CHAPTER 6: Static Optimization

1 Parametric Optimization Problems in $\mathbb{R}^n$

Let $f : X \times \Theta \longrightarrow \mathbb{R}$ be a given function, where $X \subseteq \mathbb{R}^n$ and $\Theta \subseteq \mathbb{R}^m$. A constrained maximization problem is written as

$$\max_x \{f(x;\, \theta) \mid x \in C(\theta)\}. \qquad (P_1)$$

That is, given some value of θ, we look for the value of x that maximizes the function $f(\cdot;\, \theta)$ over the set $C(\theta)$. Similarly, a constrained minimization problem is written as

$$\min_x \{f(x;\, \theta) \mid x \in C(\theta)\}. \qquad (P_1')$$

The function $f(.;\, \theta)$ is called the *objective function* and the set $C(\theta) \subseteq X$ is called the *constraint set* or *feasible set* which depends on $\theta \in \Theta$. The vector $x = (x_1, \ldots, x_n) \in X$ is a vector of decision or choice variables, and $\theta = (\theta_1, \ldots, \theta_m) \in \Theta$ is a vector of parameters. Intuitively, let Θ represent all possible "environments" in which an agent may find herself, and let X be the set of all "actions" available to her. Given a value of θ, the agent will find her choices restricted to some subset of X. Changes in the parameters will result in changes in the constraint set, as described by the *constraint correspondence* $C : \Theta \rightarrow\rightarrow X$. Given the objective function, $f(x;\, \theta)$ gives her *payoff* when she faces an environment θ and chooses an action x. The set of optimal actions is described by the *decision rule* or *best-response correspondence* $S : \Theta \rightarrow\rightarrow X$ that assigns to each θ a subset $S(\theta)$ of X. The set

$$S(\theta) := \arg\max_x \{f(x;\, \theta) \mid x \in C(\theta)\} = \{x \in C(\theta) \mid f(x;\, \theta) \geq f(y;\, \theta) \text{ for all } y \in C(\theta)\}$$

is the set of maximizers of f on $C(\theta)$ whose elements x^* solve the problem (P_1). When $S(\theta)$ is singleton, the best-response correspondence becomes a function, and we write $x^* = x(\theta)$. The payoff accruing to the agent is given by the *(maximum) value function* $V : \Theta \longrightarrow \mathbb{R}$, defined by

$$V(\theta) = \max_x \{f(x;\, \theta) \mid x \in C(\theta)\} = f(x^*;\, \theta), \text{ where } x^* \in S(\theta).$$

Given a value of the parameter vector θ, $V(\theta)$ is the maximum attainable payoff to the agent.

Example **1 (Utility maximization)** An agent consumes non-negative quantities of n commodities. The utility function is given by $u(x)$ where $x = (x_1, \ldots, x_n)$ is the vectors of quantities consumed. She has an income $m \geq 0$, and faces a price vector $p = (p_1, \ldots, p_n)$, where $p_i \geq 0$ denotes the price of the i-th commodity. Her budget set, the constraint set, is given by

$$B(p, m) = \{x \in \mathbb{R}^n_+ \mid p \cdot x \leq m\}.$$

The utility maximization problem is written as

$$\max_x \{u(x) \mid x \in B(p, m)\}.$$

Suppose there is a unique maximizer $x^* = x(p, m) = (x_i(p, m))_{i=1}^n$ of the utility maximization problem. The value function is given by $V(p, m) = u(x(p, m))$. The function $x_i(p, m)$ is called the demand function for commodity i, and $V(p, m)$ is called the indirect utility function. ■

Example 2 (Cost minimization) A firm's cost minimization problem is to identify the combination of n inputs $z = (z_1, \ldots, z_n)$ that minimizes its total cost of producing at least y units of output, given the production function $f : \mathbb{R}_+^n \longrightarrow \mathbb{R}_+$ and the input price vector $w = (w_1, \ldots, w_n) \in \mathbb{R}_+^n$, where w_i denotes the price of the i-th input. The firm's feasible production set is given by

$$F(y) = \{z \in \mathbb{R}_+^n \mid f(z) \geq y\},$$

and the problem is to solve

$$\min_z \{w \cdot z \mid z \in F(y)\}.$$

Suppose there is a unique minimizer $z^* = z(w, y) = (z_i(w, y))_{i=1}^n$ of the cost minimization problem. The minimum value function is given by $C(w, y) = w \cdot z(w, y) = \sum_{i=1}^n w_i z_i(w, y)$. The function $z_i(w, y)$ is called the conditional demand for input i, and $C(w, y)$ is called the cost function. ■

In what follows, we will develop the analysis when the optimization problem is given by (P_1). The analysis of a problem of type (P_1') is analogous because of the fact that $\arg\min_x \{f(x; \theta) \mid x \in C(\theta)\} = \arg\max_x \{-f(x; \theta) \mid x \in C(\theta)\}$. We state a useful result in the following lemma.

Lemma 1 *Let $f : X \longrightarrow \mathbb{R}$ where $X \subseteq \mathbb{R}^n$ and $h : \mathbb{R} \longrightarrow \mathbb{R}$ be strictly increasing. Then x^* is a maximum of f on $C(\theta)$ if and only if x^* is a maximum of $h \circ f$ on $C(\theta)$, and z^* is a minimum of f on $C(\theta)$ if and only if z^* is a minimum of $h \circ f$ on $C(\theta)$.*

Proof. Suppose that $x^* \in S(\theta) = \arg\max_x \{f(x; \theta) \mid x \in C(\theta)\}$. Pick any $x \in C(\theta)$. Then, for any $\theta \in \Theta$, $f(x^*; \theta) \geq f(x; \theta)$, and since h is strictly increasing, $h(f(x^*; \theta)) \geq h(f(x; \theta))$. Since x is an arbitrary point in $C(\theta)$, the inequality holds for all $x \in C(\theta)$, and hence x^* is a maximizer of $h \circ f$ over $C(\theta)$.

To show that converse, suppose that x^* maximizes $h \circ f$ over $C(\theta)$, i.e., $h(f(x^*; \theta)) \geq h(f(x; \theta))$ for all $x \in C(\theta)$, but does not maximize f over $C(\theta)$. Then there exists some $y \in C(\theta)$ such that $f(y; \theta)) > f(x^*; \theta))$ for any $\theta \in \Theta$. Since h is strictly increasing, we must have $h(f(y; \theta)) > h(f(x^*; \theta))$, which is a contradiction. □

One important question is under what conditions the set $S(\theta)$ is non-empty. The answers to this question depend obviously on the nature of the objective function and the constraint set. For example, if the objective function is continuous and the constraint set is compact, then $S(\theta)$ is non-empty. This is the Weierstrass theorem we have studied in Chapter 2. In Chapter 2, we have also studied several fixed point theorems that address such problem of existence. In this chapter, our objective is twofold. First, we would analyze the conditions an optimal solution must satisfy, keeping aside the problem of existence. Second, we would like to analyze how changes in the value of the parameter vector θ change the set of maximizers $S(\theta)$ and the value function $V(\theta)$, which are called the *comparative statics* problems.

2 Optimality Conditions

2.1 Convex Constraint Sets

Consider the problem

$$\max_x \{f(x) \mid x \in C\}, \tag{1}$$

where C is a convex set in $\mathbb{R}^n$, and $f : X \longrightarrow \mathbb{R}$ is a $\mathscr{C}^2$ function where $X \subseteq \mathbb{R}^n$. For the time being, we are omitting the parameter θ since we are interested in the solution to the maximization problem given in (1) for a fixed value of θ. Consider a special case when $C = [-1, 2]$ and $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined as $f(x) = 2x^3 - 3x^2 + 2$. Notice that, on C, $f'(0) = f'(1) = 0$, but neither 0 nor 1 is a maximizer. The maximum is achieved at $x = 2$ (we have $f(2) = 6$), but $f'(2) \neq 0$. We have seen that x^* is a maximizer of f implies that $f'(x^*) = 0$. Such a result fails here because the maximum may be achieved at the boundary of the constraint set. The idea is that if x^* maximizes f on C, then as we move away from this point in any *feasible direction* the value of the function must decrease. In what follows, we state this result formally.

Definition 1 (Feasible direction) *Consider the problem stated in (1), where C is a convex set. Take any point $x \in C$ and a direction vector $h \in \mathbb{R}^n$. We say that h is a feasible direction from x if there exists some $\delta > 0$ such that $x + th \in C$ for all $t \in (0, \delta)$.*

The above definition asserts that h is a feasible direction if any small movement away from x in the direction of h still leaves us inside the feasible set.

Theorem 1 (First order necessary conditions for a local maximum) *Assume that f is a $\mathscr{C}^1$ function, and let x^* be a solution to problem (1). Then*

$$Df(x^*) \cdot h \leq 0 \tag{2}$$

for every feasible direction h from x^*.

Proof. Let x^* be a solution to (1), and h be an arbitrary direction vector feasible from x^*. Then there exists a $\delta > 0$ such that $x^* + th \in C$ for all $t \in (0, \delta)$. Because any feasible movement away from x^* reduces the value of f, we have

$$f(x^* + th) - f(x^*) \leq 0.$$

for all t such that $x^* + th \in C$. Dividing the above by $t > 0$ and taking the limit at $t \to 0^+$, we have

$$\lim_{t \to 0^+} \frac{f(x^* + th) - f(x^*)}{t} = Df(x^*; h) = Df(x^*) \cdot h \leq 0.$$

Because f is $\mathscr{C}^1$, the directional derivative exists, and can be written as the scalar product of the derivative and the direction vector. This completes the proof. □

If C is an open set, all points x in C are by definition interior points, and given any x in C, all directions are feasible from it. In this case the inequality $Df(x^*) \cdot h \leq 0$ holds for all h only if all first order partials of f are zero at x^*. Otherwise, it is possible to increase the value of the function by moving in the direction of (or opposite to) the coordinate vector corresponding to the non-zero partial. For example, suppose that $f_k(x^*) > 0$ and $f_j(x^*) = 0$ for all $j \neq k$, and choose the direction vector h such that $h_k > 0$ and $h_j = 0$ for all $j \neq k$. Then

$$Df(x^*) \cdot h = f_k(x^*)h_k > 0,$$

which contradicts the above theorem. Thus,

Corollary 1 *Assume that f is a $\mathscr{C}^1$ function and C is open, and let x^* be a solution to problem (1). Then*

$$Df(x^*) \cdot h = 0.$$

Notice that the result of the above theorem is also a necessary condition for a local maximum of a function on an open set. Now, for an arbitrary direction vector h in $\mathbb{R}^n$, $Df(x^*) \cdot h = 0$ can hold only if $Df(x^*) = 0$. Obviously, the fact that $Df(x^*) = 0$ does not imply that x^* is a local maximum. To see this, consider the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = x^3$. Notice that $f'(0) = 0$, but $x = 0$ is not a maximum of f on $\mathbb{R}$. For a $\mathscr{C}^1$ function $f : X \longrightarrow \mathbb{R}$, if we have that $Df(x) = 0$ for $x \in X$, then x is said to a *critical point* of f. For the above example, $x = 0$ is only a critical point of f. In fact, for this example $x = 0$ is neither a maximum nor a minimum. Such a critical point is called a *saddle point* of the function. For example, for $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by $f(x, y) = x^2 - y^2$, $(0, 0)$ is a saddle point.

Theorem 2 (Second order necessary conditions for a local maximum) *Assume that f is a $\mathscr{C}^2$ function. If f has a local maximum at x^*, then the Hessian matrix of f at x^* is negative semi-definite, i.e.,*

$$h^T D^2 f(x^*) h \leq 0 \tag{3}$$

for all feasible direction $h \in \mathbb{R}^n$.

Proof. Fix an arbitrary $h \in \mathbb{R}^n$. For any given $\alpha > 0$ we can use Taylor's theorem to write

$$f(x^* + \alpha h) - f(x^*) = Df(x^*)(\alpha h) + \frac{\alpha^2}{2} h^T D^2 f(x^* + \lambda_\alpha \alpha h) h$$

for some $\lambda_\alpha \in (0, 1)$. As x^* is a local maximizer, it is a critical point of f, i.e., $Df(x^*) = 0$, and the above equation reduces to

$$f(x^* + \alpha h) - f(x^*) = \frac{\alpha^2}{2} h^T D^2 f(x^* + \lambda_\alpha \alpha h) h.$$

Moreover, it must be true that, for sufficiently small α, we have $f(x^* + \alpha h) - f(x^*) \leq 0$, and hence

$$h^T D^2 f(x^* + \lambda_\alpha \alpha h) h \leq 0$$

Taking the limit of the above expression as $\alpha \to 0$, we get the desired result. □

The following theorem gives sufficient conditions for a strict local maximum of a function on an open set. That is, $f(x^*) > f(x)$ for all $x \in B_r(x^*)$ for some $r > 0$. A point that satisfies the conditions of the following theorem is said to be a regular maximizer of f on C.

Theorem 3 (Sufficient conditions for a strict local maximum) *Assume that f is a $\mathscr{C}^2$ function with the constraint set C being open and convex, and let x^* be a critical point of f on C, i.e., $Df(x^*) = 0$. If the Hessian matrix of f at x^* is negative definite, then f has a strict local maximum at x^*.*

Proof. Fix an arbitrary $h \in \mathbb{R}^n$. By the convexity and openness of C, there exists some $\delta > 0$ such that $x^* + \alpha h \in C$ for all $\alpha \in (0, \delta)$. Fixing some α in this interval, both x^* and $x^* + \alpha h$ lie in C, and Taylor's theorem gives

$$f(x^* + \alpha h) - f(x^*) = Df(x^*)(\alpha h) + \frac{1}{2}(\alpha h)^T D^2 f(x^* + \lambda_\alpha \alpha h)(\alpha h)$$

for some $\lambda_\alpha \in (0, 1)$. Given that $Df(x^*) = 0$, the above equation reduces to

$$f(x^* + \alpha h) - f(x^*) = \frac{\alpha^2}{2} h^T D^2 f(x^* + \lambda_\alpha \alpha h)h = \frac{\alpha^2}{2} Q(\alpha),$$

where $Q(\alpha) \equiv h^T D^2 f(x^* + \lambda_\alpha \alpha h)h$ is a quadratic form for a given h. It can be shown that $Q(\alpha)$ is continuous at $\alpha = 0$. Since $D^2 f(x^*)$ is negative definite by assumption, we have

$$Q(0) = h^T D^2 f(x^*)h < 0.$$

Then continuity of $Q(\cdot)$ at $\alpha = 0$ implies that, for sufficiently small values of α, we have $Q(\alpha) < 0$. Then it follows that

$$f(x^* + \alpha h) - f(x^*) < 0.$$

Because h was chosen arbitrarily, any sufficiently small movement away from x^* reduces the value of f, and hence x^* is a strict local maximum. $\square$

Unfortunately, the above theorem cannot be strengthen to a condition of the sort: "if x^* is a critical point of f such that $D^2 f(x^*)$ is negative semi-definite, then f achieves a local maximum at x^*." To see this, consider $C = \mathbb{R}$ and $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ defined as $f(x) = x^4$ and $g(x) = -x^4$. First notice that $f'(0) = g'(0) = 0$, i.e., $x = 0$ is a critical point of both the functions. We also have $f''(0) = g''(0) = 0$. However at $x = 0$, f reaches a global minimum and g reaches a global maximum. The following theorem gives conditions for the uniqueness of a maximum.

Theorem 4 (Unique maximum) *Let x^* be an optimal solution for the problem in (1) with C convex. If f is strictly quasiconcave, then x^* is unique.*

Proof. Suppose there are two optimal solutions x' and x'', i.e., $f(x') = f(x'')$. Convexity of C implies that, for any $\lambda \in (0, 1)$, $x^\lambda = \lambda x' + (1 - \lambda)x'' \in C$. And strict quasiconcavity of f implies that $f(x^\lambda) > f(x') = f(x'')$ which is a contradiction. $\square$

Example 3 (Derivation of factor demands) Consider a competitive firm that produces a single output in quantity y using two inputs in quantities x_1 and x_2. The firm's production technology is described by a Cobb-Douglas production function

$$y = f(x_1, x_2) = x_1^\alpha x_2^\beta, \text{ where } \alpha, \beta > 0 \text{ and } \alpha + \beta < 1.$$

The firm takes as given the price of output p, and the prices of inputs w_1 and w_2 in order to maximize the profit which is given by

$$\pi(x_1, x_2) = p x_1^\alpha x_2^\beta - w_1 x_1 - w_2 x_2 .$$

The factor demand functions are given by

$$\begin{aligned} x_1(p, w_1, w_2) &= \left[\frac{\alpha p}{w_1}\left(\frac{\beta w_1}{\alpha w_2}\right)^{\beta}\right]^{\frac{1}{1-\alpha-\beta}}, \\ x_2(p, w_1, w_2) &= \left[\frac{\beta p}{w_2}\left(\frac{\alpha w_2}{\beta w_1}\right)^{\alpha}\right]^{\frac{1}{1-\alpha-\beta}}. \end{aligned}$$

In an exercise, you are asked solve the above problem. ■

2.2 Inequality Constraints: The Karush-Kuhn-Tucker Theorem

2.2.1 Necessary Conditions for Optimality

In this section we will consider the problem of maximizing a non-linear real-valued function subject to a finite collection of non-linear constraints. Let $M = \{1, \ldots, m\}$ be an index set. For each $i \in M$ we have a $\mathscr{C}^2$ function $g^i : \mathbb{R}^n \longrightarrow \mathbb{R}$ which will give rise to a constraint. The objective function will be a $\mathscr{C}^2$ function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. Consider the following maximization problem

$$\max\{f(x) \mid g^i(x) \geq 0 \text{ for } i \in M\}. \tag{KKT}$$

Recall that x is a local maximum for problem (KKT) if there is some $r > 0$ such that $f(x) \geq f(x')$ for all $x' \in B_r(x) \cap C$. Since f is a $\mathscr{C}^1$ function, by Taylor's theorem we have the following:

$$f(x + \alpha h) = f(x) + (\alpha h) \cdot \triangledown f(x) + R(\alpha),$$

for $\alpha > 0$. For sufficiently small values of α, $h \cdot \triangledown f(x) > 0$ implies that $f(x + \alpha h) > f(x)$. We will often make use of the above result. Theorem 7 in Chapter 4, which is known as "Gordon's theorem of alternatives" is very useful in establishing the existence of the Lagrange multipliers for constrained optimization problems. Following theorem, known as the Fritz John conditions, is one of the earliest results in the theory of constrained optimization.

Theorem 5 (Fritz John) *Suppose $M \neq \varnothing$ and x^* is a local maximum for problem (KKT). Then there exists a set of non-negative multipliers $\{\mu_0, \mu_1, \ldots, \mu_m\}$ not all zero such that*

$$\mu_0 \triangledown f(x^*) + \sum_{i \in M} \mu_i \triangledown g^i(x^*) = 0, \tag{4}$$

$$\mu_i g^i(x^*) = 0, \text{ for } i \in M. \tag{5}$$

Proof. Let $z^i := \triangledown g^i(x^*)$ for $i \in M$ and $z^0 := \triangledown f(x^*)$. If the theorem were true we can divide Equation (4) throughout by $\sum_{i \in M \cup \{0\}} \mu_i$, and Equation (4) can then be interpreted as saying that 0 belongs to the convex hull of z^i's. This is what we intend to prove. Suppose the contrary, i.e., $0 \notin co(\{z^0, z^1, \ldots, z^m\})$. Recall that the convex hull of a finite number of vectors is closed. Then, by the Separating Hyperplane Theorem, there is a vector h in $\mathbb{R}^n$ such that $h \cdot z^i > 0$ for all $i = 0, 1, \ldots, m$. Since each g^i is differentiable, by Taylor's theorem we have for $i \in M$ and for any $\alpha \in (0, 1)$

$$g^i(x^* + \alpha h) = g^i(x^*) + (\alpha h) \cdot \triangledown g^i(x^*) + R(\alpha) = (1 - \alpha) g^i(x^*) + \alpha g^i(x^*) + (\alpha h) \cdot \triangledown g^i(x^*) + R(\alpha),$$

where the error term $R(\alpha)$ is quadratic in α. Since $g^i(x^*) \geq 0$ for $i \in M$, the above expression reduces to

$$g^i(x^* + \alpha h) \geq \alpha[g^i(x^*) + h \cdot \nabla g^i(x^*)] + R(\alpha).$$

Since $h \cdot z^i = h \cdot \nabla g^i(x^*) > 0$ and $g^i(x^*) \geq 0$, for $\alpha > 0$ we have that

$$\frac{g^i(x^* + \alpha h)}{\alpha} \geq g^i(x^*) + h \cdot \nabla g^i(x^*) + \frac{R(\alpha)}{\alpha} > 0, \text{ for all } i \in M. \tag{6}$$

The above inequality implies that $x^* + \alpha h$ is feasible for α sufficiently small. On the other hand, $h \cdot z^0 > 0$ implies that $f(x^* + \alpha h) > f(x^*)$ for $\alpha > 0$ and sufficiently small, which contradicts the local optimality of x^*. Now we prove the second part of the theorem. Consider the following system of strict inequalities.

$$\begin{aligned} &g^i(x^*) + h \cdot \nabla g^i(x^*) > 0, \text{ for all } i \in M, \\ &h \cdot \nabla f(x^*) > 0 \end{aligned}$$

The first inequality is implied by condition (6) for small values of α, and the second one is equivalent to $h \cdot z^0 > 0$. If $0 \in co(\{z^0, z^1, \ldots, z^m\})$, i.e., if x^* is a local maximum, then there is no $h \in \mathbb{R}^n$ that solves the above system of inequalities. Since the functions $h \cdot \nabla f(x^*)$ and $g^i(x^*) + h \cdot \nabla g^i(x^*)$ for $i \in M$ are affine functions, by Gordon's theorem of the alternative (in Chapter 4) we can find non-negative multipliers $\mu_0, \mu_1, \ldots, \mu_m$ not all zero such that

$$h \cdot \left[\mu_0 \nabla f(x^*) + \sum_{i \in M} \mu_i \nabla g^i(x^*)\right] + \sum_{i \in M} \mu_i g^i(x^*) \leq 0, \text{ for all } h \in \mathbb{R}^n.$$

Choosing $h = \mu_0 \nabla f(x^*) + \sum_{i \in M} \mu_i \nabla g^i(x^*)$, the above inequality reduces to

$$\|h\|^2 + \sum_{i \in M} \mu_i g^i(x^*) \leq 0.$$

Given that $\|h\|^2 \geq 0$, we must have $\sum_{i \in M} \mu_i g^i(x^*) \leq 0$. On the other hand, $g^i(x^*) \geq 0$ and $\mu_i \geq 0$ for all $i \in M$ imply that $\sum_{i \in M} \mu_i g^i(x^*) \geq 0$. Thus, $\sum_{i \in M} \mu_i g^i(x^*) = 0$. Given that the multipliers are non-negative, we have $\mu_i g^i(x^*) = 0$ for each $i \in M$. □

The above lemma may fail if M is not a finite set since we cannot guarantee the existence of an $\alpha > 0$ such that $g^i(x^* + \alpha h) > 0$ for all $i \in M$. We also may not be able to characterize the optimal solution x^* if $\mu_0 = 0$. In this case, Equation (4) reduces to

$$\sum_{i \in M} \mu_i \nabla g^i(x^*) = 0.$$

It is possible to encounter non-linear programing problem of the sort described in the above equation. To see this, consider the following example.

Example 4 Let $f(x_1, x_2) = x_2$, and the constraints given by $g^1(x_1, x_2) = x_1 \geq 0$ and $g^2(x_1, x_2) = -x_1 - x_2^2 \geq 0$. Notice that the only feasible solution is $(x_1^*, x_2^*) = (0, 0)$. Now $\nabla f(0, 0) = (0, 1)$, $\nabla g^1(0, 0) = (1, 0)$ and $\nabla g^2(0, 0) = (-1, 0)$. Thus, Equation (4) yields:

$$\mu_0(0, 1) + \mu_1(1, 0) + \mu_2(-1, 0) = 0.$$

All non-negative solutions of the above equation have $\mu_0 = 0$, and hence $\nabla f(0, 0)$ cannot lie in the convex hull of $\nabla g^1(0, 0)$ and $\nabla g^2(0, 0)$. Hence, without additional assumptions we cannot guarantee that $\mu_0 > 0$. These additional assumptions are called *constraint qualifications*, which says that at a local maximum x^* for the problem (KKT), the vectors in $\{\nabla g^i(x^*)\}_{i \in M}$ are linearly independent. ■

Theorem 6 (Karush-Kuhn-Tucker) *Suppose $M \neq \varnothing$ and x^* is a local maximum for problem (KKT). If the vectors in $\{\nabla g^i(x^*)\}_{i\in M}$ are linearly independent, then there exists a set of non-negative multipliers $\{\lambda_1, \ldots, \lambda_m\}$ not all zero such that*

$$\nabla f(x^*) + \sum_{i\in M} \lambda_i \nabla g^i(x^*) = 0, \tag{7}$$

$$\lambda_i g^i(x^*) = 0, \text{ for } i \in M. \tag{8}$$

Proof. Apply Theorem 5. If $\mu_0 > 0$, then divide Equations (4) and (5) by μ_0 to obtain the result where $\lambda_i = \mu_i/\mu_0$. If $\mu_0 = 0$, then $\sum_{i\in M} \mu_i \nabla g^i(x^*) = 0$. However, the linear independence of the vectors in $\{\nabla g^i(x^*)\}_{i\in M}$ implies that $\mu_i = 0$ for all $i \in M$, which is a contradiction. □

Stated componentwise, Equation (7) reads

$$\frac{\partial f}{\partial x_j}(x^*) + \sum_{i\in M} \lambda_i \frac{\partial g^i}{\partial x_j}(x^*) = 0 \text{ for all } j = 1, \ldots, n.$$

Condition (8) is known as the *complementary slackness* conditions, which implies that for some $i \in M$ it cannot be the case that both λ_i and $g^i(x^*)$ are strictly positive ("slack"). Thus if $g^i(x^*) > 0$, then $\lambda_i = 0$, and if $\lambda_i > 0$, then $g^i(x^*) = 0$.

Now we consider a special case of the maximization problem (KKT) in which the inequality constraints are replaced by equality constraints.

$$\max\{f(x) \mid g^i(x) = 0 \text{ for } i \in M\}. \tag{L}$$

The theorem of Lagrange gives the first order optimality conditions for a local maximum of the above problem.

Theorem 7 (Lagrange) *Suppose $M \neq \varnothing$ and x^* is a local maximum for problem (L). If the vectors in $\{\nabla g^i(x^*)\}_{i\in M}$ are linearly independent, then there exists a set of multipliers $\{\lambda_1^*, \ldots, \lambda_m^*\}$ not all zero such that*

$$\nabla f(x^*) + \sum_{i\in M} \lambda_i^* \nabla g^i(x^*) = 0. \tag{9}$$

Proof. Given the optimality of x^* we will construct a vector of multipliers $\lambda^* \in \mathbb{R}^m$ which satisfies condition (9). Let $x = (z, y)$ where $z \in \mathbb{R}^m$ is the first m coordinate vectors of x, and $y \in \mathbb{R}^{n-m}$ is the last $n-m$ coordinate vectors of x. Thus, $x^* = (z^*, y^*)$. Denote by $g(x)$ the vector $(g^1(x), \ldots, g^m(x))$. The linear independence assumption implies that $\rho(\nabla g_z(z^*, y^*)) = m$, where $\rho(A)$ is the rank of a matrix A. Further, given a function $F(z, y)$ from $\mathbb{R}^{l+k}$ to $\mathbb{R}^l$, $\nabla F_z(z, y)$ denotes the portion of the matrix $\nabla F(z, y)$, which is an $l \times (l+k)$ matrix, corresponding to the first l variables, and $\nabla F_y(z, y)$ denotes the portion of the matrix $\nabla F(z, y)$ corresponding to the last k variables. We have to show the existence of λ^* such that

$$\nabla f_z(z^*, y^*) + \lambda^* \cdot \nabla g_z(z^*, y^*) = 0, \tag{10}$$

$$\nabla f_y(z^*, y^*) + \lambda^* \cdot \nabla g_y(z^*, y^*) = 0. \tag{11}$$

Since $\rho(\nabla g_z(z^*, y^*)) = m$, by the Implicit Function Theorem, there are an open set $V \subset \mathbb{R}^{n-m}$ containing y^* and a $\mathscr{C}^1$ function $h : V \longrightarrow \mathbb{R}^m$ such that $z^* = h(y^*)$ and $g(h(y), y) = 0$ for all $y \in V$. Therefore,

$$\nabla h(y^*) = -(\nabla g_z(z^*, y^*))^{-1} \nabla g_y(z^*, y^*). \tag{12}$$

Now define

$$\lambda^* := -(\nabla g_z(z^*, y^*))^{-1} \nabla f_z(z^*, y^*), \tag{13}$$

which immediately implies Equation (10). Now, define the function $F : V \longrightarrow \mathbb{R}$ by $F(y) = f(h(y), y)$. Since $(h(y^*), y^*)$ is local maximum of f on an open set V, we have $\nabla F(y^*) = 0$, i.e.,

$$\nabla f_z(z^*, y^*) \nabla h(y^*) + \nabla f_y(z^*, y^*) = 0. \tag{14}$$

Substituting for $\nabla h(y^*)$ as in (12) and using the definition of λ^* in (13) in the above equation, we get condition (11). □

Notice that although the maximization problem in (L) is a special case of that in (KKT), the above theorem is not a corollary to the Karush-Kuhn-Tucker theorem. In Theorem 7, the complementary slackness conditions are trivially satisfied by any $\lambda^* \in \mathbb{R}^m$ (without any restrictions on the sign) because at the optimum $g^i(x^*) = 0$ for all $i \in M$. The following theorem generalizes Theorems 6 and 7. Append to the index set M of the maximization problem (KKT) an additional set $M^= := \{j \mid g^j(x) = 0\}$. Now consider the following maximization problem.

$$\max\{f(x) \mid g^i(x) \geq 0 \text{ for all } i \in M, \text{ and } g^j(x) = 0 \text{ for all } j \in M^=\}. \tag{P}$$

Theorem 8 *Let x^* be a local maximum for problem (P). If the vectors in $\{\nabla g^k(x^*)\}$ for $k \in M^= \cup \{i \in M \mid g^i(x^*) = 0\}$ are linearly independent, then there exists a set of multipliers $\{\mu_i^*\}_{i \in M \cup M^=}$ such that*

(a) $\nabla f(x^*) + \sum_{i \in M \cup M^=} \mu_i^* \nabla g^i(x^*) = 0$,

(b) $\mu_i^* \geq 0$ *for all* $i \in M$,

(c) μ_i^* *is unrestricted for all* $i \in M^=$,

(d) $\mu_i^* g^i(x^*) = 0$ *for all* $i \in M$,

(e) $g^i(x^*) \geq 0$ *for all* $i \in M$, *and*

(f) $g^i(x^*) = 0$ *for all* $i \in M^=$.

Proof. See Vohra (2005, pp. 93). □

The above theorem gives the optimality conditions for a maximization problem when both equality and inequality constraints are involved. We should not confuse the constraints for $j \in M^=$ with those for i in $\{i \in M \mid g^i(x^*) = 0\}$, which is a subset of M such that $g^k(x^*) = 0$ for k belonging to this set. This is the set of *binding* or *active* constraints at the optimum.

2.2.2 Using the Optimality Conditions

In what follows we describe "cookbook" procedures for using the optimality conditions stated in Karush-Kuhn-Tucker and Lagrange theorems. For a detailed discussion on when such procedures work and when they fail, see Sundaram (1996).

A cookbook procedure for the KKT theorem:

Consider the maximization problem (KKT). To solve such a problem in practice, one may follow the following three steps. In the first step, we form a function $\mathscr{L}:\mathbb{R}^n\times\mathbb{R}^m\longrightarrow\mathbb{R}$, called the *Lagrangean*, which is given by

$$\mathscr{L}(x,\lambda)=f(x)+\sum_{i=1}^{m}\lambda_i g^i(x). \tag{15}$$

In the second step, find all solutions (x,λ) to the following set of equations:

$$\frac{\partial\mathscr{L}}{\partial x_j}(x,\lambda)=0,\ \text{for } j=1,\ldots,n, \tag{16}$$

$$\frac{\partial\mathscr{L}}{\partial\lambda_i}(x,\lambda)\geq 0,\ \lambda_i\geq 0,\ \lambda_i\frac{\partial\mathscr{L}}{\partial\lambda_i}(x,\lambda)=0\ \text{for } i=1,\ldots,m. \tag{17}$$

Let L be the set of solutions to the above system. In the third step, we compute the value of f at each x in the set $\{x\mid \text{there is }\lambda\text{ such that }(x,\lambda)\in L\}$. In practice, the value of x that maximizes f over this set is typically also the solution to the original maximization problem (KKT).

Example 5 (Numerical example) Consider the following maximization problem:

$$\max\{x^2-y\mid x^2+y^2\leq 1\}. \tag{18}$$

Let $f(x,y):=x^2-y$ and $g(x,y):=1-x^2-y^2$. First notice that the constraint set $C:=\{x,y\mid x^2+y^2\leq 1\}$ is the unit disc in $\mathbb{R}^2$, and hence is compact, and the objective function is continuous. Therefore by Weierstrass theorem, the is at least one maximum of f over C. Next, notice that at the points (x,y) at which the constraint is binding, we must either have $x\neq 0$ or $y\neq 0$ to have $x^2+y^2=1$. Since $\nabla g(x,y)=(-2x,-2y)$, it follows that at all such points where the constraint is binding, we must have $\rho(\nabla g(x,y))=1$. Therefore the constraint qualification is satisfied if the optimum occurs on the boundary of the disc. If the optimum occurs at a point where $g(x,y)>0$, then the set of binding constraints is empty, and the constraint qualification holds vacuously. Now write down the Lagrangean as follows.

$$\mathscr{L}(x,y,\lambda)=x^2-y+\lambda(1-x^2-y^2).$$

The optimal solutions to Problem (18) must be given by the following system:

$$2x(1-\lambda)=0, \tag{19}$$

$$-1-2\lambda y=0, \tag{20}$$

$$\lambda\geq 0,\ 1-x^2-y^2\geq 0,\ \lambda(1-x^2-y^2)=0. \tag{21}$$

Notice that λ cannot be equal to 0 in order to have Equation (20) satisfied. Hence, we must have $1-x^2-y^2=0$, i.e., the constraint must bind at the optimum. For the first equation to hold, we must have $x=0$ or $\lambda=1$. If $\lambda=1$, then Equation (20) implies that $y=-1/2$. Then the binding constraint implies that $x=\pm\sqrt{3}/2$. Thus we get two candidate solutions which are given by

$$(x,y,\lambda)=\left(-\frac{\sqrt{3}}{2},-\frac{1}{2},1\right)\ \text{and}\ \left(\frac{\sqrt{3}}{2},-\frac{1}{2},1\right).$$

The value of f at both the candidate solutions is equal to $5/4$. Next, we have to check whether $x = 0$ is a candidate solution. In this case the equation in (21) implies that we must have $y = \pm 1$. But $y = 1$ is inconsistent with the fact that $\lambda > 0$ since Equation (20) implies, in this case, that $\lambda = -1/2$. Thus, the other candidate solution is given by

$$(x, y, \lambda) = \left(0, -1, \frac{1}{2}\right).$$

At the above candidate optimum, $f(0, -1) = 1 < 5/4$, and hence we can discard this point. Since there is no other candidate solution, there are exactly two optimal solutions, namely $(-\sqrt{3}/2, -1/2)$ and $(\sqrt{3}/2, -1/2)$. Also at these two points the sufficient conditions are clearly satisfied. These two points cannot be minima as we have discarded the candidate optimum $(0, -1)$ because at this point f reaches a value equal to 1 which is lower than $5/4$. ■

A cookbook procedure for the Lagrange theorem:

The procedure is similar to that of the KKT theorem. Consider the maximization problem (L). In the first step, write down the *Lagrangean*, which is given by

$$\mathscr{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g^i(x). \tag{22}$$

In the second step, find all solutions (x, λ) to the following set of equations:

$$\frac{\partial \mathscr{L}}{\partial x_j}(x, \lambda) = 0, \text{ for } j = 1, \ldots, n, \tag{23}$$

$$\frac{\partial \mathscr{L}}{\partial \lambda_i}(x, \lambda) = 0, \text{ for } i = 1, \ldots, m. \tag{24}$$

Let L' be the set of solutions to the above system, which is given by

$$L' := \{(x, \lambda) \mid \nabla \mathscr{L}(x, \lambda) = 0\}.$$

In the third step, we compute the value of f at each x in the set $\{x \mid \text{there is } \lambda \text{ such that } (x, \lambda) \in L'\}$.

Discussion on the Lagrangean method:

We will give an intuitive interpretation of the Lagrangean method described above. Consider the following problem.

$$\max_{x_1, x_2} \{f(x_1, x_2) \mid g(x_1, x_2) = c\}. \tag{25}$$

Instead of directly forcing the agent to respect the constraint, imagine that we allow her to choose the values of x_1 and x_2 freely, but make her pay a fine λ "per unit violation" of the restriction. The agent's payoff, net of the penalty, is given by the Lagrangean function:

$$\mathscr{L}(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda[c - g(x_1, x_2)]. \tag{26}$$

The agent maximizes (26), taking λ as given. The first-order conditions are given by:

$$\frac{\partial \mathscr{L}}{\partial x_1}(x_1^*, x_2^*, \lambda) = 0, \tag{L1}$$

$$\frac{\partial \mathscr{L}}{\partial x_2}(x_1^*, x_2^*, \lambda) = 0. \tag{L2}$$

Given an arbitrary λ, there is no guarantee that the solution to the above system will be an optimal solution to problem (25). But if we pick the correct penalty λ^*, the agent will choose to respect the constraint even if in principle she is free not to do so. Hence, apart from (L1) and (L2), the optimal solution $(x_1^*, x_2^*, \lambda^*)$ must also satisfy the constraint which can be written as

$$\frac{\partial \mathscr{L}}{\partial \lambda}(x_1^*, x_2^*, \lambda^*) = g(x_1^*, x_2^*) - c = 0. \tag{F}$$

Thus, the Lagrangean method is a way in which one transforms a constrained optimization problem to an unconstrained one.

2.2.3 Sufficient Conditions for Optimality

The conditions given in Theorems 6 and 7 are only the necessary conditions that an optimal solution x^* must satisfy. These conditions only assert that x^* is a critical point of the Lagrangean, but do not say anything about whether x^* is a minimum or a maximum. If the objective and the constraint functions are twice continuously differentiable, then the second order conditions sometimes give the sufficient conditions for optimality. As we have mentioned above that the maximization of the Lagrangean function is an unconstrained maximization problem, we can make use of the theorems on sufficiency in the previous section along with our knowledge of concavity and quasiconcavity to get the sufficient conditions in this context. We state two important theorems. The first one guarantees a strict local maximum of a the Lagrange problem, and the second theorem gives sufficient conditions for uniqueness for the KKT problem.

Theorem 9 (Sufficient condition for a strict local maximum) *Let $x^* \in \mathbb{R}^n$ be a feasible point for the problem (L) for some $\lambda^* \in \mathbb{R}^m$ such that the vectors in $\{\nabla g(x^*)\}_{i\in M}$ are linearly independent, and let (x^*, λ^*) satisfies condition (9) in Theorem 7. Define*

$$\mathscr{H} := \{h \in \mathbb{R}^n \mid \nabla g(x^*) \cdot h = 0\},$$

and let $\nabla_x^2 \mathscr{L}(x^*, \lambda^*) = \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 g^i(x^*)$ denote the $n \times n$ matrix of derivatives of the Lagrangean with respect to x at (x^*, λ^*). If $h^T \nabla_x^2 \mathscr{L}(x^*, \lambda^*) h < 0$ for all $h \in \mathscr{H}$ with $h \neq 0$, then x^* is a strict local maximum for the problem (L).

Proof. See Sundaram (1996, pp. 118). □

Theorem 10 (Uniqueness) *Let x^* be an optimal solution to the problem (KKT). If f is strictly quasiconcave and the constraint function g^i for all $i \in M$ are quasiconcave, then x^* is the unique optimal solution.*

Proof. Notice that the set $C^i := \{x \in \mathbb{R}^n \mid g^i(x) \geq 0\}$ is convex because g^i is quasiconcave. x^* maximizes the strictly quasiconcave function f over the convex set $C = \cap_{i\in M} C^i$. Hence, the theorem follows from Theorem 4. □

2.2.4 Application

In this subsection we analyze one problem from economics in order to illustrate the the usage of Theorems 6 and 7.

Optimal consumption choice by utility maximization:

A consumer chooses quantities $(x_1, x_2) \in \mathbb{R}^2_+$ of goods 1 and 2, given the price vector (p_1, p_2) with $p_i > 0$ for $i = 1, 2$ and given her income $m > 0$. Her utility function is given by $u(x_1, x_2) = x_1 x_2$. Formally, the consumer solves the following problem.

$$\max\{x_1 x_2 \mid p_1 x_1 + p_2 x_2 \leq m,\ x_1 \geq 0,\ x_2 \geq 0\}. \tag{27}$$

First notice that the budget set, which is given by

$$B(p_1, p_2, m) = \{(x_1, x_2) \in \mathbb{R}^2 \mid p_1 x_1 + p_2 x_2 \leq m,\ x_1 \geq 0,\ x_2 \geq 0\}, \tag{28}$$

is compact, and the utility function $u(\cdot)$ is continuous. Thus, by Weierstrass Theorem, a solution (x_1^*, x_2^*) to the maximization problem (27) exists.

Now, if either $x_1 = 0$ or $x_2 = 0$, then $u(x_1, x_2) = 0$. On the other hand, the consumption point $(\bar{x}_1, \bar{x}_2) = (m/2p_1, m/2p_2)$ is feasible and gives a utility $u(\bar{x}_1, \bar{x}_2) = m^2/4p_1p_2 > 0$. Since any solution (x_1^*, x_2^*) must be such that $u(x_1^*, x_2^*) \geq u(\bar{x}_1, \bar{x}_2)$, we must have that $x_i^* > 0$ for $i = 1, 2$. Then the complementary slackness conditions of Theorem 6 imply that the multipliers associated with the constraints $x_1 \geq 0$ and $x_2 \geq 0$ must be zero at the optimal solution. Further notice that, given the monotonicity of the preferences, the third constraint must hold with equality. Therefore, the reduced budget set is given by

$$B^*(p_1, p_2, m) = \{(x_1, x_2) \in \mathbb{R}^2 \mid p_1 x_1 + p_2 x_2 = m\}. \tag{29}$$

Now we are within the settings of the Lagrange theorem. The Lagrangean is given by:

$$\mathscr{L}(x_1, x_2, \lambda) = x_1 x_2 + \lambda[m - p_1 x_1 + p_2 x_2].$$

The critical points of $\mathscr{L}(x_1, x_2, \lambda)$ are given by

$$\begin{aligned} x_2 - \lambda p_1 &= 0, \\ x_1 - \lambda p_2 &= 0, \\ m - p_1 x_1 - p_2 x_2 &= 0. \end{aligned}$$

If $\lambda = 0$, then the above system has no solutions. Hence, we have that $\lambda \neq 0$. The first two equations imply that $\lambda = p_1/x_1 = p_2/x_2$, so $x_2 = (p_1/p_2)x_1$. Using this in the third equation, we see that the unique solution to the set of above equations is given by $x_1^* = m/2p_1$, $x_2^* = m/2p_2$ and $\lambda^* = m/2p_1p_2 > 0$. Also, the other two inequality constraints $x_i^* > 0$ for $i = 1, 2$ are also satisfied.

Now we will use the second-order conditions to show that (x_1^*, x_2^*) is a strict local maximum of u on $B^*(p_1, p_2, m)$. Notice that $\nabla g(x^*) = (-p_1, -p_2)$. Hence we have that

$$\mathscr{H} = \{h \in \mathbb{R}^2 \mid \nabla g(x^*) \cdot h = 0\} = \{h \in \mathbb{R}^2 \mid h_1 = -(p_2 h_2)/p_1\}.$$

And the second derivative of the Lagrangean at the optimum with respect to (x_1, x_2) is given by

$$\nabla_x^2 \mathscr{L}(x^*, \lambda^*) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \lambda^* \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

For $h \in \mathscr{H}$ we have $h^T \nabla_x^2 \mathscr{L}(x^*, \lambda^*)h = -2p_2h_2^2/p_1 < 0$. Hence, by Theorem 9 we have that (x_1^*, x_2^*) is a strict local maximum. Moreover, (x_1^*, x_2^*) is the unique maximum. Notice that the budget set $B(p_1, p_2, m)$ is convex and the objective function $u(\cdot)$ is strictly quasiconcave, and the claim follows from Theorem 10.

3 Comparative Statics

3.1 Comparative Statics of Smooth Optimization Problems

We consider the parametric maximization problem (P_1). In this subsection, we first analyze several properties of the value function $V(\theta)$ and of the optimal solution correspondence $S(\theta)$. Next, we study the behavior of the value function and the optimal solution with respect to the parameter θ. This is known as the *comparative static analysis*.

3.1.1 Theorem of the Maximum

The following theorem states some properties of the optimal solution to the maximization problem (P_1) and of the value function.

Theorem 11 (Berge) *Given the sets $X \subseteq \mathbb{R}^n$ and $\Theta \subseteq \mathbb{R}^m$, let $f : X \times \Theta \longrightarrow \mathbb{R}$ be a continuous function, and $C(\theta)$ is compact for each $\theta \in \Theta$. Then*

(a) *the solution correspondence $S : \Theta \rightarrow\!\!\rightarrow X$ which is given by $S(\theta) := \arg\max_x \{f(x;\theta) \mid x \in C(\theta)\}$ is non-empty, and the value function $V(\theta) = \max_x\{f(x;\theta) \mid x \in C(\theta)\}$ of the problem P_1 is continuous;*

(b) *if $f : X \times \Theta \longrightarrow \mathbb{R}$ is concave on $X \times \Theta$, and $C(\theta)$ is convex for each $\theta \in \Theta$, then $S(\theta)$ is a convex set. Moreover, if $f(x, \theta)$ is strictly concave, then the set $S(\theta)$ is singleton, i.e., $S(\theta) = \{x(\theta)\}$. The function $x(\theta)$ is a continuous function.*

Proof. Omitted. □

3.1.2 Applications

In what follows we analyze two concrete examples from economics where the above theorems readily apply.

Walrasian Equilbrium:

Consider an exchange economy $\xi = (u^i, \omega^i)_{i=1}^n$ where $u^i : \mathbb{R}_+^L \longrightarrow \mathbb{R}$ is the utility function of consumer $i = 1, \ldots, n$, which is continuous and strictly quasiconcave, and $\omega^i \in \mathbb{R}_+^L$ is her endowment vector. An allocation $x = (x^1, \ldots, x^n) \in \mathbb{R}_+^{n \times L}$ is a vector that describes the amount of each commodity consumed by each agent. An allocation is *feasible* if the consumption of each commodity does not exceed its total endowment, i.e., $\sum_{i=1}^n x^i \leq \sum_{i=1}^n \omega^i$.

Definition 2 (Walrasian equilibrium) *A Walrasian equilibrium for the exchange economy ξ is a price-allocation pair (p^*, x^*) such that (i) all agents maximize utility taking p^* as given, and (ii) the market for each commodity clears.*

First notice that, given a price vector p, the demand vector of consumer i is given by:

$$x^i(p, p \cdot \omega^i) \in \text{argmax}_{x^i} \{u^i(x^i) \mid p \cdot x^i \leq p \cdot \omega^i\}.$$

From Berge's theorem it follows that the demand function $x^i(p, p \cdot \omega^i)$ is continuous for any strictly positive price vector. Also the demand function is homogeneous of degree zero. One obtains the *aggregate demand function* by summing the individual demands across all consumers, i.e.,

$$x(p, \omega) = \sum_{i=1}^n x^i(p, p \cdot \omega^i), \tag{30}$$

and the *aggregate excess demand* is given by:

$$z(p, \omega) = \sum_{i=1}^n [x^i(p, p \cdot \omega^i) - \omega^i] = x(p, \omega) - \omega. \tag{31}$$

It is not so difficult to show that $x(p, \omega)$ and $z(p, \omega)$ are continuous functions, and are homogeneous of degree zero in prices. Now notice that the individual budget constraints must hold with equality, i.e.,

$$p \cdot x^i(p, p \cdot \omega^i) = p \cdot \omega^i.$$

Summing over all consumers in the above equation we get the *Walras' law*:

$$p \cdot z(p, \omega) = 0. \tag{32}$$

It is worthwhile to note two important implications of the above result. First, Walras' law implies that if the markets of any $L-1$ goods clear, then the remaining market must automatically clear. Second, the homogeneity of degree zero of $z(p, \omega)$ means that only relative prices matter. Formally, it allows us to normalize prices and worry about only $L-1$ markets. Hence, we do the following normalization such that $p \in \Delta$ where Δ is the unit simplex in $\mathbb{R}_+^L$, i.e., $\sum_{l=1}^L p_l = 1$. Formally,

$$\Delta := \left\{ p \in \mathbb{R}_+^L : \sum_{l=1}^L p_l = 1 \right\}.$$

It is easy to show that Δ is a compact and convex set. For simplicity we will assume that each of the individual utility functions is strictly quasiconcave which implies that the aggregate excess demand $z(p, \omega)$ is single-valued, and abuse notations to express it as $z(p)$ since the function is defined for constant endowment vectors. The following theorem asserts that a Walrasian equilibrium for the economy ξ exists.

Theorem 12 *Let $z : \Delta \longrightarrow \mathbb{R}^L$ be the aggregate excess demand function which is continuous. Then there exists a price vector $p^* \in \Delta$ such tha $z(p^*) \leq 0$.*

Proof. Define a price-adjustment rule $g : \Delta \longrightarrow \Delta$ as follows.

$$g_l(p) = \frac{p_l + \max\{0, z_l(p)\}}{1 + \sum_{l=1}^{L} \max\{0, z_l(p)\}} \quad \text{for } l = 1, \ldots, L.$$

It is easy to show that $g_l(p)$ is continuous for each l, and $g(p) = (g_1(p), \ldots, g_L(p)) \in \Delta$. It follows from the Brouwer's fixed-point theorem that there is a price vector $p^* \in \Delta$ such that $g(p^*) = p^*$. Then

$$p_l^* = \frac{p_l^* + \max\{0, z_l(p^*)\}}{1 + \sum_{l=1}^{L} \max\{0, z_l(p^*)\}} \quad \text{for } l = 1, \ldots, L. \tag{33}$$

We will show that the price vector p^* are the Walrasian prices. Notice that the above equation implies

$$\begin{aligned} & p_l^* + p_l^* \sum_{l=1}^{L} \max\{0, z_l(p^*)\} = p_l^* + \max\{0, z_l(p^*)\} \text{ for } l = 1, \ldots, L, \\ \Longrightarrow \; & p_l^* \sum_{l=1}^{L} \max\{0, z_l(p^*)\} = \max\{0, z_l(p^*)\} \text{ for } l = 1, \ldots, L, \\ \Longrightarrow \; & z_l(p^*) p_l^* \sum_{l=1}^{L} \max\{0, z_l(p^*)\} = z_l(p^*) \max\{0, z_l(p^*)\} \text{ for } l = 1, \ldots, L, \end{aligned}$$

Adding up the above L equations we get:

$$\left[\sum_{j=1}^{L} \max\{0, z_j(p^*)\}\right] \left[\sum_{l=1}^{L} p_l^* z_l(p^*)\right] = \sum_{l=1}^{L} z_l(p^*) \max\{0, z_l(p^*)\}.$$

The Walras' law implies that

$$\sum_{l=1}^{L} p_l^* z_l(p^*) = 0.$$

Hence,

$$\sum_{l=1}^{L} z_l(p^*) \max\{0, z_l(p^*)\} = \sum_{l=1}^{L} \tilde{z}_l(p^*) = 0.$$

Each term $\tilde{z}_l(p^*)$ of the above summation is greater than or equal to zero because $\tilde{z}_l(p^*) \in \{0, z_l^2(p^*)\}$. In fact, each $\tilde{z}_l(p^*)$ must equal zero, otherwise the last equality would not hold, and hence $z_l(p^*) \leq 0$ for $l = 1, \ldots, L$. In other words, p^* is a Walrasian price vector. □

Nash Equilibrium:

Let $\Gamma = \langle N, \{P_i\}_{i \in N}, \{\pi_i(\cdot)\}_{i \in N} \rangle$ be a normal-form Bertrand game, where $N = \{1, \ldots, n\}$ is the set of n firms, P_i is a non-empty subset of $\mathbb{R}_+$ for each $i \in N$ and $p_i \in P_i$ is the unit price of the product of firm i, and $\pi_i : P_1 \times \ldots \times P_n \longrightarrow \mathbb{R}$ is the profit function of firm i. A vector $p = (p_1, \ldots, p_n) = (p_i, p_{-i})$ where $p_{-i} = (p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n)$, called a *strategy profile*, is an element of $P := P_1 \times \ldots \times P_n = P_i \times P_{-i} \subseteq \mathbb{R}_+^n$ where $P_{-i} = P_1 \times \ldots \times P_{i-1} \times P_{i+1} \times \ldots \times P_n \subseteq \mathbb{R}_+^{n-1}$.

Definition 3 (Nash equilibrium) *A strategy profile $p^* = (p_1^*, \ldots, p_n^*)$ is a Nash equilibrium if it is feasible (i.e., is an element of P) and if*

$$p_i^* = \arg\max_{p_i \in P_i} \pi_i(p_i, p_{-i}^*) \text{ for each } i \in N. \tag{NE}$$

The choice p_i of price by firm i is a *best response* for this firm to the prices set by its rivals p_{-i} if it maximizes firm i's profit. Formally,

Definition 4 (Best-response correspondence) *The best-response correspondence of firm i is a mapping $\psi_i : P_{-i} \twoheadrightarrow P_i$ if*

$$\psi_i(p_{-i}) = \arg\max_{p_i \in P_i} \pi_i(p_i, p_{-i}).$$

Define by $\psi : P \twoheadrightarrow P$ the Cartesian product of the best-reply correspondences of the n firms, i.e., $\psi(p) := \psi_1(p_{-1}) \times \ldots \times \psi_n(p_{-n})$. Now assume that each π_i is strictly quasiconcave, and hence the best-response correspondence of firm i is a function. Notice that if a strategy profile p^* is a Nash equilibrium of the game Γ, then for each firm i we have that $p_i^* = \psi_i(p_{-i}^*)$. Thus, $p^* = \psi(p^*)$, i.e., p^* is a fixed point of the function ψ. Therefore, proving the existence of a Nash equilibrium is equivalent to proving the set of of fixed points of ψ is non-empty.

Theorem 13 (Existence of Nash equilibrium) *Let $\Gamma = \langle N, \{P_i\}_{i \in N}, \{\pi_i(\cdot)\}_{i \in N} \rangle$ be a Bertrand game, and assume for each firm i that*

(a) *the strategy space P_i is a non-empty, compact and convex set in $\mathbb{R}_+$, and*

(b) *the profit function $\pi_i(\cdot)$ is continuous and strictly quasiconcave in p_i given p_{-i}.*

Then the set of Nash equilibria of Γ is non-empty.

Proof. We will use Brouwer's fixed-point theorem to prove the existence of a fixed point. So we will prove that the function ψ is continuous, and that P is compact and convex. Since P is a Cartesian product of n compact and convex sets, it is compact and convex. Consider the best-response function of firm i. Since π_i is continuous and P_i is compact and convex, by Weierstrass theorem, ψ_i exists, and by Berge's theorem it is continuous. The function ψ which is the Cartesian product of n continuous functions, is itself continuous. Therefore, by Brouwer's fixed point theorem its set of fixed points is non-empty. □

3.2 Value Function and the Envelope Theorem

The value function $V : \Theta \longrightarrow \mathbb{R}$ for a maximization problem gives the maximum attainable value of the objective function for each value of the parameters:

$$V(\theta) = \max_x \{f(x; \theta) \mid x \in C(\theta)\} = f(x^*; \theta), \text{ where } x^* \in S(\theta).$$

The above expression has a nice and intuitive geometric interpretation. Consider the above maximization problem with a single parameter θ. Fix an x that is feasible, and plot the objective function f as the function of the parameter alone. The function $V(\theta)$ corresponds to the upper envelope of this family of curves. Two important observations emerge. First, the concavity or convexity of V will crucially depend on the nature of the objective function and the constraint set. Second, the value function is, in general, not differentiable. An important result, called the *envelope theorem* will give us sufficient conditions for the differentiability of the value function.

Theorem 14 (Concavity of the value function) *Consider the following maximization problem and the associated value function:*

$$V(\theta) = \max_x \{f(x;\, \theta) \mid g(x;\, \theta) \geq 0\}.$$

Suppose the objective function f is concave in $(x,\, \theta)$ and that all the constraint functions g^i for $i = 1, \ldots, m$ are quasiconcave in $(x,\, \theta)$. Then the value function is concave.

Proof. Take two arbitrary parameter vectors θ' and θ'', and let $x' = x(\theta')$ and $x'' = x(\theta'')$ be the corresponding optimal choices of x. Consider now the pair $(x^\lambda,\, \theta^\lambda)$ defined by

$$x^\lambda = (1-\lambda)x' + \lambda x'' \text{ and } \theta^\lambda = (1-\lambda)\theta' + \lambda\theta'', \text{ for } \lambda \in (0,\, 1),$$

and observe that, in principle, x^λ is not necessarily an optimal choice corresponding to θ^λ. We first show that x^λ is feasible for θ^λ. Since x' is feasible for θ' and x'' is feasible for θ'', it follows that, for each $j = 1, \ldots, m$, $g^j(x';\, \theta') \geq 0$ and $g^j(x'';\, \theta'') \geq 0$. Thus, quasiconcavity of g^j implies that

$$g^j(x^\lambda;\, \theta^\lambda) \geq \min\{g^j(x';\, \theta'),\, g^j(x'';\, \theta'')\} \geq 0,$$

for each $j = 1, \ldots, m$, and hence x^λ is feasible for θ^λ. Now,

$$\begin{aligned} V(\theta^\lambda) &= f(x(\theta^\lambda);\, \theta^\lambda) \\ &\geq f(x^\lambda;\, \theta^\lambda) \\ &\geq (1-\lambda) f(x';\, \theta') + \lambda f(x'';\, \theta'') \\ &= (1-\lambda)V(\theta') + \lambda V(\theta''). \end{aligned}$$

Thus, V is concave. □

Theorem 15 (Envelope theorem) *Let $V(\theta) = \max_x\{f(x;\, \theta) \mid g(x;\, \theta) = 0\}$, where $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ are $\mathscr{C}^2$ functions. If x^* be a regular solution, i.e., a strict local maximum of this problem for θ^0, then V is differentiable at θ^0, and the derivative is given by*

$$DV(\theta^0) = D_\theta \mathscr{L}(x^*,\, \theta^0) = D_\theta f(x^*,\, \theta^0) + \lambda^T D_\theta g(x^*,\, \theta^0).$$

Proof. Differentiating the Lagrange function

$$\mathscr{L}(x^*,\, \theta^0) = f(x^*,\, \theta^0) + \lambda^T g(x^*,\, \theta^0)$$

with respect to $(x,\, \lambda)$, for $\theta = \theta^0$, we get the first-order conditions of the problem:

$$D_x \mathscr{L}(x,\, \theta^0) = D_x f(x,\, \theta^0) + \lambda^T D_x g(x,\, \theta^0) = 0, \tag{34}$$

$$D_\lambda \mathscr{L}(x,\, \theta^0) = g(x,\, \theta^0) = 0. \tag{35}$$

Under the assumptions of the theorem, the decision rule is well-defined and is a differentiable function. Thus,

$$V(\theta) = f(x(\theta); \theta).$$

Differentiating the value function with respect to θ and using (34), we obtain

$$\begin{aligned} DV(\theta^0) &= D_\theta f(x^*, \theta^0) + D_x f(x^*, \theta^0) Dx(\theta^0) \\ &= D_\theta f(x^*, \theta^0) - \lambda^T D_x g(x^*, \theta^0) Dx(\theta^0) \end{aligned} \tag{36}$$

Substituting $x(\theta)$ in (35) and differentiating with respect to θ

$$D_x g(x^*, \theta^0) Dx(\theta^0) + D_\theta g(x^*, \theta^0) = 0.$$

Using the above expression in (36) we get the desired result. □

We will use the Envelope theorem to give an intuitive interpretation of the Lagrange multiplier. Consider the Lagrange problem with the constraints $g^i(x) + \gamma_i = 0$ for $i \in M$. The value function is given by

$$V(\gamma) = \max_x \{ f(x) \mid g(x) + \gamma = 0 \}.$$

By Envelope theorem we have $DV(\gamma^0) = \lambda^0$. Thus, the multipliers measure the sensitivity of the value function to the changes in the constants of the constraint functions. Therefore, λ can be interpreted as *shadow prices*.

Example 6 (Roy's identity) Consider the utility maximization problem defined as

$$V(p, m) = \max_x \{ u(x) \mid m - p \cdot x = 0 \}.$$

The Lagrangean for the consumer problem is given by

$$\mathscr{L}(x, \lambda; p, m) = u(x) + \lambda \left[m - \sum_{i=1}^{l} p_i x_i \right].$$

By envelope theorem

$$\begin{aligned} \frac{\partial V}{\partial p_i}(p, m) &= -\lambda^* x_i(p, m), \\ \frac{\partial V}{\partial m}(p, m) &= \lambda^*. \end{aligned}$$

Dividing the first equality by the second one we get

$$x_i(p, m) = -\frac{\partial V(p, m)/\partial p_i}{\partial V(p, m)/\partial m},$$

which is the Roy's identity. ■

3.3 Monotone Comparative Statics

Consider the maximization problem (P_1), and assume that the best response correspondence is given by $S(\theta) = \{x(\theta)\}$. We aim at finding sufficient conditions under which the optimal solution is monotone on Θ. First consider the following definition.

Definition 5 (Supermodularity) *Let $f : X \longrightarrow \mathbb{R}$ be a $\mathscr{C}^2$ function where $X \subseteq \mathbb{R}^n$. The function f is said to be (strictly) supermodular on X if the off-diagonal elements of its Hessian matrix are all (strictly) positive, i.e., $f_{ij}(x)(>) \geq 0$ for all $i, j = 1, \ldots, n$ and $i \neq j$.*

The following theorem states a simpler version of "monotone comparative statics" result of Topkis (1978).

Theorem 16 *Let $f : X \times \Theta \longrightarrow \mathbb{R}$ be a $\mathscr{C}^2$ function, where $X, \Theta \subset \mathbb{R}$. If f is (strictly) supermodular on $X \times \Theta$, then the function $x(\theta) = argmax_x\{f(x, \theta) \mid x \in C(\theta)\}$ where $C(\theta) \subset X$ is (strictly) increasing in θ.*

We omit the proof. Instead, we consider the following examples to understand the intuition of the above theorem.

Example 7 (Firm-worker assignment) Consider a continuum $[0, 1]$ of firms with a machine apiece, and a continuum $[0, 1]$ of workers who are heterogeneous with respect to productivity q. Let $F(s)$ denote the distribution of machine size and let $G(q)$ be the distribution of productivity. An worker with a given productivity level q produces a total output $y(s, q)$ if he uses a machine of size s. where $r(s)$ is the rental rate of a type s machine and $w(q)$ is the wage of a type q worker. Assume that workers are assigned to the machines according to the assignment rule $s = s(q)$ where $q \equiv s^{-1}$. A Walrasian equilibrium is a tuple $(s, w(q), r(s))$ such that

(i) (a) $q(s) = \text{argmax}_q \{y(s, q) - w(q)\}$ for each s;
    (b) $r(s) = \max_q \{y(s, q) - w(q)\}$ for each s;
    (c) $w(q) = \max_s \{y(s, q) - r(s)\}$ for each q;

(ii) If $s([q_1, q_2]) = [s_1, s_2]$ for any $[q_1, q_2], [s_1, s_2] \subseteq [0, 1]$, then $F(s_2) - F(s_1) = G(q_2) - G(q_1)$.

Suppose that the equilibrium assignment is increasing, i.e., $s'(q) > 0$. We will establish conditions under which such supposition is valid. The first order condition of the maximization problem of a type s firm implies that

$$w'(q) = \frac{\partial y(s, q)}{\partial q} \quad \text{for } s = s(q). \tag{37}$$

The above equation implies that the marginal wage of a type q worker is equal to his marginal product. Applying the envelope theorem to the above maximization problem one gets

$$r'(s) = \frac{\partial y(s, q)}{\partial s} \quad \text{for } s = s(q), \tag{38}$$

i.e., the marginal rent of a type s machine is its marginal product. In an equilibrium, the assignment of workers to machine must also satisfy the second order condition of the maximization problem:

$$\frac{\partial^2 y}{\partial q^2}(s(q), q) - w''(q) \leq 0. \tag{39}$$

Differentiating equation (37) one gets

$$w''(q) = \frac{\partial^2 y}{\partial s \partial q}(s(q), q)s'(q) + \frac{\partial^2 y}{\partial q^2}(s(q), q).$$

From the above it is clear that $s'(q) > 0$, i.e., the assignment is positively assortative if $y(s, q)$ is supermodular on $[0, 1] \times [0, 1]$. Now suppose that s and q are both distributed uniformly over the intervals $[\mu - \sigma_s, \mu + \sigma_s]$ and $[\mu - \sigma_q, \mu + \sigma_q]$, respectively. Then the second equilibrium condition implies that

$$\begin{aligned} \int_{\mu-\sigma_s}^{s} \frac{s - (\mu - \sigma_s)}{(\mu + \sigma_s - \mu - \sigma_s)} ds &= \int_{\mu-\sigma_q}^{q} \frac{q - (\mu - \sigma_q)}{(\mu + \sigma_q - \mu - \sigma_q)} dq \\ \implies s = s(q) = \mu + \frac{\sigma_s}{\sigma_q}(q - \mu). \end{aligned}$$

That is the matching function is linear whose slope depends on the relative dispersion of the workers to machines, i.e., whether $\sigma_s \gtreqless \sigma_q$. Notice that if $y(s, q) = sq$, then $w''(a) = \sigma_s/\sigma_q > 0$. Thus the equilibrium wage function is convex, i.e., the economy represents increasing wage inequality. ■

Example 8 **(Bertrand competition)** Let there be two firms 1 and 2 in a Bertrand market. They face an identical $\mathscr{C}^2$ demand function $D(p_i, p_j)$ with $D_i(p_i, p_j) < 0$, $D_{ii}(p_i, p_j) \leq 0$ and $D_j(p_i, p_j) > 0$ for $i, j = 1, 2$ and $i \neq j$. Both firms have zero cost of production. The profit of firm i is given by

$$\Pi^i(p_i, p_j) = p_i D(p_i, p_j) \tag{40}$$

which she maximizes with respect to p_i. The Nash equilibrium is given by (p_1^*, p_2^*) such that $p_1^* = \psi_1(p_2^*)$ and $p_2^* = \psi_2(p_1^*)$ where $\psi_i(p_j)$ is the best response of firm i. We will look for conditions under which the best response functions are increasing. Notice at a Nash equilibrium that

$$\Pi_i^i(p_i, p_j) = p_i D_i(p_i, p_j) + D(p_i, p_j) = 0.$$

Differentiating the above expression with respect to p_j at $p_i = \psi_i(p_j)$, we get

$$\begin{aligned} &\psi_i'(p_j)D_i(\psi_i(p_j), p_j) + \psi_i(p_j)[D_{ii}(\psi_i(p_j), p_j)\psi_i'(p_j) + D_{ij}(\psi_i(p_j), p_j)] \\ &+ D_i(\psi_i(p_j), p_j)\psi_i'(p_j) + D_j(\psi_i(p_j), p_j) = 0 \\ \implies &\psi_i'(p_j) = -\frac{\psi_i(p_j)D_{ij}(\psi_i(p_j), p_j) + D_j(\psi_i(p_j), p_j)}{2D_i(\psi_i(p_j), p_j) + \psi_i(p_j)D_{ii}(\psi_i(p_j), p_j)}. \end{aligned}$$

Given the assumption on $D(p_1, p_2)$, the best response functions are increasing if $D_{ij}(p_i, p_j) \geq 0$ for all (p_i, p_j), i.e., the demand function of any firm is supermodular in its own price and that of its rival. ■