Lecture 8

Dependence of Constrained Optima on Parameters

We continue to study the problem

\[
\begin{align*}
\max_{x \in U} & \quad f(x) \\
\text{s.t.} & \quad g_k(x) = \bar{a}_k,
\end{align*}
\]

for \( k = 1, \ldots, K, \)

where \( U \) is an open subset of \( \mathbb{R}^N \), \( f \) and each of the \( g_k \) are twice continuously differentiable

functions from \( U \) to \( \mathbb{R} \), and the \( \bar{a}_k \) are numbers. Suppose that a local optimum is achieved at \( \bar{x} \) and let \( \bar{\lambda} = (\bar{\lambda}_1, \ldots, \bar{\lambda}_K) \) be the vector of corresponding Lagrange multipliers, the existence of which is guaranteed by theorem 7.6. Suppose that we vary the vector of constraint levels, \( a = (a_1, \ldots, a_K) \) over an open set containing \( \bar{a} = (\bar{a}_1, \ldots, \bar{a}_K) \). I use the implicit function theorem to show that a local optimum \( \bar{x} \) and the corresponding vector of Lagrange multipliers \( \bar{\lambda} \) may be written as continuously differentiable functions of the \( K \)-vector \( a \). Assume that the first and second order conditions for a local maximum hold at \( \bar{x} \). That is, \( Df(\bar{x}) = \bar{\lambda}^T Dg(\bar{x}) \) and

\[
\nabla^T D^2 f(\bar{x}) (\bar{\lambda}, \bar{a}) v < 0, \quad \text{for all non-zero vectors} \quad v \in Z(\bar{x}),
\]

where

\[
\begin{align*}
g(x) &= \begin{pmatrix} g_1(x) \\ \vdots \\ g_K(x) \end{pmatrix} \\
L(\bar{x}, \bar{\lambda}, \bar{a}) &= f(\bar{x}) - \bar{\lambda}^T [g(\bar{x}) - \bar{a}] = f(\bar{x}, \ldots, \bar{x}) - \sum_{k=1}^K \bar{\lambda}_k [g_k(\bar{x}, \ldots, \bar{x}) - \bar{a}_k], \quad \text{and} \\
Z(\bar{x}) &= \{v \in \mathbb{R}^N \mid Dg(\bar{x})(v) = 0, \quad \text{for} \quad k = 1, \ldots, K\}. \quad \text{Assume also that the constraint qualification applies, so that} \quad Dg(\bar{x}) \quad \text{has rank} \quad K.
\end{align*}
\]

I begin by writing down the necessary conditions for a constrained local optimum with constraint level vector \( a \),
\[ Df(x) - \lambda^T Dg(x) = 0 \]
\[ -g(x) + a = 0. \]

These equations may be written as the single equation \( H(x, \lambda, a) = 0 \), where

\[
H(x, \lambda, a) = \left( \begin{array}{c}
(Df(x))^T - (Dg(x))^T \lambda \\
-g(x) + a
\end{array} \right) = \left( \begin{array}{c}
(Df(x))^T - \sum_{k=1}^{K} \lambda_k (Dg_k(x))^T \\
-g(x) + a
\end{array} \right)
\]

\[
= \left( \begin{array}{c}
(D\mathcal{L}(x, \lambda, a))^T \\
-g(x) + a
\end{array} \right)
\]

where \( H(x, \lambda, a) \) is an \((N+K)\times 1\) matrix and the vectors \( g(x) \), \( a \), and \( \lambda \) are to be thought of as \( K \times 1 \) matrices and the vectors \( Df(x) \) and \( Dg(x) \) as \( 1 \times N \) matrices. Because \( f \) and the \( g_k \) are twice continuously differentiable, \( H \) is continuously differentiable and we may apply the implicit function theorem 7.5 to the equation \( H(x, \lambda, a) = 0 \) and write \( (x, \lambda) \) as a function of the parameter vector \( a \), provided the \((N+K)\times(N+K)\) matrix \( D_{(x,\lambda)}H(x,\lambda,a) \) has full rank. If we calculate this derivative, we see that

\[
D_{(x,\lambda)}H(x,\lambda,a) = \left( \begin{array}{c}
D^2f(x) - \sum_{k=1}^{K} \lambda_k D^2g_k(x) \\
\lambda_k (Dg_k(x))^T
\end{array} \right)
\]

\[
= \left( \begin{array}{cc}
D^2\mathcal{L}(x,\lambda,a) - Dg(x)^T & -Dg(x)^T \\
-Dg(x) & 0
\end{array} \right)
\]

This matrix is called the bordered Hessian. It is the Jacobian at \( (x,\lambda,a) \) of \( H(x,\lambda,a) \) considered as a function of \( x \) and \( \lambda \).

I show that under the assumed conditions, the bordered Hessian is non-singular, that is, that it has rank \( N+K \). I must show that if

\[
\left( \begin{array}{cc}
D^2\mathcal{L} - (Dg)^T \\
-Dg & 0
\end{array} \right) \begin{pmatrix} v \\ w \end{pmatrix} = 0,
\]

then \( \begin{pmatrix} v \\ w \end{pmatrix} = 0 \), where \( v \in \mathbb{R}^N \) and \( w \in \mathbb{R}^K \). If
\[
\begin{pmatrix}
D^2_x \mathcal{L} & -(Dg)^T \\
-Dg & 0
\end{pmatrix}
\begin{pmatrix}
v \\
w
\end{pmatrix}
= 0, \quad (8.1)
\]

then \(-Dg(\bar{x})v = 0\) and so \(v^T(Dg(\bar{x}))^T = (Dg(\bar{x})v)^T = 0\). Therefore

\[
0 = (v^T, w^T) \begin{pmatrix}
D^2_x \mathcal{L} & -(Dg)^T \\
-Dg & 0
\end{pmatrix}
\begin{pmatrix}
v \\
w
\end{pmatrix}
= v^T(D^2_x \mathcal{L})v - v^T(Dg)^Tw - w^T(Dg)v
\]

\[
= v^T(D^2_x \mathcal{L})v. \quad (8.2)
\]

Because \(Dg(\bar{x})v = 0\), it follows that \(v \in Z(\bar{x})\) and hence the second order conditions satisfied at \(\bar{x}\) imply that \(v^T D^2_x \mathcal{L}(\bar{x}, \bar{\lambda})v < 0\), unless \(v = 0\). Therefore equation 8.2 implies that \(v = 0\).

Equation 8.1 implies that \((D^2_x \mathcal{L})v - (Dg)^Tw = 0\). Hence \((Dg)^Tw = 0\), so that

\[
w^T(Dg(\bar{x})) = [(Dg(\bar{x}))^Tw]^T = 0. \quad (8.3)
\]

The constraint qualification implies that the \(K\) rows of \(Dg(\bar{x})\) are independent and so \(w = 0\). Since I have shown that \((v, w) = 0\), it follows that the bordered Hessian has rank \(N+K\).

By the implicit function theorem, there exists an open set \(V\) in \(\mathbb{R}^K\) such that \(\bar{a} \in V\) and there exist continuously differentiable functions \(x(a)\) and \(\lambda(a)\) such that \(x(\bar{a}) = \bar{x}\) and \(\lambda(\bar{a}) = \bar{\lambda}\) and \(H(x(a), \lambda(a), a) = 0\), for all \(a \in V\). That is, \(Df(x(a)) - \sum_{k=1}^{K} \lambda_k(a) Dg(x(a)) = 0\)

\[
= 0 \quad \text{and} \quad -g(x(a)) + a = 0, \quad \text{for all} \quad a \in V. \quad (8.4)
\]

In more compact notation, \(Df(x(a)) = \lambda^T(a) Dg(x(a))\) and \(g(x(a)) = a\), for all \(a \in V\). By theorem 7.7, given the vector \(a\), sufficient conditions for a local maximum at \(x\) are that \(g(x) = a\) and for some \(K\)-vector \(\lambda\), \(Df(x) = \lambda^T Dg(x)\), and \(v^T D^2_x \mathcal{L}(x, \lambda)v < 0\), for all \(v \in \mathbb{R}^N\) such that \(v \neq 0\) and \(Dg(x)v = 0\). These conditions hold at \((\bar{x}, \bar{\lambda})\) and the conditions \(g(x(a)) = a\) and \(Df(x(a)) = \lambda(a)^T Dg(x(a))\) hold for all \(a \in V\). I show that
if $a$ is close enough to $\bar{a}$, then $v^T D^2_x L(x(a), \lambda(a)) v < 0$, for all $v \in \mathbb{R}^n$ such that $v \neq 0$ and $Dg(x(a))(v) = 0$. 

(8.3)

Assume that this statement is false. Then there is a sequence $a_n$ in $V$ such that $\lim_{n \to \infty} a_n = \bar{a}$ and a sequence $v_n$ in $\mathbb{R}^n$ such that, for all $n$, $v_n \neq 0$, $Dg(x(a_n))(v_n) = 0$, and

$v^T D^2_x L(x(a_n), \lambda(a_n)) v_n \geq 0$. We may assume that $\|v_n\| = 1$, for all $n$. By the Bolzano-Weierstrass theorem 4.12, there is a subsequence $v_{n_k}$ that converges to some $\bar{v} \in \mathbb{R}^n$. Then $\|\bar{v}\| = 1$, so that $\bar{v} \neq 0$. Since $\lim_{k \to \infty} a_{n_k} = \bar{a}$ and the function $x(a)$ is continuous, $x(a_{n_k})$ converges to $x(\bar{a}) = \bar{x}$. Similarly $\lambda(a_{n_k})$ converges to $\lambda(\bar{a}) = \bar{\lambda}$. Since $Dg(x(a))$ depends continuously on $a$,

$$Dg(x)(\bar{v}) = \lim_{k \to \infty} Dg(x(a_{n_k}))(v_{n_k}) = \lim_{k \to \infty} 0 = 0.$$ 

Similarly since $D^2_x L(x(a), \lambda(a))$ depends continuously on $a$,

$$v^T D^2_x L(\bar{x}, \bar{\lambda}) \bar{v} \geq 0,$$

contrary to hypothesis. This proves statement 8.3, so that we may assume that the open set $V$ containing $\bar{a}$ is so small that $v^T D^2_x L(x(a), \lambda(a)) v < 0$, for all $a \in V$ and for all $v \in \mathbb{R}^n$ such that $v \neq 0$ and $Dg(x(a))(v) = 0$. It follows that if $V$ is made this small, then the constrained problem has a local maximum at $x(a)$, for all $a \in V$.

I now show that the Lagrange multipliers $\lambda_k(a)$ are the marginal values of the constraint limits $a_k$, where $a_k$ is the $k$th component of $a$. Let $F(a) = f(x(a))$, for all $a \in V$. I show that $DF(a) = \lambda(a)$, so that, for all $k$, $\lambda_k(a)$ is the rate at which the locally maximized value of the
optimization problem increases as \( a \) increases. Because \( g(x(a)) = a \), for all \( a \in V \), the chain rule implies that 
\[
D_x g(x(a)) Dx(a) = I \quad \text{and so}
\]
\[
DF(a) = D_x f(x(a)) Dx(a) = \lambda^T(a) D_x g(x(a)) Dx(a) = \lambda^T(a) I = \lambda^T(a).
\]

The Envelope Theorem for Constrained Optimization

Consider the problem
\[
\max_x f(x, b) \quad \text{s.t. } g(x, c) = a,
\]
where \( b \) and \( c \) are vectors of parameters. The question addressed is how to compute the derivative of the maximum value with respect to \( a, b, \) and \( c \), where the maximum value is
\[
F(a, b, c) = \max_x \{ f(x, b) \mid g(x, c) = a \}.
\]

Assume that \( x \) varies over an open subset \( U \) of \( \mathbb{R}^N \), that \( b \) varies over an open subset \( V \) of \( \mathbb{R}^F \), and that \( c \) varies over an open subset \( W \) of \( \mathbb{R}^M \), so that \( f: U \times V \to \mathbb{R} \) and \( g: U \times W \to \mathbb{R}^K \). Assume that both \( f \) and \( g \) are twice continuously differentiable and that the \( K \times N \) matrix \( D_x g(x, c) \) has rank \( K \), for all \( x \in U \) and \( c \in W \). Let \( \bar{x} \) be a solution of problem 8.4 when \( (a, b, c) = (\bar{a}, \bar{b}, \bar{c}) \).

Then \( g(\bar{x}, \bar{c}) = \bar{a} \) and by theorem 7.6 there is \( \bar{\lambda} \in \mathbb{R}^K \) such that
\[
D_x f(\bar{x}, \bar{b}) = \bar{\lambda}^T Dg(\bar{x}, \bar{c}).
\]

Assume that the second order conditions in theorem 7.7 for a local maximum apply, so that
\[
v^T D_x^2 \mathcal{L}(\bar{x}, \bar{\lambda}, \bar{a}, \bar{b}, \bar{c}) v < 0,
\]
for all \( v \in \mathbb{R}^N \) such that \( v \neq 0 \) and \( D_x g(\bar{x}, \bar{c})(v) = 0 \), where
\[
\mathcal{L}(x, \lambda, a, b, c) = f(x, b) - \lambda^T [g(x, c) - a].
\]
The first order conditions and constraints for parameter values \( a, b, c \) are

5
\[ \mathcal{D} f(x, b) - \lambda^T \mathcal{D} g(x, c) = 0 \]

\[-g(x, c) + a = 0.\]

Let

\[ H(x, \lambda, a, b, c) = \begin{pmatrix} (\mathcal{D} f(x, b))^T - (\mathcal{D} g(x, c))^T \lambda \\ -g(x, c) + a \end{pmatrix} \]

By an argument given earlier in this lecture, \( \mathcal{D} H(\overline{x, \lambda, a, b, c}) \) is invertible, so that by the implicit function theorem 7.5 applied to the equation \( H(x, \lambda, a, b, c) = 0 \), there exist locally defined continuously differentiable functions \( x(a, b, c) \) and \( \lambda(a, b, c) \) such that \( x(a, b, c) \) is a local constrained maximum with corresponding Lagrange multipliers \( \lambda(a, b, c) \in \mathbb{R}^k \).

Also \( x(\overline{a, b, c}) = \overline{x} \) and \( \lambda(\overline{a, b, c}) = \overline{\lambda} \). Since \( x(a, b, c) \) may be a local rather than a global constrained maximum, we define \( F \) by the equation

\[ F(a, b, c) = f(x(a, b, c), b) \]

rather than equation (8.5). Then

\[ \mathcal{D} F(\overline{a, b, c}) = \mathcal{D} f(x(a, b, c), b) \bigg|_{(a, b, c) = (\overline{a, b, c})} \]

I show that

\[ \mathcal{D} F(\overline{a, b, c}) = \mathcal{D} f(x, b) - \overline{\lambda}^T (g(x, c) - a) \bigg|_{(a, b, c) = (\overline{a, b, c})} \]

\[ = \mathcal{D} (\overline{x, \lambda, a, b, c}) \bigg|_{(a, b, c) = (\overline{a, b, c})} \]

(8.6)

This equation is the envelope theorem for constrained maximization problems. I divide its proof into three steps.

**Step 1:** (The argument in this step was made earlier without the extra parameters \( b \) and \( c \).) Differentiate both sides of the equation

\[ g(x(a, b, c), c) = a \]

with respect to \( a \) at \( \overline{a} \). Then
\[ Dg(x, c) Dx(a, b, c) = I, \] (8.7)

so that

\[
\begin{align*}
DF(a, b, c) &= Df(x(a, b, c), b) \bigg|_{b=a} \\
&= Df(x, b) Dx(a, b, c) \\
&= \lambda^T Dg(x, c) Dx(a, b, c) = \lambda^T I = \lambda^T,
\end{align*}
\]

where the second to last equation follows from equation 8.7.

**Step 2:** Differentiate both sides of the equation

\[ g(x(a, b, c), c) = a \]

with respect to \( b \) at \( b = \overline{b} \). Then

\[ 0 = Dg(x(a, b, c), c) = Dx(x, c) Db(x, a, b, c), \] (8.8)

so that

\[
\begin{align*}
DF(a, b, c) &= Df(x(a, b, c), b) \bigg|_{b=b} \\
&= Df(x, b) Dx(a, b, c) + Df(x, b) \\
&= \lambda^T Dg(x, c) Dx(a, b, c) + Df(x, b) = Df(x, b),
\end{align*}
\]

where the last equation follows from equation 8.8.

**Step 3:** Differentiate both sides of the equation

\[ g(x(a, b, c), c) = a \]

with respect to \( c \) at \( c = \overline{c} \). Then

\[ 0 = Dg(x(a, b, c), c) \bigg|_{c=c} \]
\[ (8.9) \]
\[
= Dg(x, c)Dx(a, b, c) + Dg(x, c),
\]
so that
\[
Df(a, b, c) = Df(x(a, b, c), b) \bigg|_{b=c}
\]
\[
= Df(x, b)Dx(a, b, c)
\]
\[
= \lambda^TDg(x, c)Dx(a, b, c)
\]
\[
= -\lambda^TDg(x, c),
\]
where the last equation follows from equation 8.9.

These three steps together imply equation 8.6. The same equation applies if max is replaced by \( \min \) in problem 8.4 and if the second order conditions for a local minimum in theorem 7.7 apply.

**Example:** I apply the theory of constrained optimization to a simple example of consumer demand. Suppose a consumer buys \( N \) commodities and let \( x_n \) be the quantity purchased of commodity \( n \), for \( n = 1, \ldots, N \). The vector of quantities purchased, called a consumption bundle, is \( (x_1, \ldots, x_N) \). Let the consumer’s utility for such bundles be
\[
u(x_1, \ldots, x_N) = \alpha_1 \ln(x_1) + \ldots + \alpha_N \ln(x_N),
\]
where \( \alpha_n > 0 \), for all \( n \). Such a utility function is called *Cobb-Douglas*. Assume that the consumer spends a fixed amount of money on the \( N \) commodities so as to maximize her or his utility. If the consumer’s wealth is \( w \) and the price of the \( n \)th commodity is \( p_n \), for all \( n \), then the consumer solves the following constrained maximization problem.

\[
(8.10) \quad \max_{x \in \mathbb{R}^N_+} \left[ \sum_{n=1}^N \alpha_n \ln(x_n) \right]
\]
\[
\text{s.t. } p_1 x_1 + \ldots + p_N x_N \leq w,
\]
where \( \mathbb{R}^N_+ = \{(x_1, \ldots, x_N) \in \mathbb{R}^N | x_n \geq 0, \text{ for all } n\} \). It is natural to assume that \( w > 0 \) and \( p_n > 0 \), for all \( n \). The Lagrangian for this problem is
\[
\mathcal{L}(x_1, \ldots, x_N, \lambda, w, p_1, \ldots, p_N, \alpha_1, \ldots, \alpha_N)
\]
\[ = \alpha \ln(x)_{1} + \ldots + \alpha \ln(x)_{N} - \lambda p x_{1} - \ldots - \lambda p x_{N} + \lambda w. \]

If \( \lambda \leq 0 \), then for all \( n \), \( \frac{\partial L(x_{1}, \ldots, x_{N}, \lambda, p_{1}, \ldots, p_{N})}{\partial x_{n}} > 0 \), for all \( x_{1}, \ldots, x_{N} \), and so the first order conditions for optimality are nowhere satisfied. Hence \( \lambda \) must be positive.

Since \( \ln(0) = -\infty \) and \( w > 0 \), we know that \( x_{n} > 0 \), for \( n = 1, \ldots, N \), at an optimum. Since the utility function \( u(x_{1}, \ldots, x_{N}) = \alpha \ln(x)_{1} + \ldots + \alpha \ln(x)_{N} \) is increasing and the prices are all positive, we know that \( p x_{1} + \ldots + p x_{N} = w \) at an optimum. Therefore we may write the optimization problem as

\[
\max_{x_{1} > 0, \ldots, x_{N} > 0} \left[ \alpha \ln(x)_{1} + \ldots + \alpha \ln(x)_{N} \right] \\
\text{s.t. } p x_{1} + \ldots + p x_{N} = w.
\]

Theorem 7.6 applies to this form of the problem, so that we know that \( Du(x) = \lambda D(p x_{1} + \ldots + p x_{N}) \) at an optimum. That is,

\[ \frac{\partial u(x_{1}, \ldots, x_{N})}{\partial x_{n}} = \lambda p_{n}, \]

for all \( n \), so that

\[ \frac{\alpha_{n}}{x_{n}} = \lambda p_{n}, \]

for all \( n \). Hence

\[ p_{n} x_{n} = \alpha_{n} \lambda^{-1}, \quad (8.11) \]

for all \( n \). Since the constraint is

\[ \sum_{n=1}^{N} p_{n} x_{n} = w, \]

if we add the \( N \) equations 8.11, we obtain

\[ \lambda^{-1} \sum_{n=1}^{N} \alpha_{n} = \sum_{n=1}^{N} p_{n} x_{n} = w. \]
That is,

\[ \lambda^{-1} = \frac{W}{\sum_{k=1}^{N} \frac{\alpha_n}{\alpha_k}}. \]

Substituting this equation into equation 8.11, we obtain

\[ p_n x_n = \frac{\alpha_n}{\sum_{k=1}^{N} \frac{\alpha_k}{\alpha_n}} w, \]

for all \( n \). That is, the fraction of wealth the consumer spends on commodity \( n \) is constant and is proportional to \( \frac{\alpha_n}{\alpha_k} \), and these fractions sum to 1. We see that

\[ x_n = \frac{\alpha_n}{\sum_{k=1}^{N} \frac{\alpha_k}{\alpha_n}} \frac{w}{p_n}, \]

for all \( n \), and

\[ \lambda = \frac{\sum_{k=1}^{N} \frac{\alpha_k}{\alpha_n}}{w}. \]

Let the maximized value of the utility function be

\[ F(p_1, \ldots, p_N, w, \alpha_1, \ldots, \alpha_N) \]

\[ = \max \{ \alpha_1 \ln(x_1) + \ldots + \alpha_N \ln(x_N) \mid x \geq 0, \text{ for all } n, \text{ and } \sum_{n=1}^{N} p_n x_n = w \}. \]

Since

\[ \lambda = \frac{\partial}{\partial w} L(x_1, \ldots, x_N, \lambda, w, p_1, \ldots, p_N), \]

the envelope theorem for constrained maximization implies that

\[ \lambda = \frac{\partial}{\partial w} F(p_1, \ldots, p_N, w, \alpha_1, \ldots, \alpha_N). \]

For this reason, \( \lambda \) is called the marginal utility of wealth or income. Similarly
\[ \frac{\partial F}{\partial p_n} = \frac{\partial L}{\partial p_n} = -\lambda x_n = -\frac{\alpha}{p_n} \quad \text{and} \]
\[ \frac{\partial F}{\partial \alpha_n} = \frac{\partial L}{\partial \alpha_n} = \ln(x_n), \]

for all \( n \).

**Convex Analysis**

We now turn from necessary and sufficient conditions for local maxima and minima to conditions for global optima. These conditions have to do with the shapes of the domains and graphs of the objective and constraint functions.

**Definition:** If \( f: A \to R \), a **global maximum** of \( f \) is achieved at a point \( a \in A \) if \( f(a) \geq f(b) \), for all \( b \in A \). A **global minimum** of \( f \) is achieved at a point \( a \) in \( A \) if \( f(a) \leq f(b) \), for all \( b \in A \).

**Definition:** A subset \( A \) of \( R^n \) is **convex** if whenever \( a \) and \( b \) belong to \( A \) and \( \alpha \) is a number such that \( 0 \leq \alpha \leq 1 \), then \( \alpha a + (1 - \alpha)b \) belongs to \( A \).

Since \( \alpha a + (1 - \alpha)b = b + \alpha(a - b) \), we see that \( \alpha a + (1 - \alpha)b \) lies on the straight line segment going from \( b \) to \( a \). The set shown on the left below is convex and that on the right is not.

**Definition:** If the subset \( A \) of \( R^n \) is convex and \( f: A \to R \), then \( f \) is said to be **concave** if for every pair of points \( a \) and \( b \) in \( A \) and every number \( \alpha \) such that \( 0 \leq \alpha \leq 1 \),

\[ f(\alpha a + (1 - \alpha)b) \geq \alpha f(a) + (1 - \alpha)f(b). \]

The function \( f \) is **strictly concave**, if for every pair \( a \) and \( b \) in \( A \) such that \( a \neq b \) and for every number \( \alpha \) such that \( 0 < \alpha < 1 \),
\[ f(\alpha a + (1 - \alpha)b) > \alpha f(a) + (1 - \alpha)f(b). \]

For our purposes, the most important thing about concavity is that it allows us to dispense with the distinction between local and global maximality.

**Theorem 8.1:** If \( A \) is a convex subset of \( \mathbb{R}^n \) and \( f: A \to \mathbb{R} \) is concave, then a local maximum of \( f \) is a global maximum.

**Proof:** Suppose that \( f \) achieves a local maximum at \( a \in A \). If \( f \) does not achieve a global maximum there, then there is a point \( b \in A \) such that \( f(b) > f(a) \). Because \( f \) achieves a local maximum at \( a \), there is a positive number \( \varepsilon \) such that \( f(x) \leq f(a) \), if \( ||x - a|| < \varepsilon \). Because \( A \) is convex, if \( 0 < \alpha < 1 \), then \( \alpha a + (1 - \alpha)b \in A \). If in addition, \( \alpha \) is sufficiently close to 1, then \( ||(\alpha a + (1 - \alpha)b) - a|| = (1 - \alpha)||b - a|| < \varepsilon \), so that \( f(\alpha a + (1 - \alpha)b) \leq f(a) \). Because \( f \) is concave, \( f(\alpha a + (1 - \alpha)b) \geq \alpha f(a) + (1 - \alpha)f(b) > f(a) \), if \( \alpha < 1 \). This contradiction proves that \( f \) achieves a global maximum at \( a \).

The importance to us of strict concavity is that it implies that maximizers are unique.

**Theorem 8.2:** If \( A \) is a convex subset of \( \mathbb{R}^n \) and \( f: A \to \mathbb{R} \) is strictly concave, then \( f \) achieves a global maximum at at most one point.

**Proof:** Suppose that \( f \) achieves a maximum at points \( a \) and \( b \) in \( A \), where \( a \neq b \). Then \( f(a) = f(b) \). Since \( A \) is convex, \( a/2 + b/2 \in A \). Because \( f \) is strictly concave,

\[ f(a/2 + b/2) > f(a)/2 + f(b)/2 = f(a) \]

This contradicts the hypothesis that \( f \) achieves a global maximum at \( a \).

For functions that are concave as well as differentiable, the first order condition for a local maximum is a necessary and sufficient condition for a global maximum.

**Theorem 8.3:** If \( A \) is a convex and open subset of \( \mathbb{R}^n \) and \( f: A \to \mathbb{R} \) is concave and differentiable, then \( f \) achieves a global maximum at \( a \in A \) if and only if \( Df(a) = 0 \).
Proof: Since a global maximum is also a local maximum, we know from theorem 6.8 that \( Df(a) = 0 \), if \( f \) achieves a global maximum at \( a \).

I show that if \( Df(a) = 0 \), then \( f \) achieves a global maximum at \( a \). Suppose not. Then there is a \( b \in A \) such that \( f(b) > f(a) \). Because \( f \) is concave,

\[
f(a + \alpha(b - a)) = f((1 - \alpha)a + \alpha b) \\
\geq (1 - \alpha)f(a) + \alpha f(b) = f(a) + \alpha[f(b) - f(a)],
\]

for \( \alpha \) such that \( 0 < \alpha < 1 \). Hence

\[
\frac{f(a + \alpha(b - a)) - f(a)}{\alpha} \geq f(b) - f(a),
\]

and so

\[
\frac{d}{d\alpha} f(a + \alpha(b - a)) \bigg|_{\alpha = 0} = \lim_{\alpha \to 0, \alpha > 0} \frac{f(a + \alpha(b - a)) - f(a)}{\alpha} \geq f(b) - f(a) > 0.
\]

However

\[
\frac{d}{d\alpha} f(a + \alpha(b - a)) \bigg|_{\alpha = 0} = Df(a)(b - a) = 0.
\]

This contradiction proves that there is no \( b \in A \) such that \( f(b) > f(a) \) and hence that \( f \) achieves a global maximum at \( a \).

Corresponding concepts and assertions apply to minima.

Definition: If the subset \( A \) of \( \mathbb{R}^n \) is convex and \( f: A \to \mathbb{R} \), then \( f \) is said to be convex if for every pair of points \( a \) and \( b \) in \( A \) and every number \( \alpha \) such that \( 0 \leq \alpha \leq 1 \),

\[
f(\alpha a + (1 - \alpha)b) \leq \alpha f(a) + (1 - \alpha)f(b).
\]

The function \( f \) is strictly convex, if for every pair \( a \) and \( b \) in \( A \) such that \( a \neq b \) and for every number \( \alpha \) such that \( 0 < \alpha < 1 \),

\[
f(\alpha a + (1 - \alpha)b) < \alpha f(a) + (1 - \alpha)f(b).
\]

A function \( f: A \to \mathbb{R} \) is convex if and only if the function \(-f\) is concave and is strictly convex if and only if \(-f\) is strictly concave. Because of these assertions, the following theorems require no proof.

Theorem 8.4: If \( A \) is a convex subset of \( \mathbb{R}^n \) and \( f: A \to \mathbb{R} \) is convex, then a local minimum
of $f$ is a global minimum.

**Theorem 8.5:** If $A$ is a convex subset of $\mathbb{R}^n$ and $f: A \to \mathbb{R}$ is strictly convex, then $f$ achieves a global minimum at at most one point.

**Theorem 8.6:** If $A$ is a convex and open subset of $\mathbb{R}^n$ and $f: A \to \mathbb{R}$ is convex and differentiable, then $f$ achieves a global minimum at $a \in A$ if and only if $Df(a) = 0$.

If $f$ is twice differentiable, then the second derivative of $f$ may be used to determine whether it is concave or convex.

**Theorem 8.7:** Let $A$ be a convex and open subset of $\mathbb{R}^n$ and let $f: A \to \mathbb{R}$ be everywhere twice differentiable. Then

1) $f$ is concave if and only if $D^2f(x)$ is negative semi-definite, for all $x$,

2) $f$ is strictly concave if $D^2f(x)$ is negative definite, for all $x$,

3) $f$ is convex if and only if $D^2f(x)$ is positive semi-definite, for all $x$, and

4) $f$ is strictly convex if $D^2f(x)$ is positive definite, for all $x$.

**Proof:** I prove only the sufficiency part of (1). I first of all prove this statement for the case $N = 1$. Suppose that $D^2f(x)$ is negative semi-definite. If $f$ is not concave, then there exist numbers $a$ and $b$ in $A$, such that $a < b$, and there is a number $c = \alpha a + (1 - \alpha) b$, where $0 < \alpha < 1$ and $f(c) < \alpha f(a) + (1 - \alpha) f(b)$. By the mean value theorem 5.10, there is a number $c_i$ such that $a < c_i < c$ and

$$
\frac{df(c_i)}{dx} = \frac{f(c) - f(a)}{c - a} = \frac{f(c) - f(a)}{\alpha a + (1 - \alpha) b - a} < \frac{\alpha f(a) + (1 - \alpha) f(b) - f(a)}{(1 - \alpha) (b - a)}
$$

$$
= \frac{(1 - \alpha) [f(b) - f(a)]}{(1 - \alpha) (b - a)} = \frac{f(b) - f(a)}{b - a}.
$$

Similarly there is a number $c_2$ such that $c < c_2 < b$ and

$$
\frac{df(c_2)}{dx} = \frac{f(b) - f(c)}{b - c}
$$

$$
= \frac{f(b) - f(c)}{b - \alpha a - (1 - \alpha) b}.
$$
\[ \frac{f(b) - \alpha f(a) - (1 - \alpha) f(b)}{\alpha (b - a)} = \frac{\alpha (f(b) - f(a))}{b - a} = \frac{f(b) - f(a)}{b - a}. \]

Therefore

\[ \frac{df(c_2)}{dx} > \frac{f(b) - f(a)}{b - a} > \frac{df(c_1)}{dx}. \quad (8.12) \]

Since the second derivative of \( f \) is non-positive, the first derivative \( df(x)/dx \) is a non-increasing function of \( x \). (This last assertion is again a consequence of the mean value theorem.) Therefore, since \( c_2 > c_1 \),

\[ \frac{df(c_2)}{dx} \leq \frac{df(c_1)}{dx}. \quad (8.13) \]

The contradiction between inequalities 8.12 and 8.13 implies that \( f \) is concave.
The argument is illustrated in the diagram on the previous page.

I now take up the case $N > 1$. Let
\[ g(\alpha) = f(\alpha a + (1 - \alpha)b) = f(b + \alpha(a - b)), \]
where $a$ and $b$ belong to $A$ and $a \neq b$. Then
\[ \frac{dg(\alpha)}{d\alpha} = Df(\alpha a + (1 - \alpha)b)(a - b) \]
and
\[ \frac{d^2g(\alpha)}{d\alpha^2} = (a - b)^2 D^2f(\alpha a + (1 - \alpha)b)(a - b) \leq 0, \]
where the inequality applies because $D^2f(x)$ is negative semi-definite, for all $x$. Therefore by what has been proved for the case $N = 1$,
\[ f(\alpha a + (1 - \alpha)b) = g(\alpha) = g((1 - \alpha)0 + \alpha(1)) \geq (1 - \alpha)g(0) + \alpha g(1) \]
\[ = \alpha f(a) + (1 - \alpha)f(b), \]
and so $f$ is concave.

**Example:** This example shows that $D^2f(x)$ being everywhere negative definite is not a necessary condition for $f$ to be strictly concave. Let $f: R \to R$ be defined by $f(x) = -x^4$. The function $f$ is clearly strictly concave, though $\frac{d^2f(0)}{dx^2} = 0$. Similarly $g(x) = x^4$ is clearly strictly convex, though $\frac{d^2g(0)}{dx^2} = 0$, so that $D^2g(x)$ being everywhere positive definite is not a necessary condition for $g$ to be strictly convex.

In the discussion of the envelope theorem in lecture 7, it was assumed that $D^2f(x, b)$ was negative definite, for all $(x, b)$. This assumption implies that $f$ is a concave function of $x$ and hence that a local maximum with respect to $x$ is a global maximum, a fact mentioned parenthetically in lecture 7.

**Remark:** An affine function is both concave and convex.
Kuhn-Tucker Theory

Kuhn-Tucker theory extends some of the ideas of the theory of Lagrange multipliers to constrained optimization problems in which the functions involved are not necessarily differentiable, but are concave or convex. The theory applies to the problem

$$\begin{align*}
\max_{x \in C} & \quad f(x) \\
\text{s.t.} & \quad g_k(x) \leq \ \underline{a}_k, \text{ for } k = 1, \ldots, K,
\end{align*}$$

(8.14)

where $C$ is a non-empty convex subset of $\mathbb{R}^n$, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is concave, and, for $k = 1, \ldots, K$, $g_k: \mathbb{R}^n \rightarrow \mathbb{R}$ is convex and $\underline{a}_k$ is a number. The theory also applies to the problem

$$\begin{align*}
\min_{x \in C} & \quad f(x) \\
\text{s.t.} & \quad g_k(x) \geq \ \underline{a}_k, \text{ for } k = 1, \ldots, K,
\end{align*}$$

where $C$ is a non-empty convex subset of $\mathbb{R}^n$, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, and, for $k = 1, \ldots, K$, $g_k: \mathbb{R}^n \rightarrow \mathbb{R}$ is concave and $\underline{a}_k$ is a number. The second problem may be transformed into the first by replacing $f$ by $-f$, each of the $g_k$ by $-g_k$, and each of the $\underline{a}_k$ by $-\underline{a}_k$. I focus on the first problem.

An $N$-vector that satisfies all the constraints of problem 8.14 is said to be feasible.

Definition: The vector $x \in \mathbb{R}^n$ is feasible if $x \in C$ and $g_k(x) \leq \underline{a}_k$, for all $k$.

Kuhn-Tucker Theorem: 8.8: Suppose that the feasible vector $\underline{x}$ and that $\lambda \in \mathbb{R}^K_+$ are such that, for $k = 1, \ldots, K$, $\lambda_k = 0$ if $g_k(\underline{x}) < \underline{a}_k$, and $\underline{x}$ solves the problem

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

(8.15)

Then $\underline{x}$ solves problem 8.14.

Suppose that $\underline{x}$ solves problem 8.14 and that the following constraint qualification is satisfied. There exists a vector $\underline{x} \in C$ such that $g_k(\underline{x}) < \underline{a}_k$, for all $k$. Then there exists $\lambda \in \mathbb{R}^K_+$ such that, for all $k$, $\lambda_k = 0$ if $g_k(\underline{x}) < \underline{a}_k$ and $\underline{x}$ solves problem 8.15.
**Terminology:** The conditions that \( \lambda_k = 0 \) if \( g_k(x) < a_k \), for all \( k \), are called the **complementary slackness conditions**. The function

\[
\mathcal{L}(x, \lambda) = f(x) - \sum_{k=1}^{K} \lambda_k g_k(x)
\]
is called the **Lagrangian**. The numbers \( \lambda_1, \ldots, \lambda_K \) are called **Kuhn-Tucker coefficients**. The \( N \)-vector \( x \) and the \( K \)-vector \( \lambda \) are said to satisfy the **Kuhn-Tucker conditions** if \( x \) is feasible, \( \lambda \geq 0 \), for all \( k \), \( x \) and \( \lambda \) satisfy the complementary slackness conditions, and \( x \) maximizes the Lagrangian, that is, it solves problem 8.15.

**Remark:** Notice that problem 8.15 is constrained only to the extent that \( x \) must belong to \( C \). In this sense, the Kuhn-Tucker theorem converts a constrained maximization problem to an unconstrained one.

**Example:** I apply the Kuhn-Tucker theorem to a consumer's maximization problem. Let the consumer's utility function be \( u: \mathbb{R}^N \to \mathbb{R} \), the consumer's wealth be \( w \), and the price of the \( n \)-th commodity be \( p_n \). The consumer's problem is

\[
\begin{align*}
\max_{x \in \mathbb{R}^N} & \quad u(x_1, \ldots, x_N) \\
\text{s.t.} & \quad p_1 x_1 + \ldots + p_N x_N \leq w,
\end{align*}
\tag{8.16}
\]

which may be written as

\[
\begin{align*}
\max_{x \in \mathbb{R}^N} & \quad u(x) \\
\text{s.t.} & \quad p.x \leq w.
\end{align*}
\]

Assume that \( u \) is concave and increasing and that \( w \) and each of the \( p_n \) are positive. Because \( w > 0 \), the problem satisfies the constraint qualification. (Let \( \lambda = 0 \).) Hence the Kuhn-Tucker theorem implies that if \( x \) solves problem 8.16, then there exists a non-negative number \( \lambda \).
such that $\bar{x}$ solves the problem

$$\max_{x \in \mathbb{R}^n} [u(x) - \lambda p.x].$$

Because $u$ is increasing, problem 8.17 has no solution if $\lambda = 0$, so that $\lambda$ must be positive. The complementary slackness condition then implies that $p.x = w$. The units of $\lambda$ are utiles per dollar, where utiles are units of utility. The units of $u(x) - \lambda p.x$ must then be utiles, and this quantity is consumer's surplus measured in utiles. This finishes the example.

In general, if the maximization problem is

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

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

then the Lagrangian, $\mathcal{L}(x, \lambda) = f(x) - \sum_{k=1}^{K} \lambda_k g_k(x)$, is a surplus, and the quantity $\lambda g_k(x)$ measures the cost of the $k$th constraint.

**Example:** This example demonstrates the need for the constraint qualification. Let $N = 2 = K, C = \mathbb{R}^2, \bar{a} = (0, 0), f(x_1, x_2) = x_1, g_1(x_1, x_2) = -x_1 + x_2, g_2(x_1, x_2) = x_1 + x_2$. The maximization problem is

$$\max_{x \in \mathbb{R}^2} x_1$$

$$\text{s.t. } -x_1 + x_2^2 \leq 0$$

$$x_1 + x_2^2 \leq 0.$$ 

This problem does not satisfy the constraint qualification, because the only feasible vector, $x = 0$, does not satisfy the constraints with strict inequality. Because 0 is the only feasible point, it is also optimal. There are no non-negative numbers $\lambda_1$ and $\lambda_2$ such that $x = 0$ maximizes the Lagrangian at these values for the Kuhn-Tucker coefficients, for suppose there were. Then $x = 0$ would solve the problem

$$\max_{x \in \mathbb{R}^2} [x - \lambda_1 (-x_1 + x_2^2) - \lambda_2 (x_1 + x_2^2)].$$

19
and the derivative of the objective function of this problem would be zero at $x = 0$. Setting the partial derivative with respect to $x_2$ equal to zero at $x = 0$, we find that

$$1 - 2 (\lambda_1 + \lambda_2) x_2 = 0,$$

which is impossible.

The figure on the next page should make clear why the Kuhn-Tucker theorem does not apply to this example. The region where $-x_1 + x_2^2 \leq 0$ is labeled as $g_1(x) \leq 0$, and the region where $x_1 + x_2^2 \leq 0$ is labeled as $g_2(x) \leq 0$. The Lagrangian is

$$f(x) - \lambda_1 g_1(x) - \lambda_2 g_2(x),$$

and its derivative is zero at $x = 0$, if it is maximized there. That is,

$$Df(0) - \lambda_1 Dg_1(0) - \lambda_2 Dg_2(0) = 0.$$ 

The derivative $Df(0)$ points straight upward, and $Dg_1(0)$ and $Dg_2(0)$ are horizontal. Since a vertical vector cannot be a linear combination of horizontal vectors, the Lagrangian is not maximized at the optimal value of $x$.

It is enlightening to see why the Kuhn-Tucker theorem applies when the resource bounds, $a_1$ and $a_2$, are increased in the above example so that the constraint qualification applies. Consider the problem

20
\[
\begin{align*}
\text{max } x \\
\text{s.t. } & -x_1 + x_2^2 \leq 1 \\
\text{and } & x_1 + x_2^2 \leq 1.
\end{align*}
\]

The solution to this problem is \( x = (0, 1) \). If \( \lambda_1 = \lambda_2 = 1/4 \), then this solution maximizes the Lagrangian,

\[
x - \frac{1}{4} (-x + x^2) - \frac{1}{4} (x + x^2).
\]

The derivative of the Lagrangian is zero at \( x = (0, 1) \), because

\[
Df(1, 0) = (0, 1) = \frac{1}{4} (-1, 2) + \frac{1}{4} (1, 2) = \frac{1}{4} Dg_1(0, 1) + \frac{1}{4} Dg_2(0, 1).
\]

The constraints are pictured in the figure on the next page, which shows vectors pointing in the direction of the derivatives, \( Dg_1(0, 1) \) and \( Dg_2(0, 1) \), of the constraint functions at the optimum point \((0, 1)\) as well as a vertical vector corresponding to the derivative \( Df(0, 1) \) of the objective function. The derivatives \( Dg_1(0, 1) \) and \( Dg_2(0, 1) \) are orthogonal at \((0, 1)\) to the boundaries of the corresponding constraint sets, \( \{x \mid g_1(x) \leq 1\} \) and \( \{x \mid g_2(x) \leq 1\} \). It should be apparent that \( Df(0, 1) \) may be written as a linear combination of \( Dg_1(0, 1) \) and \( Dg_2(0, 1) \) with positive coefficients, as is required by maximization of the Lagrangian at \( x = (0, 1) \).
In theorem 7.6 on Lagrange multipliers, there was a constraint qualification that is closely related to the constraint qualification of the Kuhn-Tucker theorem. In order to see the connection, recall that the constraint qualification of theorem 7.6 is that \( \text{D}_g(x), \ldots, \text{D}_g(x) \) be independent. Suppose that this constraint applies. Recall that in theorem 7.6, the constraint functions are defined on an open set \( U \), so that \( x \) must belong to the interior of \( U \). I show the constraint qualification of the Kuhn-Tucker theorem must apply if \( C = U \). Let \( g(x) = (g_1(x), \ldots, g_k(x)) \). Since \( \text{D}_g(x) \) has rank \( k \), the linear transformation \( \text{D}_g(x) \) from \( \mathbb{R}^n \) to \( \mathbb{R}^k \) is onto, and there exists a vector \( v \) in \( \mathbb{R}^n \) such that \( \text{D}_g(x)(v) = (-1, \ldots, -1) \). If \( \varepsilon \) is a sufficiently small positive number, then \( x + \varepsilon v \in U \), and, for all \( k \),

\[
g_k(x + \varepsilon v) = g_k(x) + \varepsilon \text{D}_g_k(x)(v) + o(\varepsilon) = a_k - \varepsilon + o(\varepsilon) < a_k,
\]

where \( \lim_{\varepsilon \to 0} o(\varepsilon)/\varepsilon = 0 \), so that the inequality applies for \( \varepsilon \) sufficiently small. If we let \( x = x - \varepsilon v \), for sufficiently small \( \varepsilon \), then \( x \) satisfies the requirements of the constraint qualification of the Kuhn-Tucker theorem.