Optimization - MA Math Camp 2023

Andrea Ciccarone

August 30, 2023

General Setup

Maximization Problem

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

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

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 $\max_{x \in X} \{f(x) : x \in D\}$

Do we always have a maximizer?

The maximum does not need to exist in general, i.e the set of maximizers can be empty. Consider for example the function :

$$f : (0,1) \to (0,1)$$
$$x \mapsto x$$

This function does not have a maximum because for every point $x \in (0,1)$ I can find a point $x' \in (0,1)$ such that $f(x') > f(x)$ (by moving arbitrarily close to 1). In other words, the set $f((0,1))$ does not have a maximum.

(Still) on Maximizers

- By anti-simmetry of the partial order on Y, the maximum - if it exists - is a well defined (unique) element of Y

$$\max_{x \in D} f(x) \in Y$$

- The set of maximizers, by contrast, is a *subset* of D in general :

$$\arg\max_{x \in D} f(x) \subset D$$

- If $(Y, \leq)$ has the least upper bound property, then we know that the following supremum always exist :

$$\sup_{x \in D} f(x).$$

Then, the question of whether a maximum exists can be interpreted as whether this supremum is *attained* by a point in D.

Types of Optimization Problems

- "Unconstrained optimization": D is an open set of a metric space.
- "Optimization under equality constraint": D is of the form $D = \{x \in X, \forall i \in I, g_i(x) = 0\}$, where $g_i : X \to \mathbb{R}$ for all $i \in I$.
- "Optimization under inequality constraint": D is of the form $D = \{x \in X, \forall i \in I, g_i(x) \leq 0\}$, where $g_i : X \to \mathbb{R}$ for all $i \in I$.

- We say that $x_0 \in D$ is a *global maximum* if $f(x_0) \geq f(x)$ for all $x \in D$.
- We say that $x_0 \in D$ is a *local maximum* if there exists a neighborhood of x_0 in D such that $f(x_0) \geq f(x)$ for all x in this neighborhood.

Examples

1. **Choice Problem**

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

2. **Best Response**

$$br_i(a_{-i}) = \arg\max_{a_i \in A} u_i(a_i, a_{-i})$$

3. **Least Squares**

$$\min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} |y_i - \beta x_i|^2$$

Existence of Maximizers

The first issue about maximization problems is the existence of maximizers. Remember that Weierstrass theorem states that a continuous real-valued function on a compact set must achieve its maximum/minimum.

Weierstrass

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

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

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

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

i.e. the maximization problem has a solution.

Uniqueness

- Once we establish existence, we can think about uniqueness
- Use (quasi) concavity $\rightarrow$ strict quasi concavity implies uniqueness, as with two maxima we could take a convex combination and do better.
- Formally:

Proposition 2.1 - Uniqueness

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

Uniqueness (ctd.)

- 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 $\arg\max_{x \in X} \{f(x) : x \in D\}$ is a convex set in V.

- These are just guidelines... showing existence and uniqueness of maximizers will require ad-hoc strategies

- We now focus on functions $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$
- Unconstrained optimization problem: the set over which the optimization problem is defined is an open set in $\mathbb{R}^n$
- We first consider single variable functions.
- The next theorem provides the *necessary first order condition* and the *necessary second order condition* for an interior maximizer.
- The result does not require that D is open but restricts the attention to interior maximizers which is equivalent (we ignore maximizers on the boundary).

FOC and SOC

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

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

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 $f''(x^*) \leq 0$.

FOC and SOC

- Openness plays such a crucial role: we can consider derivatives at every point in the interior, which gives us information about local variations of the function
- These are **necessary** conditions for x^* to be a maximizer and equivalently characterize local maximizers (which is a necessary condition to be a global maximizer)

FOC and SOC for 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

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

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.

FOC

- To maximize f, in practice, we take partials of f and set them equal to 0. This is the **(necessary) first order condition (FOC)** of the maximization problem.
- Suppose x is a maximizer and take Taylor approximation:

$$f(x+h) = f(x) + \nabla f(x) \cdot h + o(||h||)$$

hence $\nabla f(x) \cdot h \leq 0$ for h small enough as $f(x) \geq f(x+h)$

- Now, set $h = t\nabla f(x)$ for t small enough:

$$\nabla f(x) \cdot (t\nabla f(x)) = t||\nabla f(x)||^2 \leq 0$$

Which is only possible if $\nabla f(x) = 0$.

SOC

- Negative semi-definite $H_f(x^*)$ is sometimes called the **necessary second order condition (necessary SOC)** of the maximization problem.
- Suppose x is a candidate maximizer, so that $\nabla f(x) = 0$:

$$\begin{aligned} f(x+h) &= f(x) + \nabla f(x) \cdot h + \frac{1}{2} h^T \mathcal{H}_f(x) h + o(||h||^2) \\ &= f(x) + \frac{1}{2} h^T \mathcal{H}_f(x) h + o(||h||^2) \end{aligned}$$

hence when h small :

$$f(x+h) - f(x) \approx \frac{1}{2} h^T \mathcal{H}_f(x) h$$

so if $\mathcal{H}_f(x)$ is negative semi-definite, $f(x+h) - f(x) \leq 0$ for all h small enough, i.e, $f(x') \leq f(x)$ for x' in some small ball around x, in other words x is a local maximum.

Summarizing

- Partials equal to 0 are a necessary (not sufficient condition)
- The second derivative can give local sufficient conditions

General recipe:

- Solve for all solutions to FOC. If there are many, SOC can help filter
- Find the point with the highest value
- Check if the function takes on a higher value along the boundary

Concave Functions Are Cool!

- Concave functions really simplify our job...

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

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

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

- When our function is concave, having partials equal to 0 is a sufficient condition!

Quasi-Concave Functions Are Also Cool (but not quite)!

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

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

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

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.

Optimization under Equality Constraints

- We now know to deal with interior points: this provided us with a method to deal with open sets and some non-open sets by considering "individually" all boundary and non-differentiability points as candidates.
- What if we have "a lot" of boundary points?
- We consider **equality constraints**: all admissible points are boundary points so we cannot use the interior characterization directly.

Optimization under Equality Constraints

- Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^k$, with $g(x) = (g_1(x), ..., g_k(x))$ and $c = (c_1, ..., c_k) \in \mathbb{R}^k$. The set $\{x \in \mathbb{R}^n, g(x) = c\}$ is called *a level set* of g, and is pinned down by the choice of the constant c. We consider the problem of optimizing f on a level set of g :

$$\max_{x \in \{x, g(x) = c\}} f(x)$$

which we rewrite equivalently in the constrained form :

$$\max_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t. } g(x) = c$$

Where $g(x) = c$ explicitly rewrites as $g_i(x) = c_i$ for all i :

$$\begin{cases} g_1(x) & = c_1 \\ \vdots & \\ g_k(x) & = c_k \end{cases}$$

Parametrization

- Sometimes we can rewrite the level set as a parametrized region, where the parameter belongs to an open set, so we can rewrite the whole problem as an unconstrained problem.
- For example:

$$\max_{c_1,c_2} \ln c_1 + \alpha \ln c_2$$

$$\text{s.t. } p_1 c_1 + p_2 c_2 = M$$

- We can write:

$$c_1 = \frac{M - p_2 c_2}{p_1}$$

and solve

$$\max_{c_2} \ln \frac{M - p_2 c_2}{p_1} + \alpha \ln c_2$$

Lagrange

- Sometimes we can't parametrize, but we can develop tools to deal with equality constraints directly when the functions f and g are differentiable.

Theorem 4.2 - 1 constraint

Let $f, g : D \subset \mathbb{R}^n \to \mathbb{R}$ and $x^* \in int(D)$. If x^* is a local extremum of f under the constraint $g = c$, if f is differentiable at x^*, g is differentiable in a neighborhood of x^* and if $\nabla g(x^*) \neq 0$, then there exists $\lambda \in \mathbb{R}$ such that :

$$\nabla f(x^*) = \lambda \nabla g(x^*)$$

λ is called the Lagrange multiplier associated to the constraint.

Interpretation of Lagrange

- At an extremum, the gradient of the objective function must be *colinear* to the gradient of the constraint
- We can see with one constraint that the gradient of f at an optimum has to be orthogonal to the line tangent to the space $\{g(x) = c\}$ at that point.
- This captures the idea that "otherwise, we could move a little bit while staying in the constraint space and improve f".

Lagrangian function

- Define the *Lagrangian of the problem* as the function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ such that :

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

- The previous theorem rewrites as follows : if x^* is an extremum of f under the constraint $g = c$, then there exists λ such that (x^*, λ) is a critical point of $\mathcal{L}$, i.e :

$$\nabla\mathcal{L}(x^*, \lambda) = 0 \Leftrightarrow \begin{cases} \frac{\partial\mathcal{L}}{\partial x}(x^*, \lambda) = 0 \\ \frac{\partial\mathcal{L}}{\partial\lambda}(x^*, \lambda) = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \nabla f(x^*) - \lambda\nabla g(x^*) = 0 \\ g(x^*) - c = 0 \end{cases}$$

Example

- An example from your PS: Find the maximum and minimum of $f(x, y) = x^2 - y^2$ on the unit circle $x^2 + y^2 = 1$ using the Kuhn-Tucker method. Using the substitution $y^2 = 1 - x^2$ solve the same problem as a single variable unconstrained problem. Do you get the same results? Why or why not?

Extending to more than one constraint

Theorem 4.3

Let $f, g_1, ..., g_k : D \subset \mathbb{R}^n \to \mathbb{R}$ and $c = (c_1, ..., c_k) \in \mathbb{R}^k$. If $x^* \in int(D)$ is a local extremum of f under the constraints $g_i = c_i$ for all i and if

1. f is differentiable at x^*
2. g is C^1 in a neighborhood of x^*
3. the family $(\nabla g_1(x^*), ..., \nabla g_k(x^*))$ is independent

then there exists $(\lambda_1, ..., \lambda_k) \in \mathbb{R}^k$ such that :

$$\nabla f(x^*) = \sum_{i=1}^{k} \lambda_i \nabla g_i(x^*)$$

Extending to more than one constraint

- The Lagrangian with several constraints is defined as :

$$\mathcal{L}(x,\lambda) = f(x) - \sum_{i=1}^{k} \lambda_i (g_i(x) - c_i) = f(x) - \lambda \cdot (g(x) - c)$$

- If x^* is an extremum of f under the constraint $g = c$, then there exists $\lambda \in \mathbb{R}^k$ such that (x^*, λ) is a critical point of $\mathcal{L}$, i.e :

$$\begin{aligned} \nabla \mathcal{L}(x^*,\lambda) = 0 &\Leftrightarrow \begin{cases} \frac{\partial \mathcal{L}}{\partial x_j}(x^*,\lambda) = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_i}(x^*,\lambda) = 0 \end{cases} \\ &\Leftrightarrow \begin{cases} \nabla f(x^*) - \sum_{i=1}^{k} \lambda_i \nabla g_i(x^*) = 0 \\ g_i(x^*) - c_i = 0 \quad \forall i \end{cases} \end{aligned}$$

Inequality Constraints: KKT

- Inequality constraints are more complex in nature because they might not bind : if we have a constraint $g(x) \leq c$, and it turns out that at the optimum $g(x) < c$, then x is essentially an interior point and it is "as if" the constraint was not there locally.
- On the other if we saturate the constraint, i.e $g(x) = c$ at the optimum, then we need a machinery similar to that we just introduced to deal with an extra equality constraint – given the admissible directions of increase are locally reduced.

Constraint Qualification

Definition 5.1

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

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

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

The Problem

- We are studying 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:

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

- We define the Lagrangian function 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**.

KKT

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

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

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)**:

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

KKT

Kuhn-Tucker

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

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

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

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

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

KKT

- 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 $i \in \{1, \ldots, n\}$

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

KKT

- 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:

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

- 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.

KKT

- 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**.

CQ failure

- 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.
- Example:

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

s.t.

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

KKT Example

- We are going to apply KT theorem by solving this problem together:

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

s.t.

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

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

Sufficient Conditions

- The K-T provides a condition that is necessary for maximizers at which CQ holds, and it is by no means a sufficient condition.
- However...

Sufficiency KKT

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

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
(2) CSC
(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.

Sufficiency (ctd.)

- 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.
- There are other theorems for sufficiency (all involving quasi-concavity and or concavity) - see Th 5.6, 5.7 on the lecture notes

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.

Envelope Thm

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

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

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

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

Envelope

Envelope

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. $\forall\ \alpha \in B_\varepsilon(\alpha^*)$, $(\arg\max P(\alpha)) \cap B_r(x^*) \neq \emptyset$.
Then the value function f^* is differentiable at α^*, and

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

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

What??

- (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 theorem is basically saying that instead of deriving the value function and then compute derivative, we can simply take the derivative the Lagrangian wrt α at the optimum - this is often simpler

That's it!

Last slide of math camp! Thank you guys!!