Optimization

1 Some Concepts and Terms

The general mathematical problem studied here is how to choose some variables, collected into a vector $\mathbf{x} = (x_1, x_2, \ldots x_n)$, to maximize, or in some situations minimize, an *objective function* $f(\mathbf{x})$, often subject to some *equation constraints* for the type $g(\mathbf{x}) = c$ and/or some *inequality constraints* of the type $g(\mathbf{x}) \leq c$. (In this section I focus on maximization with $\leq$ constraints. You can, and should as a good exercise to improve your understanding and facility with the methods, obtain similar conditions for problems of minimization, or ones with inequality constraints of the form $g(\mathbf{x}) \geq c$, merely by changing signs.)

I will begin with the simplest cases and proceed to more general and complex ones. Many ideas are adequately explained using just two variables. Then instead of a vector $\mathbf{x} = (x_1, x_2)$ I will use the simpler notation (x, y).

A warning: The proofs given below are loose and heuristic; a pure mathematician would disdain to call them proofs. But they should suffice for our applications-oriented purpose.

First some basic ideas and terminology.

An $\mathbf{x}$ satisfying all the constraints is called *feasible*. A particular feasible choice $\mathbf{x}$, say $\mathbf{x}^* = (x_1^*, x_2^*, \ldots x_n^*)$, is called *optimum* if no other feasible choice gives a higher value of $f(\mathbf{x})$ (but other feasible choices may tie this value), and a *strict optimum* if all other feasible choices give a lower value of $f(\mathbf{x})$ (no ties allowed). An optimum $\mathbf{x}^*$ is called *local* if the comparison is restricted to other feasible choices within a sufficiently small neighborhood of $\mathbf{x}^*$ (using ordinary Euclidean distance). If the comparison holds against all other feasible points, no matter how far distant, the optimum is called *global*.

Every global optimum is a local optimum but not vice versa. There may be two or more global maximizers $\mathbf{x}^{*a}$, $\mathbf{x}^{*b}$ etc. but they must all yield equal values $f(\mathbf{x}^{*a}) = f(\mathbf{x}^{*b}) = \ldots$ of the objective function. There can be multiple local maximizers with different values of the function. A function can have at most one strict global maximizer; it may not have any (the optimum may fail to exist) if the function has discontinuities, or if it is defined only over an open interval or an infinite interval and keeps increasing without reaching a maximum.

We will look for conditions to locate optima. These conditions take the form of mathematical statements about the functions, or their derivatives. Consider any such statement $\mathcal{S}$. We say that $\mathcal{S}$ is a *necessary condition* for $\mathbf{x}^*$ to be an optimum if, starting with the premise that $\mathbf{x}^*$ is an optimum, the truth of $\mathcal{S}$ follows by logical deduction. We say that $\mathcal{S}$ is a *sufficient condition* for $\mathbf{x}^*$ to be an optimum if the optimality of $\mathbf{x}^*$ follows as a logical deduction from the premise that $\mathcal{S}$ is true.

If a function is sufficiently differentiable, its *regular maxima* are characterized by conditions on derivatives. Other types of maxima are called *irregular* by contrast. We also classify the conditions according to the order of the derivatives of the functions: first-order, second-order etc. Then we abbreviate the label of a condition by its order and type; for example, FONC stands for first-order necessary condition.

Finally, maxima may occur at an *interior point* in the domain of definition of the functions or at a *boundary point.* A point $\mathbf{x}^*$ is called an interior point of a set D if, for some positive real number δ, all points x within (Euclidean) distance δ of $\mathbf{x}^*$ are also in D.

2 One Variable, No Constraints

Notation: R will denote the real line, $[a, b]$ a closed interval (includes end-points) of R, and $]a, b[$ an open interval (excludes end-points). We consider a function, $f : [a, b] \mapsto R$.

2.1 Interior, Regular, Local Maxima

In each of the arguments in this section, we suppose that $f(x)$ is sufficiently differentiable.

To test whether a particular x^* in the interior of $[a, b]$, that is, in the open interval $]a, b[$, gives a local maximum of $f(x)$, we consider the effect on $f(x)$ of moving x "slightly" away from x^*, to $x = x^* + \Delta x$. We use the Taylor expansion:

$$f(x) = f(x^*) + f'(x^*)\,\Delta x + \tfrac{1}{2}\,f''(x^*)\,(\Delta x)^2 + \ldots . \tag{1}$$

Then we have

FONC: If an interior point x^* of the domain of a differentiable function f is a (local or global) maximizer of $f(x)$, then $f'(x^*) = 0$.

Intuitive statement or sketch of proof: For Δx sufficiently small, the leading term in (1) is the first order, that is,

$$f(x^* + \Delta x) \approx f(x^*) + f'(x^*)\,\Delta x, \qquad \text{or} \qquad f(x^* + \Delta x) - f(x) \approx f'(x^*)\,\Delta x\,.$$

Since x^* is interior $(a < x^* < b)$, we can choose the deviation Δx positive or negative. If $f'(x^*)$ were non-zero, then we could make $f(x^* + \Delta x) - f(x^*)$ positive by choosing Δx to have the same sign as $f'(x^*)$. Then x^* would not be a local maximizer. We have shown that $f'(x^*) \neq 0$ implies that x^* cannot be a maximizer (not even local, and therefore certainly not global). Therefore if x^* is a maximizer (local or global), we must have $f'(x^*) = 0$. Note how, instead of the direct route of proving that "optimum implies condition-true", we took the indirect route "condition-false implies no-optimum." (The two implications are logically equivalent, or mutually "contrapositive" in the jargon of formal logic.) Such "proofs by contradiction" are often useful.

Exercise: Go through a similar argument and show that the FONC for $x^* \in]a, b[$ to be a local minimizer is also $f'(x^*) = 0$.

Now, taking FONC as satisfied, turn to second order conditions. First we have:

SONC: If an interior point x^* of the domain of a twice-differentiable function f is a (local or global) maximizer of $f(x)$, then $f''(x^*) \leq 0$.

Sketch of proof: The FONC above tells us $f'(x^*) = 0$. Then for Δx sufficiently small, the Taylor expansion in (1) yields

$$f(x^* + \Delta x) \approx f(x^*) + \tfrac{1}{2}\,f''(x^*)\,\Delta x^2, \qquad \text{or} \qquad f(x^* + \Delta x) - f(x^*) \approx \tfrac{1}{2}\,f''(x^*)\,\Delta x^2$$

2

If $f''(x^*) > 0$, then for Δx sufficiently small, $f(x^* + \Delta x) > f(x^*)$, so x^* cannot be a maximizer (not even local and certainly not global). Therefore, if x^* is a maximizer, local or global, we must have $f''(x^*) \le 0$. (Again a proof by contradiction.)

SOSC: If x^* is an interior point of the domain of a twice differentiable function f, and $f'(x) = 0$ and $f''(x^*) < 0$, then x^* yields a strict local maximum of $f(x)$.

Sketch of proof: Using the same expression, for Δx sufficiently small, $f''(x^*) < 0$ implies $f(x^* + \Delta x) < f(x^*)$. (A direct proof.)

A twice-differentiable function f is said to be *(weakly) concave at x^** if $f''(x^*) \le 0$, and *strictly concave at x^** if $f''(x^*) < 0$. ("Weakly" is the default option, intended unless "strictly" is specified.) Thus (given $f'(x^*) = 0$) concavity at x^* is necessary for x^* to yield a local or global maximum, and strict concavity at x^* is sufficient for x^* to yield a strict local maximum. Soon I will define concavity in a more general and more useful way.

What if the FONC $f'(x^*) = 0$ holds, but $f''(x^*) = 0$ also? Thus the SONC is satisfied, but the SOSC is not (for either a maximum or a minimum). Any x^* with $f'(x^*) = 0$ is called a *stationary point* or *critical point*. Such a point may be a local *extreme point* (maximum or minimum), but need not be. It could be a point of inflexion, like 0 for $f(x) = x^3$. To test the matter further, we must carry out the Taylor expansion (1) to higher-order terms. The general rule is that the first non-zero term on the right hand side should be of an even power of Δx, say $2k$. With the first $(2k - 1)$ derivatives zero at x^*, the necessary condition for a local maximum is that the $2k$-th order derivative at x^*, written $f^{(2k)}(x^*)$, should be ≤ 0; the corresponding sufficient condition is that it be < 0. Even this may not work; the function f defined by $f(0) = 0$ and $f(x) = \exp(-1/x^2)$ for $x \ne 0$ is obviously globally minimized at 0, but *all* its derivatives $f^{(k)}(0)$ at 0 equal 0. Luckily such complications requiring checking for derivatives of third or higher order are very rare in economic applications.

2.2 Irregular Maxima

2.2.1 Non-differentiatbility (kinks and cusps)

Here suppose f is not differentiable at x^*, but has left and right handed derivatives, denoted by $f'(x^*_-)$ and $f'(x^*_+)$ respectively, that may not be equal to each other.

FONC: If an interior point x^* in the domain of a function $f(x)$ is a local or global maximizer, and f has left and right first-order derivatives at x^*, then $f'(x^*_-) \ge 0 \ge f'(x^*_+)$

In the special case where f is differentiable at x^*, the two weak inequalities collapse to the usual $f'(x^*) = 0$.

The proof works by looking at one-sided Taylor expansions:

$$f(x^* + \Delta x) - f(x^*) \approx \begin{cases} f'(x^*_-)\, \Delta x & \text{if } \Delta x < 0 \\ f'(x^*_+)\, \Delta x & \text{if } \Delta x > 0 \end{cases}$$

and using the same kinds of arguments as in the proof for the regular FONC separately for positive and negative deviations Δx.

The intuition is that the function should be increasing from the left up to x^*, and then decreasing to the right of x^*. If the derivatives are finite we have a *kink*; if infinite, a *cusp*.

3

Now we also have, somewhat unusually, a first-order sufficient condition, making it unnecessary to go to any (left and right handed) second-order derivatives:

FOSC: If f has left and right first-order derivatives at an interior point x^* in its domain, and $f'(x^*_-) > 0 > f'(x^*_+)$, then x^* is a local maximizer of $f(x)$.

2.2.2   End-Point Maxima

For a function defined over the interval $[a, b]$, the boundaries of its domain are just the end-points a and b. A maximum, local or global, can occur at either. We test this in the same way as before, but only one-sided deviations are feasible. Therefore the first-order conditions become one-sided inequalities:

FONC: If a is a local or global maximizer of a function defined over $[a, b]$ and differentiable at a, then, $f'(a) \leq 0$. If b is a local or global maximizer of such a function, then $f'(b) \geq 0$.

Sketch of proof: From the Taylor expansion, for Δx sufficiently small

$$f(a + \Delta x) - f(a) \approx f'(a) \; \Delta x$$

If $f'(a) > 0$, then for $\Delta x > 0$ we get $f(a + \Delta x) > f(a)$ and a cannot give a local maximum. So $f'(a) \leq 0$ is a necessary condition. Similar argument at b.

Intuitively, this just says that the function should not start increasing immediately to the right of a. If it actually started falling, we could be sure of a local maximum. Thus we have an unusual

FOSC: Suppose a function f defined over $[a, b]$ is differentiable at a, and $f'(a) < 0$. Then a yields a local maximum of $f(x)$. If the function is differentiable at b, and $f'(b) > 0$, then b yields a local maximum of $f(x)$.

Exercise: Prove this. Note that at an end-point, differentiability simply means the appropriate one-sided differentiability.

If $f'(a) = 0$ or $f'(b) = 0$, the FONC is satisfied at that end-point but the FOSC is not, and we must turn to second-order conditions. These are the same as for interior maxima.

In economic applications, many variables are intrinsically non-negative, and the optimum choice for some of them may be 0. Then the usual FONC $f'(0) = 0$ need not be true and the more general FONC $f'(0) \leq 0$ must be used.

Figure 1 sheds a slightly different and useful light on the idea of end-point maxima. Suppose the function could be extrapolated to the left beyond the lower end-point a of its domain, as shown by the thinner curve. It might go on increasing and attain a maximum at an $\tilde{x}$ as shown there; it might even go on increasing for ever ($\tilde{x} = -\infty$). At a the function is already on its downward-sloping portion to the right of such an $\tilde{x}$, so $f'(a) < 0$. A limiting case of this is where the maximum on the enlarged domain happens to coincide with a. Then we have $f'(a) = 0$, which satisfies the FONC but not the FOSC, so for checking sufficiency we have to go to second-order conditions as in the basic theory of regular maxima.
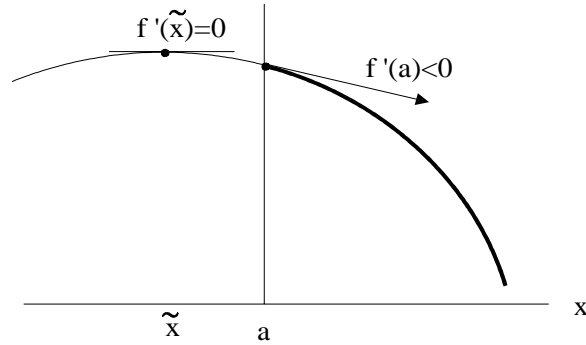
Figure 1: Interpreting end-point maxima

2.2.3 A General Example

Here is an example with many local, irregular and boundary maxima/minima, to collect a lot of possibilities for easy comparison and remembering. Consider the real-valued function f defined over the interval [-3,3] by

$$f(x) = \begin{cases} 3\,x + 8 & \text{for } -3 \le x \le -2 \\ (x^4 - 2\,x^2)/4 & \text{for } -2 \le x \le 2 \\ -x^2 + 6\,x - 6 & \text{for } 2 \le x \le 3 \end{cases}$$
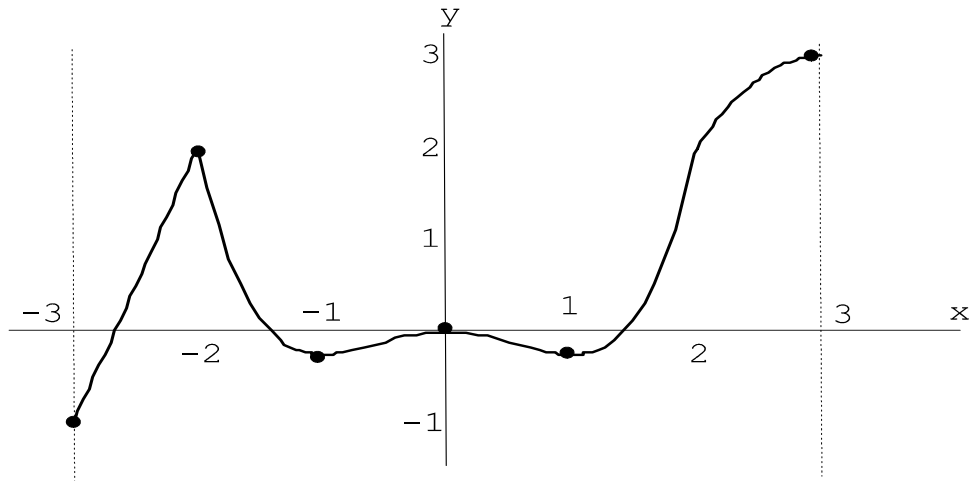
Figure 2 graphs this.

Figure 2: Various kinds of maxima and minima

This has three regular critical points, $x = -1$, 0, and 1. The first and the third are local minima, with tied values $f(\pm 1) = 0.25$. There is a local maximum at $x = 0$, with $f(0) = 0$. There is a local maximum with a kink at $x = -2$, with $f(-2) = 2$. There $f'(-2_-) = 3 > 0$ and $f'(-2_+) = -6 < 0$, so the left- and right- FONC and FOSC are satisfied. The left end-point is a local minimum, with $f(-3) = -1$ and $f'(-3) = 3 > 0$ satisfies the FONC

5

and FOSC. It is also the global minimum, by direct comparison with other local maxima. At the other end-point $x = 3$, we have a local maximum, with $f(3) = 3$. There $f'(3) = 0$, satisfying the end-point FONC but not the FOSC. It is also the global maximum, by direct comparison with other local maxima.

2.2.4 Global Maxima

Here I develop sufficient conditions for a critical point to yield a global maximum, generalizing the concept of concavity.

Definition: A twice-differentiable function f defined on the domain $[a, b]$ is *(globally, weakly) concave* if $f''(x) \leq 0$ for all $x \in [a, b]$. If $f''(x) < 0$ for all $x \in [a, b]$, we will call f *(globally) strictly concave*. ("Global" and "weak" are the "default options," understood unless a more restricted domain or strictness is specified. Thus a linear function is an extreme case of a (weakly) concave function.

SOSC: If f is concave over $[a, b]$, then any x^* satisfying an FONC is a global maximizer. If further, f is globally strictly concave, then such an x^* is the unique and strict global maximizer.

Sketch of proof: Taylor's theorem based on intermediate points says that for any x, we can find an $\tilde{x}$ between x^* and x such that

$$f(x) = f(x^*) + (x - x^*)f'(x^*) + \tfrac{1}{2}\,(x - x^*)^2\,f''(\tilde{x}).$$

By the FONC, $f'(x^*) = 0$. And by concavity, $f''(\tilde{x}) \leq 0$. Therefore $f(x) \leq f(x^*)$. If the function is strictly concave, $f''(\tilde{x}) < 0$ and then $f(x) < f(x^*)$.

Actually concavity on an interval can be defined much more simply without requiring second-order derivatives and in a sense that is valid for functions that are not twice (or even once) differentiable. There are two ways to do this:

Curve lies below any of its tangents: Call a once-differentiable function f concave on $[a, b]$ if, for any x^* and x^{**} in $[a, b]$,

$$f(x^{**}) \leq f(x^*) + (x^{**} - x^*)\,f'(x^*)$$

Strict concavity requires the inequality to be strict whenever $x^{**} \neq x^*$.

The left hand side of Figure 3 shows this; the inequality above appears there in the fact that the point C lies below the point B.

Any chord lies below the curve: For any x^*, x^{**} in $[a, b]$ and any $\theta \in [0, 1]$, writing $x^{***} = \theta\,x^* + (1 - \theta)\,x^{**}$, then

$$f(x^{***}) \geq \theta\,f(x^*) + (1 - \theta)\,f(x^{**})$$

Strict concavity requires the inequality to be strict whenever $x^{**} \neq x^*$ and $\theta \neq 0$ or 1.

The right hand side of Figure 3 shows this; the inequality above appears there in the fact that the point G lies below the point F. There x^{***} is shown two-thirds of the way from x^* to x^{**}. Mathematically, it is a weighted average of x^* and x^{**} with a weight of $\tfrac{1}{3}$ on x^* and $\tfrac{2}{3}$ on x^{**}, so $\theta = \tfrac{1}{3}$.

6

f(x)

B

C

A

x

x*   x**

f(x)

F
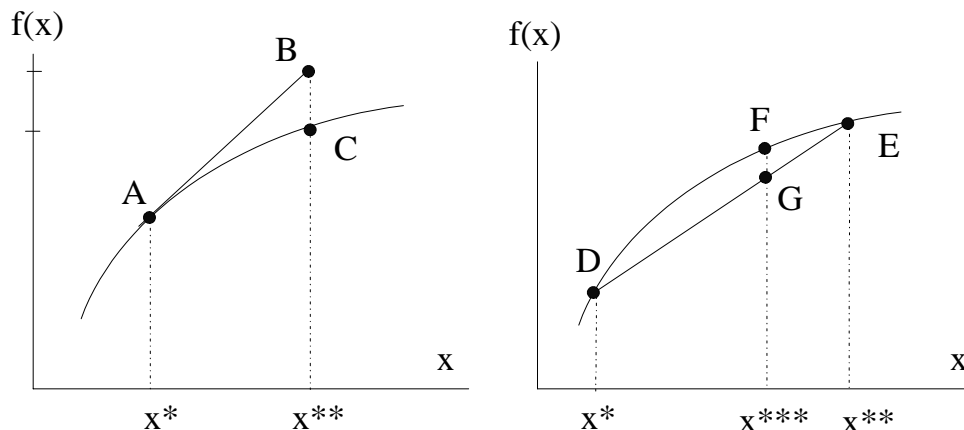
E

G

D

x

x*   x***   x**

Figure 3: Two alternative definitions of concavity

If f is concave, then any local maximum is also the global maximum. In fact any point satisfying one of the above FONCs – for a regular maximum, one at a point of non-differentiability, or at an end-point – is also the global maximum. For interior stationary points, the tangent is horizontal, so the proof is immediate using the "curve below tangent" definition. This also works for end-points; the tangent itself is falling as one moves from the end-point into the interior of the domain, and the graph of the function is below the tangent. For the non-differentiable case, the left and right tangents are both falling as one moves away from the point in question, and the graph lies below the tangent on each side.

2.2.5   Procedure for Finding Maxima

We can now summarize the above analysis into a set of rules. Begin by using the necessary conditions to narrow down your search. Find all regular stationary points, that is, solutions of the equation $f'(x) = 0$, in (a, b). Then see if f fails to be differentiable at any point of (a, b); if so check if the kink/cusp FONCs are satisfied at these points. Finally, check whether the FONCs at a and b are satisfied. (With experience, you will be able to omit some of these steps if you know they are not going to be relevant for the function at hand. For example, if the function is defined by a formula involving polynomial, exponential etc. functions, it is differentiable. If the function is defined by separate such formulas over separate intervals, then at the points where the formula changes, you should suspect non-differentiability. If the formula involves logarithms or fractions, problems can arise where the argument of the logarithm or the denominator in the fraction becomes zero. And so on.)

If the function is concave, the necessary conditions should have led you to one and only one point, which is the global maximum. Otherwise, for each of the points that satisfy the necessary conditions, check the sufficient conditions. These give you the local maxima.

If there is only one local maximum, it must be the global maximum. If there are several local maxima, the global maximum must be found by actually comparing the values of the function at all of them. (Again, experience will provide some short-cuts but it is not possible to lay down very general rules in advance.)

2.3 Many Variables, No Constraints

Now I consider $f : D \mapsto R$ where the domain of definition D is a set in R^n, the n-dimensional real space. A typical D in economic applications will consist of vectors whose components are all non-negative, but other possibilities also arise. Recall the notation for vectors introduced on pages 1 and 2. Remember also that when it suffices to develop the ideas using the special case of $n = 2$, I will do so, and then will denote the general point by (x, y), a particular point by (x^*, y^*) and so on. This notation is now supplemented with notation for partial derivatives etc.

If f is differentiable, its partial derivatives $\partial f / \partial x_i$ will be denoted by f_i. These are themselves functions, and $f_i(x)$ will mean $\partial f / \partial x_i$ evaluated at $\mathbf{x}$. (Note that in general f_i can be a function of the whole vector $\mathbf{x}$, not just the component x_i. That special case arises if, and only if, f is additively separable in its arguments.) The partial derivatives form a vector that is usually called the *gradient vector* of the function, and denoted by $\nabla f(\mathbf{x})$. If f is twice differentiable, the second-order partial derivatives $\partial^2 f / \partial x_i \partial x_j$ will be denoted by f_{ij}, and again $f_{ij}(x)$ will indicate evaluation at x. (Notes: [1] For a twice-differentiable function, by Young's Theorem we have $f_{ij}(x) = f_{ji}(x)$ for all x. [2] When $i = j$ the partials become the second-order own partial derivatives: $f_{ii}(x) \equiv \partial^2 f / \partial x_i^2$.) With two independent variables, the notation will be f_x, f_y, f_{xy} and so on.

The various steps of the analysis follow closely upon those in the one-variable case.

2.3.1 Interior, Regular, Local Maxima

Here we test an $\mathbf{x}^*$ for optimality by considering a vector deviation $\Delta \mathbf{x} = (\Delta x_1 . \Delta x_2 . \ldots \Delta x_n)$. We use the multi-variable Taylor expansion:

$$
\begin{aligned}
f(\mathbf{x}^* + \Delta \mathbf{x}) \;=\;& f(\mathbf{x}^*) + \sum_{i=1}^{n} f_i(\mathbf{x}^*)\,(\Delta x_i) \\
& + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(\mathbf{x}^*)\,\Delta x_i\,\Delta x_j + \ldots .
\end{aligned}
\tag{2}
$$

This leads to

FONC: If $\mathbf{x}^*$ is an interior point of D, f is differentiable at $\mathbf{x}^*$, and $\mathbf{x}^*$ yields a maximum (local or global) of $f(x)$, then $f_i(\mathbf{x}^*) = 0$ for all i.

Sketch of proof: Since $\mathbf{x}^*$ is interior, a deviation in any direction is feasible. For each i, consider a deviation $(\Delta \mathbf{x})_i$ defined so that its i-th component Δx_i non-zero and all other components zero. For Δx_i sufficiently small,

$$
f(\mathbf{x}^* + (\Delta \mathbf{x})_i) - f(\mathbf{x}^*) \approx f_i(\mathbf{x}^*)\,\Delta x_i .
$$

If $f_i(\mathbf{x}^*) \neq 0$, then we can make $f(\mathbf{x}^* + (\Delta \mathbf{x})_i) > f(\mathbf{x}^*)$ by taking Δx_i to be of the same sign as $f_i(\mathbf{x}^*)$. This contradiction establishes the necessary condition.

The intuition is simply that since the function is maximized as a whole at $\mathbf{x}^*$, it must be maximized with respect to each component of $\mathbf{x}$, and then the standard one-dimensional

argument applies for the necessary conditions. (The converse is not true; dimension-by-dimension maximization does not guarantee maximization as a whole. Therefore sufficent conditions are more than a simple compilation of those with respect to the variables x_1, x_2, ..., x_n taken one at a time.)
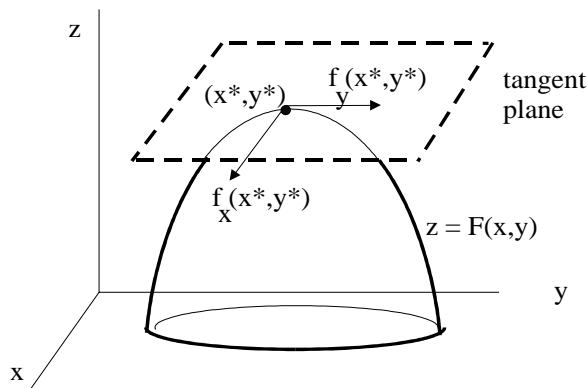

Figure 4: Function of two variables – 3D view

Figure 4 illustrates this for a function of two variables. The maximum occurs at (x^*, y^*). The graph of the function is like a hill with a peak at this point. At (x^*, y^*), the tangent plane to the hill is flat. The slope of the hill at this point is zero in every direction, and in particular it is zero as we move parallel to either of the two axes. Therefore the partial derivatives of f with respect to x and y are both equal to zero at (x^*, y^*).

The function falls as one moves farther away from this point in every direction. In other words, the graph is concave at this point. The algebraic conditions for this are complicated and I will mention them later, but we will rely on intuitive and geometric notions of concavity most of the time when we check second-order conditions.

A different geometric visualization proves very useful in economics. Figure 5 gives a "top-down" view of the hill. I have shown several countours or equal-height curves, each labeled by the common height of all of its points (x, y). This two-dimensional view can be related to the three-dimensional view of Figure 4: each 2-D contour is a horizontal slice of the hill in the 3-D figure, cut at the appropriate height. Contours at successively greater heights are successively smaller, eventually shrinking to the point (x^*, y^*); in the figure, $f(x^*, y^*) = 6$.

Hold y constant at y^* and increase x so the point (x, y^*) moves to the right, passing through (x^*, y^*). This takes us to successively higher contours of f until x reaches x^*, and then successively lower ones. At x^* the highest feasible contour is attained, and the rate of climb (the derivative) is zero. Similary with respect to y.

More generally, the contour map indicates how fast or slowly the function changes with respect to (x, y) at any point, that is, the steepness of the gradient of the hill in the 3-D picture can be visualized from the 2-D map. If the contours are packed close together, as happens in Figure 5 to the south-west side of (x^*, y^*), then each unit increase in the value of $f(x, y)$ comes about for a small change in (x, y). Therefore the partial derivatives of f are large, and the gradient is steep. To the north-east of (x^*, y^*), the contours are spaced farther apart, so each unit increase in the value of $f(x, y)$ requires a larger change in (x, y).
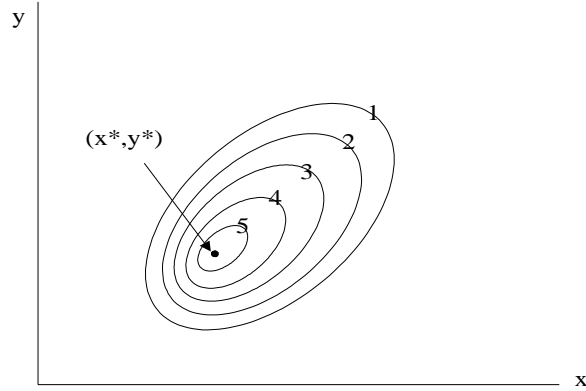
9

Figure 5: Function of two variables – contours (2D view)

Therefore the partial derivatives of f are smaller in numerical values, and the gradient is flatter.

2.3.2 Irregular and Boundary Maxima

In multiple dimensions, these can arise in a variety of ways. They are best handled in each specific context, and a general discussion is not very helpful.

There is one boundary case that deserves special attention. In economic applications, the choice variables are often required to be non-negative. An optimum on the boundary of the feasible set can then arise when some of the variables are zero and others positive. Then the objective function should not increase as a zero variable is made slightly positive. That is, the partial derivative of the objective function with respect to this variable should be negative, or at worst zero, at the optimum. If some other variables are positive at the optimum, the usual zero-partial-derivative conditions should continue to hold for them. A demonstration for a function of two variables makes the point:

FONC : If (x, y) are required to be non-negative, and $(0, y^*)$ (where $y^* > 0$) is a (local or global) maximizer of $f(x, y)$, then $f_x(0, y^*) \leq 0$, $f_y(0, y^*) = 0$.

Sketch of proof: We have the usual Tayor expansion for a small deviation $(\Delta x, \Delta y)$:

$$f(\Delta x, y^* \Delta y) - f(0, y^*) \approx f_x(0, y^*)\, \Delta x + f_y(0, y^*)\, \Delta y\,.$$

Consider a feasible deviation in the x-direction. Thus $\Delta y = 0$, and since x must be non-negative, we must have $\Delta x > 0$. Then the premise that $(0, y^*)$ gives a maximum rules out only $f_x(0, y^*) > 0$. Therefore we have the necessary condition $f_x(0, y^* \leq 0$. Next consider a feasible deviation in the y-direction. Here Δy can be of either sign. The same argument as for a function of one variable goes through, and rules out $f_y(0, y^*) \neq 0$. Therefore $f_y(0, y^*) = 0$ is a necessary condition.

Figure 6 illustrates this. The figure is similar to Figure 5, except that the position of the y-axis has been shifted, and the points to its left are not allowed. The contours of f in the feasible region (non-negative (x, y)) are shown thicker. If the function could be extrapolated to the region of negative x, its maximum would occur at $(\tilde{x}, \tilde{y})$. Given the non-negativity
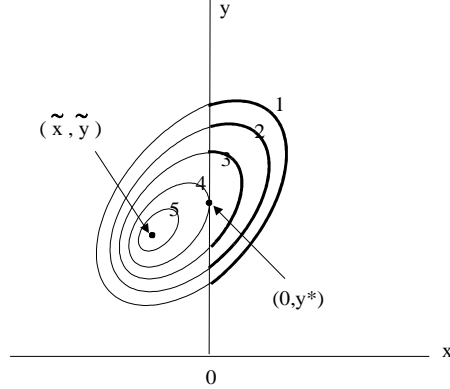
Figure 6: Boundary optimum with one variable zero

restriction, the highest attainable contour is at height 4, and the maximum occurs at $(0, y^*)$. Observe the interaction between the two variables: when the non-negativity condition for one of them matters, this can change the value that is best for the other.

Starting at $(0, y^*)$, if we move slightly the right (increase x from 0 to δx, say), we move to lower contours at a non-zero speed (negative gradient); thus $f_x(0, y^*) < 0$. But if we consider moves in the y direction, $(0, y^*)$ is a point where we stop going to successively higher contours and start going to successively lower ones, so at that point the gradient in the y direction just flattens out: $f_y(0, y^*) = 0$.

Unlike in the one-variable case, $f_x(0, y^*) < 0$ is not a sufficient condition (FOSC) even in conjunction with $f_y(0, y^*) = 0$, because the y-condition is perfectly compatible with a minimum or a point of inflection in the y direction. Second-order conditions are needed to establish sufficiency.

2.3.3   Second-Order Conditions

As with functions of one variable, there can be stationary points that are neither local maxima nor local minima. With several variables, a new situation can arise – a stationary point may yield a maximum of f with respect to one subset of variables and a minimum with respect to the rest. Then the graph of $f(\mathbf{x})$ looks like a saddle. Exercise: Draw a picture.

Concavity is relevant in the second-order conditions for unconstrained maximization of functions of several variables just as it was for functions of one variable. Once again we have three definitions, which are equivalent if the function is twice differentiable.

Definitions for twice-differentiable functions: f is *(weakly) concave* at $\mathbf{x}^*$ if $(f_{ij}(\mathbf{x}^*))$ is the matrix of a negative semi-definite quadratic form; that is,

$$\sum_i \sum_j f_{ij}(\mathbf{x}^*)\, z_i\, z_j \leq 0 \quad \text{for all vectors } \mathbf{z} = (z_1, z_2, \ldots z_n).$$

It is *strictly concave* at $\mathbf{x}^*$ if $(f_{ij}(\mathbf{x}^*))$ is the matrix of a negative definite quadratic form, that is,

$$\sum_i \sum_j f_{ij}(\mathbf{x}^*)\, z_i\, z_j < 0 \quad \text{for all vectors } \mathbf{z} \neq 0.$$

11

f is *(weakly) globally concave over D* if $(f_{ij}(\mathbf{x}))$ is the matrix of a negative semi-definite quadratic form for all $\mathbf{x} \in D$; *globally strictly concave over D* if that quadratic form is negative definite for all $\mathbf{x} \in D$.

Definition for once-differentiable functions: *Curve lies below any of its tangent hyper-planes:* f is (weakly) concave on D if, for any vectors $\mathbf{x}^*$ and $\mathbf{x}^{**}$ in D,

$$f(\mathbf{x}^{**}) \le f(\mathbf{x}^*) + \sum_{i=1}^{n} (x_i^{**} - x_i^*) \; f_i(\mathbf{x}^*) \equiv \nabla f(\mathbf{x}^*) \cdot (\mathbf{x}^{**} - \mathbf{x}^*)$$

The last term on the right hand side is the inner product of the two vectors $\nabla f(\mathbf{x}^*)$ (the gradient vector of f) and $(\mathbf{x}^{**} - \mathbf{x}^*)$.

Strict concavity requires the inequality to be strict whenever $\mathbf{x}^{**} \ne \mathbf{x}^*$.

Definition without differentiability requirement: *Any chord lies below the curve:* f is *concave on D* if, for any $\mathbf{x}^*$, $\mathbf{x}^{**}$ in D and any $\theta \in [0, 1]$, on writing $\mathbf{x}^{***} = \theta \, \mathbf{x}^* + (1 - \theta) \, \mathbf{x}^{**}$, we have:

$$f(\mathbf{x}^{***}) \ge \theta \; f(\mathbf{x}^*) + (1 - \theta) \; f(\mathbf{x}^{**})$$

Strict concavity requires the inequality to be strict whenever $\mathbf{x}^{**} \ne \mathbf{x}^*$ and $\theta \ne 0$ or 1. For this definition to be useful, the domain D should be a convex set, that is, for any $\mathbf{x}^*$, $\mathbf{x}^{**}$ in D and any $\theta \in [0, 1]$, the vector $\mathbf{x}^{***} = \theta \, \mathbf{x}^* + (1 - \theta) \, \mathbf{x}^{**}$ should also be in D.

The concavity conditions in terms of second derivatives are much more complex for fuc-ntions of several variables than those for functions of one variable. But the "curve below tangent" and "chord below curve" definitions for functions of several variables are straight-forward generalizations of those for functions of one variable, and the geometry is essentially the same. Therefore these definitions are much simpler to visualize, and to ensure that they hold in specific applications.

Using these definitions, we have the conditions

SONC: Weak concavity of f at $\mathbf{x}^*$ is a necessary condition for this point to be a local maximum of $f(\mathbf{x})$.

Local **SOSC:** Strict concavity of f at $\mathbf{x}^*$ is sufficient for that point to yield a strict local maximum of $f(\mathbf{x})$.

Global **SOSC:** If f is concave on D, then any point satisfying an FONC is a global maximizer. If f is strictly concave on D, then at most one point can satisfy its FONC, and it is the unique and strict global maximizer.

2.3.4 Procedure for Finding Maxima

The rules are similar to those with one variable: first narrow down the search using the FONCs. For interior regular maxima, the critical points are defined by $f_i(x) = 0$ for $i = 1$, 2, ... n. These are n simultaneous equations in the n unknowns $(x_1, x_2, \ldots x_n)$. The equations are non-linear and may have no or multiple solutions, but in almost all cases there will be a finite number of solutions. Thus the search is narrowed down.

Next, either use local SOSCs to identify local maxima and select the largest among them for the global maximum, or if possible use concavity directly to identify the global maximum.

But SOSCs in terms of second-order derivatives are often hard. To check a quadratic form for negative definiteness (semi or full) requires the computation of the determinants of all the leading principal minors of its matrix. In practice, a little geometric imagination saves a lot of work. By drawing (or visualizing) the graph of $f(\mathbf{x})$, one can often quickly identify which stationary points are maxima and which ones minima or saddle-points or the like. In this course, if there is a difficulty with second-order conditions in a many-variable optimization problem, you will be alerted explicitly.

Irregular and boundary maxima are often best identified by geometric visualization than by algebra. Of course doing this successfully needs practice. We will generally not need boundary solutions except for the case when variables are required to be non-negative and the optimum for one or more of them is at zero; I will alert you explicitly in such cases.

**** Sections on constrained optimization to come. ****

Further and Alternative Readings

Binger, Brian R. and Elizabeth Hoffman. 1998. *Microeconomics with Calculus.* Second edition. New York: Addison-Wesley. Chapter 2 gives some better 3-D pictures and economic applications.

Nicholson, Walter. 2002. *Microeconomic Theory: Basic Principles and Extensions.* Eighth edition. Mason, OH: South=Western. Chapter 2, pp. 22–34 contain similar material but do not consider end-point and boundary optima.