,1]$, we have

$$\begin{aligned}
\frac{f(x^* + \lambda z) - f(x^*)}{\lambda} &= \frac{f\left(\lambda \hat{x} + (1-\lambda) x^*\right) - f(x^*)}{\lambda} \\
&\geq \frac{\lambda f(\hat{x}) + (1-\lambda) f(x^*) - f(x^*)}{\lambda} \\
&= f(\hat{x}) - f\left(x^*\right) > 0
\end{aligned}$$

Because f is differentiable at x^* and $\nabla f(x^*) = 0$, we have

$$\lim_{\lambda \to 0} \frac{f(x^* + \lambda z) - f(x^*)}{\lambda} = \frac{d}{d\lambda} f\left(x^* + \lambda z\right)\Big|_{\lambda=0} = \nabla f\left(x^*\right) \cdot z = 0$$

which contradicts

$$\frac{f(x^* + \lambda z) - f(x^*)}{\lambda} \geq f(\hat{x}) - f\left(x^*\right) > 0$$

for all $\lambda \in (0,1]$. □

If we replace the concavity assumption in the theorem above by quasi-concavity, the sufficiency result does not hold. For example, consider quasi-concave function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x^3$. Clearly $0 \in int(\mathbb{R})$ and $f'(0) = 0$, but 0 is not a maximizer on $D = \mathbb{R}$.

However, if we further assume that the function is C^2 in some open ball around x^*, and that $H_f(x^*)$ is negative definite, and we can restore the sufficiency.

Theorem 3.4. *Let X be a convex set in $\mathbb{R}^n$, and $D \subset X$. Let $f : X \to \mathbb{R}$ be a quasi-concave function, and consider the problem*

$$\max_{x \in X} f(x) \ \ s.t. \ x \in D$$

Suppose that

(1) f is differentiable at $x^* \in int(X) \cap D$, $\nabla f(x^*) = 0$, and

(2) f is C^2 in some open ball around x^*, and $H_f(x^*)$ is negative definite.

Then x^* is a maximizer of the problem.

4 Kuhn-Tucker Theorem

This section discusses the Kuhn-Tucker Theorem, which is a crucial result for constrained optimization. You may refer to FMEA Chapter 3.3 - 3.10.

First, let's define the problem we study in this section, and introduces the concept: *constraint qualification (CQ).*

Definition 4.1. *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. Consider the problem*

$$\max_{x \in X} f(x) \text{ s.t. } g(x) \geq 0 \text{ and } h(x) = 0$$

*For a feasible point $\hat{x} \in X$, the inequality constraint $g_j(x) \geq 0$ is said to be **binding at** $\hat{x}$ iff $g_j(\hat{x}) = 0$.*

*We say that the **constraint qualification (CQ) holds at** $\hat{x}$ iff the derivatives of all binding constraints*

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

*in $\mathbb{R}^n$ are linearly independent; otherwise we say that the **constraint qualification (CQ) fails at** $\hat{x}$.*

As stated in the definition above, we study the problem

$$\max_{x \in X} f(x) \text{ s.t. } x \in D$$

where the constraint set D is described by a set of k weak inequalities and a set of m equalities:

$$D := \left\{x \in X : g(x) \geq 0 \text{ and } h(x) = 0\right\}$$

In the problem, we require the domain X to be open, and the inequalities to be weak[4].

In practice, we often define the **Lagrangian function** of the maximization problem as

$$\begin{aligned} \mathcal{L}(x, \lambda, \mu) &= f(x) + \lambda^T g(x) + \mu^T h(x) \\ &= f(x) + \sum_{j=1}^k \lambda_j g_j(x) + \sum_{l=1}^m \mu_l h_l(x) \end{aligned}$$

and λ_j's and μ_l's are called the **Lagrangian multipliers**.

Now let's state Kuhn-Tucker theorem.

---

[4]If we have strict inequality in some of the constraints, then those constraints should be considered as a part of the definition of the open set X. If X is not open by the nature of the problem, then we should consider the "closed boundary" of X as a weak inequality constraint. For example, if the problem is

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

s.t.

$$\begin{cases} x_2 > 1 \\ x_1 + x_2 \leq 4 \end{cases}$$

then we should rewrite it as

$$\max_{(x_1,x_2)\in\mathbb{R}\times(1,+\infty)} x_1 x_2$$

s.t.

$$\begin{cases} x_1 \geq 0 \\ 4 - x_1 - x_2 \geq 0 \end{cases}$$

Theorem 4.2 (Kuhn-Tucker). *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. Consider the problem*

$$\max_{x \in X} f(x) \;\; s.t. \;\; g(x) \geq 0 \text{ and } h(x) = 0$$

If x^* is a maximizer of the problem above, and CQ 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):***

$$\nabla f\left(x^*\right) + \lambda^T g'\left(x^*\right) + \mu^T h'\left(x^*\right) = 0$$

(2) ***Complementary slackness condition (CSC):***

$$h_l(x^*) = 0$$

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

$$\lambda_j \geq 0, \; g_j\left(x^*\right) \geq 0, \text{ and } \lambda_j g_j\left(x^*\right) = 0$$

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

Clearly, when there is no constraint ($k = m = 0$), the problem in the theorem above becomes $\max_{x \in X} f(x)$, and the FOC reduces to $\nabla f(x^*) = 0$, which is the necessary condition we saw in the previous section for an interior maximizer x^* at which f is differentiable. Under the assumptions of this theorem, x^* is automatically an interior point of X because X is open, and f is differentiable at x^* because f is differentiable everywhere.

The FOC in the theorem above is essentially

$$\nabla f\left(x^*\right) + \sum_{j=1}^{k} \lambda_j \nabla g_j\left(x^*\right) + \sum_{l=1}^{m} \mu_l \nabla h_l\left(x^*\right) = 0$$

or equivalently, for each $i \in \{1, \ldots, n\}$

$$\frac{\partial f}{\partial x_i}\left(x^*\right) + \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i}\left(x^*\right) + \sum_{l=1}^{m} \mu_l \frac{\partial h_l}{\partial x_i}\left(x^*\right) = 0$$

which is essentially setting the partials of the Lagrangian $\mathcal{L}(x, \lambda, \mu)$ w.r.t. x_i, $i = 1, 2, ..., n$ to zero:

$$\left.\frac{d}{dx}\mathcal{L}\left(x, \lambda, \mu\right)\right|_{x = x^*} = 0$$

or equivalently, for each $i \in \{1, \ldots, n\}$

$$\frac{\partial \mathcal{L}}{\partial x_i}\left(x^*, \lambda, \mu\right) = 0$$

Simply put, Kuhn-Tucker theorem states that if x^* is a maximizer and satisfies CQ, then there exist λ and μ s.t. (x^*, λ, μ) satisfies FOC + CSC.

In practice, we often write down the following system of conditions

$$\begin{cases} x \in X \\ \frac{\partial f}{\partial x_i}(x) + \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i}(x) + \sum_{l=1}^{m} \mu_l \frac{\partial h_l}{\partial x_i}(x) = 0, \;\forall\, i = 1, \ldots n \\ h_l(x) = 0, \;\forall\, l = 1, \ldots m \\ \lambda_j \geq 0, \; g_j(x) \geq 0, \text{ and } \lambda_j g_j(x) = 0, \;\forall\, j = 1, \ldots, k \end{cases}$$

which is sometimes known as the **Kuhn-Tucker condition**. Then we can solve for all solutions (x, λ, μ) to this system. Kuhn-Tucker theorem states that if x^* is a maximizer and satisfies CQ, x^* must be a part of some solution (x, λ, μ) to the K-T condition, and therefore, we must be able to find this x^* by solving for all solutions to the K-T condition.

Notice that the theorem only works for maximizer x^*'s at which CQ holds. If CQ fails at x^*, then there may not exist (λ, μ) s.t. (x^*, λ, μ) satisfies FOC and CSC, even if x^* is a maximizer of the problem. Therefore, we may never be able to find such maximizers by solving the K-T condition. An example is given below.

Example 4.3. *Consider the problem*

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

s.t.

$$x_1^2 - x_2^3 = 0$$

Because $x_2^3 = x_1^2 \geq 0$*, and so* $x_2 \geq 0$*, and clearly the unique maximizer of this problem is* $\left(x_1^*, x_2^*\right) = (0, 0)$*. However, if we write down the Lagrangian*

$$L\left(x_1, x_2, \lambda\right) = -x_2 + \lambda\left(x_1^2 - x_2^3\right)$$

and consider the FOC

$$\begin{cases} \frac{\partial L}{\partial x_1}\left(x_1, x_2, \lambda\right) = 2\lambda x_1 = 0 \\ \frac{\partial L}{\partial x_2}\left(x_1, x_2, \lambda\right) = -1 - 3\lambda x_2^2 = 0 \end{cases}$$

clearly there exists no $\lambda \in \mathbb{R}$ *s.t.* $(0, 0, \lambda)$ *satisfies the FOC above. Therefore, we will never find the correct maximizer* $\left(x_1^*, x_2^*\right) = (0, 0)$ *by solving the FOC.*

This is not a violation of Kuhn-Tucker theorem because CQ fails at $(0, 0)$*, and so K-T theorem is silent about the maximizer* $\left(x_1^*, x_2^*\right) = (0, 0)$*. To see why CQ fails at* $(0, 0)$*, let* $h\left(x_1, x_2\right) := x_1^2 - x_2^3$ *and we have* $\nabla h\left(0, 0\right) = (0, 0)$*, which is a not linearly independent when considered as a set of only one vector.*

The right way to use K-T theorem to find the maximizers of a problem is the following:

First, we collect all x's that appears in some solution (x, λ, μ) to K-T condition, and consider them as "type 1" candidates for the maximizers. Then we collect all x's at which CQ fails, and consider them as "type 2" candidates. Then we combine the two types of candidates and examine them carefully. It is possible that the problem does not have a solution at all, in which case no candidate is a maximizer. However, if we know that the problem has a maximizer, possibly by Weierstrass theorem, then we know that it must be among the candidates we have found. Then the maximizers are exactly those candidates that give us the highest value among all candidates.

Our procedure above involves solving the K-T condition

$$\begin{cases} x \in X \\ \frac{\partial f}{\partial x_i}\left(x\right) + \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i}\left(x\right) + \sum_{l=1}^{m} \mu_l \frac{\partial h_l}{\partial x_i}\left(x\right) = 0,\ \forall\ i = 1, \ldots n \\ h_l\left(x\right) = 0,\ \forall\ l = 1, \ldots m \\ \lambda_j \geq 0,\ g_j\left(x\right) \geq 0, \text{ and } \lambda_j g_j\left(x\right) = 0,\ \forall\ j = 1, \ldots, k \end{cases}$$

and one may wonder how we can solve this system in practice. This problem is difficult in general, and we can only analytically solve this system when the functions take very simple forms. Notice that the third line implies that either $g_j\left(x\right) = 0$ or $g_j\left(x\right) > 0$,which implies $\lambda_j = 0$, and so we have two cases to discuss for each j. In total, we have 2^k cases to discuss. In each case, we have

$n+m+k$ equations and k weak inequalities for $n+m+k$ unknowns. Also, the number 2^k of cases we need to discuss increases very fast as k increases, and this is another difficulty for this kind of problems in general.

In many economic applications, we might be able to use our economic intuitions to guess which constraints are binding at optimum and which are not. You can verify your guess that the constraint $g_j(x) \geq 0$ is binding at optimum by showing that there is no solution to the K-T condition with $\lambda_j = 0$.

4.1 Applying Kuhn-Tucker Theorem: an Example

Sometimes, K-T theorem cannot be directly applied to solve a problem, and we need to analyze the problem carefully, and try to transform the original problem into another to which K-T theorem applies. Now let's carefully analyze a specific maximization problem as an example.

Consider the problem

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

s.t.

$$p_1x_1 + p_2x_2 \leq m$$

where $\alpha \in (0,1)$, $p_1, p_2 \in \mathbb{R}_{++}$, and $m \in \mathbb{R}_+$ are parameters.

The first issue we should consider is the existence of maximizers. Because power function is continuous, and the objective function $x_1^\alpha x_2^{1-\alpha}$ is a product of two power functions, and so it is continuous. If we can also show that the constraint set

$$D(p,m) := \left\{(x_1,x_2) \in \mathbb{R}^2_+ : p_1x_1 + p_2x_2 \leq m\right\}$$

is a nonempty compact set, then by Weierstrass theorem the problem must have a solution. Clearly we have $(0,0) \in D$, and so we only need to show compactness

Claim 4.4. *The constraint set $D(p,m)$ is compact in $\left(\mathbb{R}^2, d_2\right)$, for any $p \in \mathbb{R}^2_{++}$ and $m \in \mathbb{R}_+$.*

Proof. By Heine-Borel, it is sufficient to show that D is closed and bounded in $\left(\mathbb{R}^2, d_2\right)$.

(1) Closedness

Take any sequence (x^n) in D s.t. $x^n \to x^0 \in \mathbb{R}^2$. WTS: $x^0 \in D$ (Recall that what we use here is the sequential definition of closed sets.)

Because $x^n \to x^0$, we have $x_1^n \to x_1^0$, and $x_2^n \to x_2^0$. Because for each n, we have $x^n \in D$, and so $x_1^n \geq 0$ and $x_2^n \geq 0$, and so $x_1^0 \geq 0$ and $x_2^0 \geq 0$ (weak inequality is preserved under limit), and therefore $x^0 \in \mathbb{R}^2_+$. Because $p_1x_1 + p_2x_2$ is a continuous function in x, we have $p_1x_1^n + p_2x_2^n \to p_1x_1^0 + p_2x_2^0$. Because $p_1x_1^n + p_2x_2^n \leq m$ for each n, we have $p_1x_1^0 + p_2x_2^0 \leq m$. As a result, we have $x^0 \in D$.

(2) Boundedness

Take any $x \in D$, we have $p_1x_1 \leq m$ and $p_2x_2 \leq m$, and so

$$d_2(x,0) = \sqrt{x_1^2 + x_2^2} \leq \sqrt{\left(\frac{m}{p_1}\right)^2 + \left(\frac{m}{p_2}\right)^2} = m\sqrt{p_1^{-2} + p_2^{-2}}$$

Let $r := m\sqrt{p_1^{-2} + p_2^{-2}} + 1$, and we have $D \subset B_r(0)$. □

In fact, this claim is also true in $\mathbb{R}^n$, i.e. $D(p,m) := \left\{\mathbf{x} \in \mathbb{R}^n_+ : p \cdot \mathbf{x} \le m\right\}$ is compact in $(\mathbb{R}^n, d_2)$ for any $p \in \mathbb{R}^n_{++}$ and $m \in \mathbb{R}_+$.

According to the claim above, we know that the problem always has a solution by Weierstrass theorem.

The difficulty to apply K-T theorem is that the objective function has the domain $\mathbb{R}^2_+$ by its nature, because the power function z^α is only defined on $\mathbb{R}_+$ for $\alpha \in (0,1)$ in general. Clearly the domain $\mathbb{R}^2_+$ is not open in $\mathbb{R}^2$, and so it does not satisfy the assumption of K-T theorem. Although it is possible to smoothly extend the objective function $x_1^\alpha x_2^{1-\alpha}$ to $\mathbb{R}^2$, we still cannot apply K-T theorem because the objective function is not differentiable on the two axes. Therefore, we need to approach the problem in another way.

If $m = 0$, then the only feasible point is $(x_1, x_2) = (0,0)$, and the problem becomes trivial: it has a unique maximizer $(0,0)$, and the maximum is 0.

If $m > 0$, consider the feasible point $\hat{x} = \left(m/(2p_1), m/(2p_2)\right) \in \mathbb{R}^2_{++}$. At this point, the objective takes a strictly positive value. However, whenever $x_1 = 0$ or $x_2 = 0$, the objective takes the value 0. Therefore, there cannot be any maximizer on the two axes, and it is without loss of optimality to focus on the domain $\mathbb{R}^2_{++}$. Consider the new problem

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

s.t.

$$p_1 x_1 + p_2 x_2 \le m$$

This new problem has the same set of maximizers as the original problem, and so we can solve this new problem instead. To see this, take any maximizer x^* of the original problem, and then we have $f(x^*) \ge f(\hat{x}) > 0$, and so $x^* \in \mathbb{R}^2_{++}$, and so it is a maximizer of the new problem. On the other hand, take any maximizer x^* of the new problem. Because $f(x^*) \ge f(\hat{x}) > 0$, and $f(x) = 0$ for any $x \in \mathbb{R}^2_+ \backslash \mathbb{R}^2_{++}$, we know that x^* is a maximizer of the original problem.

In this new problem, the domain $\mathbb{R}^2_{++}$ is an open set in $\mathbb{R}^2$. Also, the objective function $x_1^\alpha x_2^{1-\alpha}$ is C^1 on the entire domain $\mathbb{R}^2_{++}$, and so K-T theorem applies. Define $g : \mathbb{R}^2_{++} \to \mathbb{R}$ as

$$g(x) := m - p_1 x_1 - p_2 x_2$$

and $x \in \mathbb{R}^2_{++}$, we have $\nabla g(x) = (-p_1, -p_2) \ne 0$, which is linearly independent when considered as a set of only one vector. Therefore, CQ holds at all feasible points.

Write down the Lagrangian

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

and then the K-T condition

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

By the two FOCs, we have $\lambda > 0$, and so by CSC we have $m - p_1 x_1 - p_2 x_2 = 0$. Also, comparing the two FOCs gives us

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

i.e.
$$\frac{p_1 x_1}{p_2 x_2} = \frac{\alpha}{1-\alpha}$$
and so we have
$$(x_1, x_2, \lambda) = \left(\frac{\alpha m}{p_1}, \frac{(1-\alpha)\, m}{p_2}, \frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}{p_1^{\alpha} p_2^{1-\alpha}} \right)$$
as the unique solution to the K-T condition. So $(x_1, x_2) = \big(\alpha m / p_1, (1-\alpha)\, m / p_2\big)$ is the unique type 1 candidate in this problem.

Because CQ holds at all feasible point, there is no type 2 candidate at all. Because the problem has a solution by Weierstrass, we know that the unique type 1 candidate
$$(x_1, x_2) = \left(\frac{\alpha m}{p_1}, \frac{(1-\alpha)\, m}{p_2} \right)$$
must be the unique maximizer of the problem.

This conclusion also applies to the case $m = 0$, and so we can unify the two cases. Therefore, for any $\alpha \in (0,1)$, $p_1, p_2 \in \mathbb{R}_{++}$, and $m \in \mathbb{R}_{+}$, the problem has a unique maximizer $\big(x_1^*, x_2^*\big) = \big(\alpha m / p_1, (1-\alpha)\, m / p_2\big)$.

4.2 Sufficient Conditions

The K-T theorem we have studied provides a condition that is necessary for maximizers at which CQ holds, and it is by no means a sufficient condition. However, the theorem below provides a sufficient condition for an x^* being a maximizer of a constrained maximization problem.

Theorem 4.5. *Let X be an open and convex 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. Consider the problem*
$$\max_{x \in X} f(x) \;\; s.t. \;\; g(x) \geq 0 \text{ and } h(x) = 0$$
If x^* is feasible, and there exists $(\lambda, \mu) \in \mathbb{R}_+^k \times \mathbb{R}^m$ s.t. the following three conditions hold

(1) FOC:
$$\nabla f\left(x^*\right) + \lambda^T g'\left(x^*\right) + \mu^T h'\left(x^*\right) = 0$$
(2) CSC:
$$\lambda_j \geq 0,\; g_j\left(x^*\right) \geq 0, \text{ and } \lambda_j g_j\left(x^*\right) = 0$$
for each $j \in \{1, \ldots, k\}$, and

(3) The Lagrangian $L_{\lambda,\mu} : X \to \mathbb{R}$ defined as
$$L_{\lambda,\mu}(x) := f(x) + \lambda^T g(x) + \mu^T h(x)$$
is a concave function,

then x^* is a maximizer of this problem.

In the theorem above, condition (1) and (2) are FOC and CSC in the K-T theorem. The additional requirement (3) requires the Lagrangian function to be concave in x. According to this theorem, when we solve the K-T condition for type 1 candidates, if we happen to find a solution $\left(\hat{x}, \hat{\lambda}, \hat{\mu}\right)$ to K-T condition s.t. under this $\left(\hat{\lambda}, \hat{\mu}\right)$ the Lagrangian is a concave function in x, then we can immediately conclude that $\hat{x}$ is a maximizer of the problem. But keep in mind that there might be other maximizers, since the theorem is silent about uniqueness.

As a special case of the theorem above, when there is no constraint at all, i.e. $k = m = 0$, the FOC reduces to $\nabla f(x^*) = 0$, and the concavity of the Lagrangian reduces to the concavity of the objective function f. This is consistent with Theorem 3.3.

The next theorem provides yet another sufficient condition for an x^* being a maximizer.

Theorem 4.6. *Let X be an open and convex 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, and f is concave. Consider the problem*

$$\max_{x \in X} f(x) \;\; s.t. \;\; g(x) \geq 0 \;\; and \;\; h(x) = 0$$

If x^* is feasible, and there exists $(\lambda, \mu) \in \mathbb{R}^k_+ \times \mathbb{R}^m$ s.t. the following three conditions hold

(1) FOC:

$$\nabla f\left(x^*\right) + \lambda^T g'\left(x^*\right) + \mu^T h'\left(x^*\right) = 0$$

(2) CSC:

$$\lambda_j \geq 0, \; g_j\left(x^*\right) \geq 0, \; and \; \lambda_j g_j\left(x^*\right) = 0$$

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

(3) $\lambda_j g_j$ is quasi-concave for each $j = 1, \ldots, k$, and $\mu_l h_l$ is quasi-concave for each $l = 1, \ldots, m$, then x^* is a maximizer of this problem.

This theorem requires the objective f to be concave, g_j to be quasi-concave if $\lambda_j > 0$, h_l to be quasi-concave (quasi-convex) if $\mu_l > 0$ ($\mu_l < 0$). There is no restriction on g_j (h_l) if λ_j (μ_l) is zero.

We often deal with objective functions which are quasiconcave, rather than concave. The following result gives conditions under which the Kuhn-Tucker conditions are sufficient for a maximum, when f is quasiconcave:

Theorem 4.7. *Let X be an open and convex 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, and f is quasi-concave. Consider the problem*

$$\max_{x \in X} f(x) \;\; s.t. \;\; g(x) \geq 0 \;\; and \;\; h(x) = 0$$

If x^* is feasible, and there exists $(\lambda, \mu) \in \mathbb{R}^k_+ \times \mathbb{R}^m$ s.t. the following three conditions hold

(1) FOC:

$$\nabla f\left(x^*\right) + \lambda^T g'\left(x^*\right) + \mu^T h'\left(x^*\right) = 0$$

(2) CSC:

$$\lambda_j \geq 0, \; g_j\left(x^*\right) \geq 0, \; and \; \lambda_j g_j\left(x^*\right) = 0$$

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

(3) $\nabla f(x^*) \neq 0$, $\lambda_j g_j$ is quasi-concave for each $j = 1, \ldots, k$, and $\mu_l h_l$ is quasi-concave for each $l = 1, \ldots, m$,

then x^* is a maximizer of this problem.

4.3 Comparative Statics

Let's consider the parameterized optimization problem $P(\alpha)$:

$$\max_{x \in X} f(x, \alpha) \;\; \text{s.t.} \;\; g(x, \alpha) \geq 0 \text{ and } h(x, \alpha) = 0$$

where the parameter α is taken from some set A. For each α, if the problem $P(\alpha)$ has a solution, then we can calculate the maximum value of the problem $P(\alpha)$, and define it as $f^*(\alpha)$. Then it might be interesting to study how the value function $f^*(\alpha)$ changes as the parameter α changes.

Theorem 4.8 (Envelope). *Let X be an open set in $\mathbb{R}^n$, and A be an open set of parameters in $\mathbb{R}^s$. Let $f : X \times A \to \mathbb{R}$, $g : X \times A \to \mathbb{R}^k$, and $h : X \times A \to \mathbb{R}^m$ be C^1 functions. For each parameter $\alpha \in A$, define the problem $P(\alpha)$ as*

$$\max_{x \in X} f(x, \alpha) \ \ s.t. \ \ g(x, \alpha) \geq 0 \ \ and \ \ h(x, \alpha) = 0$$

Let $\hat{A} := \{\alpha \in A : \arg\max P(\alpha) \neq \emptyset\}$, and define the value function $f^* : \hat{A} \to \mathbb{R}$ as

$$f^*(\alpha) := \max_{x \in X} \left\{ f(x, \alpha) : g(x, \alpha) \geq 0 \ \ and \ \ h(x, \alpha) = 0 \right\}$$

For parameter $\alpha^* \in A$, suppose:
(1) In the problem $P(\alpha^*)$, there is a unique maximizer x^*, and CQ holds at x^*.
(2) There exists $\varepsilon > 0$ and $r > 0$ s.t. $\forall\ \alpha \in B_\varepsilon(\alpha^*)$, $\big(\arg\max P(\alpha)\big) \cap B_r(x^*) \neq \emptyset$.
Then the value function f^* is differentiable at α^*, and

$$\begin{aligned} f^{*\prime}(\alpha^*) &= \frac{d}{d\alpha} L(x^*, \lambda^*, \mu^*, \alpha)\Big|_{\alpha=\alpha^*} \\ &= \frac{d}{d\alpha} f(x^*, \alpha)\Big|_{\alpha=\alpha^*} + \lambda^{*T} \frac{d}{d\alpha} g(x^*, \alpha)\Big|_{\alpha=\alpha^*} + \mu^{*T} \frac{d}{d\alpha} h(x^*, \alpha)\Big|_{\alpha=\alpha^*} \end{aligned}$$

where λ^* and μ^* are the unique Lagrangian multipliers found by K-T theorem for the problem $P(\alpha^*)$.

In the theorem above, condition (1) guarantees that K-T theorem applies to the problem $P(\alpha^*)$, and so we can find a unique λ^* and μ^* s.t. (x^*, λ^*, μ^*) satisfies FOC and CSC. Condition (2) implies that $f^*(\alpha)$ is well-defined for any $\alpha \in B_\varepsilon(\alpha^*)$, and so we can talk about differentiability of f^* at α^*.

The proof of this theorem is not straightforward. However, let's provide a heuristic "proof", assuming away some technical aspects of the problem. Assume that for each $\alpha \in B_\varepsilon(\alpha^*)$, we can find $x(\alpha) \in \arg\max P(\alpha)$ s.t. $x(\alpha)$ is differentiable at α^*. Also, assume that for each $\alpha \in B_\varepsilon(\alpha^*)$, in the problem $P(\alpha)$, CQ holds at $x(\alpha)$. By K-T theorem, there exists $\lambda(\alpha)$ and $\mu(\alpha)$ s.t. $\big(x(\alpha), \lambda(\alpha), \mu(\alpha)\big)$ satisfies FOC and CSC for the problem $P(\alpha)$. Assume that $\lambda(\alpha)$ and $\mu(\alpha)$ are differentiable at α^*.

By definition of f^*, we have $f^*(\alpha) = f\big(x(\alpha), \alpha\big)$ for any $\alpha \in B_\varepsilon(\alpha^*)$, and therefore

$$\begin{aligned} f^{*\prime}(\alpha^*) &= \frac{d}{d\alpha} f\big(x(\alpha), \alpha\big)\Big|_{\alpha=\alpha^*} \\ &= \frac{d}{d\alpha} \Big[f\big(x(\alpha), \alpha\big) + \lambda(\alpha)^T g\big(x(\alpha), \alpha\big) + \mu(\alpha)^T h\big(x(\alpha), \alpha\big) \Big]\Big|_{\alpha=\alpha^*} \end{aligned}$$

The second equality is because CSC implies that $\lambda(\alpha)^T g\big(x(\alpha), \alpha\big)$ and $\mu(\alpha)^T h\big(x(\alpha), \alpha\big)$ are constantly 0 for any $\alpha \in B_\varepsilon(\alpha^*)$. Then by chain rule, we have

$$\begin{aligned} f^{*\prime}(\alpha^*) = {} & \frac{d}{dx} f(x, \alpha^*)\Big|_{x=x^*} \cdot x'(\alpha^*) + \frac{d}{d\alpha} f(x^*, \alpha)\Big|_{\alpha=\alpha^*} \\ & + \lambda^{*T} \left(\frac{d}{dx} g(x, \alpha^*)\Big|_{x=x^*} \cdot x'(\alpha^*) + \frac{d}{d\alpha} g(x^*, \alpha)\Big|_{\alpha=\alpha^*} \right) \\ & + \mu^{*T} \left(\frac{d}{dx} h(x, \alpha^*)\Big|_{x=x^*} \cdot x'(\alpha^*) + \frac{d}{d\alpha} h(x^*, \alpha)\Big|_{\alpha=\alpha^*} \right) \\ & + g(x^*, \alpha^*)^T \cdot \lambda'(\alpha^*) + h(x^*, \alpha^*)^T \cdot \mu'(\alpha^*) \end{aligned}$$

By feasibility, the last term $h(x^*,\alpha^*)^T \cdot \mu'(\alpha^*) = 0 \cdot \mu'(\alpha^*) = 0$. In the second last term, if $g_j(x^*,\alpha^*) = g_j\big(x(\alpha^*),\alpha^*\big) > 0$, we have $g_j\big(x(\alpha),\alpha\big) > 0$ when α is in some open ball around α^*. By CSC, we have $\lambda_j(\alpha) = 0$ when α is in this open ball around α^*, and so $\lambda_j'(\alpha^*) = 0$. Therefore, the second last term $g(x^*,\alpha^*)^T \cdot \lambda'(\alpha^*) = 0$. Therefore, we have

$$\begin{aligned}
& f^{*\prime}(\alpha^*) \\
&= \left(\frac{d}{dx} f(x,\alpha^*)\bigg|_{x=x^*} + \lambda^{*T} \frac{d}{dx} g(x,\alpha^*)\bigg|_{x=x^*} + \mu^{*T} \frac{d}{dx} h(x,\alpha^*)\bigg|_{x=x^*} \right) \cdot x'(\alpha^*) \\
&+ \frac{d}{d\alpha} f(x^*,\alpha)\bigg|_{\alpha=\alpha^*} + \lambda^{*T} \frac{d}{d\alpha} g(x^*,\alpha)\bigg|_{\alpha=\alpha^*} + \mu^{*T} \frac{d}{d\alpha} h(x^*,\alpha)\bigg|_{\alpha=\alpha^*}
\end{aligned}$$

By FOC, we have

$$\frac{d}{dx} f(x,\alpha^*)\bigg|_{x=x^*} + \lambda^{*T} \frac{d}{dx} g(x,\alpha^*)\bigg|_{x=x^*} + \mu^{*T} \frac{d}{dx} h(x,\alpha^*)\bigg|_{x=x^*} = 0$$

and therefore we have

$$\begin{aligned}
f^{*\prime}(\alpha^*) &= \frac{d}{d\alpha} f(x^*,\alpha)\bigg|_{\alpha=\alpha^*} + \lambda^{*T} \frac{d}{d\alpha} g(x^*,\alpha)\bigg|_{\alpha=\alpha^*} + \mu^{*T} \frac{d}{d\alpha} h(x^*,\alpha)\bigg|_{\alpha=\alpha^*} \\
&= \frac{d}{d\alpha} L(x^*,\lambda^*,\mu^*,\alpha)\bigg|_{\alpha=\alpha^*}
\end{aligned}$$

which is the envelope result we want to show[5].

With the envelope theorem, we do not need to derive the value function $f^*(\alpha)$ explicitly to analyze how it responds to changes to the parameter α. The derivative of the Lagrangian are usually simpler, because in many cases the constraints are linear in the parameters (e.g. the budget constraint is linear in endowment and prices).

Now let's verify the equation in the envelope theorem for the eample given in Section 4.4. The solution to the Kuhn-Tucker conditions is given as

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

The value function of the maximization problem is

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

Taking its first order derivative w.r.t. (p_1,p_2,m) we have:

$$\frac{\partial v}{\partial p_1} = -\frac{m\alpha^{1+\alpha}(1-\alpha)^{1-\alpha}}{p_1^{\alpha+1} p_2^{1-\alpha}} = -\lambda^* x_1^*$$

$$\frac{\partial v}{\partial p_2} = -\frac{m\alpha^{\alpha}(1-\alpha)^{2-\alpha}}{p_1^{\alpha} p_2^{2-\alpha}} = -\lambda^* x_2^*$$

[5] In the heuristic calculation above, we assumed that $x(\alpha)$, $\lambda(\alpha)$, and $\mu(\alpha)$ are all differentiable at α^*. However, the result is still true even when this assumption does not hold.

$$\frac{\partial v}{\partial m} = \frac{\alpha^\alpha (1-\alpha)^{1-\alpha}}{p_1^\alpha p_2^{1-\alpha}} = \lambda^*$$

From the envelope theorem, we have:
$\frac{\partial \mathcal{L}}{\partial p_1}(x_1^*, x_2^*, \lambda^*, p_1, p_2, m) = -\lambda^* x_1^*$,
$\frac{\partial \mathcal{L}}{\partial p_2}(x_1^*, x_2^*, \lambda^*, p_1, p_2, m) = -\lambda^* x_2^*$,
$\frac{\partial \mathcal{L}}{\partial m}(x_1^*, x_2^*, \lambda^*, p_1, p_2, m) = \lambda^*$.

4.4 Interpretation of Lagrangian Multipliers*

We can use envelope theorem to obtain an interpretation of the Lagrangian multipliers.

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. Consider the parameterized problem $P\left(a, b\right)$

$$\max_{x \in X} f\left(x\right)$$

s.t.

$$\begin{cases} g\left(x\right) + a \geq 0 \\ h\left(x\right) + b = 0 \end{cases}$$

where $(a, b) \in \mathbb{R}^k \times \mathbb{R}^m$ are parameters. If the problem $P\left(a, b\right)$ has a solution, define $f^*\left(a, b\right)$ as the maximum value of the problem $P\left(a, b\right)$.

When we move (a, b) around $(a^*, b^*) = (0, 0)$, we are considering perturbations around the original problem $P\left(0, 0\right)$

$$\max_{x \in X} f\left(x\right)$$

s.t.

$$\begin{cases} g\left(x\right) \geq 0 \\ h\left(x\right) = 0 \end{cases}$$

A small positive a_j can be viewed as a slight relaxation of the constraint $g_j\left(x\right) \geq 0$, which might make the feasible set slightly larger, which in turn might make the maximum value slightly higher. We are interested in how such a slight relaxation of the constraint $g_j\left(x\right) \geq 0$ will affect the maximum value, i.e. we are interested in the partial derivative $\frac{\partial f^*}{\partial a_j}\left(0, 0\right)$.

If in the original problem $P\left(0, 0\right)$ there is a unique maximizer x^*, CQ holds at x^*, and $\exists\, \varepsilon > 0$ and $r > 0$ s.t. $\forall\ (a, b) \in B_\varepsilon\left(0, 0\right)$, $\exists\ x \in \big(\arg\max P\left(a, b\right)\big) \cap B_r\left(x^*\right)$, then we can invoke the envelope theorem at $(a^*, b^*) = (0, 0)$, and we have

$$\begin{aligned}
&\frac{d}{d\left(a, b\right)} f^*\left(a, b\right)\Big|_{(a,b)=(0,0)} \\
&= \frac{d}{d\left(a, b\right)} f\left(x^*\right)\Big|_{(a,b)=(0,0)} + \lambda^{*T}\ \frac{d}{d\left(a, b\right)} \Big(g\left(x^*\right) + a\Big)\Big|_{(a,b)=(0,0)} \\
&+ \mu^{*T}\ \frac{d}{d\left(a, b\right)} \Big(h\left(x^*\right) + b\Big)\Big|_{(a,b)=(0,0)} \\
&= 0 + \lambda^{*T} \cdot \left[I_k | 0_{k\times m}\right] + \mu^{*T} \cdot \left[0_{m\times k} | I_m\right] \\
&= \left(\lambda_1, \ldots, \lambda_k, \mu_1, \ldots \mu_m\right)
\end{aligned}$$

Therefore, we have

$$\frac{\partial f^*}{\partial a_j}\left(0, 0\right) = \lambda_j$$

for each $j = 1, \ldots, k$, and

$$\frac{\partial f^*}{\partial b_l}(0,0) = \mu_l$$

for each $l = 1, \ldots, m$.

Therefore, the Lagrangian multiplier λ_j corresponding to the inequality constraint $g_j(x) \geq 0$ measures the marginal increase in the maximum value under a marginal relaxation of the constraint $g_j(x) \geq 0$. As a consequence, λ_j is sometimes called the **shadow price of the constraint** $g_j(x) \geq 0$.

In a firm's maximization problem, the objective function is usually the firms profit function, and a constraint $g_j(x) \geq 0$ usually represents the requirement that total usage of some resource (labor/capital/electricity/...) is weakly less than the total amount of this resource available to the firm. Then λ_j can be called the **shadow price of this resource** (labor/capital/electricity/...), and by envelope theorem, it measures the marginal increase in profit by marginally increasing the total amount of this resource available to the firm. In other words, λ_j is the price the firm is willing to pay for an additional unit of this resource.

By K-T theorem, the Lagrangian multiplier λ_j corresponding to the weak inequality constraint $g_j(x) \geq 0$ is required to be nonnegative. This is consistent with our interpretation of λ_j as the marginal gain by slightly relaxing the constraint $g_j(x) \geq 0$, because a relaxation of a constraint never decreases the maximum value. Also, CSC in K-T theorem states that if the constraint $g_j(x) \geq 0$ is not binding at optimum, i.e. $g_j(x^*) > 0$, where x^* is the unique maximizer, then we must have $\lambda_j = 0$, i.e. we will gain nothing by slightly relaxing the constraint. On the other hand, if there is a strictly positive marginal gain by slightly relaxing the constraint $g_j(x) \geq 0$, i.e. $\lambda_j > 0$, then there is no reason not to fully exploit the constraint in the optimization, i.e. the constraint must be binding at optimum.

Notice that CSC only requires at least one of λ_j and $g_j(x)$ is zero, and in fact they could be both zero. In other words, it is possible for some constraint to be binding, while slightly relaxing this constraint does not increase the maximum. For example, consider the problem

$$\max_{x \in \mathbb{R}} -x^2 \text{ s.t. } x \geq 0$$

in which case both the Lagrangian multiplier is 0 and the constraint is binding.

The Lagrangian multiplier μ_l corresponding to the equality constraint $h_l(x) = 0$ measures the marginal change in the maximum value under a marginal perturbation of the equality constraint $h_l(x) = 0$. By its nature, it may be positive or negative, which is consistent with the assumption on μ_l in K-T theorem.

5 A Brief Introduction to Dynamic Programming

Consider the infinite horizon inequality-constrained maximization problem (one-sector optimal growth problem):

$$\begin{aligned}
\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} & \sum_{t=0}^{\infty} \beta^t U(c_t) \\
s.t.\ & c_t + k_{t+1} \leq f(k_t) \\
& c_t, k_{t+1} \geq 0, t = 0, 1, ... \\
& k_0 > 0 \ \ given.
\end{aligned}$$

where U and f are strictly increasing. Here the choice variables are a sequence $\{c_t, k_{t+1}\}_{t=0}^{\infty}$. Since k_0 is a parameter, we can find the value function $v : \mathbb{R}_+ \to \mathbb{R}$ which gives the maximized value of the object function given k_0:

$$\begin{aligned} v(k_0) = \max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} & \sum_{t=0}^{\infty} \beta^t U(c_t) \\ s.t.\ & c_t + k_{t+1} \leq f(k_t) \\ & c_t, k_{t+1} \geq 0, t = 0, 1, ... \\ & k_0 > 0 \ \ given. \end{aligned}$$

Since the problem is time-independent, $v(k_1)$ would be the maximized value of the object function if the program starts in period $t+1$. $\beta v(k_1)$ is this value discounted at period $t = 0$. So we can write the problem in $t = 0$ as

$$\begin{aligned} & \max_{c_0, k_1} U(c_0) + \beta v(k_1) \\ s.t.\ & c_0 + k_1 \leq f(k_0) \\ & c_0 \geq 0, k_1 \geq 0 \\ & k_0 > 0 \ \ given. \end{aligned}$$

By definition of v we substitute out c_0:

$$v(k_0) = \max_{0 \leq k1 \leq f(k_0)} U\big(f(k_0) - k_1\big) + \beta v(k_1)$$

Since subscript does not matter, we can write

$$v(k) = \max_{0 \leq y \leq f(k)} U\big(f(k) - y\big) + \beta v(y)$$

Now the unknown is not a variable but the function v. We call this a **functional equation**.

Theorem 5.1 (Blackwell's sufficient condition for contraction)**.** *Let $X \subset \mathbb{R}^k$ and $B(X)$ a real vector space of bounded functions $f : X \to \mathbb{R}$, with norm defined as $||f|| = \sup_{x \in X} |f(x)|$. Let $T : B(X) \to B(X)$ be an operator satisfying*

(1) (monotonicity) If $f, g \in B(X)$ and $f(x) \leq g(x)$ for $\forall x \in X$, then $(Tf)(x) \leq (Tg)(x)$ for $\forall x \in X$.

(2) (discounting) There exists some $\beta \in (0, 1)$ s.t.

$$(T(f + a))(x) \leq (Tf)(x) + \beta a, for\ \forall f \in B(X), a \geq 0, x \in X,$$

where $(f + a)$ is defined as $(f + a)(x) = f(x) + a$.

Then T is a contraction with modulus β.

In dynamic programming, Blackwell's sufficient conditions are often easy to verify. In the problem above, we define an operator T as

$$(Tv)(k) = \max_{0 \leq y \leq f(k)} U\big(f(k) - y\big) + \beta v(y)$$

We want to find a fixed point of operator T, i.e. a function v s.t. $Tv = v$. We can verify that T is a contraction using Blackwell, and by contraction mapping theorem we know that such a fixed point exists. To find it we iteratively apply T, starting with an arbitrary function v_0 until convergence (under certain specified convergence criteria).