NOTE ON MAXIMIZATION

The one variable case

Characterizing the solution to a maximization problem is especially easy if all we have to do is examine the slope of the function. Intuitively, this must be the case as long as the graph of the function \( f(x) \) “bends forward,” that is, the slope, \( \frac{df}{dx}(x) \), is decreasing. For then, if the slope is zero at \( x^* \), it must be positive for smaller \( x \) and negative for larger \( x \). A function with this property is said to be concave.

**Remark:** When we say that the slope is decreasing we mean “not increasing.” Thus for \( x^2 > x^1 \), \( \frac{df}{dx}(x^1) \geq \frac{df}{dx}(x^2) \). If the inequality is always strict we will say that the slope is strictly decreasing. Similarly the statement “\( x \) is positive” means \( x \geq 0 \) and the statement “\( x \) is strictly positive” means \( x > 0 \).

A second way of expressing the idea that a function “bends forward” is that any tangent line (that is, a line that touches the curve at some point \( x^0 \)) must lie above the curve. This is depicted below. The function \( f \) has a slope \( \frac{df}{dx}(x^0) \) at \( x^0 \) therefore the tangent line is \( y = f(x^0) + \frac{df}{dx}(x^0)(x-x^0) \). This line lies above the curve if

\[
f(x) \leq f(x^0) + \frac{df}{dx}(x^0)(x-x^0)
\]

![Fig. 1.3-1: Concave Function](image)
A third way of expressing the idea that a function “bends forward” is that the chords joining any two points on the curve must lie below the curve. This is depicted below.

Consider any two points $x^0$ and $x^1$ on the real line. Any point $\hat{x}$ lying between them is a weighted average or “convex combination” of $x^0$ and $x^1$.

$$\hat{x} = (1-\lambda)x^0 + \lambda x^1, \ 0 < \lambda < 1.$$  

Another way of writing this is $\hat{x} - x^0 = \lambda(x^1 - x^0)$. Thus the distance between the point $\hat{x}$ and the point $x^0$ is a fraction $\lambda$ of the distance between $x^1$ and $x^0$.

The chord AEB lies below the curve AFB if FD exceeds ED. Note that BC = $f(x^1) - f(x^0)$. Then, since $AD = \lambda(x^1 - x^0)$, $ED = \lambda(f(x^1) - f(x^0))$. Also $FD = f(\hat{x}) - f(x^0)$ Therefore $f$ is concave if

$$f(\hat{x}) - f(x^0) \geq \lambda(f(x^1) - f(x^0)), \ 0 < \lambda < 1.$$  

(1.3-1)
Rearranging this expression, 
\[ f(\hat{x}) \geq (1 - \lambda) f(x^0) + \lambda f(x^1). \]
For any differentiable function these three definitions are in fact equivalent.

**Def\(^n\): Concave and Strictly Concave Function**

A differentiable function \( f \) is concave on the interval \([a, b]\) if

For any \( x^0 \) and \( x^1 \) in the interval \([a, b]\)

(i) if \( x^0 < x < x^1 \) then \( \frac{df}{dx}(x^0) \geq \frac{df}{dx}(x) \)

(ii) \( f(x^1) \leq f(x^0) + \frac{df}{dx}(x^0)(x^1 - x^0) \)

(iii) for any convex combination \( \hat{x} = (1 - \lambda)x^0 + \lambda x^1, \ 0 < \lambda < 1, \)

\[ f(\hat{x}) \geq \lambda f(x^1) + (1 - \lambda) f(x^0). \]

The function is strictly concave if the inequalities are always strict.

**Maximization with one variable**

Suppose that a function is concave and you have found an interior point satisfying

the “First Order Condition” \( \frac{df}{dx}(x^*) = 0. \) Since the slope is everywhere decreasing, it

must be non-negative for \( x < x^* \) and non-positive for \( x > x^* \), as depicted below.

---

Fig. 1.3-3: Single turning point
Thus for a concave function, if \( x^0 \) satisfies the First Order Condition (FOC) it must be a maximizing value of \( x \).

Another way of showing this is to appeal to Definition (ii). That is, for any \( x^* \) and \( x \),
\[
 f(x) \leq f(x^*) + \frac{df}{dx}(x^*)(x - x^*) .
\]
Thus if the slope is zero at \( x^* \), then, for all other \( x \),
\[
f(x) \leq f(x^*).\]

Next suppose that \( x \in \mathbb{R}^+ \) so that the problem to be maximized is
\[
\max_{x} \{ f(x) \mid x \geq 0 \} .
\]
If the slope is zero for \( x^* > 0 \) we can argue exactly as before. The other possibility is that the slope is negative at the boundary, that is \( \frac{df}{dx}(0) \leq 0 \). Appealing to (ii)
\[
f(x) \leq f(0) + \frac{df}{dx}(0)x \leq f(0)
\]
since \( x \geq 0 \). Then again the “First Order Condition” yields the maximum.

**Two Variables**

Rather than leap directly to the analysis of multi-variable optimization problems we begin with the two variable case. Let \( x = (x_1, x_2) \) be an ordered pair (or ordered couple) of real numbers. Then a function \( f \) maps ordered pairs onto the real line. For the one variable case the graph of the function is a curve in the 2-dimensional space of real numbers \( \mathbb{R}^2 \). In the two variable case, the graph of the function is a surface in 3-dimensional space \( \mathbb{R}^3 \).

As with one variable, linear and quadratic functions play an important role.

**Linear Function**

\[
f(x) = a_0 + a_1x_1 + a_2x_2
\]

The graph of the linear function \( y = a_0 + a_1x_1 + a_2x_2 \) is a plane in 3 dimensional space.
An example is depicted below.
**Quadratic Function**

\[ f(x) = a_0 + \sum_{i=1}^{2} a_i x_i + \sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} x_i x_j \]

Note that \( f \) is the sum of a linear function and the function

\[ q(x) = \sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} x_i x_j = a_{11} x_1^2 + (a_{12} + a_{21}) x_1 x_2 + a_{22} x_2^2. \]  
(1.4-1)

This function is called a quadratic form. Since the second term on the right hand side is the sum of two terms, there is no loss in generality in assuming that \( a_{12} = a_{21} \)

The figures below depict some of the possible graphs of a quadratic form.

Both the graphs in the top row depict functions that have a minimum at \( x = 0 \). On the bottom left we have a maximum and on the bottom right a saddle point.

**Example:** Quadratic form with a saddle point

Consider the quadratic form \( q(x) = -x_1^2 + 4x_1 x_2 - x_2^2 \). Then \( q(x_1, 0) = -x_1^2 \) and \( q(0, x_2) = -x_2^2 \). Thus, taking one variable at a time, it appears that the quadratic form has a maximum at \( x = 0 \). However, setting \( x_1 = x_2 = z \) we have \( q(z, z) = 2z^2 \). Therefore the quadratic form has the saddle shape depicted in the bottom right of Fig. 1.4-2.
As a preliminary step in obtaining the necessary conditions for a maximum, we provide the conditions that ensure that a quadratic function has a maximum. The second of these propositions is proved in the Appendix. The proof of the first is very similar.

**Propn : Sufficient Conditions for** $q(x)$ **to be strictly negative for all** $x \neq 0$

If (i) $a_{11} < 0$ and (ii) $a_{11}a_{22} - a_{12}a_{21} > 0$ then $q(x) < 0$ for all $x \neq 0$

**Propn : Necessary and Sufficient conditions for** $q(x)$ **to be non-positive.**

$q(x) \leq 0$ if and only if (i) $a_{11}, a_{22} \leq 0$ and (ii) $a_{11}a_{22} - a_{12}a_{21} \geq 0$. 
Maximization

Suppose that the function \( f \) takes on its maximum over \( \mathbb{R}^2 \) at \( x^0 = (x_1^0, x_2^0) \). Taking one variable at a time we know that the first partial derivatives must be zero and that the second partial derivatives with respect to each variable must be less than or equal to zero. To obtain a further necessary condition we consider the change in \( f \) as \( x \) changes from \( x^0 \) in the direction of some other vector \( x^1 \). We do this by considering weighted averages of \( x^0 \) and \( x^1 \).

Define \( z \) to be the difference vector \( x^1 - x^0 \) and the weighted average

\[
x^\mu = (1 - \mu)x^0 + \mu x^1 = x^0 + \mu z.
\]

Note that we do not restrict ourselves to convex combinations so \( \mu \) can be any real number.

Next define \( g(\mu) = f(x^\mu) = f(x^0 + \mu z) \). Since \( g(0) = f(x^0) \) and \( f \) takes on its maximum at \( x^0 \), it must be the case that for all \( \mu \), \( g(\mu) \leq g(0) \). Thus \( g \) takes on its maximum at \( \mu = 0 \). From the previous section, the following first and second order conditions must hold.
\[
\frac{dg}{d\mu}(0) = 0 \quad \text{and} \quad \frac{d^2g}{d\mu^2}(0) \leq 0.
\]

Differentiating \(g\),
\[
\frac{dg}{d\mu}(\mu) = z_1 \frac{\partial f}{\partial x_1}(x^0 + \mu z) + z_2 \frac{\partial f}{\partial x_2}(x^0 + \mu z)
\]
and
\[
\frac{d^2g}{d\mu^2}(\mu) = z_1^2 \frac{\partial^2 f}{\partial x_1^2}(x^\mu) + 2z_1z_2 \frac{\partial^2 f}{\partial x_1 \partial x_2}(x^\mu) + z_2^2 \frac{\partial^2 f}{\partial x_2^2}(x^\mu)
\]
Note that the right hand side of the last equation is a quadratic form. Appealing to the second Proposition we have the following theorem.

**Propn**: Second Order Conditions for a Maximum

If \(f\) takes on its maximum over \(\mathbb{R}^2\) at \(x^0\), then

\[
(i) \quad \frac{\partial^2 f}{\partial x_i \partial x_i}(x^0) \leq 0, \quad i = 1, 2 \quad \text{and} \quad (ii) \quad \frac{\partial^2 f}{\partial x_i \partial x_j}(x^0) \cdot \frac{\partial^2 f}{\partial x_k \partial x_l}(x^0) - \frac{\partial^2 f}{\partial x_i \partial x_j}(x^0) \cdot \frac{\partial^2 f}{\partial x_k \partial x_l}(x^0) \geq 0.
\]

Just as in the one variable case we now show that if \(f\) is concave, these necessary conditions are also sufficient.

Of the three definitions for the one variable case, it is only the third that carries over directly to the multi-variable case. Just as before, for any 2-dimensional vectors \(x^0\) and \(x^1\) we define the convex combination
\[
x^\lambda = (1-\lambda)x^0 + \lambda x^1 = x^0 + \lambda(x^1 - x^0), \quad 0 < \lambda < 1.
\]

**Defn**: Concave Function

The function \(f\) is concave on \(X \subset \mathbb{R}^2\) if for any \(x^0, x^1 \in X\) and any \(\lambda \in (0,1)\)
\[
f(x^\lambda) \geq (1-\lambda)f(x^0) + \lambda f(x^1)
\]

**Propn**: Sufficient Conditions for a maximum

If \(f\) is concave, the necessary conditions are also sufficient.
This is an important result so it is well worth seeing if you can follow each step of the proof.

**Proof:** Consider any vectors $x^0$ and $x^1$. Note that $f(x^1) = f(x^0 + \lambda(x^1 - x^0))$. We can therefore think of this as being a function of the parameter $\lambda$, that is, we define
\[
g(\lambda) = f(x^1) = f(x^0 + \lambda(x^1 - x^0))\.
\]

From the definition of a concave function,
\[
g(\lambda) \geq (1 - \lambda)g(0) + \lambda g(1) \quad \lambda \in (0,1)\,.
\]

Rearranging this expression,
\[
\frac{g(\lambda) - g(0)}{\lambda} \geq g(1) - g(0), \quad \lambda \in (0,1)\,.
\]

Taking the limit as $\lambda \downarrow 0$, it follows that
\[
\frac{dg}{d\lambda}(0) \geq g(1) - g(0) \quad (1.4-4)
\]

From (1.4-2), the slope is
\[
\frac{dg}{d\lambda} = \frac{\partial f}{\partial x_1}(x^1_1 - x^0_1) + \frac{\partial f}{\partial x_1}(x^1_2 - x^0_2) \,.
\]  
\[
(1.4-5)
\]

Suppose that the Necessary Conditions for a maximum holds at $x^0$. That is, the partial derivatives of $f$ are both zero at $x^0$. Then from (1.4-5), $\frac{dg}{d\lambda}(0) = 0$.

Hence, from (1.4-4), $g(1) \leq g(0)$, that is $f(x^1) \leq f(x^0)$

\[
\text{QED}
\]

Finally, we note that using arguments similar to those made in the proof of the above Proposition, it can be shown that the following. The first step is to note that $f$ is concave if and only if $g(\mu) = f(x^\mu)$ is a concave function of $\mu$ for all $x^0$ and $x^1$. Then we have the three equivalent definitions of a concave function of one variable. Thus the slope $\frac{dg}{d\lambda}(\lambda)$ must be decreasing. From (1.4-3), the slope is decreasing $\frac{d^2 g}{d\lambda^2}(\lambda) \leq 0$ if and only if the quadratic form of the second derivatives of $f$ is be everywhere negative semi-definite. This yields the following result.
Alternative definition of a concave function

A twice continuously differentiable function $f$ is concave if and only if

\[
(i) \quad \frac{\partial^2 f}{\partial x_i \partial x_i} (x) \leq 0, \quad i = 1, 2 \quad \text{and} \quad (ii) \quad \frac{\partial^2 f}{\partial x_i \partial x_1} (x) \frac{\partial^2 f}{\partial x_2 \partial x_2} (x) - \frac{\partial^2 f}{\partial x_i \partial x_2} (x) \frac{\partial^2 f}{\partial x_2 \partial x_i} (x) \geq 0
\]

Appendix

**Prop** : Necessary and Sufficient conditions for $q(x)$ to be non-positive.

$q(x) \leq 0$ if and only if (i) $a_{11}, a_{22} \leq 0$ and (ii) $a_{11}a_{22} - a_{12}a_{21} \geq 0$.

**Proof**: Setting $x_2 = 0$ in (1.4-1) we have,

\[
q(x) = a_{11}x_1^2
\]

(1.4-6)

Thus for $q(x)$ to be non-positive a necessary condition is that $a_{11} \leq 0$. A symmetric argument establishes that it is also necessary that $a_{22} \leq 0$.

Suppose first that $a_{11} = a_{22} = 0$. Then $q(x) = 2a_{12}x_1x_2$. Since can choose $x_1$ and $x_2$ so that their product is either positive or negative, a necessary condition is $a_{12} = 0$. (Note that with $a_{11} = 0$ the condition $a_{11}a_{22} - a_{12}a_{21} \geq 0$ is equivalent to the condition $a_{12} = 0$.)

Next suppose that $a_{11} < 0$. Completing the square, (1.4-1) can be rewritten as

\[
q(x) = a_{11} (x_1 + \frac{a_{12}}{a_{11}} x_2)^2 + \frac{1}{a_{11}} (a_{11}a_{22} - a_{12}a_{21})x_2^2
\]

Actually, since $a_{12} = a_{21}$ we can write this more conveniently as

\[
q(x) = a_{11} (x_1 + \frac{a_{12}}{a_{11}})^2 + \frac{1}{a_{11}} (a_{11}a_{22} - a_{12}a_{21})x_2^2
\]

(1.4-7)

Suppose we choose $x_i$ so that the first term on the right hand side of (1.4-7) is zero. Since $a_{11} < 0$, a necessary condition for the right hand side to be non-positive is

\[
a_{11}a_{22} - a_{12}a_{21} \geq 0
\]

(1.4-8)
A symmetrical argument holds for the case in which $a_{22} < 0$.

We have therefore established the necessity of conditions (i) and (ii). Sufficiency follows almost immediately. In the first case each of the $a_{ij}$ terms is zero so $q(x) = 0$. In the second case it follows immediately from (1.4-7) that $q(x) \leq 0$. 