1 Optimization

Mathematical programming refers to the basic mathematical problem of finding a maximum to a function, f, subject to some constraints.[1] In other words, the objective is to find a point, x^*, in the domain of the function such that two conditions are met:

i) x^* satisfies the constraint (i.e. it is **feasible**).

ii) There is no other feasible x with $f(x) > f(x^*)$.

Whenever all points in the domain of f are feasible, the problem is called **unconstrained**. A more general type of mathematical programming which will be referred to as **classical (or Lagrangian) programming** problem is to maximize a given function subject to a set of equality constraints.

The general programming problem is the **nonlinear programming** problem where a given function is maximized subject to a set of inequality constraints.

A special case, important in itself, is the **linear programming** problem which seeks to maximize a function of a given linear form subject to a set of linear inequality constraints.

We restrict our attention to optimization problems in which the feasible set is a given subset of $\mathbb{R}^n$, and we refer to these problems as **static optimization** problems. In **dynamic optimization** problems, the feasible set of future periods is affected by the choices today.

[1] All problems will be stated subsequently as one of maximization. A problem of minimization can be treated as one of maximization simply by changing the sign of the function to be minimized.

The general form of the mathematical programming problem can be stated

$$\max f(x) \qquad \text{subject to } x \in X$$

where $x \in \mathbb{R}^n$, $X \subset \mathbb{R}^n$ and $f : X \to \mathbb{R}$

It will generally be assumed that X is not empty, that is, that there exists a feasible vector $x \in X$. In economics, the vector x is frequently called the vector of **instruments**, the function $f(x)$ is frequently called the **objective function**, and the set X of feasible instruments is frequently called the **constraint set** or the **opportunity set** or the **feasible set**.

The basic economic problem of allocating scarce resources among competing ends can then be represented as one of mathematical programming, where a particular resource allocation is represented by the choice of a particular vector of instruments; the scarcity of resources is represented by the opportunity sets, reflecting constraints on the instruments; and the competing ends are represented by the objective function, which gives the values attached to each of the alternative allocations.

2 Unconstrained Optimization

2.1 Quadratic forms and Taylor's theorem

For convenience and later reference, we represent here the matrix notation used further on as well as several characterizations of quadratic forms.

Suppose A is an $n \times n$ symmetric matrix of the following form

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{12} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{13} & a_{23} & a_{33} & \dots & a_{3n} \\ \dots & & & & \dots \\ a_{1n} & a_{2n} & a_{3n} & \dots & a_{nn} \end{pmatrix}$$

Definition 1 *A **quadratic form** is a function $Q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by*

$$Q_A(y) = y \cdot Ay = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} y_i y_j$$

where A is an $n \times n$ symmetric matrix and $y \in \mathbb{R}^n$.

Definition 2 *Suppose that A is an $n \times n$ symmetric matrix and that $Q_A(y) = y \cdot Ay$ is the quadratic form associated with A. Then A and Q_A are called:*

1. **positive semidefinite** if $Q_A(y) = y \cdot Ay \geq 0$ for all $y \in \mathbb{R}^n$;

2. **positive definite** if $Q_A(y) = y \cdot Ay > 0$ for all $y \in \mathbb{R}^n, y \neq 0$;

3. **negative semidefinite** if $Q_A(y) = y \cdot Ay \leq 0$ for all $y \in \mathbb{R}^n$;

4. **negative definite** if $Q_A(y) = y \cdot Ay < 0$ for all $y \in \mathbb{R}^n, y \neq 0$;

5. **indefinite** if $Q_A(y) = y \cdot Ay < 0$ for some $y \in \mathbb{R}^n$ and $Q_A(y) > 0$ for other $y \in \mathbb{R}^n$.

Suppose A is a $n \times n$ symmetric matrix. Define Δ_k to be the determinant of the upper left-hand corner $k \times k$ submatrix of A for $1 \leq k \leq n$. The determinant Δ_k is called the k-th **leading principal minor** of A with $\Delta_1 = a_{11}$

$$\Delta_2 = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$$

and finally $\Delta_n = \det A$.

Theorem 3 *If A is an $n \times n$ symmetric matrix and if Δ_k is the k-th* leading principal minor *of A for $1 \leq k \leq n$, then*

1. *A is **positive definite** if and only if $\Delta_k > 0$ for $k = 1, 2, ..., n$;*

2. *A is **negative definite** if and only if $(-1)^k \Delta_k > 0$ for $k = 1, 2, ..., n$ (that is, the* leading principal minors *alternate in sign with $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0$,etc.).*

To test semidefiniteness we have to evaluate all *principal minors*, and therefore we have to examine a greater number of determinants. If A is a quadratic matrix of order n and we wipe out (arbitrary) r of the rows and the **corresponding** r columns as well, the resulting $(n - r) \times (n - r)$ submatrix is called a **principal minor** $\tilde{\Delta}_k$ is the k-th principal minor of A with $k = n - r$. We have the following theorem:

Theorem 4 *If A is an $n \times n$ symmetric matrix and if $\tilde{\Delta}_k$ is the k-th principal minor of A for $1 \leq k \leq n$, then*

1. *A is **positive semidefinite** if and only if $\tilde{\Delta}_k \geq 0$ for all principal minors of dimension k and $k = 1, 2, ..., n$;*

2. *A is **negative semidefinite** if and only if $(-1)^k \tilde{\Delta}_k \geq 0$ for all principal minors of dimension k and for $k = 1, 2, ..., n$ (that is, the principal minors alternate in sign with $\tilde{\Delta}_1 \leq 0, \tilde{\Delta}_2 \geq 0, \tilde{\Delta}_3 \leq 0$, etc.).*

3. *A positive (negative) semidefinite matrix A is positive (negative) defnite if and only if A is a nonsingular matrix.*

Let f be a function defined on an open set $B \subset \mathbb{R}^n$, whose first and second partial derivatives exist at x. The vector of the first partial derivatives is called the **gradient**, and it's denoted by $gradf(x)$ or $\nabla f(x)$. Formally

$$\nabla f(x^*) = \left(\frac{\partial(f(x^*)}{\partial x_1}, \frac{\partial f(x^*)}{\partial x_2}, ..., \frac{\partial f(x^*)}{\partial x_n}\right)$$

The symmetric $n \times n$ matrix of second partial derivatives $\nabla^2 f(x) \equiv H(x)$,evaluated at point $x \in B$ is called the **Hessian matrix** of f, that is

$$\nabla^2 f(x) \equiv H(x) = \begin{bmatrix} \partial^2 f(x)/\partial x_1 \partial x_1 & ... & \partial^2 f(x)/\partial x_1 \partial x_n \\ ... & & ... \\ ... & & ... \\ \partial^2 f(x)/\partial x_n \partial x_1 & ... & \partial^2 f(x)/\partial x_n \partial x_n \end{bmatrix}$$

Recall that Taylor's formula can be extended to real-valued functions f defined on an open subset B of $\mathbb{R}^n$. In order to state the general theorem in a form which resembles the one-dimensional case, recall the special symbols

$$f'(x;t), f'(x;t)\ , ..., f^{\ (m)}(x;t),$$

that we introduced for Taylor's formula. These play the role of higher-order directional derivatives, and they are defined as follows:
If x is a point in $\mathbb{R}^n$ where all second-order partial derivatives of f exist, and if $t = (t_1, ..., t_n)$ is an arbitrary point in $\mathbb{R}^n$, we write

$$f\ ''(x;t) = \sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij} f(x) t_j t_i$$

We also define

$$f\ '''(x;t) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} D_{ij,k}\, f(x) t_k t_j t_i$$

if all the third order partial derivatives exist at x. The symbol $f^{(m)}(x;t)$ is similarly defined if all m-th order partials exists.

Note that these sums are analogous to the formula

$$f'(x;t) = \sum_{i-1}^{n} D_1 f(x) t_i$$

for the directional derivative of a function which is differentiable at x.

Theorem 5 ***(Taylor's formula):*** *Assume that f and all its partial derivatives of order $< m$ are differentiable at each point of an open set B in $\mathbb{R}^n$. If a and b are two points in S such that $L(a,b) \subseteq B$, then there is a point z on the line segment $L(a,b)$ such that*

$$f(b) - f(a) = \sum_{k=1}^{m-1} \frac{1}{k!} f^{\ (k)}(a; b-a) \ \ + \ \ \frac{1}{m!} f^{(m)}(z; b-a).$$

Now that we are armed with Taylor's formula for functions of several variables, we can return to our primary objective-to develop tests for maximizers (and minimizers) among the critical points of a function.

2.2 Basic definitions and existence

Definition 6 *The* ***r-neighborhood of x,*** *$B_r(x)$ is the set of all vectors y in $\mathbb{R}^n$ whose distance from x is less than r, that is*

$$B_r(x) = \{ y \in \mathbb{R}^n : d(x,y) < r \},$$

where $d(\cdot,\cdot)$ is the Euclidean distance.

Definition 7 *Suppose that $f(x)$ is a real-valued function defined on a subset $C \subset \mathbb{R}^n$. A point x^* in C is*

1. *a* ***global maximizer*** *for $f(x)$ if $f(x^*) \geq f(x)$ for all $x \in C$;*

2. *a* ***strict global maximizer*** *for* $f(x)$ *on* C *if* $f(x^*) > f(x)$ *for all* $x \in C$ *such that* $x \neq x^*$;

3. *a* ***local maximizer*** *for* $f(x)$ *if there is a positive number* δ *such that* $f(x^*) > f(x)$ *for all* $x \in C$ *for which* $x \in B_\delta(x^*)$;

4. *a* ***strict local maximizer*** *for* $f(x)$ *if there is a positive number* δ *such that* $f(x^*) > f(x)$ *for all* $x \in C$ *for which* $x \in B_\delta(x^*)$ *and* $x \neq x^*$;

5. *a* ***critical point*** *for* $f(x)$ *if the first partial derivatives of* $f(x)$ *exists at* x^* *and*

$$(\partial f/\partial x_i)(x^*) = 0, \quad i = 1, 2..., n.$$

Before we assert how to characterize and identify extrema, we have to be concerned with the existence of extrema per se.

Theorem 8 ***(Weierstrass or extreme value theorem)***
Suppose that $f(x)$ *is a continuous function defined on* C, *which is compact (i.e. closed and bounded) in* $\mathbb{R}^n$. *Then there exists a point* $x^* \in C$ *at which* f *has a maximum, and there exists a point* $x_* \in C$ *at which* f *has a minimum. Thus*

$$f(x_*) \leq f(x) \leq f(x^*)$$

for all $x \in C$.

2.3 Local and global extrema. Saddlepoints

2.3.1 Functions of one variable

Theorem 9 ***(Necessary condition for maximum in*** $\mathbb{R}$)
Suppose that $f(x)$ *is a differentiable function on an interval* I. *If* x^* *is a local maximizer of* $f(x)$, *then either* x^* *is an endpoint of* I *or* $f'(x^*) = 0$.

Theorem 10 ***(Second order sufficient conditions for a maximum in*** $\mathbb{R}$***)***

Suppose that $f(x), f'(x), f''(x)$ *are all continuous on an interval in* I *and that* $x^* \in I$ *is a critical point of* $f(x)$*.*

1. *If* $f''(x) \leq 0$ *for all* $x \in I$*, then* x^* *is a global maximizer of* $f(x)$ *on* I*.*
2. *If* $f''(\dot{x}) < 0$ *for all* $x \in I$*, such that* $x \neq x^*$*, then* x^**is a strict global maximizer of* $f(x)$ *on* I*.*
3. *If* $f''(x^*) < 0$ *then* x^* *is a strict local maximizer of* $f(x)$*.*

2.3.2 Functions of several variables

Theorem 11 ***(First order necessary conditions for a maximum in*** $\mathbb{R}^n$)

Suppose that $f(x)$ *is a real-valued function for which all first partial derivatives of* $f(x)$ *exists on an open subset* $B \subset \mathbb{R}^n$*. If* x^* *is a local maximizer of* $f(x)$*, then* x *is a critical point of* $f(x)$*, that is*

$$(\partial f/\partial x_i)(x^*) = 0, \quad i = 1, 2, ..., n.$$

Theorem 12 ***(Second order necessary conditions for a local maximum in*** $\mathbb{R}^n$)

Suppose, that $f(x)$ *is a real-valued function for which all first and second partial derivatives of* $f(x)$ *exists on an open subset* $B \subset \mathbb{R}^n$*. If* x^* *is a local maximizer of* $f(x)$*, then at* $x^*, H(f(x^*))$ *is negative semidefinite.*

Theorem 13 ***(Second order sufficient conditions for a strict local maximum in*** $\mathbb{R}^n$***)***

Suppose that $f(x)$ *is a real-valued function for which all first and second partial derivatives of* $f(x)$ *exists on an open subset* $B \subset \mathbb{R}^n$*and* x^**is a critical* *point of f. If $Hf(x^*)$ is negative definite, then f achieves a strict local maximum at x^*.*

So far we have described sufficient conditions for local extrema. We pursue the analysis now one step further and consider sufficient conditions for global extrema.

Theorem 14 *(Second order sufficient conditions for a (strict global maximum in $\mathbb{R}^n$)*
Suppose that x^* is a critical point of a function $f(x)$ with continuous first and second order partial derivatives on $\mathbb{R}^n$. Then:

1. x^* is a global maximizer for $f(x)$ if $(x - x^*) \cdot Hf(z) \cdot (x - x)^* \leq 0$ for all $x \in \mathbb{R}^n$ and all $z \in [x^*, x]$;

2. x^* is a strict global maximizer for $f(x)$if $(x - x) \cdot Hf(z) \cdot (x - x) < 0$ for all $x \in \mathbb{R}^n$ such that $x \neq x^*$and for all $z \in [x^*, x]$.

   In the terminology quadratic forms, we hence have

Theorem 15 *Suppose that x^* is a critical point of a function $f(x)$ with continuous first and second order partial derivatives on $\mathbb{R}^n$ and that $Hf(x)$ is the Hessian of $f(x)$. Then x^* is :*

1. A global maximizer for $f(x)$ if $Hf(x)$ is negative semidefinite on $\mathbb{R}^n$;

2. A strict global maximizer for $f(x)$ if $Hf(x)$ is negative definite on $\mathbb{R}^n$.

3 Concave and convex functions

Definition 16 *A function f defined on the convex set $C \subset \mathbb{R}^n$ is called **concave** if for every $x_1, x_2 \in C$ and $0 \leq \lambda \leq 1$, we have*

$$f(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

Definition 17 *A function f defined on the convex set $C \subset \mathbb{R}^n$ is called* ***strictly concave*** *if for every $x_1 \neq x_2$,and $0 < \lambda 1$, we have*

$$f(\lambda x_1 + (1-\lambda)x_2) > \lambda f(x_1) + (1-\lambda) f(x_2).$$

Remark 18 *If f is (strictly) concave then $g \equiv -f$ is a (strictly) convex function. We will henceforth concentrate on concave functions. All the results will also obtain with the obvious modifications for convex functions.*

4 Concave functions of one variable

Theorem 19 ***(Continuity of concave functions)***
Let f be concave function on the convex set $C \subset \mathbb{R}$. Then f is continuous on the interior of C.

Theorem 20 *Let f be a differentiable function on the open convex set $C \subset \mathbb{R}$. It is concave if and only if for every $x_0, x \in C$, we have*

$$f(x) \leq f(x_0) + f'(x_0)(x - x_0).$$

It is strictly concave if and only if the inequality is strict for $x \neq x_0$.

Theorem 21 *Let f be a differentiable function on the open convex set $C \subset \mathbb{R}$. It is concave (strictly concave) if and only if f' is a nonincreasing (decreasing) function.*

Theorem 22 ***(Concavity for C^2 functions)***
Let f be a function on the open convex set $C \subset \mathbb{R}$. Suppose that $f"$ exists on C. Then f is a concave function if and only if $f"(x) \leq 0$ for every $x \in C$. If $f"(x) < 0$ for every $x \in C$, then f is strictly concave.

5 Concave function of several variables

Definition 23 *Let f be a function defined on a convex set $C \subset \mathbb{R}^n$. The set*

$$H(f) = \{(x, \alpha) : x \in C,\ \alpha \in \mathbb{R}\,,\ f(x) \le \alpha\}$$

in $\mathbb{R}^{n+1}$ is called the **hypograph** of f. Similarly, the **epigraph** of f is the set

$$E(f) = \{(x, \alpha) : x \in C, \alpha \in \mathbb{R},\ f(x) \ge \alpha\}.$$

Theorem 24 *Let f be a function defined on a convex set $C \subset \mathbb{R}^n$. Then f is concave if and only if its hypograph $H\,(f)$ is a convex set. Similarly, f is convex if and only if its epigraph $E(f)$ is a convex set.*

The initial definition and the equivalent characterization in terms of its hypograph are communicating the same geometric concept, that, for any two points above the curve (surface) the line segment joining the two points lies entirely below the curve (surface) between the two points.

Definition 25 *For any function f on C and any $\alpha \in \mathbb{R}$ the set $U(f, \alpha)$ defined by*

$$U(f, \alpha) = \{x : x \in C, f(x) \ge \alpha\}$$

is called the ***upper level*** *(or contour set of. The set*

$$L(f, \alpha) = \{x : x \in C, f(x) \le \alpha\}$$

is called the ***lower level*** *(or contour) set of f. The set*

$$Y(f, \alpha) = \{x : x \in C, f(x) = \alpha\}$$

is called the ***level surface*** *of f at α.*

Corollary 26 *Let f be a concave (convex) function on $C \subset \mathbb{R}^n$. Then its upper (lower) level sets are convex sets for every real number α.*

Theorem 27 (***Concavity for*** $\mathbf{C}^1$***functions***)
Let f be a differentiable function on the open convex set $C \subset \mathbb{R}^n$. It is concave if and only if for every $x_0, x \in C$ we have

$$f(x) \leq f(x_0) + (x - x_0)^T \nabla f(x_0).$$

It is strictly concave if and only if the inequality is strict for $x \neq x_0$.

Theorem 28 ***(Concavity for*** C^2 ***functions)***
Let f be a twice differentiable function on an open convex set $C \subset \mathbb{R}^n$.Then f is concave if and only if its Hessian matrix is negative semidefinite for every $x \in C$. That is, for every $x \in C$ and $y \in \mathbb{R}^n$, we have

$$y^T H(x) y \leq 0$$

If $H(x)$ is negative definite for every $x \in C$, then f is strictly concave.

Now we state a few results on the closedness of concave functions under functional operations.

Theorem 29 *Let f be concave function and let λ be a nonnegative number. Then $F(x) = \lambda f(x)$ is also a concave function. Let $f_1$and f_2 be concave functions. Then $F(x) = f_1(x) + f_2(x)$ is also concave.*

6 Extrema of concave functions

The importance of concave (and convex) functions from the optimization point of view lies in some properties of concave functions with regard to their extrema.

Theorem 30 ***Local-global property of the maximum***
Let f be a concave function defined on a convex set $C \subset \mathbb{R}^n$. Then every local maximum of f at $x^* \in C$ is a global maximum of f over all C.

Theorem 31 *The set of points at which a concave function f attains its maximum over C is a convex set.*

Corollary 32 *Let f be a strictly concave function, defined on the convex set $C \subset \mathbb{R}^n$. If f attains its maximum at $x^* \in C$, this maximizing point is unique.*

Theorem 33 ***Sufficiency condition for global extrema***
Let f be a differentiable concave (strictly concave) function on the convex set C. If

$$\nabla f(x^*) = 0$$

at a point $x^* \in C$, then f attains its maximum (unique maximum at x^*).

7 Quasi-concave and quasi-convex functions

We begin generalizing concave functions by recalling from the preceding chapter that the upper-level sets of concave functions are convex sets. Concavity of a function is a sufficient condition for this property, but not a necessary one. We define a family of functions by the convexity of their upper-level sets. Such functions are called quasi-concave functions. They are generalized concave functions, since it is easy to show that every concave function is quasiconcave, but not conversely.

7.1 Quasi-concave functions

Definition 34 *Let f be defined on the convex set $C \subset \mathbb{R}^n$. It is said to be* ***quasiconcave*** *if its upper-level sets*

$$U(f, \alpha) = \{x : x \in C, f(x) \geq \alpha\}$$

are convex sets for every real α. Similarly, f is said to be ***quasiconvex*** *if its lower-level set*

$$L(f, \alpha) = \{x : x \in C, f(x) \leq \alpha\}$$

are convex sets for every real α.

Let us mention here a family of functions that can be viewed as a generalization of single variable monotonic functions on the one hand, and of affine functions, on the other hand.

Definition 35 *A function is said to be* ***quasimonotonic*** *if it is both quasiconcave and quasiconvex.*
For such a function, clearly, both its upper- and lower-level sets are convex, and for every x_1, x_2 in the convex domain of f and for every $0 \leq \lambda \leq 1$

$$\max[f(x_1), f(x_2)] \geq f(\lambda x_1 + (1-\lambda)x_2) \geq \min[f(x_1), f(x_2)].$$

Theorem 36 *Let f be defined on the convex set $C \subset IR^n$. It is a quasiconcave function if and only if*

$$f(\lambda x_1 + (1-\lambda)x_2) \geq \min[f(x_1), f(x_2)]$$

for every $x_1, x_2 \in C$,and $0 \leq \lambda \leq 1$.

Note that the preceding characterization of quasiconcavity is identical to the following one

$$f(x_1) \geq f(x_2) \Rightarrow f(\lambda x_1 + (1-\lambda)x_2) \geq f(x_2)$$

It is also clear that a concave function is also quasiconcave, since

$$f(\lambda x_1 + (1-\lambda)x_2) \geq f(x_1) + (1-\lambda)f(x_2) \geq \min\left[f(x_1), f(x_2)\right].$$

The most important property of quasiconcave functions for microeconomic theory is, however, the following:

Theorem 37 *If $f : \mathbb{R}^n \to \mathbb{R}$ is quasiconcave and if $g : \mathbb{R} \to \mathbb{R}$ is strictly increasing, then $g(f(x))$ is also quasiconcave.*

The theorem above is in contrast to the behavior of concave functions. Its importance for microeconomics stems from the fact that in consumer theory, preferences of a consumer identify the level sets of any utility function representing the preferences, but not the numerical values of the utility levels. Convexity of preference is equivalent to quasiconcavity of the utility representation and hence the theorem above states that any increasing function of a given utility function is a representation of the same preferences.

Theorem 38 ***(Local-global property of the maximum)***
Let f be a strictly quasiconcave function defined on the convex set $C \subset \mathbb{R}^n$. If $x^* \in C$ is a local maximum of f, then x^* is also a strict global maximum of f on C. The set of points at which f attains its global maximum over C is a convex set.

7.2 Differentiable quasi-concave functions

Theorem 39 ***(Quasiconcavity for C^1 functions)***
Let f be differentiable on the open convex set $C \subset \mathbb{R}^n$. Then f is quasiconcave if and only if for every $x_1, x_2 \in C$

$$f(x_1) \geq f(x_2) \Rightarrow (x_1 - x_2)^T \nabla f(x_2) \geq 0.$$

Theorem 40 *Let f be a twice differentiable quasiconcave function on the open convex set $C \subset \mathbb{R}^n$. If $x_0 \in C, v \in \mathbb{R}^n$,*

$$v^T \nabla f(x_0) = 0 \Rightarrow v^T \nabla^2 f(x_0) v \leq 0.$$

The following strengthening of the theorem allows a complete characterization of quasiconcave function for case that $\nabla f(x) \neq 0$ for every $x \in C$.

Theorem 41 *Let f be a twice differentiable quasiconcave function on the open convex set $c \subset \mathbb{R}^n$ and suppose that $\nabla f(x) \neq 0$ for every $x \in C$. Then f is quasiconcave if and only if $x \in C, v \in \mathbb{R}^n$,*

$$v^T \nabla f(x_0) = 0 \Rightarrow v^T \nabla^2 f(x_0) v \leq 0.$$

We conclude by mentioning some more necessary and sufficient conditions for quasiconcavity of twice differentiable functions - this time in terms of "bordered determinants" or "bordered Hessians."

Theorem 42 ***(Quasiconcavity for C^2 functions)***

1. *Let f be a twice differentiable function on the open convex set $C \subset \mathbb{R}^n_+$ (the non-negative orthant). Define the determinants $\mathcal{D}_k(x), k+1, ..., n$ by*

$$\mathcal{D}_k(x) = \begin{vmatrix} 0 & \frac{\partial f}{\partial x_i} & \cdots & \frac{\partial f}{\partial x_k} \\ \frac{\partial f}{x_I} & \frac{\partial^2 f}{\partial s_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_k} \\ \cdots & & & \cdots \\ \frac{\partial f}{\partial x_k} & \frac{\partial^2 f}{\partial x_k \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_k \partial x_k} \end{vmatrix}.$$

   A necessary condition for f to be quasi-concave is that $(-1)^k \mathcal{D}_k(x) \geq 0$ for all $k = 1, ..., n$ and all $x \in C$.

2. *A sufficient condition for f to be quasi-concave is that $(-1)^k \mathcal{D}_k(\dot{x}) > 0$ for all $k = 1, ..., n$ and all $x \in C$.*

7.3 Strict Quasiconcavity

Definition 43 *Let f be defined on the convex set $C \subset \mathbb{R}^n$. It is said to be strictly quasiconcave if*

$$f(\lambda x_1 + (1-\lambda) x_2)^{\cdot} > \min[f(x_1), f(x_2)]$$

for every $x_1, x_2 \in C, x_1 \neq x_2$, and $0 < \lambda < 1$. If f is strictly quasiconcave, then $g \equiv -f$ is a strictly quasiconvex function.[2]

What is the difference between quasiconcave and strictly quasiconcave functions. A function that is quasiconcave, but not strictly quasiconcave is constant on some interval of its domain of definition. Note also that strict quasiconcavity is not a proper generalization of concavity, but only of strict concavity.

Theorem 44 *Let f be a continuous function, defined on $\mathbb{R}^n$. If f is strictly quasiconcave, then its upper level sets are strictly convex.*

An important property shared by concave and strictly quasiconcave functions is that every local maximum is a global one. This property however holds for more general families of functions as well. The family of functions we will be concerned with now lies, subject to some continuity requirements, between the families of quasiconcave and strictly quasiconcave functions.

Let us recall that we can define quasiconcave functions by the condition

$$f(x_1) \geq f(x_2) \Rightarrow f(\lambda x_1 + (1 - \lambda)x_2) > f(x_2)$$

for every $0 \leq \lambda \leq 1$. Similarly, the condition for a strictly quasiconcave function can be written as

$$f(x_1) \geq f(x_2) \Rightarrow f(\lambda x_1 + (1 - \lambda)x_2) > f(x_2)$$

for every $x_1 \neq x_2$ and $0 < \lambda < 1$.

[2] Currently there are several competing definitions of strict quasiconcavity. This one is the one most commonly used in economics.

8 Constrained Optimization

9 Introduction to nonlinear programming

The solution of a nonlinear programming problem consists of finding an optimal solution vector x^*. Recognizing an optimal x^* and studying its properties form the central theme of this part of the course.

We shall see that if a vector a candidate for an optimal solution, it must satisfy certain **necessary conditions** of optimality. There may however be vectors other than the optimal ones that also satisfy these conditions. Hence, necessary conditions are primarily useful in the negative sense: if a vector x does not satisfy them, it cannot be an optimal solution. To verify optimality, we may, therefore, look for **sufficient conditions** of optimality that, if satisfied together with the necessary ones, give a clear indication of the nature of particular solution vector under consideration.

10 First-order necessary conditions for inequality constrained extrema

We begin by stating the most general mathematical program to be discussed in this section:

$$\max f(x) \tag{1}$$

subject to constraints

$$g_i(x) = 0 \qquad i = 1, ..., k$$

$$h_j(x) \leq 0 \qquad j = 1, ..., m$$

The functions $f, g_1, ..., g_k, h_1, ..., h_m$ are assumed to be defined and differentiable on some open set $B \subseteq \mathbb{R}^n$. Let $X \subseteq B$ denote the feasible set for problem (1). If the feasible set is nonempty, the program is called a **consistent** program.

Definition 45 *A vector $z \neq 0$ is called a **feasible direction vector** from x^* if there exists a number $\delta > 0$ such that $(x^* + \alpha z) \in X \bigcap N_\delta(x^*)$for all $0 \leq \alpha \leq \delta / \|z\|$.*

Let us characterize the feasible direction vectors in terms of the constraint functions g_i and h_j. Define

$$I(x^*) = \{i : h_i(x^*) = 0\}$$

Define furthermore

$$Z(x^*) = \{z : z\nabla g_j(x^*) = 0, j \in 1, ...k, z\nabla h_j(x^*) \leq 0,\ i \in I(x^*).$$

We can show that for z to be a feasible direction vector from x^*, then $z \in Z(x^*)$. Note that the set $Z(x^*)$ is a cone. It is also called the linearizing cone of X at x^*, since it is generated by linearizing the constraint functions at x^*. We will say that for any x for which $Z(x) \neq \{0\}$, the **Kuhn-Tucker constraint qualification (KTQC)** is satisfied.[3] We can now state the main result of this section, which is a direct extension of the Kuhn-Tucker necessary conditions.

Theorem 46 ***(Generalized Kuhn-Tucker necessary conditions)*** *Let x^* be a solution of problem (1) and suppose that x^* satisfies the KTCQ. Then there exist vectors $\lambda^* = (\lambda_1^*, ..., \lambda_m^*)$ and $\mu^* = (\mu_1^*, ..., \mu_p^*)$ such that*

$$\nabla_x L(x^*, \lambda^*, \mu^*) \equiv \nabla f(x^*) - \sum_{i=1}^{k} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*) = 0,$$

[3]The condition mentioned here is not the original Kuhn-Tucker constraint qualification since the original covered only inequality constraints, here we are covering mixed constraints, i.e. inequality as well as equality constraints.

and

$$\mu_j^* h_j(x^*) = 0, \qquad i = 1, ..., m, \qquad \mu^* \geq 0.$$

No specific assumptions, except differentiability, were made so far on the type of functions involved in (1). Further assumptions on these functions lead to special forms of the Kuhn-Tucker conditions. We present one special case. In many applications (e.g. general equilibrium theory), the variables x_j are required to be nonnegative. Suppose that in addition to the constraints incorporated in g_i and h_j, we also require

$$x \geq 0 \tag{2}$$

The necessary conditions for this case can be stated as follows:

Theorem 47 ***(Generalized Kuhn-Tucker nec. conditions with non-negativity constraints)***
Let x^* be a solution of problem (1) *and the added constraint* (2)*. Suppose that x^* satisfies the KTCQ. Then there exist vectors $\lambda^* = (\lambda_1^*, ..., \lambda_k^*)$ and $\mu^* = (\mu_1^*, ..., \mu_m^*)$ such that*

$$\nabla_x L(x^*, \lambda^*, \mu^*) \equiv \nabla f(x^*) - \sum_{i=1}^{k} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*) \leq 0,$$

$$\mu_j^* h_j(x^*) = 0, \qquad \mu^* \geq 0,$$

and

$$(x^*)\left[\nabla f(x^*) - \sum_{i=1}^{k} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*)\right] = 0.$$

11 Concave programming

Here we consider concave programs, a special case of the general nonlinear program (1). The optimality conditions derived there become simpler for concave programs. Consider, then, the following nonlinear program, called a **concave program**.

$$\max f(x) \tag{P}$$

subject to the constraints

$$g_i(x) = 0 \qquad i = 1, ..., k.$$

$$h_j(x) \leq 0 \qquad j = l, ..., m.$$

The functions $f, g_1, ..., g_k, h_1, ..., h_m$ are assumed to be defined and differentiable on some open set $B \subseteq \mathbb{R}^n$, where f is a concave function on IR^n, the h_j are quasiconvex functions and the h_j are affine functions of the form

$$g_i(x) = \sum_{k=1}^{n} a_{ik} x_k - b_i.$$

Note that the feasible set X for problem (P) is convex since the set of all $x \in \mathbb{R}^n$ satisfying the inequalities and the equations is a convex set. Note that generally the set of $x \in \mathbb{R}^n$ satisfying an equation of $h(x) = 0$, where h is a nonlinear convex or concave function, is not a convex set. We shall now show that the Kuhn-Tucker necessary conditions for optimality are also sufficient when applied to a concave program.

Theorem 48 ***(Kuhn-Tucker sufficient conditions for optimality)*** *Suppose that the functions f and $g_1, ..., g_k$ are real-valued, concave and quasiconvex functions on $\mathbb{R}^n$, respectively, and let $h_1, ..., h_m$ be linear. If there exist vectors $x^*, \lambda^*, \mu^*, \quad with\, x^*, \lambda^*, \mu^*$, with $x^* \in X$ and*

$$\nabla_x L(x^*, \lambda^*, \mu^*) \equiv \nabla f(x^*) - \sum_{i=1}^{k} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{m} \mu_j^* \nabla h_j\,(x^*) = 0$$

$$\mu_j^* h_j(x^*) = 0 \qquad i \quad = \quad 1, ..., m \qquad \mu^* \geq 0$$

then x^* is a global optimum of (P).

The Kuhn-Tucker constraint qualification assumed in the necessary conditions for optimality in the general case can be replaced in convex programs by a simpler and easier computable, although stronger condition. Suppose all the constraint are inequality constraints. Then if there exists a point $\hat{x} \in C$ such that

$$h_j(\hat{x}) < 0 \qquad j = 1, ..., m$$

then the program is called strongly consistent. The strong consistency condition is also known as **Slater's constraint qualification.**

12 Interpretation of the Multipliers

12.1 Envelope Theorem

Let $x(b)$ solve the equality constrained programming problem:

$$\max f(x, b)$$

$$s.t.\ g(x, b) = 0$$

where f, g are C^1 functions $f, g : \mathbb{R}^{n+k} \to \mathbb{R}^m$.

Define the value function to the problem to be the function $v : \mathbb{R}^k \to \mathbb{R}$ defined by $v(b) = f(x(b), b)$. An application of the chain rule to $v(b) = f(h(b))$ where $h(b) = (x(b), b)$ yields the following expressions for the partial derivatives of v :

$$\frac{\partial v}{\partial b_i} = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j} \cdot \frac{\partial x_j}{\partial b_i} + \frac{\partial f}{\partial b_i}. \tag{3}$$

From the first order conditions for $x(b)$ we know that:

$$\frac{\partial f\left(x\left(b\right),b\right)}{\partial x_j}=\sum_{k=i}^{m}\lambda_k\frac{\partial g_k\left(x\left(b\right),b\right)}{\partial x_j}. \tag{4}$$

Since $x\left(b\right)$ is feasible for all b,

$$\sum_{j=1}^{n}\frac{\partial g_k\left(x\left(b\right),b\right)}{\partial x_j}\cdot\frac{\partial x_j}{\partial b_i}=-\frac{\partial g_k\left(x\left(b\right),b\right)}{\partial b_i}. \tag{5}$$

Combining**??** we get:

$$\frac{\partial v}{\partial b_i}=\sum_{j=1}^{n}\sum_{k=i}^{m}\lambda_k\frac{\partial g_k\left(x\left(b\right),b\right)}{\partial x_j}\frac{\partial x_j}{\partial b_i}+\frac{\partial f}{\partial b_i}.$$

Interchanging the order of summation and using 5, we get:

Theorem 49 *(The Envelope Theorem)*

$$\frac{\partial v\left(x\left(b\right),b\right)}{\partial b_i}=\sum_{k=1}^{m}-\lambda_k\frac{\partial g_k\left(x\left(b\right),b\right)}{\partial b_i}+\frac{\partial f}{\partial b_i}.$$

Remark 50 *In words, the theorem tells that only direct effects on the constraints and the objective function from the change in a parameter produce first order changes in the value function.*

Envelope theorem allows us to interpret the multipliers of the problem in an economically meaningful way:

Suppose that the maximization problem is of the following special form:

$$\max f\left(x\right)$$

$$s.t.\ g_k\left(x\right)=b_k \text{ for } k=1,...,m.$$

Then an application of the envelope theorem yields:

$$\frac{\partial v\left(x\left(b\right),b\right)}{\partial b_i} = \lambda_k.$$

In words, the multiplier equals the marginal addition to the value function as one of the constraints is relaxed. If we take each constraint to represent the amount of a scarce resource that the decision maker has available to him, λ_k has a natural interpretation as a shadow price of resource k. If the decision maker was facing an unconstrained optimization problem, where he could choose both the resources and the optimal x's, then at prices λ_k for the resources, his optimal choice would coincide with his optimal choice in the original constrained problem.