ECONOMICS 594: LECTURE NOTES

CHAPTER 2: CONVEX SETS AND CONCAVE FUNCTIONS

W. Erwin Diewert January 31, 2008.

1. Introduction

Many economic problems have the following structure: (i) a linear function is minimized subject to a nonlinear constraint; (ii) a linear function is maximized subject to a nonlinear constraint or (iii) a nonlinear function is maximized subject to a linear constraint. Examples of these problems are: (i) the producer's cost minimization problem (or the consumer's expenditure minimization problem); (ii) the producer's profit maximization problem and (iii) the consumer's utility maximization problem. These three constrained optimization problems play a key role in economic theory.

In each of the above 3 problems, linear functions appear in either the objective function (the function being maximized or minimized) or the constraint function. If we are maximizing or minimizing the linear function of x, say $\sum_{n=1}^{N} p_n x_n \equiv p^T x$, where $p \equiv (p_1,\ldots,p_N)$ [1] is a vector of prices and $x \equiv (x_1,\ldots,x_N)$ is a vector of decision variables, then after the optimization problem is solved, the optimized objective function can be regarded as a function of the price vector, say G(p), and perhaps other variables that appear in the constraint function. Let F(x) be the nonlinear function which appears in the constraint in problems (i) and (ii). Then under certain conditions, the optimized objective function G(p) can be used to reconstruct the nonlinear constraint function F(x). This correspondence between F(x) and G(p) is known as *duality theory*. In the following chapter, we will see how the use of dual functions can greatly simplify economic modeling.

However, the mathematical foundations of duality theory rest on the theory of convex sets and concave (and convex) functions. Hence, we will study a few aspects of this theory in the present chapter before studying duality theory in the following chapter.

2. Convex Sets

Definition: A set S in R^N (Euclidean N dimensional space) is *convex* iff (if and only if):

(1) $x^1 \in S, x^2 \in S, 0 < \lambda < 1$ implies $\lambda x^1 + (1-\lambda)x^2 \in S$.

Thus a set S is convex if the line segment joining any two points belonging to S also belongs to S. Some examples of convex sets are given below.

Example 1: The *ball of radius 1* in R^N is a convex set; i.e., the following set B is convex:

[1] Our convention is that in equations, vectors like p and x are regarded as column vectors and p^T and x^T denote their transposes, which are row vectors. However, when defining the components of a vector in the text, we will usually define p more casually as $p \equiv (p_1,\ldots,p_N)$.

(2) $B \equiv \{x: x \in R^N; (x^T x)^{1/2} \leq 1\}$.

To show that a set is convex, we need only take two arbitrary points that belong to the set, pick an arbitrary number λ between 0 and 1, and show that the point $\lambda x^1 + (1-\lambda)x^2$ also belongs to the set. Thus let $x^1 \in B$, $x^2 \in B$ and let λ be such that $0 < \lambda < 1$. Since $x^1 \in B$, $x^2 \in B$, we have upon squaring, that x^1 and x^2 satisfy the following inequalities:

(3) $x^{1T}x^1 \leq 1$; $x^{2T}x^2 \leq 1$.

We need to show that:

(4) $(\lambda x^1 + (1-\lambda)x^2)^T(\lambda x^1 + (1-\lambda)x^2) \leq 1$.

Start off with the left hand side of (4) and expand out the terms:

(5) $(\lambda x^1 + (1-\lambda)x^2)^T(\lambda x^1 + (1-\lambda)x^2) = \lambda^2 x^{1T}x^1 + 2\lambda(1-\lambda)x^{1T}x^2 + (1-\lambda)^2 x^{2T}x^2$
$\leq \lambda^2 + 2\lambda(1-\lambda)x^{1T}x^2 + (1-\lambda)^2$
where we have used (3) and $\lambda^2 > 0$ and $(1-\lambda)^2 > 0$
$\leq \lambda^2 + 2\lambda(1-\lambda)(x^{1T}x^1)^{1/2}(x^{2T}x^2)^{1/2} + (1-\lambda)^2$
using $\lambda(1-\lambda) > 0$ and the Cauchy Schwarz inequality
$\leq \lambda^2 + 2\lambda(1-\lambda) + (1-\lambda)^2$ using (3)
$= [\lambda + (1-\lambda)]^2$
$= 1$

which establishes the desired inequality (4).

The above example illustrates an important point: in order to prove that a set is convex, it is often necessary to verify that a certain *inequality* is true. Thus when studying convexity, it is useful to know some of the most frequently occurring inequalities, such as the Cauchy Schwarz inequality and the Theorem of the Arithmetic and Geometric Mean.

Example 2: Let $b \in R^N$ (i.e., let b be an N dimensional vector) and let b_0 be a scalar. Define the set

(6) $S \equiv \{x : b^T x = b_0\}$.

If N = 2, the set S is *straight line*, if N = 3, the set S is a *plane* and for N > 3, the set S is called a *hyperplane*. Show that S is a convex set.

Let $x^1 \in S$, $x^2 \in S$ and let λ be such that $0 < \lambda < 1$. Since $x^1 \in S$, $x^2 \in S$, we have using definition (6) that

(7) $b^T x^1 = b_0$; $b^T x^2 = b_0$.

We use the relations (7) in (8) below. Thus

(8) $b_0 = \lambda b_0 + (1-\lambda)b_0$
$= \lambda b^T x^1 + (1-\lambda)b^T x^2$ using (7)
$= b^T[\lambda x^1 + (1-\lambda)x^2]$ rearranging terms

and thus using definition (6), $[\lambda x^1 + (1-\lambda)x^2] \in S$. Thus S is a convex set.

Example 3: Let $b \in R^N$ and let b_0 be a scalar. Define a *halfspace* H as follows:

(9) $H \equiv \{x : b^T x \leq b_0\}$.

A halfspace is equal to a hyperplane plus the set which lies on one side of the hyperplane. It is easy to prove that a halfspace is a convex set: the proof is analogous to the proof used in Example 2 above except that now we also have to use the fact that $\lambda > 0$ and $1-\lambda > 0$.

Example 4: Let S^j be a convex set in R^N for $j = 1,\ldots,J$. Then assuming that the intersection of the S^j is a nonempty set, this *intersection set* is also a convex set; i.e.,

(10) $S \equiv \cap_{j=1}^{J} S^j$

is a convex set.

To prove this, let $x^1 \in \cap_{j=1}^{J} S^j$ and let $x^2 \in \cap_{j=1}^{J} S^j$ and let $0 < \lambda < 1$. Since S^j is convex for each j, $[\lambda x^1 + (1-\lambda)x^2] \in S^j$ for each j. Therefore $[\lambda x^1 + (1-\lambda)x^2] \in \cap_{j=1}^{J} S^j$. Hence S is also a convex set.

Example 5: The *feasible region* for a linear programming problem is the following set S:

(11) $S \equiv \{x: Ax \leq b \, ; x \geq 0_N\}$

where A is an M by N matrix of constants and b is an M dimensional vector of constants.

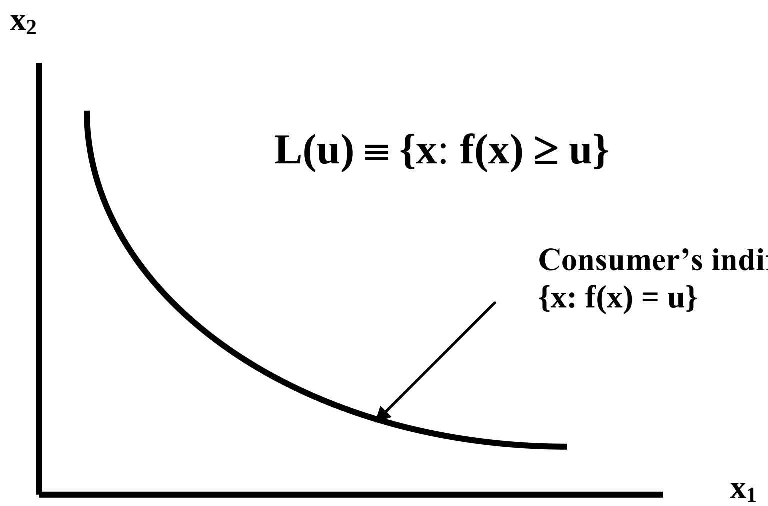
It is easy to show that the set S defined by (11) is a convex set, since the set S can be written as follows:

(12) $S = [\cap_{m=1}^{M} H^m] \cap \Omega$

where $H^m \equiv \{x: A_{m\cdot}x \leq b_m\}$ for $m = 1,\ldots,M$ is a halfspace ($A_{m\cdot}$ is the mth row of the matrix A and b_m is the mth component of the column vector b) and $\Omega \equiv \{x: x \geq 0_N\}$ is the nonnegative orthant in N dimensional space. Thus S is equal to the intersection of M+1 convex sets and hence using the result in Example 4, is a convex set.

Example 6: Convex sets occur in economics frequently. For example, if a consumer is maximizing a utility function f(x) subject to a budget constraint, then we usually assume that the upper level sets of the function are convex sets; i.e., for every utility level u in the range of f, we usually assume that the *upper level set* $L(u) \equiv \{x: f(x) \geq u\}$ is convex. This upper level set consists of the consumer's indifference curve (or surface if $N > 2$) and the set of x's lying above it.

Figure 1: A Consumer's Upper Level Set L(u)



It is also generally assumed that production functions f(x) have the property that the set of inputs x that can produce at least the output level y (this is the upper level set $L(y) \equiv \{x: f(x) \geq y\}$) is a convex set for every output level y that belongs to the range of f. Functions which have this property are call *quasiconcave*. We will study these functions later in this chapter.

3. The Supporting Hyperplane Theorem for Closed Convex Sets

The results in this section are the key to duality theory in economics as we shall see later.

Definition: A point $x^0 \in S$ is an *interior point* of S iff there exists $\delta > 0$ such that the open ball of radius δ around the point x^0, $B_\delta(x^0) \equiv \{x: (x-x^0)^T(x-x^0) < \delta^2\}$, also belongs to S; i.e., $B_\delta(x^0) \subset S$.

Definition: A set S is *open* iff it consists entirely of interior points.

Definition: A set S is *closed* iff for every sequence of points, $\{x^n: n = 1,2,\ldots,\}$ such that $x^n \in S$ for every n and $\text{limit}_{n\to\infty} x^n \equiv x^0$ exists, then $x^0 \in S$ as well.

Definition: The *closure* of a set S, Clo S, is defined as the set $\{x: x = \text{limit}_{n\to\infty} x^n, x^n \in S, n = 1,2,\ldots\}$. Thus the closure of S is the set of points belonging to S plus any additional points that are limiting points of a sequence of points that belong to S. Note that for every set S, Clo S is a closed set and $S \subset \text{Clo } S$.

Definition: x^0 is a *boundary point* of S iff $x^0 \in \text{Clo } S$ but x^0 is not an interior point of S.

Definition: Int S is defined to be the *set of interior points* of S. If there are no interior points of S, then Int $S \equiv \emptyset$, the empty set.

Theorem 1: Minkowski's (1911) *Theorem*: Let S be a closed convex set in R^N and let b be a point which does not belong to S. Then there exists a nonzero vector c such that

$$(13)\ c^Tb < \min_x \{c^Tx : x \in S\} ;$$

i.e., there exists a hyperplane passing through the point b which lies entirely below the convex set S.

Proof: Since $(x-b)^T(x-b) \geq 0$ for all vectors x, it can be seen that $\min_x \{(x-b)^T(x-b) : x \in S\}$ exists. Let x^0 be a boundary point of S which attains this minimum; i.e.,

$$(14)\ (x^0-b)^T(x^0-b) = \min_x \{(x-b)^T(x-b) : x \in S\}.$$

Now pick an arbitrary $x \in S$ and let $0 < \lambda < 1$. Since both x and x^0 belong to S, by the convexity of S, $\lambda x + (1-\lambda)x^0 \in S$. Now use (14) to conclude that

$$\begin{aligned}(15)\ (x^0-b)^T(x^0-b) &\leq ([\lambda x + (1-\lambda)x^0]-b)^T([\lambda x + (1-\lambda)x^0]-b) \\ &= (x^0-b + \lambda[x-x^0])^T(x^0-b + \lambda[x-x^0]) \\ &= (x^0-b)^T(x^0-b) + 2\lambda(x^0-b)^T(x-x^0) + \lambda^2(x-x^0)^T(x-x^0).\end{aligned}$$

Define the vector c as

$$(16)\ c \equiv x^0 - b .$$

If c were equal to 0_N, then we would have $b = x^0$. But this is impossible because $x^0 \in S$ and b was assumed to be exterior to S. Hence

$$(17)\ c \neq 0_N.$$

The inequality (17) in turn implies that

$$(18)\quad 0 < c^Tc = (x^0-b)^T(x^0-b) = c^T(x^0-b) \text{ or}$$
$$(19)\ c^Tb < c^T x^0.$$

Rearranging terms in the inequality (15) leads to the following inequality that holds for all $x \in S$ and $0 < \lambda < 1$:

(20) $0 \le 2\lambda(x^0-b)^T(x-x^0) + \lambda^2(x-x^0)^T(x-x^0)$.

Divide both sides of (20) by $2\lambda > 0$ and take the limit of the resulting inequality as λ tends to 0 in order to obtain the following inequality:

(21) $0 \le (x^0-b)^T(x-x^0)$ for all $x\in S$
$= c^T(x-x^0)$ for all $x\in S$ using definition (16) or
(22) $c^Tx^0 \le c^Tx$ for all $x\in S$.

Putting (22) and (19) together, we obtain the following inequalities, which are equivalent to the desired result, (13):

(23) $c^Tb < c^T x^0 \le c^Tx$ for all $x\in S$. Q.E.D.

Theorem 2: Minkowski's Supporting Hyperplane Theorem: Let S be a convex set and let b be a boundary point of S. Then there exists a hyperplane through b which supports S; i.e., there exists a nonzero vector c such that

(24) $c^Tb = \min_x \{c^Tx : x\in \text{Clo } S\}$.

Proof: Let $b^n \notin \text{Clo } S$ for $n = 1,2,\ldots$ but let the limit $_{n\to\infty}$ $b^n = b$; i.e., each member of the sequence of points b^n is exterior to the closure of S but the limit of the points b^n is the boundary point b. By Minkowski's Theorem, there exists a sequence of nonzero vectors c^n such that

(25) $c^{nT}b^n < \min_x \{c^{nT}x : x\in \text{Clo } S\}$; $n = 1,2,\ldots$.

There is no loss of generality in normalizing the vectors c^n so that they are of unit length. Thus we can assume that $c^{nT}c^n = 1$ for every n. By a Theorem in analysis due to Weierstrass,[2] there exists a subsequence of the points $\{c^n\}$ which tends to a limit, which we denote by c. Along this subsequence, we will have $c^{nT}c^n = 1$ and so the limiting vector c will also have this property so that $c \ne 0_N$. For each c^n in the subsequence, (25) will be true so that we have

(26) $c^{nT}b^n < c^{nT}x$; for every $x\in S$.

Thus

(27) $c^Tb = \text{limit}_{n\to\infty \text{ along the subsequence}}\ c^{nT}b^n$
$\le \text{limit}_{n\to\infty \text{ along the subsequence}}\ c^{nT}x$ for every $x\in S$ using (26)
$= c^Tx$ for every $x\in S$

which is equivalent to the desired result (24). Q.E.D.

[2] See Rudin (1953; 31).

The halfspace $\{x: c^T b \le c^T x\}$ where b is a boundary point of S and c was defined in the above Theorem is called a *supporting halfspace* to the convex set S at the boundary point b.

Theorem 3: Minkowski's Theorem Characterizing Closed Convex Sets: Let S be a closed convex set that is not equal to R^N. Then S is equal to the intersection of its supporting halfspaces.

Proof: If S is closed and convex and not the entire space R^N, then by the previous result, it is clear that S is contained in each of its supporting halfspaces and hence is a subset of the intersection of its supporting halfspaces. Now let $x \notin S$ so that x is exterior to S. Then using the previous two Theorems, it is easy to see that x does not belong to at least one supporting halfspace to S and thus x does not belong to the intersection of the supporting halfspaces to S. Q.E.D.

Problems

1. Let $A = A^T$ be a positive definite N by N symmetric matrix. Let x and y be N dimensional vectors. Show that the following generalization of the Cauchy Schwarz inequality holds:

(a) $(x^T A y)^2 \le (x^T A x)(y^T A y)$.

Hint: You may find the concept of a *square root matrix* for a positive definite matrix helpful. From matrix algebra, we know that every symmetric matrix has the following eigenvalue-eigenvector decomposition with the following properties: there exist N by N matrices U and Λ such that

(b) $U^T A U = \Lambda$;
(c) $U^T U = I_N$

where Λ is a diagonal matrix with the eigenvalues of A on the main diagonal and U is an orthonormal matrix. Note that U is the inverse of U^T. Hence premultiply both sides of (b) by U and postmultiply both sides of (b) by U^T in order to obtain the following equation:

(d) $A = U\Lambda U^T$
$= U\Lambda^{1/2}\Lambda^{1/2}U^T$ where we use the assumption that A is positive definite and we define $\Lambda^{1/2}$ to be a diagonal matrix with diagonal elements equal to the positive square roots of the diagonal elements of Λ (which are the positive eigenvalues of A, $\lambda_1,\ldots,\lambda_N$.
$= U\Lambda^{1/2}\, U^T U\Lambda^{1/2}U^T$ using (c)
$= B^T B$

where the N by N *square root* matrix B is defined as

(e) $B \equiv U\Lambda^{1/2} U^T$.

Note that B is symmetric so that

(f) $B = B^T$

and thus we can also write A as

(g) $A = BB$.

2. Let $A = A^T$ be a positive definite N by N symmetric matrix. Let x and y be N dimensional vectors. Show that the following generalization of the Cauchy Schwarz inequality holds:

$(x^Ty)^2 \leq (x^TAx)(y^TA^{-1}y)$.

3. Let $A = A^T$ be a positive definite symmetric matrix. Define the set $S \equiv \{x : x^TAx \leq 1\}$. Show that S is a convex set.

4. A set S in R^N is a *cone* iff it has the following property:

(a) $x \in S$, $\lambda \geq 0$ implies $\lambda x \in S$.

A set S in R^N is a convex *cone* iff it both a convex set and a cone. Show that S is a convex cone iff it satisfies the following property:

(b) $x \in S$, $y \in S$, $\alpha \geq 0$ and $\beta \geq 0$ implies $\alpha x + \beta y \in S$.

5. Let S be a nonempty closed convex set in R^N (that is not the entire space R^N) and let b be a boundary point of S. Using Minkowski's supporting hyperplane theorem, there exists at least one vector $c^* \neq 0_N$ such that $c^{*T}b \leq c^{*T}x$ for all $x \in S$. Define the *set of supporting hyperplanes* S(b) to the set S at the boundary point b to be the following set:

(a) $S(b) \equiv \{c : c^Tb \leq c^Tx \text{ for all } x \in S\}$.

Show that the set S(b) has the following properties:

(b) $0_N \in S(b)$;
(c) S(b) is a cone;
(d) S(b) is a closed set;
(e) S(b) is a convex set and
(f) S(b) contains at least one *ray*, a set of the form $\{x : x = \lambda x^*,\ \lambda \geq 0\}$ where $x^* \neq 0_N$.

6. If X and Y are nonempty sets in R^N, the set $X - Y$ is defined as follows:

(a) $X - Y \equiv \{x - y : x \in X \text{ and } y \in Y\}$.

If X and Y are nonempty convex sets, show that X − Y is also a convex set.

7. *Separating hyperplane theorem between two disjoint convex sets* Fenchel (1953; 48-49): Let X and Y be two nonempty, convex sets in R^N that have no points in common; i.e., $X \cap Y = \emptyset$ (the empty set). Assume that at least one of the two sets X or Y has a nonempty interior. Prove that there exists a hyperplane that separates X and Y; i.e., show that there exists a nonzero vector c and a scalar α such that

(a) $c^T y \le \alpha \le c^T x$ for all $x \in X$ and $y \in Y$.

Hint: Consider the set $S \equiv X - Y$ and show that 0_N does not belong to S. If 0_N does not belong to the closure of S, apply Minkowski's Theorem 1. If 0_N does belong to the closure of S, then since it does not belong to S and S has an interior, it must be a boundary point of S and apply Minkowski's Theorem 2.

4. Concave Functions

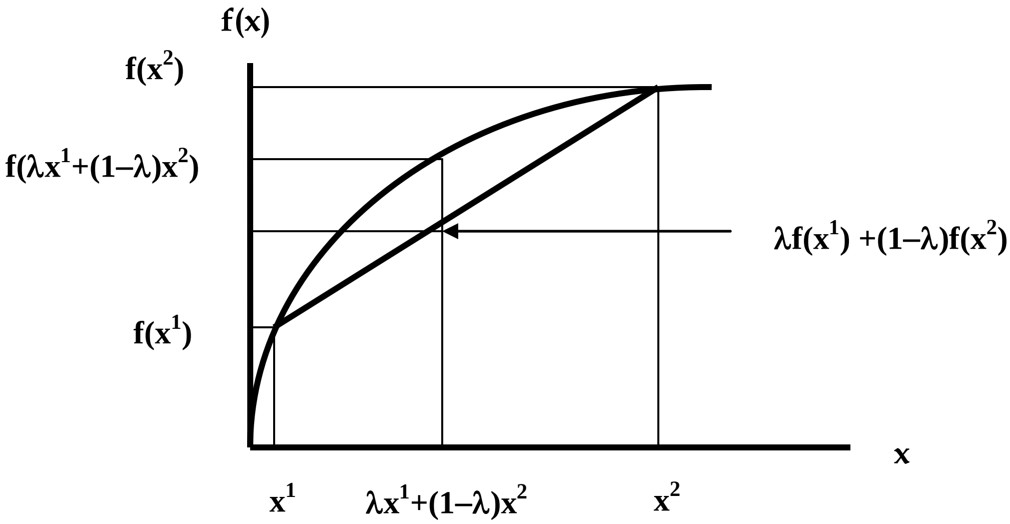
Definition: A function f(x) of N variables $x \equiv [x_1,\ldots,x_N]$ defined over a convex subset S of R^N is *concave* iff for every $x^1 \in S$, $x^2 \in S$ and $0 < \lambda < 1$, we have

(28) $f(\lambda x^1+(1-\lambda)x^2) \ge \lambda f(x^1)+(1-\lambda)f(x^2)$.

In the above definition, f is defined over a convex set so that we can be certain that points of the form $\lambda x^1+(1-\lambda)x^2$ belong to S if x^1 and x^2 belong to S. Note that when λ equals 0 or 1, the weak inequality (28) is automatically valid as an equality so that in the above definition, we could replace the restriction $0 < \lambda < 1$ by $0 \le \lambda \le 1$.

If N = 1, a geometric interpretation of a concave function is easy to obtain; see Figure 2 below. As the scalar λ travels from 1 to 0, $f(\lambda x^1+(1-\lambda)x^2)$ traces out the value of f between x^1 and x^2. On the other hand, as λ travels from 1 to 0, $\lambda f(x^1)+(1-\lambda)f(x^2)$ is a linear function of λ which joins up the point $(x^1, f(x^1))$ to the point $(x^2, f(x^2))$ on the graph of f(x); i.e., $\lambda f(x^1)+(1-\lambda)f(x^2)$ traces out the chord between two points on the graph of f. The inequality (28) says that if the function is concave, then the graph of f between x^1 and x^2 will lie above (or possibly be coincident with) the chord joining these two points on the graph. This property must hold for any two points on the graph of f.

Figure 2: A Concave Function of One Variable



For a general N, the interpretation of the concavity property is the same as in the previous paragraph: we look at the behavior of the function along the line segment joining x^1 to x^2 compared to the straight line segment joining the point $[x^1,f(x^1)]$ in R^{N+1} to the point $[x^2,f(x^2)]$. This line segment must lie below (or be coincident with) the former curve. This property must hold for any two points in the domain of definition of f.

Concave functions occur quite frequently in economic as we shall see in subsequent chapters.

One very convenient property that a concave function possesses is given by the following result.

Theorem 4: Local Maximum is a Global Maximum; Fenchel (1953; 63): Let f be a concave function defined over a convex subset S of R^N. If f attains a local maximum at the point $x^0 \in S$, then f attains a global maximum at x^0; i.e., we have

(29) $f(x^0) \geq f(x)$ for all $x \in S$.

Proof: Since f attains a local maximum at x^0, there exists a $\delta > 0$ such that

(30) $f(x^0) \geq f(x)$ for all $x \in S \cap B(x^0,\delta)$

where $B(x^0,\delta) \equiv \{x : (x-x^0)^T(x-x^0) < \delta^2\}$ is the open ball of radius δ around the point x^0. *Suppose* there exists an $x^1 \in S$ such that

(31) $f(x^1) > f(x^0)$.

Using the concavity of f, for $0 < \lambda < 1$, we have

$$
\begin{aligned}
(32)\ f(\lambda x^1+(1-\lambda)x^0) &\geq \lambda f(x^1)+(1-\lambda)f(x^0) \\
&> \lambda f(x^0)+(1-\lambda)f(x^0) \qquad \text{using } \lambda > 0 \text{ and (31)} \\
&= f(x^0).
\end{aligned}
$$

But for λ close to 0, $\lambda x^1+(1-\lambda)x^0$ will belong to $S \cap B(x^0,\delta)$ and hence for λ close to 0, (32) will contradict (30). Thus our *supposition* must be false and (29) holds. Q.E.D.

It turns out to be very useful to have several alternative characterizations for the concavity property. Our first characterization is provided by the definition (28).

Theorem 5: Second Characterization of Concavity; Fenchel (1953; 57): (a) f is a concave function defined over the convex subset S of R^N iff (b) the set $H \equiv \{(y,x) : y \leq f(x), x \in S\}$ is a convex set in R^{N+1}.[3]

Proof: (a) implies (b): Let $(y^1,x^1) \in H$, $(y^2,x^2) \in H$ and $0 < \lambda < 1$. Thus

(33) $y^1 \leq f(x^1)$ and $y^2 \leq f(x^2)$

with $x^1 \in S$ and $x^2 \in S$. Since f is concave over S,

$$
\begin{aligned}
(34)\ f(\lambda x^1+(1-\lambda)x^2) &\geq \lambda f(x^1)+(1-\lambda)f(x^2) \\
&\geq \lambda y^1+(1-\lambda)y^2 \qquad \text{using } \lambda > 0,\ (1-\lambda) > 0 \text{ and (33).}
\end{aligned}
$$

Using the definition of S, (34) shows that $[\lambda y^1+(1-\lambda)y^2, \lambda x^1+(1-\lambda)x^2] \in H$. Thus H is a convex set.

(b) implies (a): Let $x^1 \in S$, $x^2 \in S$ and $0 < \lambda < 1$. Define y^1 and y^2 as follows:

(35) $y^1 \equiv f(x^1)$ and $y^2 \equiv f(x^2)$.

The definition of H and the equalities in (35) show that $(y^1,x^1) \in H$ and $(y^2,x^2) \in H$. Since H is a convex set by assumption, $[\lambda y^1+(1-\lambda)y^2, \lambda x^1+(1-\lambda)x^2] \in H$. Hence, by the definition of H,

$$
\begin{aligned}
(36)\ f(\lambda x^1+(1-\lambda)x^2) &\geq \lambda y^1+(1-\lambda)y^2 \\
&= \lambda f(x^1)+(1-\lambda)f(x^2) \qquad \text{using (35)}
\end{aligned}
$$

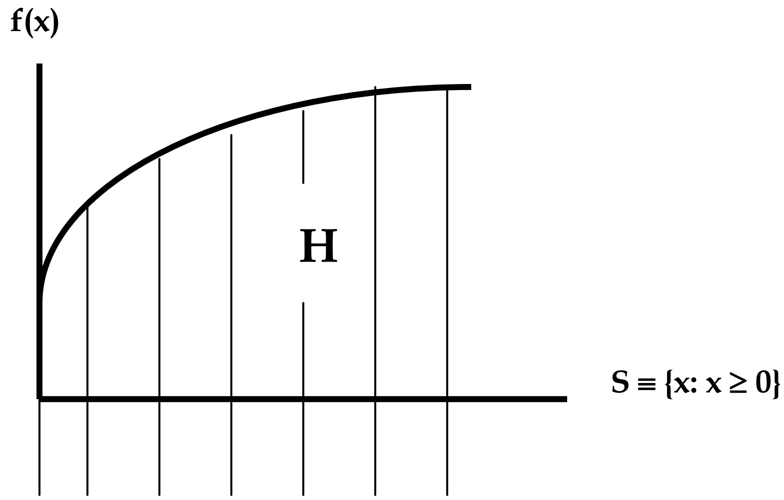
which establishes the concavity of f. Q.E.D.

A geometric interpretation of the second characterization of concavity when N = 1 is illustrated in Figure 3.

---

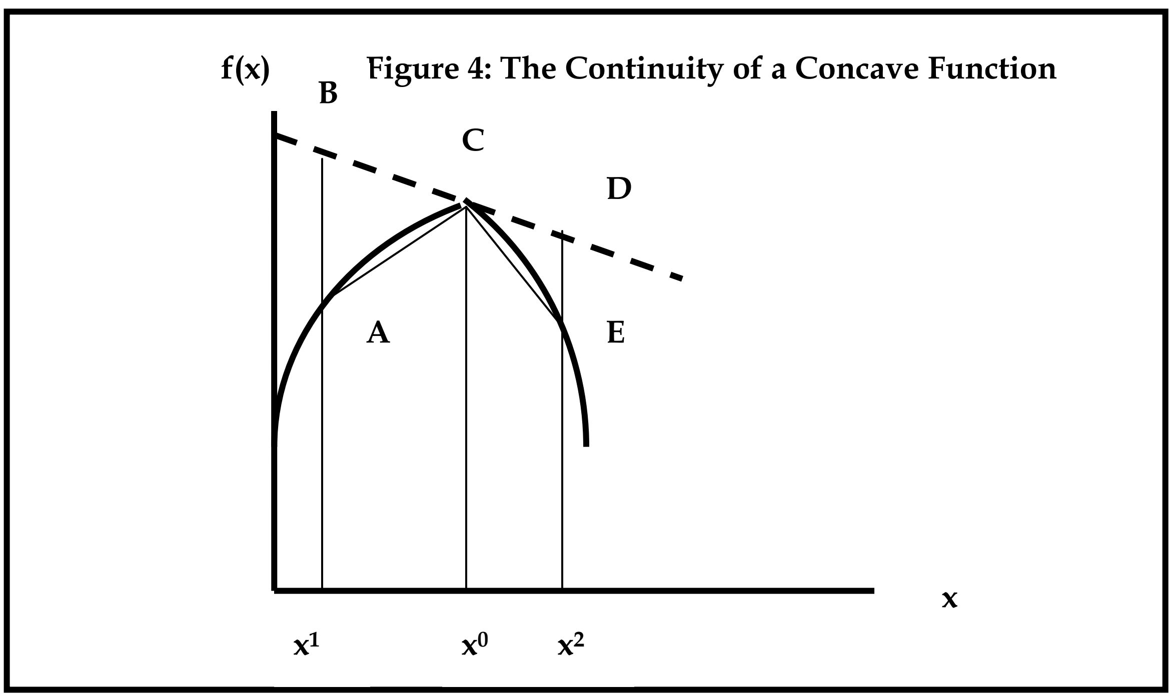
[3]The set H is called the *hypograph* of the function f; it consists of the graph of f and all of the points lying below it.

Figure 3: The Second Characterization of Concavity



The second characterization of concavity is useful because it shows us why concave functions are *continuous* over the interior of their domain of definition.

Figure 4: The Continuity of a Concave Function



Let the function of one variable, f(x), be concave. In Figure 4, the point x^0 is in the interior of the domain of definition set S and the point $[f(x^0), x^0]$ is on the boundary of the convex set hypograph H of f. By Theorem 2, there is at least one supporting hyperplane to the boundary point $[f(x^0), x^0]$ of H (this is the point C in Figure 4) and we have drawn one of these hyperplanes (which are lines in this case) as the dashed line BCD.[4] Note that this dashed line serves as an *upper bounding* approximating function to f(x). Now consider some point x^1 to the left of x^0. Since f is a concave function, it can be seen that the straight line joining the points $[f(x^1), x^1]$ and $[f(x^0), x^0]$, the line segment AC in Figure 4, is a *lower bounding* approximating function to f(x) to the left of x^0. Now consider some point x^2 to the right of x^0. Again using the concavity of f, the straight line joining the points $[f(x^0), x^0]$ and $[f(x^2), x^2]$, the line segment CE in Figure 4, is a *lower bounding* approximating function to f(x) to the right of x^0. Thus f(x) is sandwiched between two linear functions that meet at the point C both to the left and to the right of x^0 and it can be seen that f(x) must be continuous at the interior point x^0. The same conclusion can be derived for any interior point of the domain of definition set S and hence we conclude that a concave function f is continuous over Int S.[5]

The example shown in Figure 4 shows that there can be more than one supporting hyperplane to a boundary point on the hypograph of a concave function. Consider the following definition and Theorem:

Definition: A vector b is a *supergradient* to the function of N variables f defined over S at the point $x^0 \in S$ iff

(37) $f(x) \leq f(x^0) + b^T(x - x^0)$ for all $x \in S$.

Note that the function on the right hand side of (37) is a linear function of x which takes on the value $f(x^0)$ when $x = x^0$. This linear function acts as an upper bounding function to f.

Theorem 6; Rockafellar (1970; 217): If f is a concave function defined over an open convex subset S of R^N, then for every $x^0 \in S$, f has at least one supergradient vector b^0 to f at the point x^0. Denote the set of all such supergradient vectors as $\partial f(x^0)$. Then $\partial f(x^0)$ is a nonempty, closed convex set.[6]

Proof: Define the hypograph of f as $H \equiv \{(y,x) : y \leq f(x), x \in S\}$. Since f is concave, by Theorem 5, H is a convex set. Note also that $[f(x^0), x^0]$ is a boundary point of H. By Theorem 2, there exists $[c_0, c^T] \neq 0_{N+1}$ such that

(38) $c_0 f(x^0) + c^T x^0 \leq c_0 y + c^T x$ for every $[y, x] \in H$.

Suppose c_0 in (38) were equal to 0. Then (38) becomes

(39) $c^T x^0 \leq c^T x$ for every $x \in S$.

---

[4]Note that the graph of f is kinked at the point C and so there is an entire set of supporting hyperplanes to the point C in this case.

[5] The argument is a bit more complex when N is greater than 1 but the same conclusion is obtained. We cannot extend the above argument to boundary points of S because the supporting hyperplane to H may be vertical. See Fenchel (1953; 74) for a general proof.

[6] Rockafellar shows that $\partial f(x^0)$ is also a bounded set.

Since $x^0 \in \text{Int } S$, (39) cannot be true. Hence our *supposition* is false and we can assume $c_0 \neq 0$. If $c_0 > 0$, then (38) is not satisfied if $y < f(x^0)$ and $x = x^0$. Thus we must have $c_0 < 0$. Multiplying (38) through by $1/c_0$ yields the following inequality:

(40) $f(x^0) - b^{0T}x^0 \geq y - b^{0T}x$ for every $[y, x] \in H$

where the vector b^0 is defined as

(41) $b^0 \equiv - c/c_0$.

Now let $x \in S$ and $y = f(x)$. Then $[y, x] \in H$ and (40) becomes

(42) $f(x) \leq f(x^0) + b^{0T}(x - x^0)$ for all $x \in S$.

Using definition (37), (42) shows that b^0 is a supergradient to f at x^0 and hence $\partial f(x^0)$ is nonempty.

To show that $\partial f(x^0)$ is a convex set, let $b^1 \in \partial f(x^0)$, $b^2 \in \partial f(x^0)$ and $0 < \lambda < 1$. Then

(43) $f(x) \leq f(x^0) + b^{1T}(x - x^0)$ for all $x \in S$;
$f(x) \leq f(x^0) + b^{2T}(x - x^0)$ for all $x \in S$.

Thus

(44) $f(x) = \lambda f(x) + (1-\lambda) f(x)$ for all $x \in S$;
$\leq \lambda[f(x^0) + b^{1T}(x - x^0)] + (1-\lambda) f(x)$ using $\lambda > 0$ and (43)
$\leq \lambda[f(x^0) + b^{1T}(x - x^0)] + (1-\lambda)[f(x^0) + b^{2T}(x - x^0)]$ using $1-\lambda > 0$ and (43)
$= f(x^0) + [\lambda b^{1T} + (1-\lambda) b^{2T}](x - x^0)$ for all $x \in S$

and so $[\lambda b^1 + (1-\lambda) b^2] \in \partial f(x^0)$. Thus $\partial f(x^0)$ is a convex set.

The closedness of $\partial f(x^0)$ follows from the fact that the vector b enters the inequalities (37) in a linear fashion. Q.E.D.

Corollary: If f is a concave function defined over a convex subset S of R^N, $x^0 \in \text{Int } S$ and the first order partial derivatives of f evaluated at x^0 exist,[7] then $\partial f(x^0) = \{\nabla f(x^0)\}$; i.e., if f has first order partial derivatives, then the set of supergradients reduces to the gradient vector of f evaluated at x^0.

Proof: Since $x^0 \in \text{Int } S$, $\partial f(x^0)$ is nonempty. Let $b \in \partial f(x^0)$. Using definition (37) of a supergradient, it can be seen that the function g(x) defined by (45) below is nonpositive:

---

[7] Thus the vector of first order partial derivatives $\nabla f(x^0) \equiv [\partial f(x^0)/\partial x_1, \ldots, \partial f(x^0)/\partial x_N]^T$ exists.

(45) $g(x) \equiv f(x) - f(x^0) - b^T(x-x^0) \le 0$ for all $x \in S$.

Since $g(x^0) = 0$, (45) shows that g(x) attains a global maximum over the set S at $x = x^0$. Hence, the following first order necessary conditions for maximizing a function of N variables will hold at x^0:

(46) $\nabla g(x^0) = \nabla f(x^0) - b = 0_N$ or
(47) $b = \nabla f(x^0)$.

Hence we have shown that if $b \in \partial f(x^0)$, then $b = \nabla f(x^0)$. Hence $\partial f(x^0)$ is the single point, $\nabla f(x^0)$. Q.E.D.

Theorem 7; Third Characterization of Concavity: Roberts and Varberg (1973; 12): Let f be a function of N variables defined over an open convex subset S of R^N. Then (a) f is concave over S iff (b) for every $x^0 \in S$, there exists a b^0 such that

(48) $f(x) \le f(x^0) + b^{0T}(x-x^0)$ for all $x \in S$.

Proof: (a) implies (b). This has been done in Theorem 6 above.

(b) implies (a). Let $x^1 \in S$, $x^2 \in S$ and $0 < \lambda < 1$. Let b be a supergradient to f at the point $\lambda x^1+(1-\lambda)x^2$. Thus we have:

(49) $f(x) \le f(\lambda x^1+(1-\lambda)x^2) + b^T(x-[\lambda x^1+(1-\lambda)x^2])$ for all $x \in S$.

Now evaluate (49) at $x = x^1$ and then multiply both sides of the resulting inequality through by $\lambda > 0$. Evaluate (49) at $x = x^2$ and then multiply both sides of the resulting inequality through by $1-\lambda > 0$. Add the two inequalities to obtain:

$$
\begin{aligned}
(50)\ \lambda f(x^1) + (1-\lambda)f(x^2) &\le f(\lambda x^1+(1-\lambda)x^2) + \lambda\, b^T(x^1-[\lambda x^1+(1-\lambda)x^2]) \\
&\quad + (1-\lambda)\, b^T(x^2-[\lambda x^1+(1-\lambda)x^2]) \\
&= f(\lambda x^1+(1-\lambda)x^2)
\end{aligned}
$$

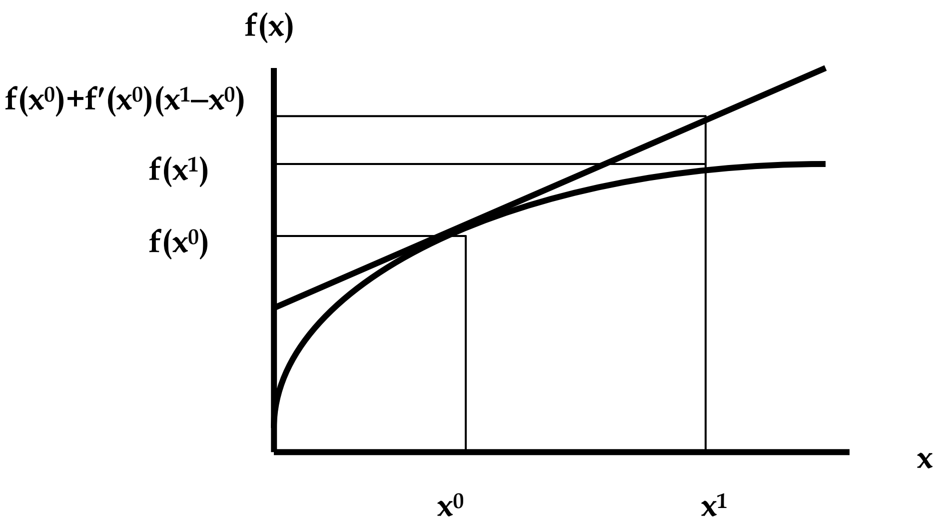
which shows that f is concave over S. Q.E.D.

Corollary: Third Characterization of Concavity in the Once Differentiable Case; Mangasarian (1969; 84): Let f be a once differentiable function of N variables defined over an open convex subset S of R^N. Then (a) f is concave over S iff

(51) $f(x^1) \le f(x^0) + \nabla f(x^0)^T(x^1-x^0)$ for all $x^0 \in S$ and $x^1 \in S$.

Proof: (a) implies (51). This follows from Theorem 6 and its corollary. (51) implies (a). Use the proof of (b) implies (a) in Theorem 7. Q.E.D.

Figure 5: The Third Characterization of Concavity



Thus in the case where the function is once differentiable and defined over an open convex set, then a necessary and sufficient condition for the function to be concave is that the first order linear approximation to f(x) around any point $x^0 \in S$, which is $f(x^0) + \nabla f(x^0)^T(x-x^0)$, must lie above (or be coincident with) the surface of the function.

Our final characterization of concavity is for functions of N variables, f(x), that are twice continuously differentiable. This means that the first and second order partial derivative functions exist and are continuous. Note that in this case, Young's Theorem from calculus implies that

(52) $\partial^2 f(x_1,\ldots,x_N)/\partial x_i \partial x_k = \partial^2 f(x_1,\ldots,x_N)/\partial x_k \partial x_i \quad$ for $1 \leq i < k \leq N$;

i.e., the matrix of second order partial derivatives, $\nabla^2 f(x) \equiv [\partial^2 f(x_1,\ldots,x_N)/\partial x_i \partial x_k]$ is symmetric.

Theorem 8: Fourth Characterization of Concavity in the Twice Continuously Differentiable Case; Fenchel (1953; 87-88): Let f be a twice continuously differentiable function of N variables defined over an open convex subset S of R^N. Then (a) f is concave over S iff (b) $\nabla^2 f(x)$ is negative semidefinite for all $x \in S$.

Proof: (b) implies (a). Let x^0 and x^1 be two arbitrary points in S. Then by Taylor's Theorem for n = 2,[8] there exists θ such that $0 < \theta < 1$ and

[8] For a function of one variable, g(t) say, Taylor's Theorem for n = 1 is the Mean Value Theorem; i.e., if the derivative of g exists for say $0 < t < 1$, then there exists t* such that $0 < t^* < 1$ and $g(1) = g(0) + g'(t^*)[1-0]$. Taylor's Theorem for n = 2 is: suppose the first and second derivatives of g(t) exist for $0 \leq t \leq 1$. Then there exists t* such that $0 < t^* < 1$ and $g(1) = g(0) + g'(0)[1-0] + (1/2)\, g''(t^*)[1-0]^2$. To see that (53) follows from this Theorem, define $g(t) \equiv f(x^0 + t[x^1-x^0])$ for $0 \leq t \leq 1$. Routine calculations show that g′(t) =

(53) $f(x^1) = f(x^0) + \nabla f(x^0)^T(x^1-x^0) + (1/2)(x^1-x^0)^T\nabla^2 f(\theta x^0+(1-\theta)x^1)(x^1-x^0)$
$\leq f(x^0) + \nabla f(x^0)^T(x^1-x^0)$

where the inequality follows from the assumption (b) that $\nabla^2 f(\theta x^0+(1-\theta)x^1)$ is negative semidefinite and hence

(54) $(1/2)(x^1-x^0)^T\nabla^2 f(\theta x^0+(1-\theta)x^1)(x^1-x^0) \leq 0.$

But the inequalities in (53) are equivalent to (51) and hence f is concave.

(a) implies (b). We show that not (b) implies not (a). Not (b) means there exist $x^0 \in S$ and $z \neq 0_N$ such that

(55) $z^T\nabla^2 f(x^0)z > 0.$

Using (55) and the continuity of the second order partial derivatives of f, we can find a $\delta > 0$ small enough so that $x^0 + tz \in S$ and

(56) $z^T\nabla^2 f(x^0 + tz)z > 0$ for t such that $-\delta \leq t \leq \delta$.

(56.5) Define $x^1 \equiv x^0 + \delta z$.

By Taylor's Theorem, there exists θ such that $0 < \theta < 1$ and

(57) $f(x^1) = f(x^0) + \nabla f(x^0)^T(x^1-x^0) + (1/2)(x^1-x^0)^T\nabla^2 f(\theta x^0+(1-\theta)x^1)(x^1-x^0)$
$= f(x^0) + \nabla f(x^0)^T(x^1-x^0) + (1/2)(\delta z)^T\nabla^2 f(\theta x^0+(1-\theta)x^1)(\delta z)$ using (56.5)
$= f(x^0) + \nabla f(x^0)^T(x^1-x^0) + \delta^2(1/2)z^T\nabla^2 f(\theta x^0+(1-\theta)[x^0 + \delta z])z$ using (56.5)
$= f(x^0) + \nabla f(x^0)^T(x^1-x^0) + \delta^2(1/2)z^T\nabla^2 f(x^0+(1-\theta)\delta z)z$
$> f(x^0) + \nabla f(x^0)^T(x^1-x^0)$

where the inequality in (57) follows from

(58) $\delta^2(1/2)z^T\nabla^2 f(x^0+(1-\theta)\delta z)z > 0$

which in turn follows from $\delta^2 > 0$, $0 < (1-\theta)\delta < \delta$ and the inequality (56). But (57) contradicts (51) so that f cannot be concave. Q.E.D.

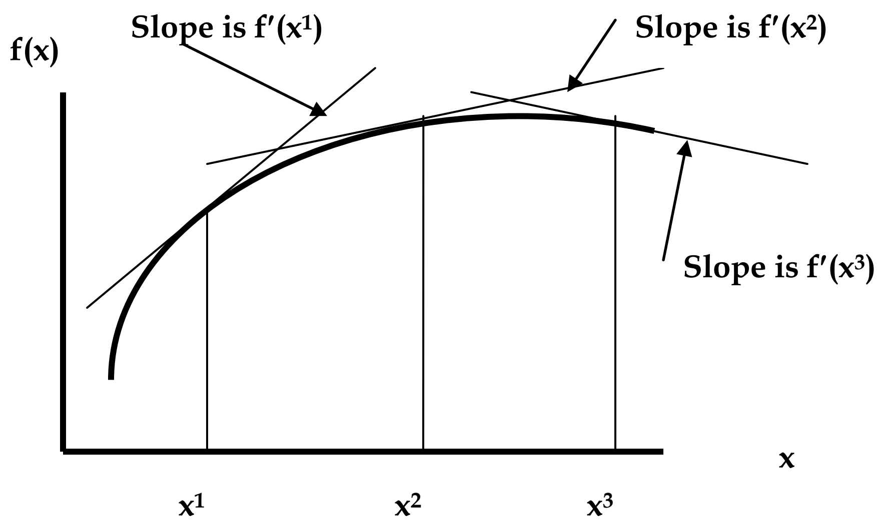
For a twice continuously differentiable function of one variable f(x) defined over the open convex set S, the fourth characterization of concavity boils down to checking the following inequalities:

$\nabla f(x^0 + t[x^1-x^0])^T[x^1-x^0]$ and $g''(t) = [x^1-x^0]^T\nabla^2 f(x^0 + t[x^1-x^0])[x^1-x^0]$. Now (53) follows from Taylor's Theorem for n = 2 with $\theta = 1-t^*$.

(59) $f''(x) \le 0$ for all $x \in S$;

i.e., we need only check that the second derivative of f is negative or zero over its domain of definition. Thus as x increases, we need the first derivative $f'(x)$ to be decreasing or constant.

Figure 6: The Fourth Characterization of Concavity



In Figure 6, as x increases from x^1 to x^2 to x^3, the slope of the tangent to the function f(x) decreases; i.e., we have $f'(x^1) > f'(x^2) > f'(x^3)$. Thus the second derivative, $f''(x)$, decreases as x increases.

Problems

8. Define the function of one variable $f(x) \equiv x^{1/2}$ for $x > 0$. Use the first, third and fourth characterizations of concavity to show that f is a concave function over the convex set $S \equiv \{x : x > 0\}$. Which characterization provides the easiest proof of concavity?

9. Let f(x) be a concave function of N variables $x \equiv [x_1,\ldots,x_N]$ defined over the open convex set S. Let $x^0 \in S$ and suppose $\nabla f(x^0) = 0_N$. Then show that $f(x^0) \ge f(x)$ for all $x \in S$; i.e., x^0 is a global maximizer for f over S. *Hint*: Use one of the characterizations of concavity. *Note*: this result can be extended to the case where S is a closed convex set with a nonempty interior. Thus the first order necessary conditions for maximizing a function of N variables are also sufficient if the function happens to be concave.

10. Prove that: if f(x) and g(x) are concave functions of N variables x defined over $S \subset R^N$ and $\alpha \ge 0$ and $\beta \ge 0$, then $\alpha f(x) + \beta g(x)$ is concave over S.

11. Fenchel (1953; 61): Show that: if f(x) is a concave function defined over the convex set $S \subset R^N$ and g is an increasing concave function of one variable defined over an interval that includes all of the numbers f(x) for $x \in S$, then $h(x) \equiv g[f(x)]$ is a concave function over S.

12. A function f(x) of N variables $x \equiv [x_1,\ldots,x_N]$ defined over a convex subset of R^N is *strictly concave* iff for every $x^1 \in S$, $x^2 \in S$ and $0 < \lambda < 1$, we have

(60) $f(\lambda x^1+(1-\lambda)x^2) > \lambda f(x^1)+(1-\lambda)f(x^2)$.

Suppose that $x^1 \in S$ and $x^2 \in S$ and $f(x^1) = f(x^2) = \max_x \{f(x) : x \in S\}$. Then show that $x^1 = x^2$. Note: This shows that if the maximum of a strictly concave function over a convex set exists, then the set of maximizers is unique.

13. Let f be a strictly concave function defined over a convex subset S of R^N. If f attains a local maximum at the point $x^0 \in S$, then show that f attains a strict global maximum at x^0; i.e., we have

(a) $f(x^0) > f(x)$ for all $x \in S$ where $x \neq x^0$.

Hint: Modify the proof of Theorem 4.

5. Convex Functions

Definition: A function f(x) of N variables $x \equiv [x_1,\ldots,x_N]$ defined over a convex subset S of R^N is *convex* iff for every $x^1 \in S$, $x^2 \in S$ and $0 < \lambda < 1$, we have

(61) $f(\lambda x^1+(1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda)f(x^2)$.

Comparing the definition (61) for a convex function with our previous definition (28) for a concave function, it can be seen that the inequalities and (28) and (61) are reversed. Thus an equivalent definition for a convex function f is: f(x) is convex over the convex set S iff $-f$ is concave over the convex set S. This fact means that we do not have to do much work to establish the properties of convex functions: we can simply use all of the material in the previous section, replacing f in each result in the previous section by $-f$ and then multiplying through the various inequalities by -1 (thus reversing them) in order to eliminate the minus signs from the various inequalities. Following this strategy leads to (61) as the *first characterization of a convex function*. We list the other characterizations below.

Second Characterization of Convexity; Fenchel (1953; 57): (a) f is a convex function defined over the convex subset S of R^N iff (b) the set $E \equiv \{(y,x) : y \geq f(x), x \in S\}$ is a convex set in R^{N+1}.[9]

[9]The set E is called the *epigraph* of the function f; it consists of the graph of f and all of the points lying above it.

Third Characterization of Convexity in the Once Differentiable Case; Mangasarian (1969; 84): Let f be a once differentiable function of N variables defined over an open convex subset S of R^N. Then f is convex over S iff

(62) $f(x^1) \geq f(x^0) + \nabla f(x^0)^T(x^1 - x^0)$ for all $x^0 \in S$ and $x^1 \in S$.

We note that a vector b is a *subgradient* to the function of N variables f defined over S at the point $x^0 \in S$ iff

(63) $f(x) \geq f(x^0) + b^T(x - x^0)$ for all $x \in S$.

Applying this definition to (62) shows that $\nabla f(x^0)$ is a subgradient to f at the point x^0.

Fourth Characterization of Convexity in the Twice Continuously Differentiable Case; Fenchel (1953; 87-88): Let f be a twice continuously differentiable function of N variables defined over an open convex subset S of R^N. Then (a) f is convex over S iff (b) $\nabla^2 f(x)$ is positive semidefinite for all $x \in S$.

The counterpart to Theorem 4 about concave functions in the previous section is Theorem 9 below for convex functions.

Theorem 9: Local Minimum is a Global Minimum; Fenchel (1953; 63): Let f be a convex function defined over a convex subset S of R^N. If f attains a local minimum at the point $x^0 \in S$, then f attains a global minimum at x^0; i.e., we have

(64) $f(x^0) \leq f(x)$ for all $x \in S$.

Problems

14. Recall that Example 5 in section 2 defined the *feasible region* for a linear programming problem as the following set S:

(a) $S \equiv \{x: Ax \leq b \; ; \; x \geq 0_N\}$

where A is an M by N matrix of constants and b is an M dimensional vector of constants. We showed in section 2 that S was a closed convex set. We now assume in addition, that S is a nonempty bounded set. Now consider the following function of the vector c:

(b) $f(c) \equiv \max_x \{c^T x : Ax \leq b \; ; \; x \geq 0_N\}$.

Show that f(c) is a convex function for $c \in R^N$.
Now suppose that we define f(c) as follows:

(c) $f(c) \equiv \min_x \{c^T x : Ax \leq b \; ; \; x \geq 0_N\}$.

Show that f(c) is a concave function for $c \in R^N$.

15. Mangasarian (1969; 149): Let f be a positive concave function defined over a convex subset S of R^N. Show that $h(x) \equiv 1/f(x)$ is a positive convex function over S. *Hint*: You may find the fact that a weighted harmonic mean is less than or equal to the corresponding weighted arithmetic mean helpful.

16. Let S be a closed and bounded set in R^N. For $p \in R^N$, define the *support function* of S as

(a) $\pi(p) \equiv \max_x \{p^T x : x \in S\}$.

(b) Show that $\pi(p)$ is a (positively) linearly homogeneous function over R^N.[10]
(c) Show that $\pi(p)$ is a convex function over R^N.
(d) If $0_N \in S$, then show $\pi(p) \geq 0$ for all $p \in R^N$.
Note: If we changed the domain of definition for the vectors p from R^N to the positive orthant, $\Omega \equiv \{x : x >> 0_N\}$, and defined $\pi(p)$ by (a), then $\pi(p)$ would satisfy the same properties (b), (c) and (d) and we could interpret $\pi(p)$ as the *profit function* that corresponds to the technology set S.

17. Define Ω as the positive orthant in R^N; i.e., $\Omega \equiv \{x : x >> 0_N\}$. Suppose f(x) is a positive, positively linearly homogeneous and concave function defined over Ω. Show that f is also increasing over Ω; i.e., show that f satisfies the following property:

(a) $x^2 >> x^1 >> 0_N$ implies $f(x^2) > f(x^1)$.

6. Quasiconcave Functions

Example 6 in section 2 above indicated why quasiconcave functions arise in economic applications. In this section, we will study the properties of quasiconcave functions more formally.

Definition: *First Characterization of Quasiconcavity*; Fenchel (1953; 117): f is a *quasiconcave function* defined over a convex subset S of R^N iff

(65) $x^1 \in S$, $x^2 \in S$, $0 < \lambda < 1$ implies $f(\lambda x^1 + (1-\lambda)x^2) \geq \min\{f(x^1), f(x^2)\}$.

The above definition asks that the line segment joining x^1 to x^2 that has height equal to the minimum value of the function at the points x^1 and x^2 lies below (or is coincident with) the graph of f along the line segment joining x^1 to x^2.

If f is concave over S, then

(66) $f(\lambda x^1 + (1-\lambda)x^2) \geq \lambda f(x^1) + (1-\lambda)f(x^2)$ using (28), the definition of concavity

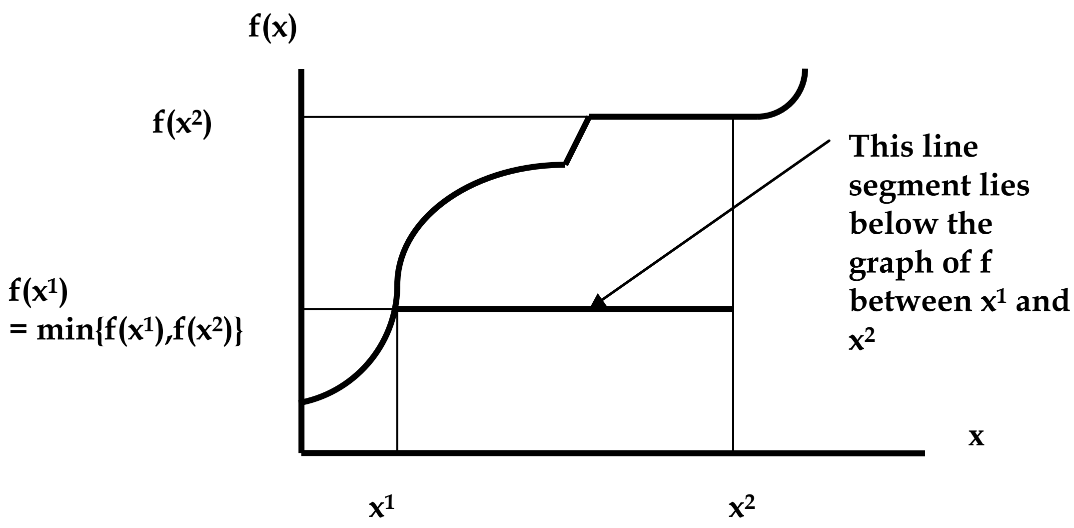
[10] This means for all $\lambda > 0$ and $p \in R^N$, $\pi(\lambda p) = \lambda\pi(p)$.

$\geq \min \{f(x^1), f(x^2)\}$

where the second inequality follows since $\lambda f(x^1) + (1-\lambda)f(x^2)$ is an average of $f(x^1)$ and $f(x^2)$. *Thus if f is concave, then it is also quasiconcave.*

A geometric interpretation of property (65) can be found in Figure 7 for the case N = 1. Essentially, the straight line segment above x^1 and x^2 parallel to the x axis that has height equal to the minimum of $f(x^1)$ and $f(x^2)$ must lie below (or be coincident with) the graph of the function between x^1 and x^2. This property must hold for any two points in the domain of definition of f.

Figure 7: The First Characterization of Quasiconcavity



If f is a function of one variable, then any *monotonic* function[11] defined over a convex set will be quasiconcave. Functions of one variable that are increasing (or nondecreasing) and then are decreasing (or nonincreasing) are also quasiconcave. Such functions need not be continuous or concave and thus quasiconcavity is a genuine generalization of concavity. Note also that quasiconcave functions can have flat spots on their graphs.

The following Theorem shows that the above definition of quasiconcavity is equivalent to the definition used in Example 6 in section 2 above.

Theorem 10: *Second Characterization of Quasiconcavity*; Fenchel (1953; 118): f is a quasiconcave function over the convex subset S of R^N iff

(67) For every $u \in$ Range f, the upper level set $L(u) \equiv \{x : f(x) \geq u ; x \in S\}$ is a convex set.

[11] A monotonic function of one variable is one that is either increasing, decreasing, nonincreasing or nondecreasing.

Proof: (65) implies (67): Let $u \in$ Range f, $x^1 \in L(u)$, $x^2 \in L(u)$ and $0 < \lambda < 1$. Since x^1 and x^2 belong to L(u),

(68) $f(x^1) \geq u$; $f(x^2) \geq u$.

Using (65), we have:

(69) $f(\lambda x^1+(1-\lambda)x^2) \geq \min\{f(x^1), f(x^2)\}$
$\geq u$

where the last inequality follows using (68). But (69) shows that $\lambda x^1+(1-\lambda)x^2 \in L(u)$ and thus L(u) is a convex set.

(67) implies (65): Let $x^1 \in S$, $x^2 \in S$, $0 < \lambda < 1$ and let $u \equiv \min\{f(x^1),f(x^2)\}$. Thus $f(x^1) \geq u$ and hence, $x^1 \in L(u)$. Similarly, $f(x^2) \geq u$ and hence, $x^2 \in L(u)$. Since L(u) is a convex set using (67), $\lambda x^1+(1-\lambda)x^2 \in L(u)$. Hence, using the definition of L(u),

(70) $f(\lambda x^1+(1-\lambda)x^2) \geq u \equiv \min\{f(x^1),f(x^2)\}$

which is (65). Q.E.D.

Figure 1 in section 2 above suffices to give a geometric interpretation of the second characterization of quasiconcavity.

The first characterization of quasiconcavity (65) can be written in an equivalent form as follows:

(71) $x^1 \in S$, $x^2 \in S$, $x^1 \neq x^2$, $0 < \lambda < 1$, $f(x^1) \leq f(x^2)$ implies $f(\lambda x^1+(1-\lambda)x^2) \geq f(x^1)$.

We now turn to our third characterization of quasiconcavity. For this characterization, we will assume that f is defined over an *open* convex subset S of R^N and that the first order partial derivatives of f, $\partial f(x)/\partial x_n$ for $n = 1,\ldots,N$, exist and are *continuous* functions over S. In this case, the property on f that will characterize a quasiconcave function is the following one:

(72) $x^1 \in S$, $x^2 \in S$, $x^1 \neq x^2$, $\nabla f(x^1)^T(x^2-x^1) < 0$ implies $f(x^2) < f(x^1)$.

Theorem 11: *Third Characterization of Quasiconcavity*; Mangasarian (1969; 147): Let f be a once continuously differentiable function defined over the open convex subset S of R^N. Then f is quasiconcave over S iff f satisfies (72).

Proof: (71) implies (72). We show that not (72) implies not (71). Not (72) means that there exist $x^1 \in S$, $x^2 \in S$, $x^1 \neq x^2$ such that

(73) $\nabla f(x^1)^T(x^2-x^1) < 0$ and

(74) $f(x^2) \geq f(x^1)$.

Define the function of one variable g(t) for $0 \leq t \leq 1$ as follows:

(75) $g(t) \equiv f(x^1+t[x^2-x^1])$.

It can be verified that

(76) $g(0) = f(x^1)$ and $g(1) = f(x^2)$.

It can be verified that the derivative of g(t) for $0 \leq t \leq 1$ can be computed as follows:

(77) $g'(t) = \nabla f(x^1+t[x^2-x^1])^T(x^2-x^1)$.

Evaluating (77) at t = 0 and using (73) shows that[12]

(78) $g'(0) < 0$.

Using the continuity of the first order partial derivatives of f, it can been seen that (78) implies the existence of a δ such that

(79) $0 < \delta < 1$ and
(80) $g'(t) < 0$ for all t such that $0 \leq t \leq \delta$.

Thus g(t) is a decreasing function over this interval of t's and thus

(81) $g(\delta) \equiv f(x^1+\delta[x^2-x^1]) = f([1-\delta]x^1+\delta x^2) < g(0) \equiv f(x^1)$.

But (79) and (81) imply that

(82) $f(\lambda x^1+(1-\lambda)x^2) < f(x^1)$

where $\lambda \equiv 1-\delta$. Since (79) implies that $0 < \lambda < 1$, (82) contradicts (71) and so f is not quasiconcave.

(72) implies (71). We show not (71) implies not (72). Not (71) means that there exist $x^1 \in S$, $x^2 \in S$, $x^1 \neq x^2$ and $0 < \lambda^* < 1$ such that

(83) $f(x^1) \leq f(x^2)$ and
(84) $f(\lambda^* x^1+(1-\lambda^*)x^2) < f(x^1)$.

Define the function of one variable g(t) for $0 \leq t \leq 1$ as follows:

---

[12]Note that g′(0) is the directional derivative of f(x) in the direction defined by x^2-x^1.

(85) $g(t) \equiv f(x^1+t[x^2-x^1])$.

Define t* as follows:

(86) $t^* \equiv 1 - \lambda^*$

and note that $0 < t^* < 1$ and

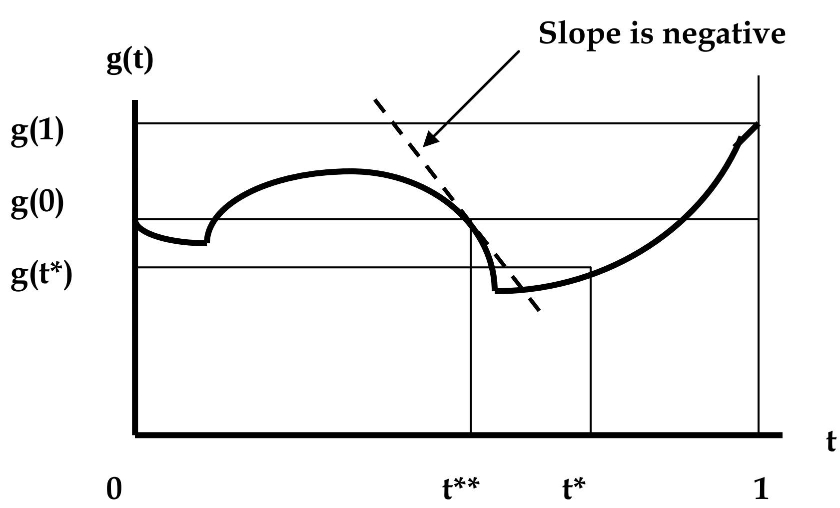
$$\begin{aligned}(87)\ g(t^*) &= f(\lambda^* x^1+(1-\lambda^*)x^2) \\ &< f(x^1) && \text{using (84)} \\ &= g(0) && \text{using definition (85).}\end{aligned}$$

The continuity of the first order partial derivatives of f implies that $g'(t)$ and $g(t)$ are continuous functions of t for $0 \le t \le 1$. Now consider the behavior of g(t) along the line segment $0 \le t \le t^*$. The inequality (87) shows that g(t) eventually decreases from g(0) to the lower number g(t*) along this interval. Thus there must exist a t** such that

(88) $0 \le t^{**} < t^*$;
(89) $g(t) \le g(0)$ for all t such that $t^{**} \le t \le t^*$ and
(90) $g(t^{**}) = g(0)$.

Essentially, the inequalities (88)-(90) say that there exists a closed interval to the immediate left of the point t*, [t**,t*], such that g(t) is less than or equal to g(0) for t in this interval and the lower boundary point of the interval, t**, is such that g(t**) equals g(0).

Figure 8: The Geometry of Theorem 11



Now *suppose* that the derivative of g is nonnegative for all t in the interval [t**, t*]; i.e.,

(91) $g'(t) \geq 0$ for all t such that $t^{**} \leq t \leq t^*$.

Then by the Mean Value Theorem, there exists t^{***} such that $t^{**} < t^{***} < t^*$ and

$$
\begin{aligned}
(92)\ g(t^*) &= g(t^{**}) + g'(t^{***})(t^*-t^{**}) \\
&\geq g(t^{**}) && \text{using (91) and } t^{**} < t^* \\
&= g(0) && \text{using (90).}
\end{aligned}
$$

But the inequality (92) contradicts $g(t^*) < g(0)$, which is equivalent to (84). Thus our *supposition* is false. Hence there exists t′ such that

(93) $t^{**} < t' < t^*$ and
(94) $g'(t') < 0$.

In Figure 8, such a point t′ is just to the right of t** where the dashed line is tangent to the graph of g(t). Using (89), we also have

(95) $g(t') \leq g(0)$.

Using definition (85), the inequalities (94) and (95) translate into the following inequalities:

$$
\begin{aligned}
(96)\ g'(t') &= \nabla f(x^1+t'[x^2-x^1])^T(x^2-x^1) < 0\ ; \\
(97)\ g(t') &= f(x^1+t'[x^2-x^1]) \leq f(x^1) \\
&\leq f(x^2) && \text{using (83).}
\end{aligned}
$$

Now define

(98) $x^3 \equiv x^1+t'[x^2-x^1]$

and note that the inequalities (93) imply that

(99) $0 < t' < 1$.

Using definition (98), we have

$$
\begin{aligned}
(100)\ x^2 - x^3 &= x^2 - \{x^1+t'[x^2-x^1]\} \\
&= (1-t')[x^2-x^1] \\
&\neq 0_N && \text{using (99) and } x^1 \neq x^2.
\end{aligned}
$$

Note that the second equation in (100) implies that

(101) $x^2 - x^1 = (1-t')^{-1}[x^2-x^3]$.

Now substitute (98) and (101) into (96) and we obtain the following inequality:

(102) $(1 - t')^{-1}\nabla f(x^3)^T(x^2 - x^3) < 0$ or
$\nabla f(x^3)^T(x^2 - x^3) < 0$ since $(1 - t')^{-1} > 0$.

(103) $f(x^3) \leq f(x^2)$.

The inequalities (102) and (103) show that (72) does not hold, with x^3 playing the role of x^1 in condition (72). Q.E.D.

Corollary: Arrow and Enthoven (1961; 780): Let f be a once continuously differentiable function defined over the open convex subset S of R^N. Then f is quasiconcave over S iff f satisfies the following condition:

(104) $x^1 \in S$, $x^2 \in S$, $x^1 \neq x^2$, $f(x^2) \geq f(x^1)$ implies $\nabla f(x^1)^T(x^2 - x^1) \geq 0$.

Proof: Condition (104) is the contrapositive to condition (72) and is logically equivalent to it. Q.E.D.

We can use the third characterization of concavity to show that the third characterization of quasiconcavity holds and hence a once continuously differentiable concave function f defined over an open convex set S is also quasiconcave (a fact which we already know). Thus let f be concave over S and assume the conditions in (72); i.e., let $x^1 \in S$, $x^2 \in S$, $x^1 \neq x^2$ and assume

(105) $\nabla f(x^1)^T(x^2 - x^1) < 0$.

We need only show that $f(x^2) < f(x^1)$. Using the third characterization of concavity, we have:

(106) $f(x^2) \leq f(x^1) + \nabla f(x^1)^T(x^2 - x^1)$
$< f(x^1)$ using (105)

which is the desired result.

It turns out that it is quite difficult to get simple necessary and sufficient conditions for quasiconcavity in the case where f is twice continuously differentiable (although it is quite easy to get sufficient conditions). In order to get necessary and sufficient conditions, we will have to take a bit of a detour for a while.

Definition: Let f(x) be a function of N variables x defined for $x \in S$ where S is a convex set. Then f has *the line segment minimum property* iff

(107) $x^1 \in S$, $x^2 \in S$, $x^1 \neq x^2$ implies $\min_t \{f(tx^1 + (1-t)x^2) : 0 \leq t \leq 1\}$ exists;

i.e., the minimum of f along any line segment in its domain of definition exists.

It is easy to verify that if f is a quasiconcave function defined over a convex set S, then it satisfies the line segment minimum property (107), since the minimum in (107) will be attained at one or both of the endpoints of the interval; i.e., the minimum in (107) will be attained at either $f(x^1)$ or $f(x^2)$ (or both points) since $f(tx^1+(1-t)x^2)$ for $0 \leq t \leq 1$ is equal to or greater than min $\{f(x^1), f(x^2)\}$ and this minimum is attained at either $f(x^1)$ or $f(x^2)$ (or both points).

Definition: Diewert, Avriel and Zang (1981; 400): The function of one variable, g(t), defined over an interval S attains a *semistrict minimum* at $t^0 \in \text{Int } S$ iff there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that $t^0 - \delta_1 \in S$, $t^0 + \delta_2 \in S$ and

(108) $g(t^0) \leq g(t)$ for all t such that $t^0 - \delta_1 \leq t \leq t^0 + \delta_2$;
(109) $g(t^0) < g(t^0-\delta_1)$ and $g(t^0) < g(t^0+\delta_2)$.

If g just satisfied (108) at the point t^0, then it can be seen that g(t) attains a *local minimum* at t^0. But the conditions (109) show that a semistrict local minimum is stronger than a local minimum: for g to attain a semistrict local minimum at t^0, we need g to attain a local minimum at t^0, but the function must eventually strictly increase at the end points of the region where the function attains the local minimum. Note that g(t) attains a *strict local minimum* at $t^0 \in \text{Int } S$ iff there exist $\delta > 0$ such that $t^0 - \delta \in S$, $t^0 + \delta \in S$ and

(110) $g(t^0) < g(t)$ for all t such that $t^0 - \delta \leq t \leq t^0 + \delta$ but $t \neq t^0$.

It can be seen that if g attains a strict local minimum at t^0, then it also attains a semistrict local minimum at t^0. Hence, a semistrict local minimum is a concept that is intermediate to the concept of a local and strict local minimum.

Theorem 12: Diewert, Avriel and Zang (1981; 400): Let f(x) be a function of N variables x defined for $x \in S$ where S is a convex set and suppose that f has the line segment minimum property (107). Then f is quasiconcave over S iff f has the following property:

(111) $x^1 \in S$, $x^2 \in S$, $x^1 \neq x^2$ implies that $g(t) \equiv f(x^1+t[x^2-x^1])$ does not attain a semistrict local minimum for any t such that $0 < t < 1$.

Proof: Quasiconcavity implies (111): This is equivalent to showing that not (111) implies not (65). Not (111) means there exists t* such that $0 < t^* < 1$ and g(t) attains a semistrict local minimum at t*. This implies the existence of t_1 and t_2 such that

(112) $0 \leq t_1 < t^* < t_2 \leq 1$;
(113) $g(t_1) > g(t^*)$; $g(t_2) > g(t^*)$.

Using the definition of g, (113) implies that

(114) $f(x^1+t^*[x^2-x^1]) < \min \{f(x^1+t_1[x^2-x^1]), f(x^1+t_2[x^2-x^1])\}$.

But (112) can be used to show that the point $x^1+t^*[x^2-x^1]$ is a convex combination of the points $x^1+t_1[x^2-x^1]$ and $x^1+t_2[x^2-x^1]$,[13] and hence (114) contradicts the definition of quasiconcavity, (65). Hence f is not quasiconcave.

(111) implies quasiconcavity (65). This is equivalent to showing not (65) implies not (111). Suppose f is not quasiconcave. Then there exist $x^1 \in S$, $x^2 \in S$, and λ^* such that $0 < \lambda^* < 1$ and

(115) $f(\lambda^* x^1+(1-\lambda^*)x^2) < \min \{f(x^1), f(x^2)\}$.

Define $g(t) \equiv f(x^1+t[x^2-x^1])$ for $0 \le t \le 1$. Since f is assumed to satisfy the line segment minimum property, there exists a t^* such that $0 \le t^* \le 1$ and

(116) $g(t^*) = \min_t \{g(t) : 0 \le t \le 1\}$

The definition of g and (115) shows that t^* satisfies $0 < t^* < 1$ and

(117) $f(x^1+t^*[x^2-x^1]) = f((1-t^*)x^1+t^*x^2) \;\; < \min \{f(x^1), f(x^2)\}$.

Thus f attains a semistrict local minimum, which contradicts (111). Q.E.D.

Theorem 13: *Fourth Characterization of Quasiconcavity*; Diewert, Avriel and Zang (1981; 401): Let f(x) be a twice continuously differentiable function of N variables x defined over an open convex subset S of R^N.[14] Then f is quasiconcave over S iff f has the following property:

(118) $x^0 \in S$, $v \neq 0_N$, $v^T\nabla f(x^0) = 0$ implies (i) $v^T\nabla^2 f(x^0)v < 0$ or (ii) $v^T\nabla^2 f(x^0)v = 0$ and $g(t) \equiv f(x^0+tv)$ does not attain a semistrict local minimum at $t = 0$.

Proof: We need only show that property (118) is equivalent to property (111) in the twice continuously differentiable case. Property (111) is equivalent to

(119) $x^0 \in S$, $v \neq 0_N$ and $g(t) \equiv f(x^0+tv)$ does not attain a semistrict local minimum at $t = 0$.

Consider case (i) in (118). If this case occurs, then g(t) attains a strict local maximum at $t = 0$ and hence cannot attain a semistrict local minimum at $t = 0$. Hence, in the twice continuously differentiable case, (118) is equivalent to (119). Q.E.D.

Note that the following condition is *sufficient* for (118) in the twice continuously differentiable case and hence if it holds for every $x^0 \in S$, f will be quasiconcave:

(120) $x^0 \in S$, $v \neq 0_N$, $v^T\nabla f(x^0) = 0$ implies $v^T\nabla^2 f(x^0)v < 0$.

---

[13] The weights are $\lambda \equiv [t_2-t^*]/[t_2-t_1]$ and $1-\lambda \equiv [t^*-t_1]/[t_2-t_1]$.

[14] These conditions are strong enough to imply the continuity of f over S and hence the line segment minimum property will hold for f.

If f satisfies (120), then the Hessian matrix of f, the matrix of second order partial derivatives $\nabla^2 f(x^0)$, is said to be *negative definite in the subspace orthogonal to the gradient vector*, $\nabla f(x^0)$.

Examination of (118) shows that the following condition is *necessary* for quasiconcavity in the twice continuously differentiable case:

(121) $x^0 \in S$, $v \neq 0_N$, $v^T \nabla f(x^0) = 0$ implies $v^T \nabla^2 f(x^0) v \leq 0$.

If f satisfies (121), then the Hessian matrix of f, the matrix of second order partial derivatives $\nabla^2 f(x^0)$, is said to be *negative semidefinite in the subspace orthogonal to the gradient vector*, $\nabla f(x^0)$.

Diewert, Avriel and Zang (1981) show how many other generalizations of concavity can be obtained by looking at the local minimum or maximum properties of a function.

7. Quasiconvex Functions

Definition: Fenchel (1953; 117): f is a *quasiconvex function* defined over a convex subset S of R^N iff

(122) $x^1 \in S$, $x^2 \in S$, $0 < \lambda < 1$ implies $f(\lambda x^1 + (1-\lambda)x^2) \leq \max\{f(x^1), f(x^2)\}$.

Comparing the definition (122) for a quasiconvex function with our previous definition (65) for a quasiconcave function, it can be seen that f satisfies (122) if and only if $-f$ satisfies (65). Thus an equivalent definition for a quasiconvex function f is: f(x) is quasiconvex over the convex set S iff $-f$ is quasiconcave over the convex set S. This fact means that we do not have to do much work to establish the properties of quasiconvex functions: we can simply use all of the material in the previous section, replacing f in each result in the previous section by $-f$ and then multiplying through the various inequalities by -1 (thus reversing them) in order to eliminate the minus signs from the various inequalities.

Problems

18. Write down counterparts to the second, third and fourth characterizations for quasiconcave functions for quasiconvex functions. Your characterizations for quasiconvexity should not contain any minus signs. (No proofs are required.)

19. Luenberger (1968): Let f be a quasiconcave function defined over a convex subset S of R^N. If f attains a strict local maximum at $x^0 \in S$, then f attains a strict global maximum over S at x^0.

20. Let f be a quasiconcave, positive function defined over the convex subset S of R^N. Show that $g(x) \equiv 1/f(x)$ is a quasiconvex function over S.

21. Berge (1963; 208): Let f be a positive, linearly homogeneous and quasiconcave function defined over the positive orthant in R^N, $\Omega \equiv \{x : x >> 0_N\}$. Show that f is concave over Ω.

22. Yaari (1977; 1184): Define $F(x,y) \equiv f(x) + g(y)$ for $x \in X$ and $y \in Y$ where X and Y are convex subsets of R^1. Suppose that f and g are continuous increasing functions over their domains of definition and F is quasiconcave over $X \otimes Y$. Show that at least one of the functions f or g must be concave.

23. Blackorby, Davidson and Donaldson (1977; 357-358): Define $F(x^1,x^2,\ldots,x^M) \equiv g[\sum_{m=1}^M f^m(x^m)]$ for $x^m \in X^m$, m = 1,2,…,M where each X^m is an open convex subset of R^{N_m} for each m. Assume that F is quasiconcave, g is a monotonically increasing twice continuously differentiable function of one variable, defined over the range of $\sum_{m=1}^M f^m(x^m)$ for $x^m \in X^m$, m = 1,2,…,M, and that the functions f^m are twice continuously differentiable. Show that each f^m is quasiconcave over X^m and at most one of the M functions, $f^1,f^2,\ldots,f^M$, can fail to be concave.

References

Arrow, K.J. and A.C. Enthoven (1961), "Quasi-Concave Programming", *Econometrica* 29, 779-800.

Berge, C. (1963), *Topological Spaces*, New York: MacMillan.

Blackorby, C., R. Davidson and D. Donaldson (1977), "A Homiletic Exposition of the Expected Utility Theorem", *Economica* 44, 351-358.

Diewert, W.E., M. Avriel and I. Zang (1981), "Nine Kinds of Quasiconcavity and Concavity", *Journal of Economic Theory* 25:3, 397-420.

Fenchel, W. (1953), "Convex Cones, Sets and Functions", Lecture Notes at Princeton University, Department of Mathematics, Princeton, N.J.

Luenberger, D.G. (1968), "Quasi-Convex Programming", *SIAM Journal of Applied Mathematics* 16, 1090-1095.

Mangasarian, O. (1969), *Nonlinear Programming*, New York: McGraw-Hill.

Minkowski, H. (1911), *Theorie der konvexen Körper,* Gesammelte Abhandlungen, Zweiter Band, Berlin.

Ponstein, J. (1967), "Seven Kinds of Convexity", *SIAM Review* 9, 115-119.

Roberts, A.W. and D.E. Varberg (1973), *Convex Functions*, New York: Academic Press.

Rockafellar, R.T. (1970), *Convex Analysis*, Princeton, N.J.: Princeton University Press.

Rudin, W. (1953), *Principles of Mathematical Analysis*, New York: McGraw-Hill.

Yaari, M.E. (1977), "A Note on Separability and Quasiconcavity", *Econometrica* 45, 1183-1186.