3 Optimization and Comparative Statics

Motivation:

- We know a set of conditions for a solution to exist: for example, for any upper-semicontinuous function defined on a compact set, it must have a solution.
- Now we are interested in the tools that allow us to find them.

How do we arrive at a sufficient condition for a maxima?

- Gradient must be zero, and the Hessian must be negative.

Application: Profit Maximization

- Let F be the firm's production function that is twice differentiable. Also assume no corner solutions.
- The firm's problem is to maximize

$$pF(L, K) - wL - rK$$

  for which we use the first-order condition to get "marginal productivity of labor (capital) must be the real wage (rental rate) at the optimum."

- It is at this point that we assume constant returns to scale: $F(rL, rK) = rF(L, K)$. Also known as *homogeneous* functions. It comes with nice properties that makes the firm profit equal to zero!

3.1 First-order Conditions for Constrained Optimization

A typical optimization problem looks like this:

$$\begin{aligned} &\max_{\mathbf{x} \in R^m} f(\mathbf{x}) \text{ s.t.} \\ &\qquad g_j(\mathbf{x}) \leq b_j, \forall j \in \{1, ..., m\} \\ &\qquad h_l(\mathbf{x}) = c_l, \forall l \in \{1, ..., k\} \end{aligned}$$

To solve something like this, we invoke the famous Lagrange method. Essentially, the necessary conditions for an optimal solution, in words, are:

1. First-order conditions for the Lagrangian must hold.
2. The optimal solution must satisfy the equality constraints.
3. The optimal solution must satisfy the inequality constraints.

4. We should be moving in the right direction of the maximum (i.e. Lagrange multipliers are of the correct sign).

5. Complementary slackness must hold.

   - Addendum: at an optimal solution, if a shadow price (dual variable) is positive, meaning that the objective function could be increased if the corresponding primal constraint was relaxed, then this primal constraint must be tight. If not, the primal objective function value could be improved (by changing the primal variables in order to make this non-binding primal constraint binding).

Note that the Lagrangian takes the following form:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := f(\mathbf{x}) - \sum_{j=1}^{m} \lambda_j (g_j(\mathbf{x}) - b_j) - \sum_{l=1}^{k} \mu_j (h_l(\mathbf{x}) - c_l)$$

It's convenient to interpret the constraints as being "costs" and the associated multipliers the price you pay for violating them. Then it's trivial to remember its form.

Furthermore, also note that the gradient vector of the objective function must be parallel to the gradient vector of the constraint functions:

- Recall that you shouldn't be able to perturb an optimal solution to get a better solution while maintaining the constraint, to the first order.

- See **here** for a simple explanation.

3.2 Second-order Conditions for Constrained Optimization

You need to check for second-order conditions in order to verify that the optimum is a maximum or a minimum. This check involves the construction of a bordered Hessian matrix: Often we would like to get around the need for second-order conditions via the quasiconcavity of the utility function. If we have enough structures, then there is no need to explicitly check for the second order conditions.

3.3 Into the Realm of Convexity

Why is convexity useful?

- Dealing with extreme points

- Separating and supporting hyperplane theorems - these are used everywhere!

Notion of convexity applies to both sets and functions:

- Set $S \subset X$ where X is a linear space is convex if any convex combination of the elements is in the set.

- A real valued function $f : X \to \mathbb{R}$ is convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

Convex functions come with nice properties:

- For convex $f : (a,b) \to \mathbb{R}$, it follows from the Chordal Slope Lemma that $|f'(x^-)| < \infty$ and $|f'(x^+)| < \infty$ for all $x \in (a,b)$.

- For convex $f : (a,b) \to \mathbb{R}$, f is Lipschitz (and therefore absolutely continuous) for any compact subinterval in its domain.

- By the above result, for any convex function on the interval $[a,b]$, the Fundamental Theorem of Calculus holds:

$$f(x) = f(a^+) + \int_a^x f(z)dz$$

- Convex functions are differentiable except at countably many points, since its derivative is increasing.

- Let $f : (a,b) \to \mathbb{R}$ be a real-valued function. f is convex if and only if it is a pointwise supremum of a class of affine functions.

Extreme points, those that cannot be written as a convex combination of two distinct points in a convex set, are important for many reasons:

- If S is a convex and compact subset of a normed linear space, then there exist an extreme point.

- *Caratheodory's Theorem:* In an n-dimensional Euclidean space, any point in the convex hull of a set S can be represented by a convex combination of at most $n+1$ points.

- *Krein-Milman Theorem:* Any convex and compact subset in a normed linear space can (almost) be described by it's extreme points.

Combining these together yields the Bauer Maximum Principle:

- The Bauer maximum principle implies that whenever the objective function is convex and upper-semicontinuous and the feasible set is compact and convex, there is at least one extreme point of the feasible set that is the solution of the maximization problem.

3.4 It's a bird... It's a plane... It's a hyperplane!

Here's the motivation, in case you are not motivated yet:

- The notion of "domination" is ubiquitous in economics. One way to characterize such concept is to find a tangent plane on the feasible set and regard the alternatives as if they are solutions to some constraint optimization problems that we are familiar with.

First, the *Separating Hyperplane Theorem:*

- For any two convex sets in a normed linear space such that one of them has nonempty interior, one can always find a hyperplane separating these two sets.
- This hyperplane is expressed as $x \in X | L(x) = \alpha$ for some linear functional L.

Second, the *Supporting Hyperplane Theorem:*

- When a convex set has a non-empty interior, for any point on the boundary of that set, there exists a hyperplane tangent to that point.

We now note an interesting result from linear algebra: for all linear functionals $L : \mathbb{R}^n \to \mathbb{R}$, there exists a unique vector $\mathbf{w} \in \mathbb{R}^n$ such that $L(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$. This is known as *duality*. And therefore we can replace the linear functional with a vector in the above two definitions.

Here's an application for characterizing Pareto optimal allocations. By using the supporting hyperplane theorem, it is possible to reduce the problem to finding solutions to a family of optimization problems:

- The punchline is that looking for an efficient allocation is equivalent to looking for the optimal allocation for a hypothetical social planner whose objective is some weighted sum of the individuals in the economy. This is an incredible result.
- Note that this characterization then implies the *second welfare theorem*, which states that any efficient allocation can be supported by an equilibrium under some exchange economy.

3.5 Duality

Duality theorem is another way to think about constrained optimization problems. For the original problem of:

$$\max_{\mathbf{x} \in R^m} f(\mathbf{x}) \text{ s.t.}$$
$$g_j(\mathbf{x}) \geq 0, \forall j \in \{1, ..., m\}$$
$$h_l(\mathbf{x}) = 0, \forall l \in \{1, ..., k\}$$

and denote the optimal value to be p^*. The dual problem is defined to be:

$$\inf_{\boldsymbol{\lambda}\in\mathbb{R}^m_+,\boldsymbol{\mu}\in\mathbb{R}^k_+} q(\boldsymbol{\lambda},\boldsymbol{\mu})$$

where

$$q(\boldsymbol{\lambda},\boldsymbol{\mu}) := \sup_{\mathbf{x}\in\mathbb{R}^n} \left[f(\mathbf{x}) + \boldsymbol{\lambda}^T g(\mathbf{x}) + \boldsymbol{\mu} h(\mathbf{x})\right]$$

and denote d^* to be the optimal value of this optimization problem. One important result is the *Weak Duality Theorem*, which states that $d^* \geq p^*$. The difference is referred to as the duality gap.

The next obvious question is when the duality gap becomes zero. If it is zero, then life is rosier since the objective of the dual problem is now convex by construction. This is known as *Slater's Condition*:

- If f is quasi concave, the feasible set is convex, and strictly feasible ($\exists \mathbf{x} \in \mathbb{R}^n$ such that $g_j(\mathbf{x}) > 0.\forall j \in \{1, ..., n\}$), then the duality gap is zero.

To see an application, consider the following two problems. First is the Marshallian problem, where we maximize the utility:

$$\max_{\mathbf{x}\in\mathbb{R}^n} u(\mathbf{x}) \text{ s.t. } \mathbf{p}^T\mathbf{x} \leq m$$

and the second is the Hicksian problem, in which we minimize the expenditure:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{p}^T\mathbf{x} \text{ s.t. } u(\mathbf{x}) \geq \bar{u}$$

We have the Slutsky's Identity that shows the problems are the same.

3.6 Application: Information Design

The spirit is to design an informational structure - incentive structure - for players who are rational and interact with each other so that it induces behavior desirable to the designer. But instead of designing rules (mechanism design), here we design an information structure.

In mechanism design, we are trying to affect the consequence of actions. Here, the designer is trying to affect the origin of your actions, since the players are rational and Bayesian updaters, creating a proper information structure allows to affect the players' beliefs and thereby lead to a subsequent course of actions.

The setup is rather quite simple:

- There is a Sender (S) and a Receiver (R).
- State of the world: Ω, $|\Omega| < \infty$

- R can take action $a \in A, |A| < \infty$
- Common prior: $p \in \Delta(\Omega), p(\omega) > 0, \forall \omega \in \Omega$
  - They know what the probability distribution on the possible states of the world
- Payoffs: $u_S : A \times \Omega \to \mathbb{R}$, $u_R : A \times \Omega \to \mathbb{R}$
- Information structure: (S, π) where $|S| < \infty$ and $\pi : \Omega \to \Delta(S)$ is a conditional probability system. Essentially, $\pi(s|\omega)$ is the probability of signal s given the state ω.

An observation: given $(S, \pi), \forall s \in S$, by Bayes' rule, the receiver's posterior is

$$q_S(\omega) = \frac{\pi(s|\omega)p(\omega)}{\sum_{\omega' \in \Omega} \pi(s|\omega')p(\omega')}$$

Let $\bar{\pi}(s) = \sum_{\omega \in \Omega} \pi(s|\omega)p(\omega)$.

- Upon seeing $s \in S$, receiver solves:

$$\max_{a \in A} \sum_{\omega \in \Omega} u_R(a, \omega) q_s(\omega)$$

- Let $a^*(s) \in \arg\max_{a \in A} \sum_{\omega \in \Omega} u_R(a, \omega) q_s(\omega)$ be the sender-preferred action.
- Sender's expected payoff given (S, π) is then

$$\sum_{\omega \in \Omega} \sum_{s \in S} u_s(a^*(s), \omega) \pi(s|\omega) p(\omega)$$

  which the sender presumably seeks to maximize.

Note that the receiver's posterior is a random variable from an ex ante perspective. If we take the average of the receiver's posterior for all $\omega \in \Omega$

$$\sum_{s \in S} q_s(\omega) \bar{\pi}(s) = \cdots = p(\omega)$$

which implies that if you have a distribution of posteriors whose mean is the same as the prior, then it can be induced by an information structure.

Now, because the receiver's action only depends on the posterior, we can further denote $a^*(s)$ as $a^*(q)$. And thus the sender's problem is now simplified to:

$$\sup_{\tau \in \Delta(\Delta(\Omega))} \int_{\Delta(\Delta(\Omega))} v(q) \tau(dq)$$

subject to

$$\int_{\Delta(\Delta(\omega))} q \tau(dq) = p$$

where $v(q) := \sum_{\omega \in \Omega} u_s(a^*(q), \omega) q(\omega)$ which is the sender's expected payoff when the receiver's posterior is q. Note that τ is a probability distribution over different posteriors, while π is a conditional distribution system over s and ω.

We introduce a concept of *concavification* where we take the convex hull of a function and take its supremum for any given q. Using this result, we can arrive at the important result that

$$W(p) = V(p)$$

meaning $W(p)$ is the value of the sender's problem is the concavified function $V(p)$. The outline of the proof is as follows:

- Consider the dual of the original sender's constrained maximization problem, which is equal to minimizing $D(\lambda)$. By weak duality, $\inf_{\lambda \in \mathbb{R}^{|\Omega|}} D(\lambda) \geq W(p)$
- Thus, if we find some distribution $\tau^* \in \Delta(\Delta(\Omega))$ that is feasible under the primal problem, and some Lagrange multiplier λ such that τ^* solves the inner maximization problem of the dual, then this value must be greater than or equal to $W(p)$. Because $W(p)$ is the value of the primal problem and τ^* is a feasible solution, it must be the case that this value must be also less than or equal to $W(p)$. THus, at τ^*, it attains the value.
- By construction, $\mathcal{V}$ is convex, and also because v is upper semicontinuous, for any point on the $V(p)$ it must be on the boundary of $\mathcal{V}$. We have a convex set with a non-empty interior and a point on the boundary of this convex set, so we use the supporting hyperplane theorem: we can find a vector such that the inner product with the point on the boundary is greater than the inner product with any other point on this convex set.

3.7 Envelope Theorem

This is examining how solutions change as the environment changes. Obviously, comparative statics is a useful tool for exploring causality. Note:

- Exogenous variables = variables that are taken as given when solving the model
- Endogenous variables = optimal solutions, equilibria

One interesting question is how the value function would change if we alter one of the parameters. The result is known as the *Envelope Theorem*.

- Consider the maximization problem:

  $$V(\theta) = \max_{x \in [a,b]} f(x, \theta)$$

  assuming f is concave, continuous with a unique solution $x^*(\theta)$.

- If you take the derivative at the optimal point, then we only need to consider the direct effect of the changing the parameter on the objective and ignore the indirect effect induced by optimal choices.

- Alternatively put, To sum up, the envelope theorem states that the derivative of the optimal value as a function of the parameter is exactly the partial derivative of the objective with respect to the parameter, evaluated at the optimal choice under the parameter.

As a corollary, we also have an analogous result for constrained optimization problems. This result neatly applies to the two problems that we visited before:

- Incentive Compatibility

- Slutsky's Identity: First fix the indifference curve and analyze; then change the income and analyze.

3.8 Implicit Function Theorem

Instead of asking how the value function changes, let's ask how the solution would change. Consider a simple exampl where we want to maximize $f(x, \theta) \in C^2$ where $\theta \in [0, 1]$ and is concave in x. Also assume that $x^*(\theta)$ is in (a, b) and x^* is differentiable. By the first order conditions,

$$\begin{aligned} f_1(x^*(\theta), \theta) &= 0, \forall \theta \\ \Rightarrow f_{11}(x^*(\theta), \theta)x^{*\prime}(\theta) + f_{12}(x^*(\theta), \theta) &= 0 \\ x^{*\prime}(\theta) &= -\frac{f_{12}(x^*(\theta), \theta)}{f_{11}(x^*(\theta), \theta)} \end{aligned}$$

Given our assumption of concavity of f, it suffices to find out the sign of f_{12} to figure out the direction of the change in solution.

Now the general implicit function theorem, explained in non-mathematical terms:

- The implicit function theorem really just boils down to this: if I can write down m (sufficiently nice!) equations in $n+m$ variables, then, near any sufficiently nice solution point, there is a function of n variables which give me the remaining m coordinates of nearby solution points.

- In other words, I can, in principle, solve those equations and get the last m variables in terms of the first n variables. But (!) in general this function is only valid on some small set and won't give you all the solutions either.