Lecture #9

Proof of the Kuhn-Tucker Theorem

The Kuhn-Tucker theorem is so central to economic thinking that its proof is worth understanding. It also provides the occasion to introduce new concepts that are widely used in economic theory. I first prove the sufficiency for optimality of the Kuhn-Tucker conditions, as this argument requires nothing new. Recall that the sufficiency part of the Kuhn-Tucker theorem is as follows.

Theorem 9.1: Let C be a non-empty convex subset of R^N, let $f: C \to R$ be concave, and for $k = 1, \ldots, K$, let $\overline{a}_k$ be a number and $g_k: C \to R$ be convex. Suppose that $\overline{x} \in C$ is such that $g_k(\overline{x}) \leq \overline{a}_k$, for $k = 1, \ldots, K$, and that $\lambda \in R^K_+$ is such that, for $k = 1, \ldots, K$, $\lambda_k = 0$ if $g_k(\overline{x}) < \overline{a}_k$. Finally suppose that $\overline{x}$ solves the problem

$$\max_{x \in C} \left[f(x) - \sum_{k=1}^{K} \lambda_k g_k(x) \right]. \tag{9.1}$$

Then $\overline{x}$ solves the problem

$$\begin{aligned} &\max_{x \in C} \; f(x) \\ &\text{s.t. } g(x_k) \leq \overline{a}_k, \text{ for } k = 1, \ldots, K. \end{aligned} \tag{9.2}$$

Proof: Notice that

$$\sum_{k=1}^{K} \lambda_k [g_k(\overline{x}) - \overline{a}_k] = 0,$$

because, for each k, $\lambda_k(g_k(\overline{x}) - \overline{a}_k) = 0$, since if $g_k(\overline{x}) - \overline{a}_k \neq 0$, then $g_k(\overline{x}) - \overline{a}_k < 0$ and so $\overline{\lambda}_k = 0$. Therefore

$$f(\overline{x}) = f(\overline{x}) - \sum_{k=1}^{K} \lambda_k [g_k(\overline{x}) - \overline{a}_k] = f(\overline{x}) - \sum_{k=1}^{K} \lambda_k g_k(\overline{x}) + \sum_{k=1}^{K} \lambda_k \overline{a}_k. \tag{9.3}$$

Because $\overline{x}$ solves problem 9.1,

$$f(\bar{x}) - \sum_{k=1}^{K} \lambda_k g_k(\bar{x}) + \sum_{k=1}^{K} \lambda_k \bar{a}_k$$

$$\geq f(x) - \sum_{k=1}^{K} \lambda_k g_k(x) + \sum_{k=1}^{K} \lambda_k \bar{a}_k = f(x) - \sum_{k=1}^{K} \lambda_k [g_k(x) - \bar{a}_k], \tag{9.4}$$

for all $x \in C$. If $x \in C$ is such that $g_k(x) \leq \bar{a}_k$, for all k, then because $\lambda_k \geq 0$, for all k,

$$f(x) - \sum_{k=1}^{K} \lambda_k [g_k(x) - \bar{a}_k] \geq f(x). \tag{9.5}$$

Equation 9.3 and inequalities 9.4 and 9.5 imply the $f(\bar{x}) \geq f(x)$, if x is feasible, that is, if $x \in C$ and $g_k(x) \leq \bar{a}_k$, for all k. Hence $\bar{x}$ solves problem 9.2. ■

The proof of the necessity for optimality of the Kuhn-Tucker conditions uses the concept of the value function for the Kuhn-Tucker problem.

Definition: The value function $V(a) = V(a_1, a_2, \ldots, a_K)$ is

$$V(a_1, \ldots, a_K) = \sup_{x \in C} \{f(x) \mid g_k(x) \leq a_k, \text{ for } k = 1, \ldots, K\}.$$

The domain of definition of V is

$$A = \{(a_1, \ldots, a_K) \mid \text{for some } x \in C, g_k(x) \leq a_k, \text{ for } k = 1, \ldots, K\}.$$

Lemma 9.2: The set A is convex and the value function $V : A \to R$ is concave.

Proof: To show that A is convex, let a and b belong to A and let α be a number such that $0 < \alpha < 1$. By the definition of A, there exist vectors x and y in C such that $g(x) \leq a$ and $g(y) \leq b$. Because C is convex, $\alpha x + (! - \alpha)y$ belongs to C. Because each function g_k is convex,

$$g_k(\alpha x + (1 - \alpha) y) \leq \alpha g_k(x) + (1 - \alpha) g_k(y) \leq \alpha a_k + (1 - \alpha) b_k, \tag{9.6}$$

for all k. This inequality implies that $\alpha a + (1 - \alpha)b$ belongs to A and hence that A is convex. I have also shown that $\alpha x + (1 - \alpha)y$ is a feasible vector for the vector of constraint levels $\alpha a + (1 - \alpha)b$.

I now show that V is concave. Let a and b belong to A, let α be such that $0 < \alpha < 1$. Assume first of all that $V(a) < \infty$ and $V(b) < \infty$. Let ε be a positive number. By the definition of V, there exist vectors x and y in C such that $g(x) \leq a$, $g(y) \leq b$, and $V(a) - \varepsilon < f(x) \leq V(a)$ and

$V(b) - \varepsilon < f(y) \leq V(b)$. Because f is concave and $\alpha x + (1 - \alpha)y$ is a feasible vector for the constraint vector $\alpha a + (1 - \alpha)b$

$$V(\alpha a + (1 - \alpha)b) \geq f(\alpha x + (1 - \alpha)y) \geq \alpha f(x) + (1 - \alpha)f(y)$$

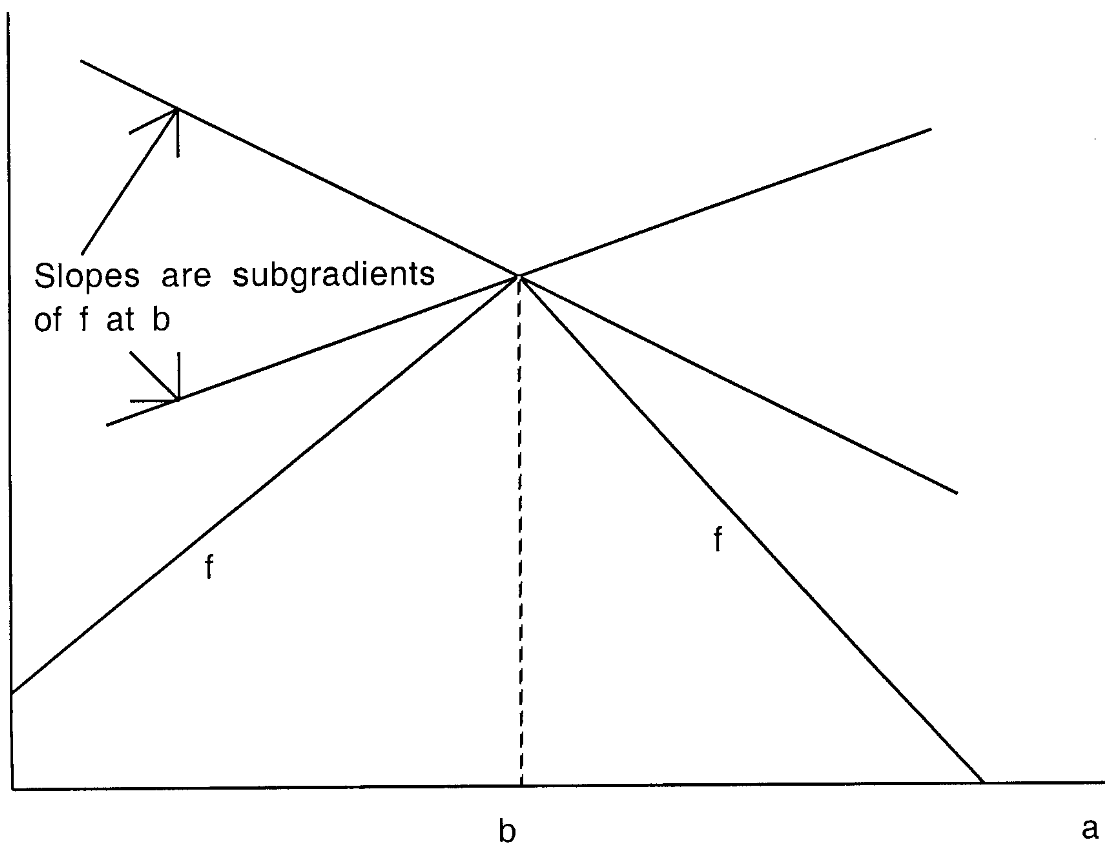
$$\geq \alpha V(a) - \alpha\varepsilon + (1 - \alpha)V(b) - (1 - \alpha)\varepsilon = \alpha V(a) + (1 - \alpha)V(b) - \varepsilon.$$

Since ε is arbitrarily small,

$$V(\alpha a + (1 - \alpha)b) \geq \alpha V(a) + (1 - \alpha)V(b).$$

This inequality proves that V is concave.

Now assume that $V(a) = \infty$. I must show that $V(\alpha a + (1 - \alpha)b) = \infty$. Let T be a positive number. Since $V(a) = \infty$, there is an x in C such that $g(x) \leq a$ and $f(x) > T$. Let y in C be such that $g(y) \leq b$. Then since $\alpha x + (1 - \alpha)y$ is feasible for $\alpha a + (1 - \alpha)b$, it follows that $V(\alpha a + (1 - \alpha)b) \geq f(\alpha x + (1 - \alpha)y) \geq f(\alpha f(x) + (1 - \alpha)f(y) > \alpha T + (1 - \alpha)f(y)$. Since T may be arbitrarily large and we may hold y fixed, it follows that $V(\alpha a + (1 - \alpha)b) = \infty$. A similar argument applies if $V(b) = \infty$. ■



Another concept required for the proof of the necessity of the Kuhn-Tucker conditions is that of a subgradient.

Definition: If A is a subset of R^K and $f: A \rightarrow R$, then a subgradient for f at $a \in A$ is a vector $\lambda \in R^K$ such that

$$f(b) \leq f(a) + \lambda(b - a),$$

for all $b \in A$.

A subgradient is a generalization of the derivative that applies to concave functions. If f is differentiable and concave, then Df(a) is a subgradient of f at a. If f is differentiable and λ is a subgradient of f at a, then $\lambda = Df(a)$. A function can have more than one subgradient at a point, as the function f has at b in the above figure.

Recall that in the theory of constrained optimization with differentiable functions and equality constraints, Lagrange multipliers are the derivatives of the value function with respect to the corresponding constraint levels. A similar assertion applies to Kuhn-Tucker coefficients, but because the value function is not necessarily differentiable but is concave, we replace the derivative with the subgradient. That is, the vector of Kuhn-Tucker coefficients is a subgradient of the value function. In this weak sense, λ_k is the marginal value of increasing the k^{th} constraint level a_k. This is the significance of the next theorem, which says that the Kuhn-Tucker conditions imply that the vector of Kuhn-Tucker coefficients is a subgradient of the value function. Since the Kuhn-Tucker theorem asserts that the Kuhn-Tucker coefficients satisfy these conditions, the vector of coefficients must be a subgradient.

Theorem 9.3: Let $V: A \rightarrow R$ be the value function for the Kuhn-Tucker maximization problem 9.2 and let f and $g_1, \ldots, g_K$ be as in that theorem. Suppose that $\bar{x} \in C$, $\bar{a} \in R^K$, and $\lambda \in R^K_+$ are such that $\bar{x}$ solves the problem

$$\max_{x \in C} \left[f(x) - \sum_{k=1}^{K} \lambda_k g_k(x) \right],$$

and for all k, $g_k(\bar{x}) \leq \bar{a}_k$ and $\lambda_k = 0$ if $g_k(\bar{x}) < \bar{a}_k$. Then λ is a subgradient of V at $\bar{a}$.

Proof: In order to show that λ is a subgradient of V at $\bar{a}$, I must show that if $a \in A$, then

$$V(a) \leq V(\bar{a}) + \sum_{k=1}^{K} \lambda_k (a_k - \bar{a}_k),$$

which is the same as

$$V(\bar{a}) - \sum_{k=1}^{K} \lambda_k \bar{a}_k \geq V(a) - \sum_{k=1}^{K} \lambda_k a_k . \qquad (9.7)$$

Because feasibility and the Kuhn-Tucker conditions are sufficient for optimality, we know that $V(\bar{a}) = f(\bar{x})$. By assumption, for all k, $g_k(\bar{x}) \leq \bar{a}_k$, and $\lambda_k = 0$, if $g_k(\bar{x}) < \bar{a}_k$. It follows that $\lambda_k g_k(\bar{x}) = \lambda_k \bar{a}_k$, for all k, and so $\sum_{k=1}^{K} \lambda_k g_k(\bar{x}) = \sum_{k=1}^{K} \lambda_k \bar{a}_k$. Putting these assertions together, we see that

$$V(\bar{a}) - \sum_{k=1}^{K} \lambda_k \bar{a}_k = f(\bar{x}) - \sum_{k=1}^{K} \lambda_k g_k(\bar{x}) . \qquad (9.8)$$

Because $\bar{x}$ maximizes $f(x) - \sum_{k=1}^{K} \lambda_k g_k(x)$ over C, we know that

$$f(\bar{x}) - \sum_{k=1}^{K} \lambda_k g_k(\bar{x}) \geq f(x) - \sum_{k=1}^{K} \lambda_k g_k(x) , \qquad (9.9)$$

if $x \in C$. Suppose that x is feasible at $a \in A$, that is, $x \in C$ and $g_k(x) \leq a_k$, for all k. Then

$$f(x) - \sum_{k=1}^{K} \lambda_k g_k(x) \geq f(x) - \sum_{k=1}^{K} \lambda_k a_k , \qquad (9.10)$$

since $\lambda_k \geq 0$, for all k. Equation 9.8 and inequalities 9.9 and 9.10 imply that

$$V(\bar{a}) - \sum_{k=1}^{K} \lambda_k \bar{a}_k = f(x) - \sum_{k=1}^{K} \lambda_k a_k ,$$

if x is feasible at a. Because V(a) is the supremum of f(x) for x that are feasible at a, inequality 9.7 follows and so the theorem is proved. ■

An important step in the proof of the necessity of the Kuhn-Tucker conditions is the converse of theorem 9.3, that is, a subgradient of the value function satisfies the Kuhn-Tucker conditions.

Theorem 9.4: Let C, f, g_1, ... , g_K be as in the Kuhn-Tucker theorem and let $V : A \to R$ be the value function. Suppose that $\bar{x}$ solves the problem

$$\max_{x \in C} f(x)$$

$$\text{s.t. } g_k(x) \leq \bar{a}_k, \text{ for } k = 1, \ldots, K.$$

If $\lambda \in R^K$ is a subgradient of V at $\bar{a} \in A$, then, for all k, $\lambda_k \geq 0$, and $\lambda_k = 0$ if $g_k(\bar{x}) < \bar{a}_k$, and $\bar{x}$ solves problem

$$\max_{x \in C} [f(x) - \sum_{k=1}^{K} \lambda_k g_k(x)].$$

<u>Proof</u>: First of all, I show that $\lambda_k \geq 0$, for all k. Because V(a) is the supremum of f(x) over x in C such that $g(x) \leq a$, it follows that if a and b belong to A and $b \leq a$, then $V(b) \leq V(a)$, since any $x \in C$ that satisfies $g(x) \leq b$ also satisfies $g(x) \leq a$. Without loss of generality, I may assume that k = 1. Let $a \in A$ be defined by $a_1 = \bar{a}_1 + 1$ and $a_k = \bar{a}_k$, for k = 2, ..., K. Since λ is a subgradient of V at $\bar{a}$,

$$V(a) \leq V(\bar{a}) + \lambda.(a - \bar{a}) = V(\bar{a}) + \lambda_1 \leq V(a) + \lambda_1,$$

where the last inequality follows because $\bar{a} \leq a$. Hence $\lambda_1 \geq 0$.

I next show that, for all k, $\lambda_k = 0$ if $g_k(\bar{x}) < \bar{a}_k$. Let $a = g(\bar{x})$. Because $a \leq \bar{a}$, $V(a) \leq V(\bar{a})$. Since $g(\bar{x}) = a$, $V(a) \geq f(\bar{x}) = V(\bar{a})$. Therefore $V(a) = f(\bar{x}) = V(\bar{a})$. Since λ is a gradient of V at $\bar{a}$,

$$V(\bar{a}) + \lambda.(a - \bar{a}) \geq V(a) = V(\bar{a}),$$

so that $\lambda.(a - \bar{a}) \geq 0$. Since $\lambda_k \geq 0$ and $a_k - \bar{a}_k \leq 0$, for all k, it follows that $\lambda.(a - \bar{a}) \leq 0$ and so $\lambda.(a - \bar{a}) = 0$. Since $\lambda_k \geq 0$ and $a_k \leq \bar{a}_k$, for all k, the equation

$$0 = \lambda.(a - \bar{a}) = \sum_{k=1}^{K} \lambda_k (a_k - \bar{a}_k)$$

implies that, for all k, $\lambda_k = 0$ if $a_k < \bar{a}_k$, that is, if $g_k(\bar{x}) < \bar{a}_k$.

The last task is to show that $\bar{x}$ solves the problem $\max_{x \in C} [f(x) - \sum_{k=1}^{K} \lambda_k g_k(x)]$. Because λ is a subgradient of V at $\bar{a}$, $V(a) \leq V(\bar{a}) + \lambda.(a - \bar{a})$, for all $a \in A$. Because, for all k, $g_k(\bar{x}) \leq \bar{a}_k$ and $\lambda_k = 0$ if $g_k(\bar{x}) < \bar{a}_k$ and , it follows that

$$\lambda.g(\bar{x}) = \sum_{k=1}^{K} \lambda_k g_k(\bar{x}) = \sum_{k=1}^{K} \lambda_k \bar{a}_k = \lambda.\bar{a}.$$

Let x be an arbitrary point in C and let a = g(x). Then

$$f(x) \leq V(a) \leq V(\bar{a}) + \lambda.(a - \bar{a}) = f(\bar{x}) + \lambda.(g(x) - g(\bar{x})),$$

so that

$$f(x) - \lambda.g(x) \leq f(\bar{x}) - \lambda.g(\bar{x}).$$

That is,

$$f(x) - \sum_{k=1}^{K} \lambda_k g_k(x) \leq f(\bar{x}) - \sum_{k=1}^{K} \lambda_k g_k(\bar{x}),$$

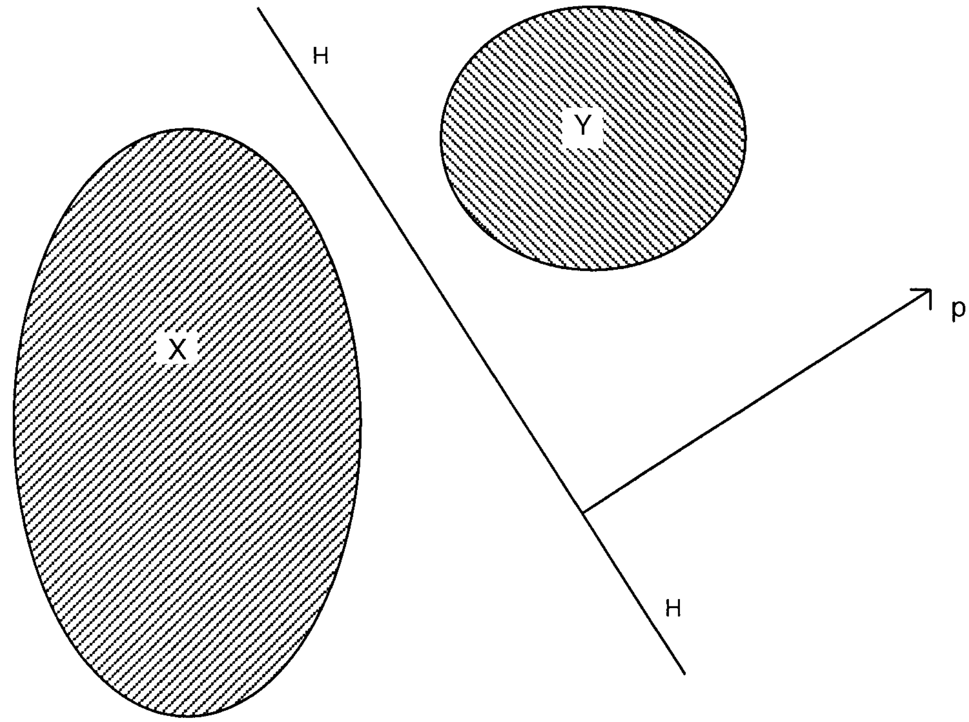
as was to be proved. ■

The necessity of the Kuhn-Tucker conditions for optimality is demonstrated by showing that if problem 9.2 has a solution and if the constraint qualification is satisfied at $\bar{a}$, then the value function V has a subgradient at $\bar{a}$. The proof uses the Minkowski separation theorem.

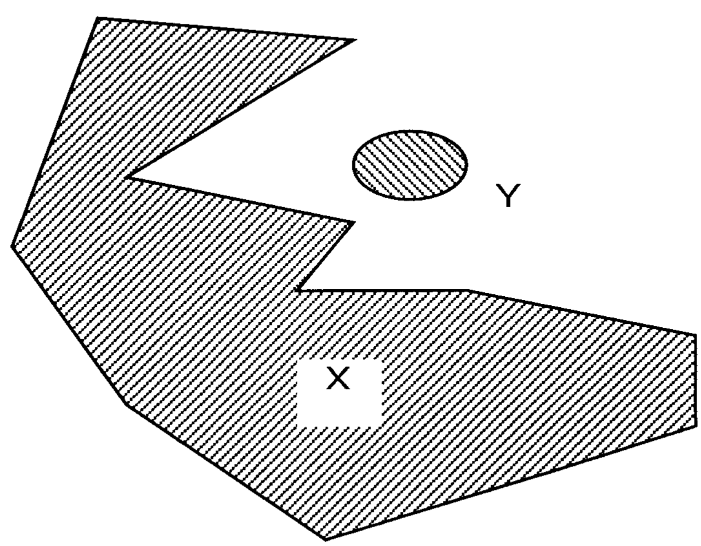
Minkowski Separation Theory: Suppose you draw two disjoint objects on a piece of paper and can cut the piece of paper into two pieces each of which contains one of the objects. In this case, you might say that you could separate the two objects. In the Minkowski theory, the separation is of this nature, except that the cut made by the scissors has to be in a straight line. If we took a vector perpendicular to that straight line, its inner or dot product with every point in one of the objects would be at least as great as its inner product with every point in the other. For this reason, separation is defined in terms of inner products with a fixed vector.

Definition: An N-vector p separates the sets of N-vectors X and Y, if $p \neq 0$ and $p.x \leq p.y$, for all $x \in X$ and $y \in Y$.

The Minkowski separation theorem gives conditions under which two sets of N-vectors may be separated by a non-zero N-vector.



In the above figure, the vector p separates the sets X and Y. The line H going between X and Y is an example of a hyperplane. A hyperplane in R^N is a set of the form a + W, where a $\in R^N$ and W is a subspace of dimension N − 1. If the vector p separates X from Y, then there is a number r such that $p.x \leq r \leq p.y$, for all $x \in X$ and $y \in Y$, and $\{x \in R^N \mid p.x = r\}$ is a hyperplane that divides R^N into two halves, one containing X and the other containing Y. For this reason, the Minkowski separation theorem is often referred to as the theorem of the separating hyperplane.



The statement of Minkowski's theorem requires the concept of the interior of a set.

Definition: If X is a set of N-vectors, the interior of X, written as int X, is

$\{x \in X \mid \text{for some } \varepsilon > 0, \|y - x\| < \varepsilon \text{ implies that } y \in X\}$.

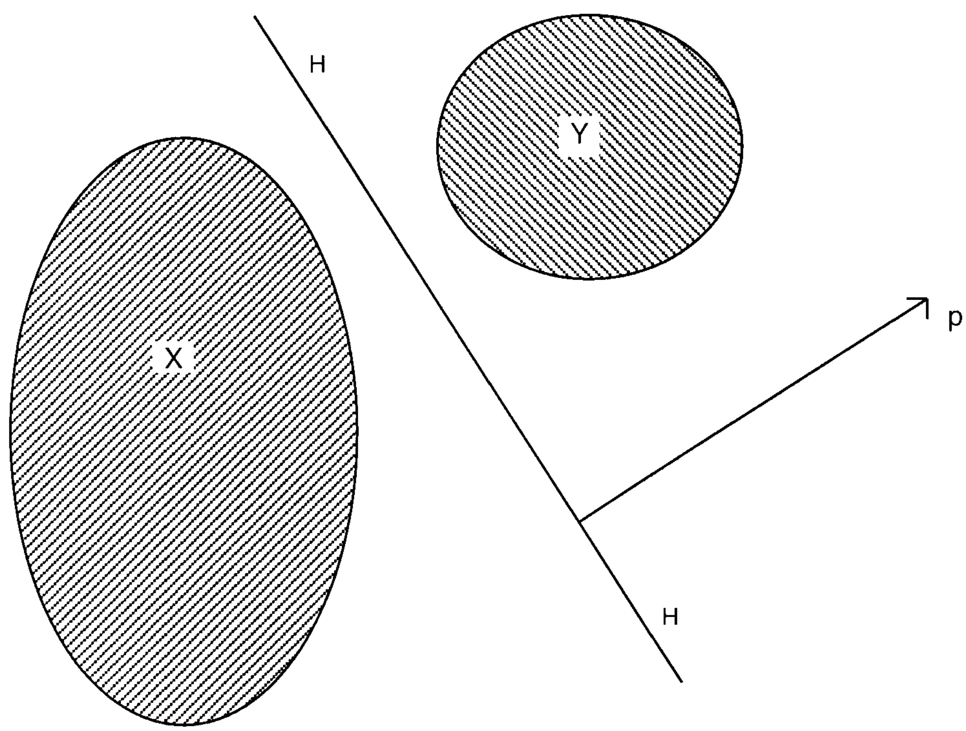
The interior of a set X is open and it is the largest open set contained in X.

Example: int [0, 1] = (0, 1).

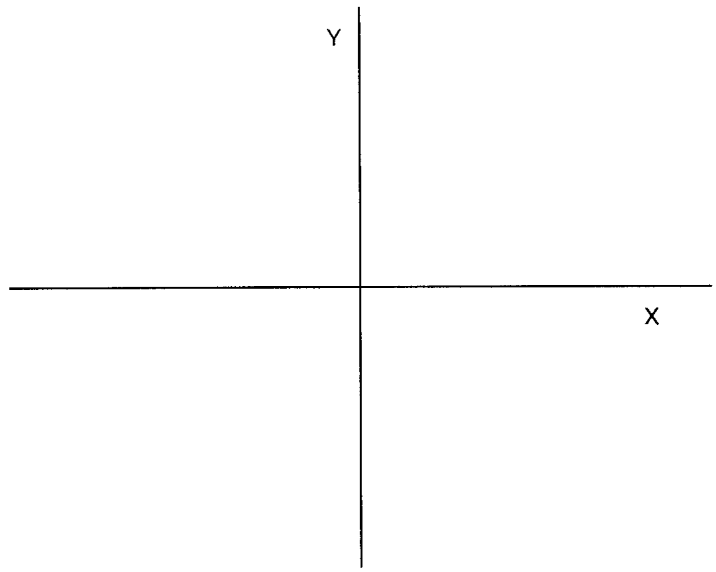
Minkowski Separation Theorem 9.5: Let X and Y be convex sets of N-vectors and suppose that int X is not empty and does not intersect Y. Then there exists a non-zero N-vector p that separates X from Y.

I do not provide a proof of this theorem.

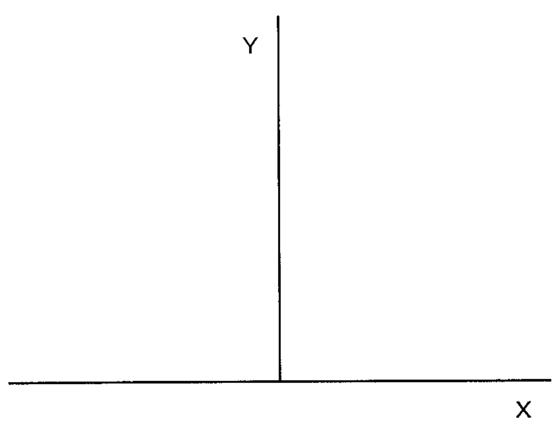
Example: The figure below illustrates why the sets X and Y are assumed to be convex in the statement of the theorem. The sets X and Y are disjoint. The set X is not convex, has non-empty interior, and clearly cannot be separated from Y.



Example: The first figure on the next page illustrates why it is assumed in the theorem that one of the sets to be separated has non-empty interior. Both X and Y are convex, but both have empty interior, so that the interior of neither set intersects the other, though the sets themselves intersect. The sets clearly cannot be separated.



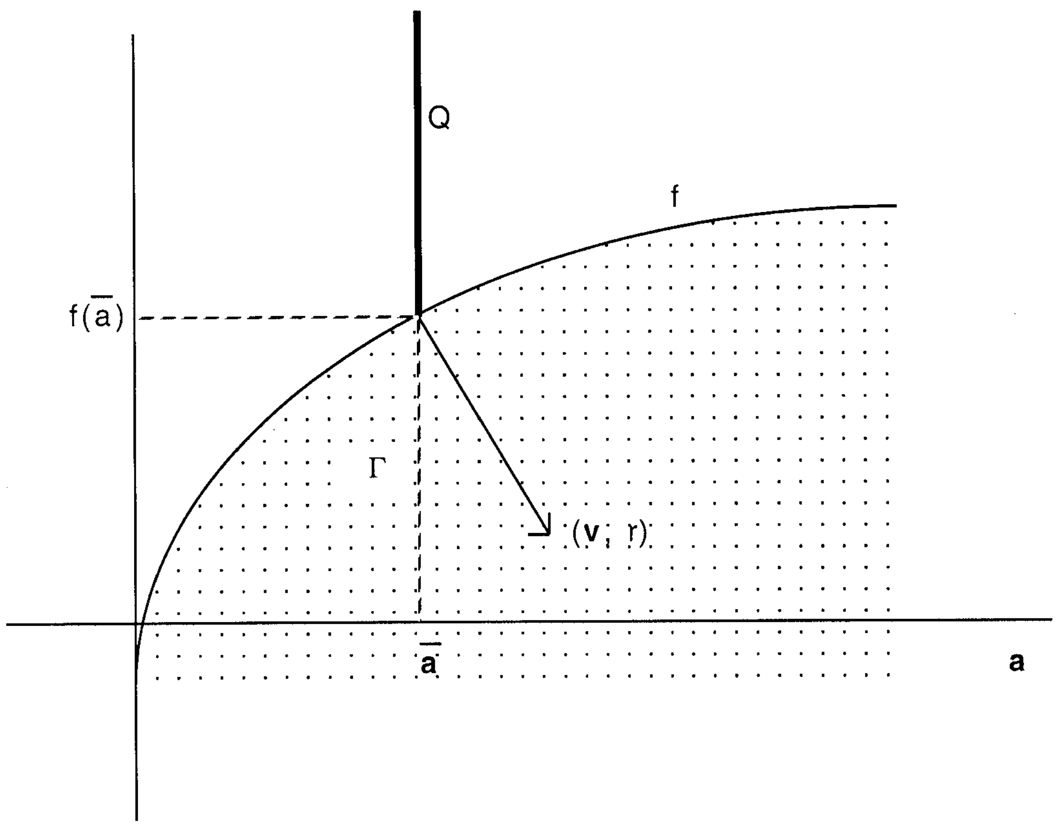
Remark: A slightly stronger version of Minkowski's theorem takes care of this example and dispenses with the assumption that the interior of X or Y be non-empty. This version replaces the interiors of X and Y with their relative interiors. The relative interior of a subset X of R^N is the interior of X in the smallest hyperspace containing X. A hyperspace in R^N is a set of the from a + V, where $a \in R^N$ and V is a subspace of R^N. If dim V = 0, then a + V is just the point a. If X is convex and non-empty, then its relative interior is non-empty. The stronger version of Minkowski's theorem says that X may be separated from Y by a non-zero N-vector p if X and Y are convex and their relative interiors do not intersect. This theorem would apply to the example pictured below. In employing this version of Minkowski's theorem, you need to be sure to obtain a separating vector p which does not have zero inner product with all the vectors in X and Y. In order to guarantee such a non-trivial p, you may proceed as follows. First replace X and Y by X – a and Y – a, where a belongs to X or Y, so that the vector 0 belongs to the union of the new sets X and Y. Then apply the stronger version of Minkowski's theorem to the smallest subspace V of R^N that contains both X and Y. Since the separating vector p is non-zero and belongs to V, it will have a non-zero inner product with some vector in X or Y. Since V is a subset of R^N, you may think of p as a vector in R^N.



The Minkowski separation theorem will be used to prove the following lemma.

Lemma 9.6: If A is a convex subset of R^N and $f: A \to R$ is concave and $\bar{a} \in$ int A, then f has a subgradient p at $\bar{a}$.

Proof: I apply the Minkowski separation theorem. Let $\Gamma = \{(a, t) \in A \times R \mid t \le f(a)\}$ and let $Q = \{(\bar{a}, t) \mid t \in R \text{ and } t \ge f(\bar{a})\}$. It should be clear that Q is convex. The set Γ is convex, because A is convex and f is concave. Since $(\bar{a}, f(\bar{a}) - 1) \in$ int Γ, Γ has non-empty interior. If $(a, t) \in$ int Γ, then $t < f(a)$, so that Q does not intersect the interior of Γ. The next figure should help you visualize the sets Q and Γ.



Since all the conditions of Minkowski's theorem apply, there exists a non-zero vector $(v, r) \in R^N \times R$ such that

$$(v, r).(a, t) \ge (v, r).(\bar{a}, s),$$

for all $(a, t) \in \Gamma$ and $(\bar{a}, s) \in Q$.

I show that $r \le 0$. Since $(\bar{a}, f(\bar{a})) \in \Gamma$ and $(\bar{a}, f(\bar{a}) + 1) \in Q$, it follows that

$$(v, r).(\bar{a}, f(\bar{a})) \geq (v, r).(\bar{a}, f(\bar{a}) + 1) = (v, r).(\bar{a}, f(\bar{a})) + r,$$

and so $r \leq 0$.

I next show that $r < 0$. Since $r \leq 0$, it is sufficient to show that $r \neq 0$. Suppose that $r = 0$. Since $(v, r) \neq 0$, we know that $v \neq 0$. Since $(\bar{a}, f(\bar{a})) \in Q$ and $(a, f(a)) \in \Gamma$, for all $x \in A$, we know that

$$(v, 0).(a, f(a)) \geq (v, 0).(\bar{a}, f(\bar{a})),$$

for all $a \in A$. That is, $v.a \geq v.\bar{a}$, for all $a \in A$. Since $\bar{a} \in \text{int } A$, $\bar{a} - \varepsilon v \in A$, for a sufficiently small positive number ε. Hence $v.\bar{a} - \varepsilon v.v = v.(\bar{a} - \varepsilon v) \geq v.\bar{a}$, which is impossible, since $v.v > 0$. Therefore $r < 0$.

Let $p = \frac{1}{-r} v$. Since $-r > 0$, the vector $(p, -1) = \frac{1}{-r}(v, r)$ separates Γ from Q. Since $(a, f(a)) \in \Gamma$, for all $a \in A$, and $(\bar{a}, f(\bar{a})) \in Q$, we see that

$$(p, -1).(a, f(a)) \geq (p, -1).(\bar{a}, f(\bar{a})),$$

for all $a \in A$. Hence

$$p.a - f(a) \geq p.\bar{a} - f(\bar{a}),$$

which is the same as

$$f(a) \leq f(\bar{a}) + p.(a - \bar{a}).$$

That is, p is a subgradient of f at $\bar{a}$. ■

We are now in a position to prove the necessity of the Kuhn-Tucker conditions for optimality. First of all, I restate the necessity part of the Kuhn-Tucker theorem.

<u>Theorem</u> 9.7: Let C be a non-empty convex subset of R^N, let $f: C \to R$ be concave, and for $k = 1, \dots, K$, let $\bar{a}_k$ be a number and $g_k: C \to R$ be convex. Suppose that $\bar{x}$ solves problem

$$\max_{x \in C} f(x) \quad \text{s.t. } g(x_k) \leq \bar{a}_k, \text{ for } k = 1, \dots, K. \tag{9.11}$$

Assume that the constraint qualification applies. That is, there exists a vector $\underline{x} \in C$ such that

$g_k(\underline{x}) < \bar{a}_k$, for all k. Then there exists $\lambda \in R^K_+$ such that, for all k, $\lambda_k = 0$, if $g_k(\bar{x}) < \bar{a}_k$, and $\bar{x}$ solves problem

$$\max_{x \in C} \, [f(x) - \sum_{k=1}^{K} \lambda_k g_k(x)] .$$

Proof: By theorem 9.4, it is sufficient to show that the value function V for problem 9.11 has a subgradient at $\bar{a}$. Observe that $\bar{a} \in$ int A. This is so because if $\underline{x}$ is as in the constraint qualification, then $g_k(\underline{x}) < \bar{a}_k$, for all k, so that

$$\{a \in R^K \,|\, a_k > g_k(\underline{x}), \text{ for all } k\}$$

is an open set contained in A that contains $\bar{a}$. Since V is concave by lemma 9.2 and $\bar{a} \in$ int A, lemma 9.6 implies that V has a subgradient at $\bar{a}$. ■

Dynamic Programming with Discrete Time and Discounting: Introduction

I begin by describing the class of problems considered. The objective is to control a sequence of states, $x_0, x_1, \ldots, x_t, \ldots$, where x_t is the state time t. Such a sequence is called a program. It is assumed that the initial state x_0 is given and that all the states belong to a subset X of R^N. Given the state x_t, the state x_{t+1} may be chosen at time t from a set $G_t(x_t)$, where G_t is a correspondence from X to X. A program $x_0, x_1, \ldots, x_t, \ldots$ is said to be feasible if $x_{t+1} \in G_t(x_t)$, for all t. The objective or utility function in period t is $u_t : A \to R$, where A is a subset of $X \times X$. The overall objective is to maximize the function

$$u_t(x_t, x_{t+1})$$

over all feasible programs, where T may be a finite integer, though I will eventually concentrate on the case where T is infinity. The starting point of dynamic programming is that if T is finite, then the problem can, conceptually at least, be solved by backwards induction over time. Let $V_t(x_t)$ be the maximized value from time t on, when the state at time t is x_t. That is,

$$V_t(x_t) = \max_{\substack{x_{s+1} \in G_s(x_s), \\ \text{for } s = t, \ldots, T}} \sum_{s=t}^{T} u_s(x_s, x_{s+1})$$

It should be intuitively clear that the following equation ought to hold, provided the maximum exists,

$$V_t(x_t) = \max_{x_{t+1} \in G_t(x_t)} [u_t(x_t, x_{t+1}) + V_{t+1}(x_{t+1})], \qquad (9.12)$$

for t = 0, ... , T − 1 and all x_t. Equation 9.12 is known as the <u>Bellman equation</u>. Using this equation repeatedly, it should be possible to maximize $\sum_{t=0}^{T} \beta^t u_t(x_t, x_{t+1})$ by first solving the problem

$$V_T(x_T) = \max_{x_{T+1} \in G_T(x_T)} u_T(x_T, x_{T+1}),$$

for each x_T, and then solving

$$V_{T-1}(x_{T-1}) = \max_{x_T \in G_{T-1}(x_{T-1})} [u_{T-1}(x_{T-1}, x_T) + V_T(x_T)],$$

for all x_{T-1}, and then continuing backwards though time to time 0. This procedure may seem tedious, but it can give insight into particular problems, as I try to show you by means of an example involving the choice of money balances or savings, which appears as an appendix to this lecture and will not be presented in class.

There are at least three approaches to the theory of dynamic programming. One is to use backward induction to exploit the repetitive structure of dynamic programming problems in order to derive properties of solutions and of the value functions. The example in the appendix is meant to illustrate this approach. Another approach is to consider in its entirety the set of programs, x_0, x_1, ... , x_t, ... , an approach known as the sequence approach. In the next lecture, I will use this approach to show that an optimal program exists. The contraction mapping approach is a third one that applies to problems where future gain or utility is discounted by a constant factor and where the structure of the problem does not change over time. For such problems, the value function satisfies the Bellman equation with the same value function on its left and right-hand sides. An optimal program exists if and only if there is a value function that satisfies this equation, and the contraction mapping theorem is used to prove that such a value function exists. I will discuss this approach in lecture 12. All three approaches are useful and give different insights into the properties of optimal programs.

Dynamic Programming with Discrete Time and Discounting Over an Infinite Horizon

I now turn to the study of discounted dynamic infinite horizon programming problems with a structure that does not change over time. I simplify the dynamic programming model by making it stationary. That is, I let the utility at time t be $\beta^t u(x_t, x_{t+1})$, where $0 < \beta < 1$, and I assume that the set of possible states in period t+1, $G(x_t)$, does not depend directly on t. So the general optimization problem is

$$\max_{\substack{x_{s+1} \in G(x_s),\\ \text{for } s = 0, 1, \ldots}} \sum_{s=t}^{\infty} \beta^s u(x_s, x_{s+1}),$$

where $0 < \beta < 1$ and x_0 is given. The states x_t belong to the subset X of R^N, $u: A \to R$, where A is a subset of $X \times X$, and G is a correspondence from X to X.

I list assumptions made about the elements of the model.

Assumption 1: X is non-empty and $G(x)$ is non-empty, for every $x \in X$.

Assumption 2: G is upper semicontinuous.

Assumption 3: X is compact.

Assumption 4: The utility function in one time period is $u : A \to R$, where A is a compact and non-empty subset of $X \times X$ such that $(x, y) \in A$, for every $x \in X$ and $y \in G(x)$. The function u is continuous.

Definition: If $x_0 \in X$, then the set of feasible programs is

$$\mathcal{F}(x_0) = \{(x_1, x_2, \ldots.) \mid x_{t+1} \in G(x_t), \text{ for } t = 0, 1, 2, \ldots.\}.$$

Formally the maximization problem studied is as follows.

Problem: Given $x_0 \in X$, solve the problem

$$\max_{(x_1, x_2, \ldots..) \in \mathcal{F}(x_0)} \sum_{t=0}^{\infty} \beta^t u(x_t, x_{t+1}) . \tag{9.13}$$

Definition: A solution of this problem is called an optimal program.

It is first necessary to check that the objective function in problem 9.13 makes sense. Because the set A is compact and u is continuous, u is bounded on A. Let $b = \max_{(x, y) \in A} |u(x, y)|$.

Since $b < \infty$,

$$\left|\sum_{t=T}^{\infty} \beta^t u(x_t, x_{t+1})\right| \le \sum_{t=T}^{\infty} \beta^t \left|u(x_t, x_{t+1})\right| \le \frac{b\beta^T}{1-\beta}, \tag{9.14}$$

for all positive integers T. Therefore $\lim_{T \to \infty} \sum_{t=T}^{\infty} \beta^t u(x_t, x_{t+1}) = 0$ and so the sequence $\sum_{t=0}^{T} \beta^t u(x_t, x_{t+1})$ is Cauchy and hence converges, and its limit, $\sum_{t=o}^{\infty} \beta^t u(x_t, x_{t+1})$, is well-defined.

Theorem 9.8: Under assumptions 1 - 4, an optimal program exists.