Optimization

Daniel N. Hauser

August 29th, 2022

This note summarizes some of the key concepts from the fourth set of lectures.

Economic "applications" of KKT conditions (beyond supply and demand)

1. Designing a wage scheme to incentivize unobservable effort.
2. Designing the optimal redistributive tax scheme
3. Supervised learning models in machine learning (e.g. SVMs)

KKT conditions

Now we'd like to solve problems of the form

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

for $f : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}^n$. We can leverage the inequality constraints here to get a fancier version of the Lagrange multiplier conditions

Definition (Karush-Kuhn-Tucker Conditions)**.** *An $x \in \mathbb{R}^m$ and a vector $\lambda \in \mathbb{R}^n$ satisfies the KKT conditions if:*

$$\nabla f(x)^T = \lambda^T Dg(x) \qquad \textit{(First Order Conditions)}$$
$$\lambda g(x)^T = 0 \qquad \textit{(Complementary Slackness)}$$
$$g(x) \leq 0 \qquad \textit{(Feasibility)}$$
$$\lambda \geq 0 \qquad \textit{(Positivity of the Multiplier)}.$$

A couple things to note about these conditions

- The multiplier effective "turns on constraints." Complementary slackness means that any g_i whose associated multiplier is non-zero must hold with equality, i.e. $\lambda_i > 0 \Rightarrow g_i(x) = 0$.
- The sign of the multiplier is important. We can rule out any feasible points that can only satisfy these gradient condition but don't have the right sign for the multipliers. Be careful with the directions of the inequalities!
- Minimizers satisfy the same conditions, but $\lambda \leq 0$.

Like with Lagrange multipliers, these constraints are necessary at any maximizer that satisfies a rank condition on Dg. Fortunately, we can find much simpler conditions for concave problems

Theorem 1. *The KKT conditions are* necessary *if f is concave, each g_i is convex and there exists an x s.t. $g_i(x) < 0$ for all i.*

The KKT conditions are sufficient *if $\nabla f(x) \neq 0$ for all feasible x, f is quasiconcave, and g_i's are quasiconvex.*

So for nice enough problems, the KKT conditions exactly identify the set of global maximizers.[1]

In general these conditions are going to be annoying (or more likely impossible) to solve without careful thought. When approach-

[1] Note that, unlike with equality constraints, the conditions for maxima and minima are different, so there was no hope that these conditions on their own could identify maximizers. Here the sign of the multiplier in well-behaved problems sidesteps the need to worry about things like second order conditions.

ing a maximization problem, before going to this step, think carefully about the problem. Are there any constraints that clearly bind (hold with equality) or clearly don't. This can make looking for solutions to the KKT conditions much more manageable. On the other hand, even if you can't explicitly solve for the maximizer, these conditions can still tell us a lot about its properties.

Correspondences

A *correspondence* is a function that maps from each element in the domain to a set of elements in the range. We consider two different continuity notions that each capture some aspects of what it meant for a function to be continuous.

Definition. *A compact valued correspondence $\Gamma : A \rightrightarrows B$ with B compact is said to be* upper hemicontinuous *iff it has a closed graph.*

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

Value functions

Let $\Theta \subseteq \mathbb{R}^m$, $X \subseteq \mathbb{R}^n$, and consider a function continuous $f : X \to \Theta \to \mathbb{R}$ and a correspondence $C : \Theta \rightrightarrows X$. We call the object

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

the value function. This function describes the maximum for different values of the exogenous parameters (the elements of θ). The corresponding arg max we call the *policy correspondence.* The value function and policy correspondence inherit some nice properties from f and c

Unsurprisingly we use value functions, policy correspondence and the envelope theorem a lot. Supply, compensated and uncompensated demand, input demand, etc. are all policy functions, while the expenditure, profit, indirect utility, and cost functions are all value functions. It will also be a surprisingly convenient technical tool:

- Dynamic maximization problems an be written recursively using the value function.
- In mechanism design, when trying to design a pricing scheme that induces specific decisions from consumers, the envelope theorem lets us characterize how decision makers respond to different menus.

Theorem 2 (Berge's Maximum Theorem)**.** *Let Θ, X be non-empty and compact, f and C are compact valued and continuous. Then the value function is continuous and the policy correspondence is non-empty, compact-valued and UHC.*

Similarly the value function and policy correspondence inherit concavity/convexity properties. We can also describe the derivative of the value function in terms of the primatives. Consider a problem of the form

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

with single valued policy function $\phi(\theta)$. Then the change in the value function with respect to the exogenous parameters is entirely determined by how f and g change with respect to the exogenous parameters.

Theorem 3 (The Envelope Theorem)**.** *Suppose f, g, ϕ are continuously differentiable and $Dg(\phi(\theta), \theta)$ has full rank. Then the value function V is differentiable and*

$$DV(\theta) = -\lambda^T D_\theta g(\phi(\theta);\theta) + D_\theta f(\phi(\theta), \theta)$$

Fixed Point Theorems

We'll use these theorems to establish existence of Nash and Walrasian equilibrium.

A problem we'll run into a lot is establishing that a system of non-linear equations has a solution. Occasionally the separating hyperplane theorem will be enough, but often we'll have to apply a fixed point theorem.

Theorem 4 (Brouwer's Fixed Point Theorem)**.** *Every continuous function that maps from a non-empty convex, compact subset of $\mathbb{R}^n$ to itself has a fixed point.*

Theorem 5 (Kakutani's Fixed Point Theorem)**.** *Every non-empty, convex valued, UHC correspondence from a non-empty, convex, compact subset of $\mathbb{R}^n$ to itself has a fixed point.*