4 Calculus Review

4.1 The Utility Maximization Problem

As a motivating example, consider the problem facing a consumer that needs to allocate a given budget over two commodities sold at (linear) prices p_1 and p_2. Since the preferences can be represented by a utility function and the budget set is described by a set of equations we can describe this as a *constrained optimization problem.*

We seek to find a bundle (x_1^*, x_2^*), which:

1. is **feasible**, or, is in the budget set, meaning that

$$p_1 x_1^* + p_2 x_2^* \leq m \text{ and } x_1^* \geq 0, x_2^* \geq 0,$$

2. is **optimal**, meaning that $u(x_1^*, x_2^*) \geq u(x_1, x_2)$ for all (x_1, x_2) in the budget set.

It is convenient to introduce some notation for this type of problems. I will write this as

$$\begin{aligned} &\max_{x_1, x_2} u(x_1, x_2) \\ \text{subj. to} \quad & p_1 x_1 + p_2 x_2 \leq m \\ & x_1 \geq 0 \qquad\qquad , \\ & x_2 \geq 0 \end{aligned}$$

which is a rather convential way of stating a *problem of constrained optimization.*

A common tendency of students is to skip the step where the problem is written down. This is a bad idea. The reason is that we will often study variants of optimization problems that differ in what seems to be small "details". Indeed, often times the difficult step when thinking about a problem is to formulate the right optimization problem. For this reason I want you to:

1. Write out the "max" in front of the utility function (the *maximand*, or, *objective function*). This clarifies that the consumer is supposed to solve an optimization problem.

2. Below the max, it is a good idea to indicate what the *choices variables* are for the consumer (x_1 and x_2 in this example). This is to clarify the difference between the variables that are under control of the decision maker and variables that the decision maker has no control over, which are referred to as *parameters*. In the application above p_1, p_2 and m are parameters.

3. Finally, it is important that it is clear what the *constraints* to the problem are. I good habit is to write "subject to" or, more concisely, s.t. and then list whatever constraints there are, as in the problem above.

4.2 Utility functions and Indifference Curves

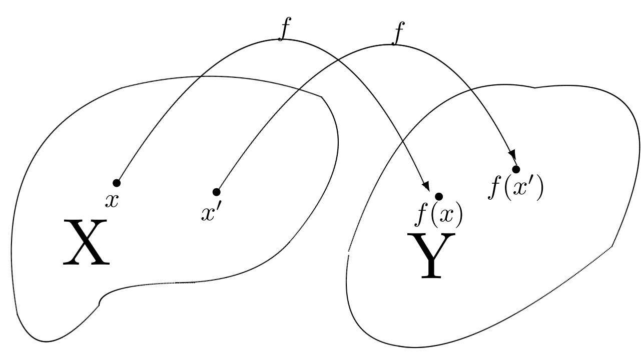

Figure 1: The Notion of a Function

In general a mathematical function assigns a point in one set (often called the "domain") to each point in another set . Think of it as a rule that "sends" each point in one set X to a point in another set Y as in Figure 1.

A utility function assigns a number (the "happiness index") to each consumption bundle (x_1, x_2). Now, the advantage with the restriction to two goods is that we can visualize such a function in terms of a three-dimensional picture as in Figure 2.

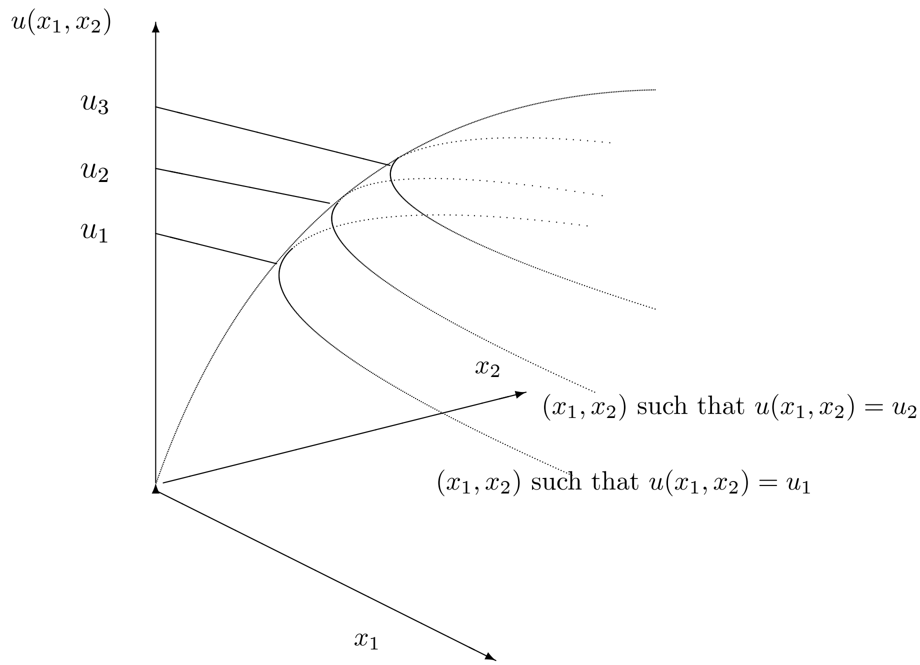

Figure 2: Utility Function with Two Arguments

The particular shape of the "mountain" depends on preferences and Figure 2 is drawn depicting the standard case (more is better and "convexity"). However, for now the important aspect is the construction of the pictures. The lines "on the mountain" is the picture can be thought of geometrically as the intersection between the "mountain" and a plane that is parallel to "the bottom" of the picture. That is, if we "cut a slice" through the "mountain" which is parallel to the bottom and at "height" u_1 from the bottom we find all combinations of x_1 and x_2 such that $u(x_1, x_2) = u_1$. We can obviously do the same for as many different utility levels as we like and the picture has three different levels of utility.

Now imagine yourself staring at the mountain straight from above so that what you see is the two-dimensional surface with x_1 and x_2 on the axes (you'll have to have good binoculars and be far away). Also imagine that you've "sliced the mountain" with a plane parallel to the $x_1 - x_2$ plane so that you only see what is above the plane. Then, what you'd see is something looking like Figure 4.

Note that everything above the plane corresponds to bundles where the utility index is higher than u_1, so bundles to the north-east in the picture are bundles which are better than

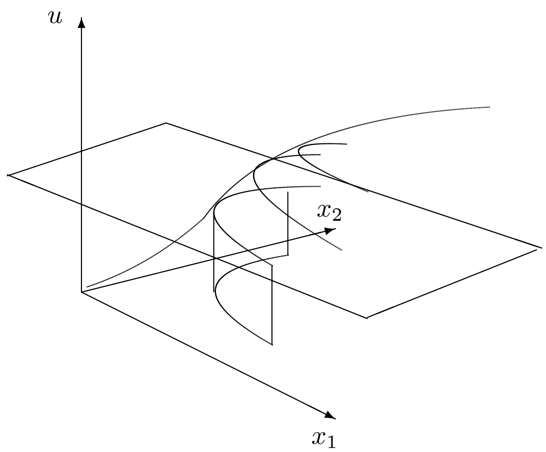

Figure 3: Projecting Utility Function to 2 Dimensions

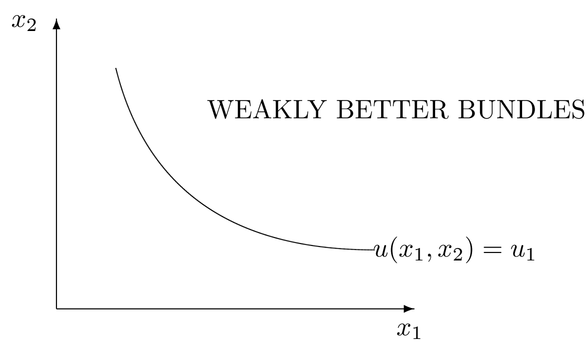

Figure 4: Indifference Curves are Projections of Level Curves

any bundle on the indifference curve. Obviously we can do the same thing for several choices of the utility index and then we generate the standard indifference curve diagram from your principles course.

4.3 Derivatives

Consider a linear function $y = h(x) = ax + b$ as the one depicted in Figure 5. Recall that the slope of a linear function is

$$\text{slope} = \frac{\text{change in } y}{\text{change in } x} = \frac{\overbrace{ax'' + b}^{f(x'')} - (\overbrace{ax' + b}^{f(x')})}{x'' - x'} = a.$$

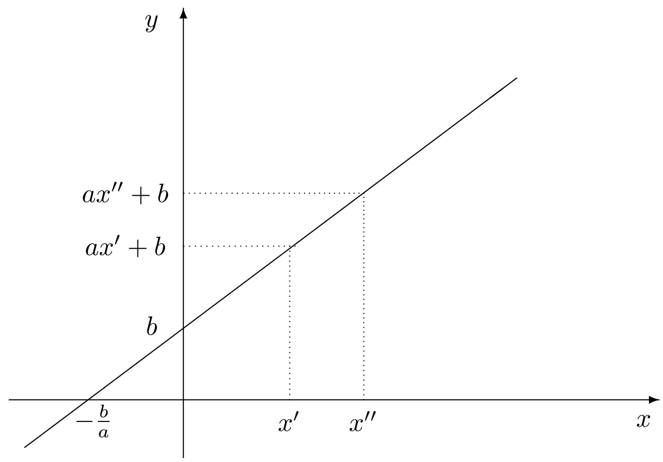

Figure 5: A Linear Function

Now, let's think about a *non-linear* function $y = f(x)$. If we just do the same thing as with the linear function and think of the "slope" as the ratio of the change in the value of the function to the change in the value of the argument.

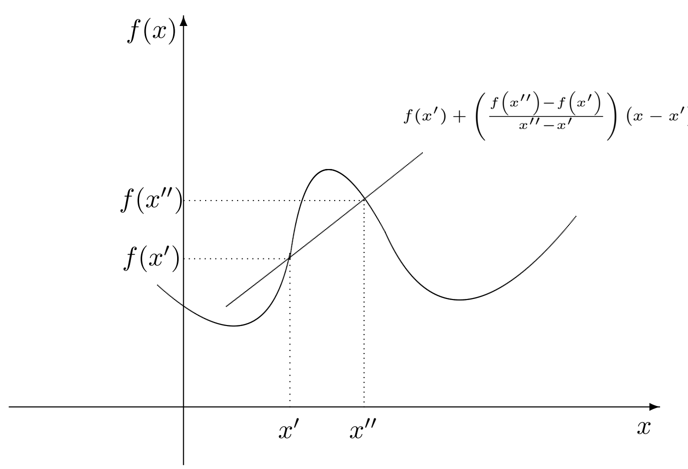

Figure 6: A Non-Linear Function

Note that

$$\begin{aligned} h(x) &= \underbrace{f(x') - \left(\frac{f(x'') - f(x')}{x'' - x'}\right) x'}_{b} + \underbrace{\left(\frac{f(x'') - f(x')}{x'' - x'}\right)}_{a} x = \\ &= f(x') + \left(\frac{f(x'') - f(x')}{x'' - x'}\right)(x - x') \end{aligned}$$

is a linear function that *assigns the same value to the function at* x' *and* x'' (to see this it is just to plug in $x = x'$ and $x = x''$ above and check that $h(x') = f(x')$ and $h(x'') = f(x'')$).

Now imagine that you do the same thing over and over again with x' fixed and x'' moved mover and mover to x' (from the right, say). Intuitively, it is rather clear that the slope of the linear function becomes "closer and closer" to what we intuitively think of at the slope of f at the point x'–*the tangent line.* Indeed, this is exactly how the derivative of a function at a particular point is *defined.*

Definition 1 *The derivative of* f *at* x'*, denoted* $\frac{df(x')}{dx}$*, is given by*

$$\frac{df(x')}{dx} = \lim_{x'' \to x'} \frac{f(x'') - f(x')}{x'' - x'}$$

You may have seen this written (as in Varian)

$$\frac{df(x')}{dx} = \lim_{\Delta x \to 0} \frac{f(x' + \Delta x) - f(x')}{\Delta x},$$

which is the same thing as you can see by letting $x'' = x' + \Delta x$. Observe that it is geometrically obvious that the *tangent line* drawn in Figure 7 and the lines between x' and x'' are getting really close to each other if you pick x'' very close to x'.

For our purposes, what is interesting about the derivative is that the *tangent line*,

$$h(x) = ax + b = f(x') + \frac{df(x')}{dx}(x - x'),$$

is "close" to the original function $f(x)$ for x "near" x'. Indeed, the tangent line is often called *the (best) linear approximation of* f *at* x'. To make sense of this at an intuitive level we observe that:

1. It is a linear function of x, with constant $b = f(x') - \frac{df(x')}{dx} x'$ and slope $a = \frac{df(x')}{dx}$

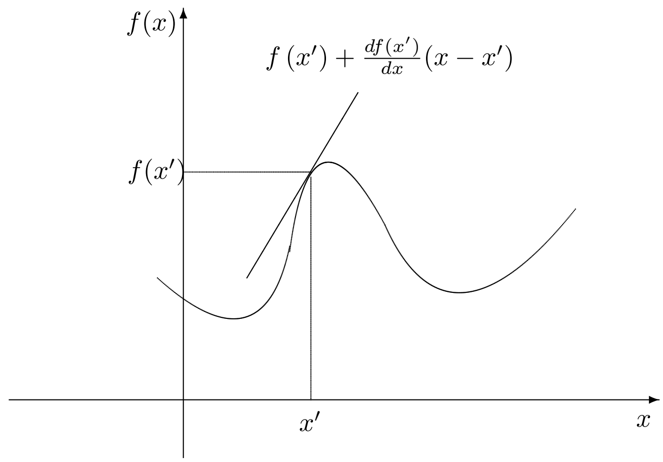

Figure 7: The Derivative of f at x'

2. Plug in $x = x'$ $\Rightarrow$the value of the linear function equals $f(x')$ $-$Thus it exactly coincides with the original function at x'.

3. $\frac{df(x')}{dx}$ is the slope of the original function at x' and also the slope coefficient of the linear function (thus, the slope for all x)

It is intuitive from the picture that the linear approximation is a good approximation of $f(x)$ near x'. In particular (which is what the optimization techniques rely on) it will be the case that the derivative is strictly positive when the function is strictly increasing and strictly negative when the function is strictly decreasing.

4.4 Optional Reading: How to Make this Rigorous?

An intuitive understanding will be sufficient for the purposes of this class, but you should not be fooled into thinking that the treatment above fully explains what a derivative is. Here, I'll only give the formal definition of a limit and discuss an example. If you want to get a better understanding, any reasonable book on calculus is a good source (the book by Sydsaeter and Hammond is excellent for example).

The hard part when talking derivatives is to give a precise meaning to

$$\lim_{x \to x'} F(x) = A.$$

Intuitively, the statement means that the difference between $F(x')$ and A can be made as small as we desire by letting the difference between x and x' be small enough. Formally, this may be stated as follows:

Definition 2 *We say that* $\lim_{x \to x'} F(x) = A$ *(in words,* $F(x)$ *tends to* A *as* x *tends to* x'*) if for every number* $\epsilon > 0$ *there exists a number* $\delta > 0$ *such that:*

$$A - \epsilon < F(x) < A + \epsilon \qquad \text{whenever} \qquad x' - \delta < x < x' + \delta$$

Since this definition makes the concept of a limit precise we know also have a precise definition of what a derivative is. Most of you probably know some rules of differentiation of certain functions and such rules can be established directly from this definition.

Example 3 *Consider* $f(x) = x^3$*. Then* $f(x'') - f(x') = (x'')^3 - (x')^3$*. Let* $x'' = x' + \Delta$ *and rewrite this as*

$$\begin{aligned}(x' + \Delta)^3 - (x')^3 &= (x' + \Delta)(x' + \Delta)(x' + \Delta) - (x')^3 \\ &= \left((x')^2 + 2\Delta x' + \Delta^2\right)(x' + \Delta) - (x')^3 \\ &= (x')^3 + (x')^2 \Delta + 2\Delta (x')^2 + 2\Delta^2 x' + \Delta^2 x' + \Delta^3 - (x')^3 \\ &= 3(x')^2 \Delta + 3x'\Delta^2 + \Delta^3\end{aligned}$$

Hence

$$\frac{f(x'') - f(x')}{x'' - x'} = \frac{f(x' + \Delta) - f(x')}{\Delta} = \frac{3(x')^2 \Delta + 3x'\Delta^2 + \Delta^3}{\Delta}$$
$$= 3(x')^2 + 3x'\Delta + \Delta^2$$

I now assert that as $\Delta \to 0$ *(same as* $x'' \to x'$*) we have that*

$$\lim_{\Delta \to 0} \frac{f(x' + \Delta) - f(x')}{\Delta} = \lim_{\Delta \to 0} 3(x')^2 + 3x'\Delta + \Delta^2 = 3(x')^2 . \tag{1}$$

Where I have just appealed to the intuitive notion of a limit (that is $3x'\Delta + \Delta^2$ *gets closer and closer to zero as* Δ *gets closer to zero).*

To see how this comes from the $\epsilon - \delta$ *definition of a limit above, consider any* $\epsilon > 0$ *and let*

$$\delta = \sqrt{\epsilon + \left(\frac{3x}{2}\right)^2} - \sqrt{\left(\frac{3x}{2}\right)^2},$$

which is an indisputably positive number for any positive ϵ*. It is a good exercise for you to verify (some algebra cranking is involved) that*

$$-\epsilon < 3x'\Delta + \Delta^2 < \epsilon \qquad \text{whenever} \qquad -\delta < \Delta < \delta,$$

which since $\delta = \sqrt{\epsilon + \left(\frac{3x}{2}\right)^2} - \sqrt{\left(\frac{3x}{2}\right)^2} > 0$ *for every* ϵ *means that the limit is as asserted in (1) above.*

4.5 Maxima

In order to say anything about the behavior of our rational consumer are interested in finding the consumption bundle that gives the highest value of the utility function of all bundles in the budget set. This is an example of an optimization problem and there are simple calculus techniques on how to handle this we can exploit.

Consider first the case where there are no constraints on x, so that x can take on any value from negative infinity to positive infinity. Then, a *maximum* is simply a point x^* such that *the value of the function at* x^* *exceeds the value of* x *for all* x. That is $f(x^*) \geq f(x)$ for all values of x. To find such a point we appeal to an absolutely crucial (and very simple) insight:

Claim If x^* is a maximum, then $\frac{df(x^*)}{dx} = 0$.

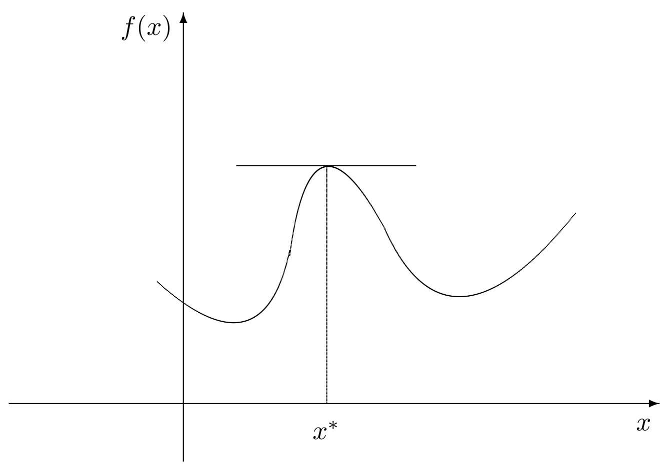

Figure 8: The Basic First Order Condition

I will not at any point test whether you understand **WHY** this is true, but it is actually very useful if you do. One way to convince yourself is to draw some graphs: a maximum must be at a place where the function "peaks", which is a point where the derivative is zero (given that the derivative is well defined at the maximum). One can also note that if you accept that the linear approximation at any point is close to the original function for "nearby" points we may argue as follows. Since the linear approximation around the maximum is good for x near x^* we have that

$$\begin{aligned} f(x) &\approx f(x^*) + \frac{df(x^*)}{dx}(x - x^*) \\ &\Leftrightarrow \\ f(x) - f(x^*) &\approx \frac{df(x^*)}{dx}(x - x^*) \end{aligned}$$

Thus,

$$\begin{aligned} \frac{df(x^*)}{dx} &> 0 \Rightarrow \text{ increase } x \text{ a little bit to increase } f(x) \\ \frac{df(x^*)}{dx} &< 0 \Rightarrow \text{ decrease } x \text{ a little bit to increase } f(x) \end{aligned}$$

4.6 Boundary Constraints

As you will see in the next lecture, the utility maximization problem for a consumer (and lots of other problems that you will see in this class) can be stated as a problem on the form

$$\max_{a \le x \le b} f(x),$$

where $a < b$. Sometimes we will use some "regularity conditions" on preferences or simply assume that the solution is interior, but in general we need to worry about "corner solutions" when we maximize a function over an interval. To understand why it is simply to observe that the logic we appealed to when we claimed that a maximum must be at a point where the derivative is zero *assumes that you can you in any direction from the candidate for a solution.* But,

1. At $x = a$ it is *impossible to decease* x further. Hence it is *possible* that we have an optimal solution a and that $\frac{df(a)}{dx} < 0$

2. At $x = b$ it is *impossible to increase* x further. Hence it is *possible* that we have an optimal solution b and that $\frac{df(b)}{dx} < 0$

For all other points except a and b the same logic as in the unconstrained case applies. We thus conclude that if we seek the maximum of $f(x)$ over an interval $[a, b]$ the only possible optimal solutions are points where the derivative is zero and the boundary points. Hence, we may use the following algoritm to solve the: maximization problem.

Step 1 Differentiate $f(x)$ and look for all values of x such that the derivative is zero. I.e., x_i^* is a "candidate solution if $\frac{df(x_i^*)}{dx} = 0$. In most problems you will see, there will be no more than one such candidate solution.

Step 2 Evaluate the function you are maximizing at all candidate solutions. In the case with a single x^* solving $\frac{df(x^*)}{dx} = 0$ that means that you compute $f(x^*), f(a)$, and $f(b)$. Then it is just to see which of these 3 numbers is the greatest, whatever x in the set $\{a, b, x^*\}$ that gives the largest value of the function is the solution to the problem.

Note that (this you will encounter) there in some cases will be no solution to $\frac{df(x)}{dx} = 0$. In this case the only candidate solutions are the boundary points a and b.

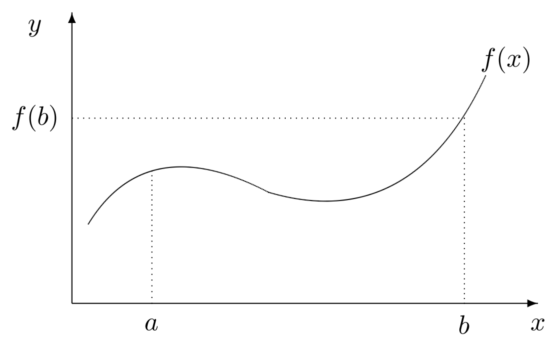

Figure 9: Example with Optimum at the Upper Boundary

4.7 Partial Derivatives

Derivatives are hard. About the only thing about them that isn't hard is to extend the concept of a derivative for a function with a single variable to partial derivatives for functions with more than one variable. Let $f(x,y)$ be a given function of x and y. The trivial step to realize is that if I set $y = y'$ one can define a new function h of a single variable (x) as

$$h(x) \equiv f(x, y')$$

Now the partial derivative of $f(x,y)$ with respect to x at a point (x', y') is simply the derivative of $h(x)$ at x'! We write $\frac{\partial f(x',y')}{\partial x}$ for the partial derivatives and what I just claimed is that

$$\frac{\partial f(x', y')}{\partial x} = \frac{dh(x')}{dx}$$

In words, the partial derivative of a function with respect to a variable is just the univariate derivative of the function you get by setting the other variable(s) to some value (corresponding to where the derivative is evaluated). In practice, this means that when you compute the partial derivative with respect to x, you simply pretend that all other variables but x are constants.

Example 4 *Consider $f(x,y) = \sqrt{x}y$. Now, to find the partial derivative of f with respect to x at some point (x', y') we treat y as a constant and differentiate with respect to y to get*

$$\frac{\partial f(x,y)}{\partial x} = \left[\frac{d}{dx}\sqrt{x}\right] y = \frac{1}{2\sqrt{x}} y$$

Now, say we want to know the partial with respect to x at $(x', y') = (1,1)$. Then it's just to plug into the expression above to get

$$\frac{\partial f(1,1)}{\partial x} = \frac{1}{2\sqrt{1}} 1 = \frac{1}{2}.$$

4.8 The Chain Rule

In examples with specific utility functions we will be able to get by without it, but to obtain a general expression we need to use the *chain rule* of differentiation to differentiate the objective of our maximization problem: (with two variables)

$$\begin{aligned} h(x) &= f(x, g(x)) \Rightarrow \\ \frac{dh(x)}{dx} &= \frac{\partial f(x, g(x))}{\partial x} + \frac{\partial f(x, g(x))}{\partial y} \frac{dg(x)}{dx} \end{aligned}$$

Intuition:

$$\frac{\partial f(x,g(x))}{\partial x} - \text{change in } f \text{ from a small change in } x$$
$$\frac{\partial f(x,g(x))}{\partial y} - \text{change in } f \text{ from a small change in } y$$
$$\frac{dg(x)}{dx} - \text{how big is the change in } y \text{ corresponding to small change in } x$$

4.9 Constrained Optimization

Consider a simple case, where there are only two choice variables, x and y. A "nonlinear programming problem" may then be written as

$$\begin{aligned} & \max f(x,y) \\ \text{s.t. } g(x,y) & \geq 0 \end{aligned} \tag{2}$$

Construct the Lagrangian

$$L(x,y,\lambda) = f(x,y) + \lambda g(x,y)$$

The basic result that underlies the actual procedure for the problem is:

Proposition 5 *Under some regularity conditions on g that we will not be careful about (they tend to be satisfied in economic applications), if (x^*, y^*) solves (2), then there exists some $\lambda^* \geq 0$ such that*

$$\begin{aligned} \frac{\partial L(x^*,y^*,\lambda^*)}{\partial x} &= \frac{\partial f(x^*,y^*)}{\partial x} + \lambda^* \frac{\partial g(x^*,y^*)}{\partial x} = 0 \\ \frac{\partial L(x^*,y^*,\lambda^*)}{\partial y} &= \frac{\partial f(x^*,y^*)}{\partial y} + \lambda^* \frac{\partial g(x^*,y^*)}{\partial y} = 0 \end{aligned} \tag{3}$$

and

$$\lambda^* g(x^*,y^*) = 0 \tag{4}$$

4.9.1 Explanation (Optional Reading)

This is not a full blown proof. However, it is still quite sophisticated and I only recommend those who are rather comfortable with calculus to read it.

Possibility 1: Suppose that $g(x^*, y^*) > 0$. In this case we say that the constraint is *non-binding*. But then it must be that there is some $\varepsilon > 0$ such that

$$f(x^*, y^*) \geq f(x, y)$$

for all (x, y) within an ε distance from (x^*, y^*). That is to say, (x^*, y^*) must be a local maximum of the function f which requires that the usual first order conditions for unconstrained optimization apply

$$\begin{aligned} \frac{\partial f(x^*, y^*)}{\partial x} &= 0 \\ \frac{\partial f(x^*, y^*)}{\partial y} &= 0 \end{aligned}$$

But, from (4), the Proposition above implies that $\lambda^* = 0$ in this case, so the first order conditions in (3) indeed reduces to the regular first order conditions for an unconstrained optimum.

Possibility 2: The only remaining possibility is that $g(x^*, y^*) = 0$. That is, the constraint is binding. To deal with this case, we observe that if $g(x^*, y^*) = 0$ at the (or a) solution to our problem (2), then it must be that (x^*, y^*) also solves

$$\begin{aligned} &\max_{x,y} f(x, y) \\ \text{s.t } g(x, y) &= 0 \end{aligned}$$

But then (this is one of the places where an actual proof gets a bit tricky) imagine that we "solve out the constraint". That is, let $h(x)$ be a function such that;

1. $y^* = h(x^*)$

2. $g(x, h(x)) = 0$ in some range $[x^* - \delta, x^* + \delta]$

3. h can be differentiated.

The mathematics that is used to validate this procedure is the implicit function theorem. If you haven't heard about that, just consider an example with $g(x, y) = p_x x + p_y y - m = 0$.

Then, $h(x) = \frac{m - p_x x}{p_y}$. Or, if $g(x,y) = x^\alpha y^{1-\alpha} - e$. Then $h(x) = \left[\frac{e}{x^\alpha}\right]^{\frac{1}{1-\alpha}}$.[1] But then, the problem is equivalent with solving

$$\max_x f(x, h(x))$$

Since we don't have a boundary constraint on this problem (such constraints are easily taken care of) we know that this requires that the derivative of $f(x, h(x))$ is zero at $x = x^*$, that is

$$\begin{aligned} \left.\frac{d}{dx}\right|_{x=x^*} [f(x, h(x))] &= 0 \Leftrightarrow \\ \frac{\partial f(x^*, y^*)}{\partial x} + \frac{\partial f(x^*, y^*)}{\partial y} h'(x^*) &= 0 \end{aligned} \tag{5}$$

But, since $h(x)$ solves out the constraint we have that (assuming that $\frac{\partial g(x^*,y^*)}{\partial y} \neq 0$)

$$\begin{aligned} g(x, h(x)) &= 0 \text{ for all } x \text{ in some range around } x^* \Rightarrow \\ \left.\frac{d}{dx}\right|_{x=x^*} [g(x, h(x))] &= 0 \Rightarrow \\ \frac{\partial g(x^*, y^*)}{\partial x} + \frac{\partial g(x^*, y^*)}{\partial y} h'(x^*) &= 0 \Rightarrow h'(x^*) = -\frac{\frac{\partial g(x^*,y^*)}{\partial x}}{\frac{\partial g(x^*,y^*)}{\partial y}} \end{aligned}$$

Now, let λ^* solve

$$\frac{\partial g(x^*, y^*)}{\partial x} \lambda^* = \frac{\partial f(x^*, y^*)}{\partial y} h'(x^*)$$

or (assuming that $\frac{\partial g(x^*,y^*)}{\partial x} \neq 0$...the other case you can treat separately)

$$\lambda^* = -\frac{\frac{\partial f(x^*,y^*)}{\partial y}}{\frac{\partial g(x^*,y^*)}{\partial y}},$$

Then the first order conditions (3) become

$$\begin{aligned} (\text{wrt } x)\ 0 &= \frac{\partial f(x^*, y^*)}{\partial x} + \lambda^* \frac{\partial g(x^*, y^*)}{\partial x} \\ &= \frac{\partial f(x^*, y^*)}{\partial x} + \frac{\partial f(x^*, y^*)}{\partial y} h'(x^*) \\ (\text{wrt } y)\ 0 &= \frac{\partial f(x^*, y^*)}{\partial y} + \lambda^* \frac{\partial g(x^*, y^*)}{\partial y} \\ &= \frac{\partial f(x^*, y^*)}{\partial y} - \frac{\frac{\partial f(x^*,y^*)}{\partial y}}{\frac{\partial g(x^*,y^*)}{\partial y}} \frac{\partial g(x^*, y^*)}{\partial y} = 0 \end{aligned}$$

[1] In general, we can't solve out explictly, but the implicit function theorem guarantees that a solution exists whenever the derivative with respect to y at (x^*, y^*) is non-zero.

That is the first order condition with respect to x is satisfied (becomes the usual tangency condition) and the first order condition with respect to y holds by construction of λ^* (i.e., the particular procedure would make this condition satisfied even if the tangency condition would fail). To sum up, we have so far verified that;

1. If the constraint binds we can let $\lambda^* = -\frac{\frac{\partial f(x^*,y^*)}{\partial y}}{\frac{\partial g(x^*,y^*)}{\partial y}}$
2. If the constraint doesn't bind one can just set $\lambda^* = 0$
3. In each case, the stated conditions are satisfied, except that we have to add an explanation for how we can be sure that the ratio of partial derivatives must be positive.

Why is $\lambda^* \geq 0$? The real advantage with inequality constraints is that we may assign the multipliers a definite sign. To understand this rewrite the first order conditions as

$$\frac{\frac{\partial f(x^*,y^*)}{\partial x}}{\frac{\partial g(x^*,y^*)}{\partial x}} = -\lambda^* = \frac{\frac{\partial f(x^*,y^*)}{\partial y}}{\frac{\partial g(x^*,y^*)}{\partial y}} \Rightarrow \frac{\frac{\partial f(x^*,y^*)}{\partial x}}{\frac{\partial f(x^*,y^*)}{\partial y}} = \frac{\frac{\partial g(x^*,y^*)}{\partial y}}{\frac{\partial g(x^*,y^*)}{\partial y}}$$

That is, despite the fact that it may look somewhat fancy, all we are doing is that we are using the Lagrange multiplier to find the point of tangency. Now, assume for the sake of argument that $\frac{\partial f(x^*,y^*)}{\partial x} > 0$ and $\frac{\partial f(x^*,y^*)}{\partial y} > 0$ (which is the case usually when we deal with utility and production functions). Then, if $\lambda^* < 0$ it would follow that $\frac{\partial g(x^*,y^*)}{\partial x} > 0$ and $\frac{\partial g(x^*,y^*)}{\partial y} > 0$. But this means that increasing either x or y will be consistent with satisfying the constraints, which is inconsistent with (x^*, y^*) being an optimum.

More generally (if you are not familiar with gradients you may ignore this), express the first order conditions as

$$\nabla f(x^*, y^*) + \lambda^* \nabla g(x^*, y^*) = 0,$$

where $\nabla f(x,y) = \left(\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y}\right)$ and $\nabla g(x,y) = \left(\frac{\partial g(x,y)}{\partial x}, \frac{\partial g(x,y)}{\partial y}\right)$ have the geometric interpretation of being vectors that are orthogonal to the level curves and pointing in the direction where the function increases. It should be intuitive that an optimum requires that $\nabla f(x^*, y^*)$ and $\nabla g(x^*, y^*)$ point in opposite directions, since otherwise there is no problem to increase the value of f and keep the constraint satisfied.

Cookbook Formula:

1. write down Lagrangian $L(x, y, \lambda) = f(x, y) + \lambda g(x, y)$

2. set the partial derivatives of Lagrangian with respect to choice variables to zero

$$\begin{aligned} \frac{\partial L(x, y, \lambda)}{\partial x} &= \frac{\partial f(x, y)}{\partial x} + \lambda \frac{\partial g(x, y)}{\partial x} = 0 \\ \frac{\partial L(x, y, \lambda)}{\partial y} &= \frac{\partial f(x, y)}{\partial y} + \lambda \frac{\partial g(x, y)}{\partial y} = 0 \end{aligned}$$

3. work with the condition that $\lambda g(x, y) = 0$ and $\lambda \geq 0$. Usually, this means that you have to think about whether it is possible that $g(x, y) > 0$ or not. This is the trickiest part of the analysis. Many times we know a priori that $g(x, y) = 0$ at the optimum, and then this step is eliminated. Regardless, it is often easy to forget the constraint $g(x, y) \geq 0$, which can sometimes lead to nonsense.

Generalization I will not provide any explanations. In fact, they are all more or less along the same lines as above (except that the question "is the constraint binding?" is replaced with "which constraints are binding?". Not surprisingly, this case is dealt with EXACTLY as the case with two variables. That is, let $x = (x_1, ..., x_n)$ and consider the problem

$$\begin{aligned} &\max_x f(x) \\ \text{s.t. } g_1(x) &\geq 0 \\ g_2(x) &\geq 0 \\ &.... \\ g_m(x) &\geq 0 \end{aligned}$$

Then the cookbook is to:

1. form the Lagrangian

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

or $L(x, \lambda) = f(x) + \lambda g(x)$ is $g(x) = (g_1(x), ..., g_m(x))$ and $\lambda = (\lambda_1, ..., \lambda_m)$ and $\lambda g(x)$ is understood as a scalar product.

2. set the partial derivatives with respect to the choice variables to zero

$$\frac{\partial L(x,\lambda)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i} + \sum_{i=1}^{m} \lambda_i \frac{\partial g(x)}{\partial x_i} = 0$$

   for $i = 1, ..., n$.

3. work with the conditions that $\lambda_i g_i(x) = 0$ and $\lambda_i \geq 0$ for every i.