# Math Camp - Optimization

# KKT conditions

Many of the problems we want to solve in economics can't be described with just equality constraints

- Consumer problem

$$\max_{x \in \mathbb{R}^m_+} u(x)$$

$$\text{s.t. } p \cdot x \leq m$$

- Cost minimization

$$\min_{(k,l) \in \mathbb{R}^2_+} rk + wl$$

$$\text{s.t. } f(k,l) \geq \bar{q}$$

# KKT conditions

$f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$.

$$\max f(x)$$
$$\text{s.t. } g(x) \leq 0$$

How can we solve this?

- Similar intuition as in older problems, want to make sure function doesn't increase in any direction we can move in.
- Now we can move in a lot more directions.
- Mechanically, a natural way to try to solve this:
  1. Assume some constraints held with equality.
  2. Ignore all other constraints.
  3. Solve the relaxed problem. If the solution is feasible, keep it.
  4. Do this for all combinations of constraints and then compare the feasible solutions.

# KKT conditions

**Definition (Karush-Kuhn-Tucker conditions)**

We say a point $x \in \mathbb{R}^n$ and a multiplier $\lambda \in \mathbb{R}^m$ satisfy the Karush-Kuhn-Tucker conditions (KKT) for a maximum if

$$
\begin{aligned}
Df(x) &= \lambda' Dg(x) \\
g_i(x) &\leq 0 \text{ for all } i \in \{1, \dots m\} \\
\lambda_i g_i(x) &= 0 \text{ for all } i \in \{1, \dots m\} \\
\lambda_i &\geq 0 \text{ for all } i \in \{1, \dots m\}
\end{aligned}
$$

# KKT conditions

There are four conditions here

$$Df(x) = \lambda' Dg(x)$$
$$g_i(x) \leq 0 \text{ for all } i \in \{1, \dots m\}$$

What we expect, the lagrange multiplier conditions + the constraints.

$$\lambda_i g_i(x) = 0 \text{ for all } i \in \{1, \dots m\}$$

This is called called complementary slackness, says either a constraint holds with equality or we can ignore it.

$$\lambda_i \geq 0 \text{ for all } i \in \{1, \dots m\}$$

The fourth condition is the extra bit of structure we get from being able to move in more directions.

- This requires something beyond our Lagrange multiplier argument. Can use the separating hyperplane theorem to show it.

# KKT conditions

Like before, we could set up the lagrangian $L(x, \lambda) = f(x) - \lambda g(x)$.

- Complementary slackness implies $f(x^*) = L(x^*, \lambda^*)$.
- There are second order conditions for this, they are annoying.
- We would like a thm that says: A point is a local max iff it satisfies the KKT conditions.
- Neither direction is true in general unfortunately

# KKT conditions-Example

$$\max xy$$
$$\text{s.t. } x^2 + y^2 \leq 1$$

Our FOCs are

$$y = \lambda 2x$$
$$x = \lambda 2y$$

- We know from earlier, $x = \pm\frac{1}{\sqrt{2}}$, $y = \pm x$ solves these conditions for some $\lambda$ and are feasible. Verify that the one's we identified as maximum's before have positive multipliers.
- Complementary slackness: Either $x^2 + y^2 = 1$ or $\lambda = 0$. If $\lambda = 0$, $x, y = 0$
- $x = y = 0$ is not a local max, even though it satisfies the KKT conditions.

# KKT conditions-Example

$$
\begin{aligned}
&\max xyz + z \\
\text{s.t.}\, &x^2 + y^2 + z \leq 6 \\
&x \geq 0 \\
&y \geq 0 \\
&z \geq 0
\end{aligned}
$$

# Example

The first order conditions are

$$\begin{aligned} yz &= \lambda_1 2x - \lambda_2 \\ xz &= \lambda_1 2y - \lambda_3 \\ xy + 1 &= \lambda_1 - \lambda_4 \end{aligned}$$

Complementary slackness

$$\lambda_1(x^2 + y^2 + z - 6) = 0$$
$$\lambda_2 x = 0,\ \lambda_3 y = 0\ \lambda_4 z = 0$$

And

$$\begin{aligned} x^2 + y^2 + z &\leq 6 \\ x, y, z, \lambda &\geq 0 \end{aligned}$$

# Example

First observe $\lambda_1 > 0$. If $\lambda_1 = 0$, then

$$xy + 1 = -\lambda_4$$

but the LHS is positive and the RHS is negative.

Are there critical points where the non-negativity constraints bind?

# Example

Recall the FOCs

$$\begin{aligned} yz &= \lambda_1 2x - \lambda_2 \\ xz &= \lambda_1 2y - \lambda_3 \\ xy + 1 &= \lambda_1 - \lambda_4 \end{aligned}$$

If $x = 0$, then by the second FOC

$$y > 0 \Rightarrow \lambda_3 > 0$$

so $y = 0$. Thus $z = 6$, $\lambda = (1, 0, 0, 0)$.

# Example

Finally, assume all three non-negativity constraints are slack, $\lambda_2 = \lambda_3 = \lambda_4 = 0$.

- $yz = 2\lambda_1 x$ and $xz = 2\lambda_1 y$ so $x = y$.
- We then know

$$z = 6 - 2x^2$$
$$z = 6 - 2(\lambda_1 - 1)$$

And from the FOC wrt to $x$

$$8 - 2\lambda_1 = 2\lambda_1$$

So $x = 1, y = 1, z = 4,\ \lambda = (2, 0, 0, 0)$ is a critical point.

# Example

$$\max_{x,y\in\mathbb{R}_+} \ x$$
$$\text{s.t. } \ y-(1-x)^3\leq 0$$

If you just graph this, the max is clearly $(1,0)$. But the FOC wrt to $x$ is

$$1=-3\lambda(1-x)^2.$$

which cannot be satisfied at $(1,0)$. Relates to the full-rank thing we had with Lagrange multipliers, note the derivative matrix of the binding constraints at $(1,0)$ is

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

# Necessity

The natural condition here, that the derivatives of binding constraints are linearly independent seems really annoying to check.

**Theorem (KKT - Necessary)**

*The KKT conditions hold at a maximizer $x^*$ if rank $Dg^*(x^*) = m^*$, where $m^*$ is the number of constraints that hold with equality and $g^*$ is the vector of $g$'s where $g(x^*) = 0$.*

Roughly, more generally, we say a maximizer satisfies constraint qualification if the KKT conditions hold at a maximizer. Some other conditions that work

- (Slater condition) $f$ concave, each $g_i$ is convex and there exists $x$ s.t. $g_i(x) < 0$ for all $i \in \{1, \ldots m\}$.
- $g(x) = Ax + b$ for some matrix $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$

Note: unlike the rank condition, these do not require us to "know" the maximum.

# KKT-Sufficiency

**Theorem (KKT sufficiency)**

*Suppose $\nabla f(x) \neq 0$ for all feasible $x$, $f$ quasiconcave, $g_i$ quasiconvex for all $i \in \{1, 2, \ldots m\}$. Then any point satisfying the KKT conditions is a global max.*

Proof:

- Fix a $(x^*, \lambda^*)$ that satisfies the KKT conditions. For any $y$

  $$\nabla f(x^*) \cdot (y - x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) \cdot (y - x^*) = 0$$

  By quasiconvexity

  $$\lambda_i^* \nabla g_i(x^*) \cdot (y - x) \leq 0$$

  since for any feasible $y$, if $g_i$ binds at $x^*$, $g_i(y) \leq g_i(x^*)$ and $g$'s are quasiconvex.
- So $\nabla f(x^*)(y - x^*) \leq 0$, and thus $f(x^*) \geq f(y)$ by quasiconcavity.

# Value functions

Think about the consumer problem:

$$\max u(x)$$
$$\text{s.t. } p \cdot x \leq m$$

This has parameters $(p, m)$. Two natural objects to think about

- $v(p, m)$ - The value of $u(x)$ at the max.
- $x(p, m)$ - The choice that maximizes $u(x)$.

More generally, $X \subset \mathbb{R}^n$, $\Theta \subset \mathbb{R}^c$. Let $f : X \times \Theta \to \mathbb{R}$ and $C : \Theta \to 2^X \setminus \{\emptyset\}$.

$$V(\theta) = \max_{x \in X} f(x;\theta)$$
$$\text{s.t. } x \in C(\theta)$$

and let $\chi(\theta)$ be the corresponding arg max.

- $V(\theta)$ is the value function
  - Note: As long as the max exists for all $\theta$, it is indeed a function.
- $\chi(\theta)$ is the policy correspondence.

# Correspondences

A correspondence is set-valued function. Both our constraints and the arg max are correspondences.

For instance, if $(x, y)$ is how many apples and bananas I buy, then the function that takes my income and the price of apples and bananas and tells me what combinations I can afford to purchase is in general going to spit out a set.

$$B(p, M) = \{(x, y) : p_1 x + p_2 y \leq m\}$$

We could define continuity directly on functions $f : \mathbb{R} \to 2^{\mathbb{R}}$ by, for instance, defining a distance between sets and then using our $\epsilon - \delta$ definition, but this seems hard (and maybe not what we want).

# Correspondences

Instead, let's think about it more heuristically. What might we want?

# Correspondences

**Definition (Upper Hemicontinuity)**

A compact-valued correspondence $\Gamma : A \rightrightarrows B$ where $B$ is compact is upper hemicontinuous iff it has a closed graph: $\forall x_n \in A$, $y_n \in \Gamma(x_n)$

$$\lim_{n\to\infty} x_n = x,\ \lim_{n\to\infty} y_n = y \Rightarrow y \in \Gamma(x)$$

**Definition (Lower Hemicontinuity)**

$\Gamma : A \rightrightarrows B$ is lower hemicontinuous iff for all $x \in X$, $x_n \to x$ and for all $y \in \Gamma(x)$, there exists a subsequence $x_{n_k}$ a sequence $y_{n_k} \in \Gamma(x_{n_k})$ such that $y_{n_k} \to y$

# UHC and LHC

- Upper-hemicontinuity preserves the property of continuous functions where sequences in the domain and the corresponding sequence in the range converge, then the limit in the range is the function evaluated at the limit in the domain.
- Lower hemicontinuity preserves the property that any point in the range can be approximated by the value of the function at nearby points.

# UHC and LHC

- A correspondence can be either, neither or both.
- If the correspondence is single valued, then both are equivalent and are equivalent to continuity.
- Very sloppily speaking, for non-empty convex + compact valued correspondences in $\mathbb{R}$:
  - If UHC you can draw a the upper + lower boundaries without lifting your pen.
  - If LHC then you can take a slice out of the correspondence that is a continuous function.
- We say a correspondence is continuous if it is UHC and LHC.

# Maximum theorem

A natural question to ask about the value function and policy correspondence is what sort of continuity properties do they inherit.

**Theorem (Berge's Maximum Theorem)**

*Let $\Theta \subseteq \mathbb{R}^m$, $X \subseteq \mathbb{R}^n$ be nonempty and compact, $f : X \times \Theta \to \mathbb{R}$ be continuous, and $C : \Theta \rightrightarrows X$ be a compact valued, continuous correspondence. Then*

- *The value function $V(\theta)$ is continuous.*
- *The policy correspondence $\chi(\theta)$ is non-empty, compact-valued and UHC.*

# Maximum Theorem

Proof:

- $V$ and $\chi$ exist and $\chi$ is non-empty is immediate from Weierstrauss theorem.

# Maximum Theorem

Proof:

- $V$ and $\chi$ exist and $\chi$ is non-empty is immediate from Weierstrauss theorem.
- $\chi(\theta)$ is closed (and thus compact).

# Maximum Theorem

Proof:

- $V$ and $\chi$ exist and $\chi$ is non-empty is immediate from Weierstrauss theorem.
- $\chi(\theta)$ is closed (and thus compact).
  - Take any convergent sequence $x_n \in \chi(\theta)$, $x_n \to x$.

# Maximum Theorem

Proof:

- $V$ and $\chi$ exist and $\chi$ is non-empty is immediate from Weierstrauss theorem.
- $\chi(\theta)$ is closed (and thus compact).
  - Take any convergent sequence $x_n \in \chi(\theta)$, $x_n \to x$.
  - Since $C(\theta)$ is compact, $x \in C(\theta)$.

# Maximum Theorem

Proof:

- $V$ and $\chi$ exist and $\chi$ is non-empty is immediate from Weierstrauss theorem.
- $\chi(\theta)$ is closed (and thus compact).
  - Take any convergent sequence $x_n \in \chi(\theta)$, $x_n \to x$.
  - Since $C(\theta)$ is compact, $x \in C(\theta)$.
  - $f(x_n, \theta) = V(\theta)$ so $f(x, \theta) = V(\theta)$.

# Maximum Theorem

Proof:

- $V$ and $\chi$ exist and $\chi$ is non-empty is immediate from Weierstrauss theorem.
- $\chi(\theta)$ is closed (and thus compact).
  - Take any convergent sequence $x_n \in \chi(\theta)$, $x_n \to x$.
  - Since $C(\theta)$ is compact, $x \in C(\theta)$.
  - $f(x_n, \theta) = V(\theta)$ so $f(x, \theta) = V(\theta)$.
- $\chi(\theta)$ is uhc.

# Maximum Theorem

Proof:

- $V$ and $\chi$ exist and $\chi$ is non-empty is immediate from Weierstrauss theorem.
- $\chi(\theta)$ is closed (and thus compact).
  - Take any convergent sequence $x_n \in \chi(\theta)$, $x_n \to x$.
  - Since $C(\theta)$ is compact, $x \in C(\theta)$.
  - $f(x_n, \theta) = V(\theta)$ so $f(x, \theta) = V(\theta)$.
- $\chi(\theta)$ is uhc.
  - Take a sequence $\theta_n \to \theta$, $x_n \to x$, $x_n \in \chi(\theta_n)$.

# Maximum Theorem

Proof:

- $V$ and $\chi$ exist and $\chi$ is non-empty is immediate from Weierstrauss theorem.
- $\chi(\theta)$ is closed (and thus compact).
  - Take any convergent sequence $x_n \in \chi(\theta)$, $x_n \to x$.
  - Since $C(\theta)$ is compact, $x \in C(\theta)$.
  - $f(x_n, \theta) = V(\theta)$ so $f(x, \theta) = V(\theta)$.
- $\chi(\theta)$ is uhc.
  - Take a sequence $\theta_n \to \theta$, $x_n \to x$, $x_n \in \chi(\theta_n)$.
  - By UHC, $x \in C(\theta)$.

# Maximum Theorem

Proof:

- $V$ and $\chi$ exist and $\chi$ is non-empty is immediate from Weierstrauss theorem.
- $\chi(\theta)$ is closed (and thus compact).
  - Take any convergent sequence $x_n \in \chi(\theta)$, $x_n \to x$.
  - Since $C(\theta)$ is compact, $x \in C(\theta)$.
  - $f(x_n, \theta) = V(\theta)$ so $f(x, \theta) = V(\theta)$.
- $\chi(\theta)$ is uhc.
  - Take a sequence $\theta_n \to \theta$, $x_n \to x$, $x_n \in \chi(\theta_n)$.
  - By UHC, $x \in C(\theta)$.
  - Take any $z \in C(\theta)$, there exists $z_n \in C(\theta_n)$, $z_n \to z$ by LHC.

# Maximum Theorem

Proof:

- $V$ and $\chi$ exist and $\chi$ is non-empty is immediate from Weierstrauss theorem.
- $\chi(\theta)$ is closed (and thus compact).
  - Take any convergent sequence $x_n \in \chi(\theta)$, $x_n \to x$.
  - Since $C(\theta)$ is compact, $x \in C(\theta)$.
  - $f(x_n, \theta) = V(\theta)$ so $f(x, \theta) = V(\theta)$.
- $\chi(\theta)$ is uhc.
  - Take a sequence $\theta_n \to \theta$, $x_n \to x$, $x_n \in \chi(\theta_n)$.
  - By UHC, $x \in C(\theta)$.
  - Take any $z \in C(\theta)$, there exists $z_n \in C(\theta_n)$, $z_n \to z$ by LHC.
  - By continuity, $f(x_n; \theta_n) \geq f(z_n; \theta_n) \Rightarrow f(x; \theta) \geq f(z; \theta)$.

# Maximum theorem

- $V(\theta)$ is continuous.

# Maximum theorem

- $V(\theta)$ is continuous.
  - Take a sequence $\theta_n \to \theta$ and a $x_n \in \chi(\theta_n)$, $x \in \chi(\theta)$.
  - Suppose $V(\theta_n)$ does not converge to $V(\theta)$. Then there is a subsequence $\theta_{n_k}, x_{n_k}$ s.t. $|V(\theta_{n_k}) - f(x;\theta)| \geq \epsilon$ for some $\epsilon > 0$.

# Maximum theorem

- $V(\theta)$ is continuous.
  - Take a sequence $\theta_n \to \theta$ and a $x_n \in \chi(\theta_n)$, $x \in \chi(\theta)$.
  - Suppose $V(\theta_n)$ does not converge to $V(\theta)$. Then there is a subsequence $\theta_{n_k}, x_{n_k}$ s.t. $|V(\theta_{n_k}) - f(x;\theta)| \geq \epsilon$ for some $\epsilon > 0$.
  - But $x_{n_k}$ has a further subsequence that converges to some $x \in \chi(\theta)$ by UHC.

# Maximum theorem

- $V(\theta)$ is continuous.
  - Take a sequence $\theta_n \to \theta$ and a $x_n \in \chi(\theta_n)$, $x \in \chi(\theta)$.
  - Suppose $V(\theta_n)$ does not converge to $V(\theta)$. Then there is a subsequence $\theta_{n_k}, x_{n_k}$ s.t. $|V(\theta_{n_k}) - f(x;\theta)| \geq \epsilon$ for some $\epsilon > 0$.
  - But $x_{n_k}$ has a further subsequence that converges to some $x \in \chi(\theta)$ by UHC.
  - Therefore $f(x;\theta) = \lim_{j\to\infty} f(x_{n_{kj}};\theta_{n_{kj}}) = V(\theta)$, which is a contradiction.

# Concavity

We can use concavity to strengthen these theorems

### Corollary

*If, in addition to the conditions of the maximum theorem, $C$ is convex-valued and $x \mapsto f(x,\theta)$ is strictly quasi-concave then $\chi$ is single-valued, and is thus a continuous function.*

### Theorem (Fancy Concave Maximum Theorem)

*Let $\Theta \subseteq \mathbb{R}^m$, $X \subseteq \mathbb{R}^n$ be nonempty and compact. $f : X \times \Theta \to \mathbb{R}$ be continuous, quasiconcave, and $C : \Theta \rightrightarrows X$ is a compact valued, continuous correspondence with convex graph. Then*

- *The value function $V(\theta)$ is continuous and quasiconcave.*
- *The policy correspondence $\chi(\theta)$ is non-empty, compact-valued, convex valued and UHC.*

# Concavity

The only thing we haven't already shown here is that $V$ is quasiconcave.

- Fix $\theta_1, \theta_2$, $x_1 \in \chi(\theta_1)$, $x_2 \in \chi(\theta_2)$.
- For any $\lambda \in (0,1)$, $\lambda x_1 + (1-\lambda)x_2 \in C(\lambda\theta_1 + (1-\lambda)\theta_2)$. Thus

$$V(\theta_\lambda) \geq f(x_\lambda; \theta_\lambda) \geq \min(f(x_1, \theta_2), f(x_2, \theta_2)) = \min(V(\theta_1), V(\theta_2))$$

The same logic shows that $V$ inherits concavity, strict concavity and strict quasiconcavity from $f$ under the conditions of the theroem.

# Envelope Theorem

Consider an optimization problem, parameterized by some $\theta \in \Theta \subseteq \mathbb{R}$, $\Theta$ open.

$$\max_{x \in \mathbb{R}} f(x;\theta)$$

Suppose this has a unique solution for each $\theta$, given by single valued and differentiable function $\chi(\theta)$. Then

$$\begin{aligned} V(\theta) &= f(\chi(\theta);\theta) \\ \frac{dV(\theta)}{d\theta} &= f_x(\chi(\theta);\theta)\chi'(\theta) + f_\theta(\chi(\theta);\theta) \\ &= f_\theta(\chi(\theta);\theta) \end{aligned}$$

# Envelope theorem

We can do the same things for constrained optimization

$$\max f(x;\theta)$$
$$\text{s.t. } g(x;\theta) = 0$$

Suppose we know that $\chi$ is single valued and differentiable. Then

$$V(\theta) = f(\chi(\theta);\theta)$$
$$DV(\theta) = D_x f(\chi(\theta);\theta) D\chi(\theta) + D_\theta f(\chi(\theta);\theta)$$

We also know

$$D_x f(\chi(\theta);\theta) = \lambda' D_x g(\chi(\theta);\theta)$$

Finally, since $g(\chi(\theta);\theta) = 0$,

$$D_x g(\chi(\theta);\theta) D\chi(\theta) = -D_\theta g(\chi(\theta);\theta)$$

So

$$DV(\theta) = -\lambda' D_\theta g(\chi(\theta);\theta) + D_\theta f(\chi(\theta);\theta)$$

This is called the envelope theorem.

# Envelope Theorem

**Theorem (Envelope theorem)**

*Suppose $f, g, \chi(\theta)$ are continuously differentiable and $Dg(\chi(\theta), \theta)$ has full rank. Then $V$ is differentiable and*

$$DV(\theta) = -\lambda' D_\theta g(\chi(\theta); \theta) + D_\theta f(\chi(\theta); \theta)$$

This theorem having conditions on $\chi$ is a bit annoying.

- We want some sort of concavity to get $\chi$ to be a function.
- We could assume more differentiability and apply the implicit function theorem to the first order conditions to get differentiability.
- Alternatively, establishing concavity/convexity of the policy function or monotonicity almost be enough.

# Envelope Theorem

This turns out to be surprisingly useful.

- It gives us a convenient tool for comparative statics.
- Gives multipliers economic meaning.

# Consumer problem

Consider the consumer problem:

$$\max u(x)$$
$$\text{s.t. } p \cdot x = m$$

Suppose this gives us single valued, differentiable demands $x(p, m)$ and value function $v(p, m)$. The envelope theorem tells us

$$\frac{\partial v(p, m)}{\partial m} = \lambda$$

the multiplier is the shadow price of income, what you lose from a decrease in wealth. Similarly

$$\frac{\partial v(p, m)}{\partial p_i} = -\lambda x_i(p, m)$$

# Cost minimization

Recall the firm's cost minimization problem

$$\min_{k,l \in \mathbb{R}^+} \; rk + wl$$
$$\text{s.t. } f(k,l) \geq \bar{q}$$

Verify for yourself that if $f(\cdot)$ is strictly concave, the policy correspondence (input demand) is single valued.

If input demand is differentiable then we know the derivative of the cost function $C(r,w,q)$ is

$$\frac{\partial C}{\partial r} = k(r,w,q)$$
$$\frac{\partial C}{\partial w} = l(r,w,q)$$

So if we know $(r,w,q)$ and a firm's costs, we can back out input demand.

# Fixed Points

We are often going to run into situations where it's unclear whether a solution to our model even exists.

For instance, are there prices where supply equals demand. Fixed point theorems give us a powerful tool to deal with this.

- A function $f : X \to X$ has a fixed point if there exists some $x \in X$ s.t. $f(x) = x$.
- Think about any continuous function $f : [0, 1] \to [0, 1]$. Does it have a fixed point.

# Fixed Points

Graphically, continuous functions look like a smooth curve with no holes, jumps, etc. So in $\mathbb{R}$, relatively easy to find fixed points of continuous functions.

### Theorem (Intermediate Value Theorem)

*Suppose $f : [a, b] \to \mathbb{R}$ is continuous. Then if $y$ lies between $f(a), f(b)$ (e.g. $f(a) \leq y \leq f(b)$) then there exists a $c \in [a, b]$ such that $f(c) = y$*

Tells us when, for instance, we can find a point where curves intersect, places where the function crosses 0, etc.

# Fixed Points

**Theorem (Brouwer's Fixed Point Theorem)**

*Every continuous function from a convex, compact subset $X$ of $\mathbb{R}^n$ to itself has a fixed point.*

**Theorem (Kakutani's Fixed Point Theorem)**

*Every non-empty, convex valued, UHC correspondence from a non-empty, convex and compact subset $X$ of $\mathbb{R}^n$ to itself has a fixed point.*

Think about what happens if you relax any of the conditions.