ECONOMICS 100A MATHEMATICAL HANDOUT

Fall 2003 Mark Machina

A. CALCULUS REVIEW[1]

Derivatives, Partial Derivatives and the Chain Rule

You should already know what a derivative is. We'll use the expressions $f'(x)$ or $df(x)/dx$ for the derivative of the function $f(x)$. To indicate the derivative of $f(x)$ evaluated at the point $x = x^*$, we'll use the expressions $f'(x^*)$ or $df(x^*)/dx$.

When we have a function of more than one variable, we can consider its derivatives with respect to each of the variables, that is, each of its **partial derivatives**. We use the expressions:

$$\partial f(x_1,x_2)/\partial x_1 \quad \text{and} \quad f_1(x_1,x_2)$$

interchangeably to denote the partial derivative of $f(x_1,x_2)$ with respect to its first argument (that is, with respect to x_1). To calculate this, just hold x_2 fixed (treat it as a constant) so that $f(x_1,x_2)$ may be thought of as a function of x_1 alone, and differentiate it with respect to x_1. The notation for partial derivatives with respect to x_2 (or in the general case, with respect to x_i) is analogous.

For example, if $f(x_1,x_2) = x_1^2 \cdot x_2 + 3x_1$, we have:

$$\partial f(x_1,x_2)/\partial x_1 \;=\; 2x_1 \cdot x_2 + 3 \qquad \text{and} \qquad \partial f(x_1,x_2)/\partial x_2 \;=\; x_1^2$$

The **normal vector** of a function $f(x_1,...,x_n)$ at the point $(x_1,...,x_n)$ is just the vector (i.e., ordered list) of its n partial derivatives at that point, that is, the vector:

$$\left(\frac{\partial f(x_1,...,x_n)}{\partial x_1},\frac{\partial f(x_1,...,x_n)}{\partial x_2},...,\frac{\partial f(x_1,...,x_n)}{\partial x_n}\right) \;=\; \big(f_1(x_1,...,x_n), f_2(x_1,...,x_n),..., f_n(x_1,...,x_n)\big)$$

Normal vectors play a key role in the conditions for unconstrained and constrained optimization.

The **chain rule** gives the derivative for a "function of a function." Thus, if $f(x) \equiv g(h(x))$, then

$$f'(x) \;=\; g'(h(x)) \cdot h'(x)$$

The chain rule also applies to taking partial derivatives. For example, if $f(x_1,x_2) \equiv g(h(x_1,x_2))$ then

$$\frac{\partial f(x_1,x_2)}{\partial x_1} \;=\; g'(h(x_1,x_2)) \cdot \frac{\partial h(x_1,x_2)}{\partial x_1}$$

Similarly, if $f(x_1,x_2) \equiv g(h(x_1,x_2),k(x_1,x_2))$ then:

$$\frac{\partial f(x_1,x_2)}{\partial x_1} \;=\; g_1(h(x_1,x_2),k(x_1,x_2)) \cdot \frac{\partial h(x_1,x_2)}{\partial x_1} \;+\; g_2(h(x_1,x_2),k(x_1,x_2)) \cdot \frac{\partial k(x_1,x_2)}{\partial x_1}$$

The **second derivative** of the function $f(x)$ is written:

$$f''(x) \quad \text{or} \quad d^2 f(x)/dx^2$$

and it is obtained by differentiating the function $f(x)$ twice with respect to x (if you want the value of $f''(\cdot)$ at some point x^*, don't substitute in x^* until *after* you've differentiated twice).

---

[1] If the material in this section is not *already* familiar to you, you may have trouble on the 100A midterms and final.

A **second partial derivative** of a function of two or more variables is analogous, i.e., we will use the expressions:

$$f_{11}(x_1,x_2) \quad \text{or} \quad \partial^2 f(x_1,x_2)/\partial x_1^2$$

to denote differentiating twice with respect to x_1 (and $\partial^2 f(x_1,x_2)/\partial x_2^2$ for twice with respect to x_2).

We get a **cross partial derivative** when we differentiate first with respect to x_1 and then with respect to x_2. We will denote this with the expressions:

$$f_{12}(x_1,x_2) \quad \text{or} \quad \partial^2 f(x_1,x_2)/\partial x_1 \partial x_2$$

Here's a strange and wonderful result: if we had differentiated in the *opposite order*, that is, first with respect to x_2 and then with respect to x_1, we would have gotten the same result. In other words, we have $f_{12}(x_1,x_2) \equiv f_{21}(x_1,x_2)$ or equivalently $\partial^2 f(x_1,x_2)/\partial x_1 \partial x_2 \equiv \partial^2 f(x_1,x_2)/\partial x_2 \partial x_1$.

Approximation Formulas for Small Changes in Functions (Total Differentials)

If $f(x)$ is differentiable, we can approximate the effect of a small change in x by:

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

where Δx is the change in x. From calculus, we know that as Δx becomes smaller and smaller, this approximation becomes extremely good. We sometimes write this general idea more formally by expressing the **total differential** of $f(x)$, namely:

$$df \;=\; f'(x)\cdot dx$$

but it is still just shorthand for saying "We can approximate the change in $f(x)$ by the formula $\Delta f \approx f'(x)\cdot\Delta x$, and this approximation becomes extremely good for very small values of Δx."

When $f(\cdot)$ is a "function of a function," i.e., it takes the form $f(x) \equiv g(h(x))$, the chain rule lets us write the above approximation formula and above total differential formula as

$$\Delta g(h(x)) \;\approx\; \frac{dg(h(x))}{dx}\cdot\Delta x \;=\; g'(h(x))\cdot h'(x)\cdot\Delta x \qquad \text{so} \qquad dg(h(x)) = g'(h(x))\cdot h'(x)\cdot dx$$

For a function $f(x_1,...,x_n)$ that depends upon several variables, the approximation formula is:

$$\Delta f \;=\; f(x_1+\Delta x_1,...,x_n+\Delta x_n) - f(x_1,...,x_n) \;=\; \frac{\partial f(x_1,...,x_n)}{\partial x_1}\cdot\Delta x_1 + ... + \frac{\partial f(x_1,...,x_n)}{\partial x_n}\cdot\Delta x_n$$

Once again, this approximation formula becomes extremely good for very small values of $\Delta x_1,\ldots,\Delta x_n$. As before, we sometimes write this idea more formally (and succinctly) by expressing the **total differential** of $f(x)$, namely:

$$df \;=\; \frac{\partial f(x_1,...,x_n)}{\partial x_1}\cdot dx_1 \;+\; ... \;+\; \frac{\partial f(x_1,...,x_n)}{\partial x_n}\cdot dx_n$$

or in equivalent notation:

$$df \;=\; f_1(x_1,...,x_n)\cdot dx_1 + \cdots + f_n(x_1,...,x_n)\cdot dx_n$$

B. ELASTICITY

Let the variable y depend upon the variable x according to some function, i.e.:

$$y = f(x)$$

How responsive is y to changes in x? One measure of responsiveness would be to plot the function $f(\cdot)$ and look at its **slope**. If we did this, our measure of responsiveness would be:

$$\text{slope of } f(x) = \frac{\text{absolute change in } y}{\text{absolute change in } x} = \frac{\Delta y}{\Delta x} \approx \frac{dy}{dx} = f'(x)$$

Elasticity is a *different* measure of responsiveness than slope. Rather than looking at the ratio of the *absolute* change in y to the *absolute* change in x, elasticity is a measure of the *proportionate* (or percentage) change in y to the *proportionate* (or percentage) change in x. Formally, if $y = f(x)$, then the **elasticity of y with respect to** x, written $E_{y,x}$, is given by:

$$E_{y,x} = \frac{\text{proportionate change in } y}{\text{proportionate change in } x} = \frac{(\Delta y/y)}{(\Delta x/x)} = \left(\frac{\Delta y}{\Delta x}\right)\cdot\left(\frac{x}{y}\right)$$

If we consider very small changes in x (and hence in y), $\Delta y/\Delta x$ becomes $dy/dx = f'(x)$, so we get that the elasticity of y with respect to x is given by:

$$E_{y,x} = \frac{(\Delta y/y)}{(\Delta x/x)} = \left(\frac{\Delta y}{\Delta x}\right)\cdot\left(\frac{x}{y}\right) \approx \left(\frac{dy}{dx}\right)\cdot\left(\frac{x}{y}\right) = f'(x)\cdot\left(\frac{x}{y}\right)$$

Note that if $f(x)$ is an increasing function the elasticity will be positive, and if $f(x)$ is a decreasing function, it will be negative.

Recall that since the *percentage change* in a variable is simply 100 times its proportional change, elasticity is also as the ratio of the *percentage change in* y to the *percentage change in* x:

$$E_{y,x} = \frac{\left(\Delta y/y\right)}{\left(\Delta x/x\right)} = \frac{100\cdot\left(\Delta y/y\right)}{100\cdot\left(\Delta x/x\right)} = \frac{\%\text{ change in } y}{\%\text{ change in } x}$$

A useful intuitive interpretation: Since we can rearrange the above equation as

$$(\%\text{ change in } y) = E_{y,x}\cdot(\%\text{ change in } x)$$

we see that $E_{y,x}$ serves as the "conversion factor" between the percentage change in x and the percentage change in y.

Although elasticity and slope are both measures of how responsive y is to changes in x, they are *different* measures. In other words, ***elasticity is not the same as slope***. For example, if $y = 7\cdot x$, the *slope* of this curve is obviously 7, but its *elasticity* is 1:

$$E_{y,x} = \frac{dy}{dx}\cdot\frac{x}{y} = 7\cdot\frac{x}{7x} \equiv 1$$

That is, if y is *exactly proportional* to x, the elasticity of y with respect to x will always be *one*, regardless of the coefficient of proportionality.

Constant Slope Functions versus Constant Elasticity Functions

Another way to see that slope and elasticity are different measures is to consider the simple function $f(x) = 3 + 4x$. Although $f(\cdot)$ has a constant slope, it does *not* have a constant elasticity:

$$E_{f(\cdot),x} \;=\; \frac{d\,f(x)}{dx}\cdot\frac{x}{f(x)} \;=\; 4\cdot\frac{x}{3+4x} \;=\; \frac{4x}{3+4x}$$

which is obviously not constant as x changes.

However, some functions *do* have a **constant elasticity** for all values of x, namely functions of the form $f(x) \equiv c\cdot x^{\beta}$, for any constants $c > 0$ and $\beta \gtreqless 0$. Since it involves taking x to a fixed power β, this function can be called a **power function.** Deriving its elasticity gives:

$$E_{f(x),x} \;=\; \frac{d\,f(x)}{dx}\cdot\frac{x}{f(x)} \;=\; \frac{\beta\cdot c\cdot x^{(\beta-1)}\cdot x}{c\cdot x^{\beta}} \;\underset{x}{\equiv}\; \beta$$

Conversely, if a function $f(\cdot)$ has a constant elasticity, it must be a power function. In summary:

linear functions all have a constant *slope*:	$f(x) \underset{x}{\equiv} \alpha+\beta\cdot x$	$\Leftrightarrow$	$\frac{d\,f(x)}{dx} \underset{x}{\equiv} \beta$
power functions all have a constant *elasticity*:	$f(x) \underset{x}{\equiv} c\cdot x^{\beta}$	$\Leftrightarrow$	$E_{f(x),x} \underset{x}{\equiv} \beta$

C. LEVEL CURVES OF FUNCTIONS

If $f(x_1,x_2)$ is a function of the two variables x_1 and x_2, a **level curve** of $f(x_1,x_2)$ is just a locus of points in the (x_1,x_2) plane along which $f(x_1,x_2)$ takes on some constant value, say the value k. The equation of this level curve is therefore simply $f(x_1,x_2) = k$, where we may or may not want to solve for x_2. For example, the level curves of a consumer's utility function are just his or her indifference curves (defined by the equation $U(x_1,x_2) = u_0$), and the level curves of a firm's production function are just the isoquants (defined by the equation $f(L,K) = Q_0$).

The **slope of a level curve** is indicated by the notation:

$$\left.\frac{dx_2}{dx_1}\right|_{f(x_1,x_2)=k} \qquad \text{or} \qquad \left.\frac{dx_2}{dx_1}\right|_{\Delta f=0}$$

where the subscripted equations are used to remind us that x_1 and x_2 must vary in a manner which keeps us on the $f(x_1,x_2) = k$ level curve (i.e., so that $\Delta f = 0$). To calculate this slope, recall the vector of changes $(\Delta x_1,\Delta x_2)$ will keep us on this level curve if and only if it satisfies the equation:

$$0 \;=\; \Delta f \;\approx\; f_1(x_1,x_2)\cdot\Delta x_1 \;+\; f_2(x_1,x_2)\cdot\Delta x_2$$

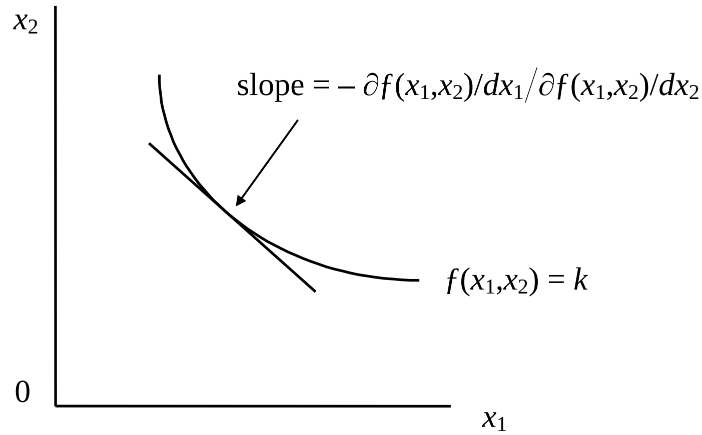
which implies that Δx_1 and Δx_2 will accordingly satisfy:

$$\left.\frac{\Delta x_2}{\Delta x_1}\right|_{f(x_1,x_2)=k} \;\approx\; -\frac{\partial f(x_1,x_2)/\partial x_1}{\partial f(x_1,x_2)/\partial x_2}$$

so that in the limit we have:

$$\left.\frac{dx_2}{dx_1}\right|_{f(x_1,x_2)=k} \quad = \quad -\frac{\partial f(x_1,x_2)/\partial x_1}{\partial f(x_1,x_2)/\partial x_2}$$

This slope gives the rate at which we can "trade off" or "substitute" x_2 against x_1 so as to leave the value of the function $f(x_1,x_2)$ unchanged. This concept will be of frequent use in this course.



An application of this result is that the slope of the indifference curve at a given consumption bundle is given by the ratio of the marginal utilities of the two commodities at that bundle. Another application is that the slope of an isoquant at a given input bundle is the ratio of the marginal products of the two factors at that input bundle.

In the case of a function $f(x_1,...,x_n)$ of several variables, we will have a **level surface** in n–dimensional space along which the function is constant, that is, defined by the equation $f(x_1,...,x_n) = k$. In this case the level surface does not have a unique tangent line. However, we can still determine the rate at which we can trade off any *pair* of variables x_i and x_j so as to keep the value of the function constant. By exact analogy with the above derivation, this rate is given by:

$$\left.\frac{dx_i}{dx_j}\right|_{f(x_1,...,x_n)=k} \quad = \quad \left.\frac{dx_i}{dx_j}\right|_{\Delta f=0} \quad = \quad -\frac{f_j(x_1,...,x_n)}{f_i(x_1,...,x_n)}$$

Given any level curve (or level surface) corresponding to the value k, its **better-than set** is the set of all points at which the function yields a higher value than k, and its **worse-than set** is the set of all points at which the function yields a lower value than k.

D. OPTIMIZATION #1: SOLVING OPTIMIZATION PROBLEMS

The General Structure of Optimization Problems

Economics is full of optimization (maximization or minimization) problems: the maximization of utility, the minimization of expenditure, the minimization of cost, the maximization of profits, etc. Understanding these is a lot easier if one knows what is systematic about such problems.

Each optimization problem has an **objective function** $f(x_1,...,x_n;\alpha_1,...,\alpha_m)$ which we are trying to either maximize or minimize (in our examples, we'll always be maximizing). This function depends upon both the **control variables** $x_1,...,x_n$ which we (or the economic agent) are able to set, as well as some **parameters** $\alpha_1,...,\alpha_m$, which are given as part of the problem. Thus a general unconstrained maximization problem takes the form:

$$\max_{x_1,...,x_n} f(x_1,...,x_n;\alpha_1,...,\alpha_m)$$

Consider the following one-parameter maximization problem

$$\max_{x_1,...,x_n} f(x_1,...,x_n;\alpha)$$

(It's only for simplicity that we assume just one parameter. *All* of our results will apply to the general case of many parameters $\alpha_1,...,\alpha_m$.) We represent the solutions to this problem, which obviously depend upon the values of the parameter(s), by the n **solution functions**:

$$\begin{aligned} x_1^* &= x_1^*(\alpha) \\ x_2^* &= x_2^*(\alpha) \\ &\vdots \\ x_n^* &= x_n^*(\alpha) \end{aligned}$$

It is often useful to ask "how well have we done?" or in other words, "how high can we get $f(x_1,...,x_n;\alpha)$, given the value of the parameter α?" This is obviously determined by substituting in the optimal solutions back into the objective function, to obtain:

$$\phi(\alpha) \equiv f(x_1^*,...,x_n^*;\alpha) \equiv f(x_1^*(\alpha),...,x_n^*(\alpha);\alpha)$$

and $\phi(\alpha)$ is called the **optimal value function**.

Sometimes we will be optimizing subject to a *constraint* on the control variables (such as the budget constraint of the consumer). Since this constraint may also depend upon the parameter(s), our problem becomes:

$$\max_{x_1,...,x_n} f(x_1,...,x_n;\alpha)$$

$$\text{subject to} \quad g(x_1,...,x_n;\alpha) = c$$

(Note that we now have an additional parameter, namely the constant c.) In this case we still define the solution functions and optimal value function in the same way – we just have to remember to take into account the constraint. Although it is possible that there could be more than one constraint in a given problem, we will only consider problems with a single constraint. For example, if we were looking at the profit maximization problem, the control variables would be the quantities of inputs and outputs chosen by the firm, the parameters would be the current input and output prices, the constraint would be the production function, and the optimal value function would be the firm's "profit function," i.e., the highest attainable level of profits given current input and output prices.

In economics we are interested both in how the optimal values of the control variables and the optimal attainable value vary with the parameters. In other words, we will be interested in differentiating both the solution functions and the optimal value function with respect to the parameters. Before we can do this, however, we need to know how to solve unconstrained or constrained optimization problems.

First Order Conditions for Unconstrained Optimization Problems

The first order conditions for the unconstrained optimization problem:

$$\max_{x_1,...,x_n} f(x_1,...,x_n)$$

are simply that each of the partial derivatives of the objective function be zero at the solution values $(x_1^*,...,x_n^*)$, i.e. that:

$$\begin{aligned} f_1(x_1^*,...,x_n^*) &= 0 \\ &\vdots \\ f_n(x_1^*,...,x_n^*) &= 0 \end{aligned}$$

The intuition is that if you want to be at a "mountain top" (a maximum) or the "bottom of a bowl" (a minimum) it must be the case that no small change in any control variable be able to move you up or down. That means that the partial derivatives of $f(x_1,...,x_n)$ with respect to each x_i must be zero.

Second Order Conditions for Unconstrained Optimization Problems

If our optimization problem is a maximization problem, the second order condition for this solution to be a local maximum is that $f(x_1, ...,x_n)$ be a weakly concave function of $(x_1,...,x_n)$ (i.e., a mountain top) in the locality of this point. Thus, if there is only one control variable, the second order condition is that $f''(x^*) < 0$ at the optimum value of the control variable x. If there are two control variables, it turns out that the conditions are:

$$f_{11}(x_1^*,x_2^*) < 0 \qquad f_{22}(x_1^*,x_2^*) < 0$$

and

$$\begin{vmatrix} f_{11}(x_1^*,x_2^*) & f_{12}(x_1^*,x_2^*) \\ f_{21}(x_1^*,x_2^*) & f_{22}(x_1^*,x_2^*) \end{vmatrix} > 0$$

When we have a minimization problem, the second order condition for this solution to be a local minimum is that $f(x_1,...,x_n)$ be a weakly convex function of $(x_1,...,x_n)$ (i.e., the bottom of a bowl) in the locality of this point. Thus, if there is only one control variable x, the second order condition is that $f''(x^*) > 0$. If there are two control variables, the conditions are:

$$f_{11}(x_1^*,x_2^*) > 0 \qquad f_{22}(x_1^*,x_2^*) > 0$$

and

$$\begin{vmatrix} f_{11}(x_1^*,x_2^*) & f_{12}(x_1^*,x_2^*) \\ f_{21}(x_1^*,x_2^*) & f_{22}(x_1^*,x_2^*) \end{vmatrix} > 0$$

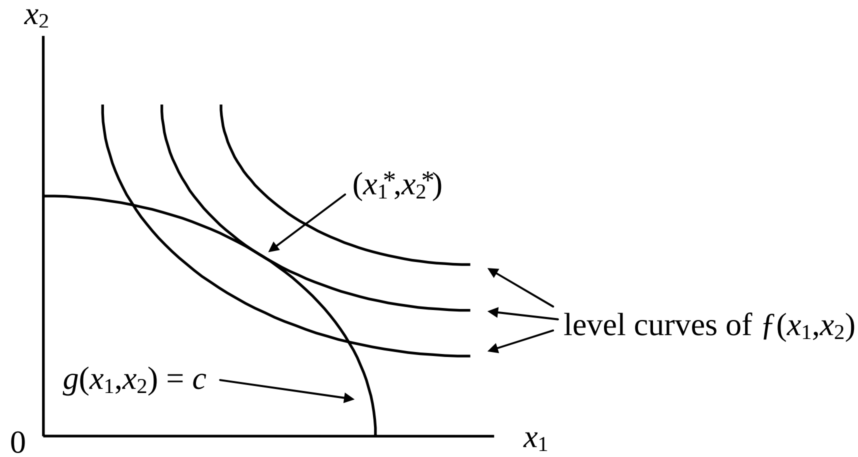
(yes, this last determinant really is supposed to be *positive*).

First Order Conditions for Constrained Optimization Problems (VERY important)

The first order conditions for the two-variable constrained optimization problem:

$$\max_{x_1,x_2} f(x_1,x_2) \qquad \text{subject to} \qquad g(x_1,x_2)=c$$

are easy to see from the following diagram



The point (x_1^*,x_2^*) is clearly not an *unconstrained* maximum, since increasing both x_1 and x_2 would move you to a higher level curve for $f(x_1,x_2)$. However, this change is not "legal" since it does not satisfy the constraint – it would move you off of the level curve $g(x_1,x_2) = c$. In order to stay on the level curve, we must jointly change x_1 and x_2 in a manner which preserves the value of $g(x_1,x_2)$. That is, we can only tradeoff x_1 against x_2 at the "legal" rate:

$$\left.\frac{dx_2}{dx_1}\right|_{g(x_1,x_2)=c} = \left.\frac{dx_2}{dx_1}\right|_{\Delta g=0} = -\frac{g_1(x_1,x_2)}{g_2(x_1,x_2)}$$

The condition for maximizing $f(x_1,x_2)$ subject to $g(x_1,x_2) = c$ is that no tradeoff between x_1 and x_2 at this "legal" rate be able to raise the value of $f(x_1,x_2)$. This is the same as saying that the level curve of the constraint function be *tangent* to the level curve of the objective function. In other words, *the tradeoff rate which preserves the value of* $g(x_1,x_2)$ *(the "legal" rate) must be the same as the tradeoff rate that preserves the value of* $f(x_1,x_2)$. We thus have the condition:

$$\left.\frac{dx_2}{dx_1}\right|_{\Delta g=0} = \left.\frac{dx_2}{dx_1}\right|_{\Delta f=0}$$

which implies that:

$$-\frac{g_1(x_1,x_2)}{g_2(x_1,x_2)} = -\frac{f_1(x_1,x_2)}{f_2(x_1,x_2)}$$

which is in turn equivalent to:

$$f_1(x_1^*,x_2^*) = \lambda\cdot g_1(x_1^*,x_2^*)$$
$$f_2(x_1^*,x_2^*) = \lambda\cdot g_2(x_1^*,x_2^*)$$

for some scalar λ.

To summarize, we have that the first order conditions for the constrained maximization problem:

$$\max_{x_1,x_2} f(x_1,x_2)$$
$$\text{subject to} \qquad g(x_1,x_2)=c$$

are that the solutions (x_1^*,x_2^*) satisfy the equations

$$\begin{aligned} f_1(x_1^*,x_2^*) &= \lambda\cdot g_1(x_1^*,x_2^*) \\ f_2(x_1^*,x_2^*) &= \lambda\cdot g_2(x_1^*,x_2^*) \\ g(x_1^*,x_2^*) &= c \end{aligned}$$

for some scalar λ. An easy way to remember these conditions is simply that the normal vector to $f(x_1,x_2)$ at the optimal point (x_1^*,x_2^*) must be a *scalar multiple* of the normal vector to $g(x_1,x_2)$ at the optimal point (x_1^*,x_2^*), i.e. that:

$$(\ f_1(x_1^*,x_2^*)\ ,\ f_2(x_1^*,x_2^*)\ ) \ = \ \lambda\cdot(\ g_1(x_1^*,x_2^*)\ ,\ g_2(x_1^*,x_2^*)\ )$$

and also that the constraint $g(x_1^*,x_2^*) = c$ be satisfied.

This same principle extends to the case of several variables. In other words, the conditions for $(x_1^*,...,x_n^*)$ to be a solution to the constrained maximization problem:

$$\max_{x_1,...,x_n} f(x_1,...,x_n)$$
$$\text{subject to} \qquad g(x_1,...,x_n)=c$$

is that no legal tradeoff between *any* pair of variables x_i and x_j be able to affect the value of the objective function. In other words, *the tradeoff rate between x_i and x_j that preserves the value of $g(x_1,...,x_n)$ must be the same as the tradeoff rate between x_i and x_j that preserves the value of $f(x_1,...,x_n)$*. We thus have the condition:

$$\left.\frac{dx_i}{dx_j}\right|_{\Delta g=0} = \left.\frac{dx_i}{dx_j}\right|_{\Delta f=0} \qquad \text{for all } i \text{ and } j$$

or in other words, that:

$$-\frac{g_j(x_1,...,x_n)}{g_i(x_1,...,x_n)} = -\frac{f_j(x_1,...,x_n)}{f_i(x_1,...,x_n)} \qquad \text{for all } i \text{ and } j$$

Again, the only way to ensure that these ratios will be equal for all *i* and *j* is to have:

$$\begin{aligned} f_1(x_1^*,...,x_n^*) &= \lambda\cdot g_1(x_1^*,...,x_n^*) \\ f_2(x_1^*,...,x_n^*) &= \lambda\cdot g_2(x_1^*,...,x_n^*) \\ &\vdots \\ f_n(x_1^*,...,x_n^*) &= \lambda\cdot g_n(x_1^*,...,x_n^*) \end{aligned}$$

To summarize: the first order conditions for the constrained maximization problem:

$$\max_{x_1,\ldots,x_n} f(x_1,\ldots,x_n)$$
$$\text{subject to} \qquad g(x_1,\ldots,x_n) = c$$

are that the solutions $(x_1^*,\ldots,x_n^*)$ satisfy the equations:

$$\begin{aligned}
f_1(x_1^*,\ldots,x_n^*) &= \lambda\cdot g_1(x_1^*,\ldots,x_n^*) \\
f_2(x_1^*,\ldots,x_n^*) &= \lambda\cdot g_2(x_1^*,\ldots,x_n^*) \\
&\vdots \\
f_n(x_1^*,\ldots,x_n^*) &= \lambda\cdot g_n(x_1^*,\ldots,x_n^*)
\end{aligned}$$

and the constraint $\qquad g(x_1^*,\ldots,x_n^*) = c$

Once again, the easy way to remember this is simply that the normal vector of $f(x_1,\ldots,x_n)$ be a *scalar multiple* of the normal vector of $g(x_1,\ldots,x_n)$ at the optimal point, i.e.:

$$(\,f_1(x_1^*,\ldots,x_n^*)\,,\,\ldots\,,\,f_n(x_1^*,\ldots,x_n^*)\,) \;=\; \lambda\cdot(\,g_1(x_1^*,\ldots,x_n^*)\,,\,\ldots\,,\,g_n(x_1^*,\ldots,x_n^*)\,)$$

and also that the constraint $g(x_1^*,\ldots,x_n^*) = c$ be satisfied.

Lagrangians

The first order conditions for the above constrained maximization problem are just a system of n+1 equations in the n+1 unknowns $x_1,\ldots,x_n$ and λ. Personally, I suggest that you get these first order conditions the direct way by simply setting the normal vector of $f(x_1,\ldots,x_n)$ to equal a scalar multiple of the normal vector of $g(x_1,\ldots,x_n)$ (with the scale factor λ). However, another way to obtain these equations is to construct the **Lagrangian function**:

$$\mathcal{L}(x_1,\ldots,x_n,\lambda) \;\equiv\; f(x_1,\ldots,x_n) + \lambda\cdot[c - g(x_1,\ldots,x_n)]$$

(where λ is called the **Lagrangian multiplier**). Then, if we calculate the partial derivatives $\partial\mathcal{L}/\partial x_1,\ldots,\partial\mathcal{L}/\partial x_n$ and $\partial\mathcal{L}/\partial\lambda$ and set them all equal to zero, we get the equations:

$$\begin{aligned}
\partial\mathcal{L}(x_1^*,\ldots,x_n^*,\lambda)\big/\partial x_1 &= f_1(x_1^*,\ldots,x_n^*) - \lambda\cdot g_1(x_1^*,\ldots,x_n^*) = 0 \\
\partial\mathcal{L}(x_1^*,\ldots,x_n^*,\lambda)\big/\partial x_2 &= f_2(x_1^*,\ldots,x_n^*) - \lambda\cdot g_2(x_1^*,\ldots,x_n^*) = 0 \\
&\vdots \\
\partial\mathcal{L}(x_1^*,\ldots,x_n^*,\lambda)\big/\partial x_n &= f_n(x_1^*,\ldots,x_n^*) - \lambda\cdot g_n(x_1^*,\ldots,x_n^*) = 0 \\
\partial\mathcal{L}(x_1^*,\ldots,x_n^*,\lambda)\big/\partial\lambda &= c - g(x_1^*,\ldots,x_n^*) = 0
\end{aligned}$$

But these equations are the same as our original n+1 first order conditions. In other words, the method of Lagrangians is nothing more than a roundabout way of generating our condition that the normal vector of $f(x_1,\ldots,x_n)$ be λ times the normal vector of $g(x_1,\ldots,x_n)$, and the constraint $g(x_1,\ldots,x_n) = c$ be satisfied.

We will not do second order conditions for constrained optimization: they are a royal pain.

E. SCALE PROPERTIES OF FUNCTIONS

A function $f(x_1,...,x_n)$ is said to exhibit **constant returns to scale** if:

$$f(\lambda\cdot x_1,...,\lambda\cdot x_n) \;\equiv\; \lambda\cdot f(x_1,...,x_n) \qquad \text{for all } x_1,...,x_n \text{ and all } \lambda > 0$$

That is, if multiplying all arguments by λ leads to the value of the function being multiplied by λ. Functions that exhibit constant returns to scale are also said to be **homogeneous of degree 1**.

A function $f(x_1,...,x_n)$ is said to be **scale invariant** if:

$$f(\lambda\cdot x_1,...,\lambda\cdot x_n) \;\equiv\; f(x_1,...,x_n) \qquad \text{for all } x_1,...,x_n \text{ and all } \lambda > 0$$

In other words, if multiplying all the arguments by λ leads to *no change* in the value of the function. Functions that exhibit scale invariance are also said to be **homogeneous of degree 0**.

Say that $f(x_1,...,x_n)$ is homogeneous of degree one, so that we have $f(\lambda\cdot x_1, ...,\lambda\cdot x_n) \equiv \lambda\cdot f(x_1,...,x_n)$. Differentiating this identity with respect to λ yields:

$$\sum_{i=1}^{n} f_i(\lambda\cdot x_1,..., \lambda\cdot x_n)\cdot x_i \;\equiv\; f(x_1,..., x_n) \qquad \text{for all } x_1,..., x_n\text{, all } \lambda > 0$$

and setting $\lambda = 1$ then gives:

$$\sum_{i=1}^{n} f_i(x_1,..., x_n)\cdot x_i \;\equiv\; f(x_1,..., x_n) \qquad \text{for all } x_1,..., x_n$$

which is called **Euler’s theorem**, and which will turn out to have very important implications for the distribution of income among factors of production.

Here’s another useful result: if a function is homogeneous of degree 1, then its partial derivatives are all homogeneous of degree 0. To see this, take the identity $f(\lambda\cdot x_1,...,\lambda\cdot x_n) \equiv \lambda\cdot f(x_1,...,x_n)$ and this time differentiate with respect to x_i, to get:

$$\lambda\cdot f_i(\lambda\cdot x_1,...,\lambda\cdot x_n) \;\equiv\; \lambda\cdot f_i(x_1,...,x_n) \qquad \text{for all } x_1,...,x_n \text{ and } \lambda > 0$$

or equivalently:

$$f_i(\lambda\cdot x_1,...,\lambda\cdot x_n) \;\equiv\; f_i(x_1,...,x_n) \qquad \text{for all } x_1,...,x_n \text{ and } \lambda > 0$$

which establishes our result. In other words, if a production function exhibits constant returns to scale (i.e., is homogeneous of degree 1), the marginal products of all the factors will be scale invariant (i.e., homogeneous of degree 0).

F. OPTIMIZATION #2: COMPARATIVE STATICS OF SOLUTION FUNCTIONS

Having obtained the first order conditions for a constrained or unconstrained optimization problem, we can now ask how the optimal values of the control variables *change* when the parameters change (for example, how the optimal quantity of a commodity will be affected by a price change or an income change).

Consider a simple maximization problem with a single control variable x and single parameter α

$$\max_{x}\; f(x;\alpha)$$

For a given value of α, recall that the solution x^* is the value that satisfies the first order condition

$$\frac{\partial f(x^*;\alpha)}{\partial x} \;=\; 0$$

Since the values of economic parameters can (and do) change, we have defined the *solution function* $x^*(\alpha)$ as the formula that specifies the optimal value x^* for each value of α. Thus, for each value of α, the value of $x^*(\alpha)$ satisfies the first order condition for that value of α. So we can basically plug the solution function $x^*(\alpha)$ into the first order condition to obtain the *identity*

$$\frac{\partial f(x^*(\alpha);\alpha)}{\partial x} \;\underset{\alpha}{\equiv}\; 0$$

We refer to this as the **identity version of the first order condition**.

Comparative statics is the study of how *changes* in a parameter affect the optimal value of a control variable. For example, is $x^*(\alpha)$ an increasing or decreasing function of α? How sensitive is $x^*(\alpha)$ to changes in α? To learn this about $x^*(\alpha)$, we need to derive its derivative $dx^*(\alpha)/d\alpha$. The easiest way to get $dx^*(\alpha)/d\alpha$ would be to solve the first order condition to get the formula for $x^*(\alpha)$ itself, then differentiate it with respect to α to get the formula for $dx^*(\alpha)/d\alpha$. But sometimes first order conditions are too complicated to solve.

Are we up a creek? No: there is another approach, **implicit differentiation**, which always gets the formula for the derivative $dx^*(\alpha)/d\alpha$. In fact, it can get the formula for $dx^*(\alpha)/d\alpha$ even when we can't get the formula for the solution function $x^*(\alpha)$ itself!

Implicit differentiation is straightforward. Since the solution function $x^*(\alpha)$ satisfies the identity

$$\frac{\partial f(x^*(\alpha);\alpha)}{\partial x} \;\underset{\alpha}{\equiv}\; 0$$

we can just totally differentiate this identity with respect to α, to get

$$\frac{\partial^2 f(x^*(\alpha);\alpha)}{\partial x^2}\cdot\frac{dx^*(\alpha)}{d\alpha} \;+\; \frac{\partial^2 f(x^*(\alpha);\alpha)}{\partial x\,\partial\alpha} \;\underset{\alpha}{\equiv}\; 0$$

and solve to get
$$\frac{dx^*(\alpha)}{d\alpha} \;\underset{\alpha}{\equiv}\; -\,\frac{\partial^2 f(x^*(\alpha);\alpha)}{\partial x\,\partial\alpha}\bigg/\frac{\partial^2 f(x^*(\alpha);\alpha)}{\partial x^2}$$

For example, let's go back to that troublesome problem max $\alpha\cdot x^2 - e^x$, with first order condition $2\cdot\alpha\cdot x^* - e^{x^*} = 0$. Its solution function $x^* = x^*(\alpha)$ satisfies the first order condition identity

$$2\cdot\alpha\cdot x^*(\alpha) \;-\; e^{x^*(\alpha)} \;\underset{\alpha}{\equiv}\; 0$$

So to get the formula for $dx^*(\alpha)/d\alpha$, totally differentiate this identity with respect to α:

$$2\cdot x^*(\alpha) \;+\; 2\cdot\alpha\cdot\frac{dx^*(\alpha)}{d\alpha} \;-\; e^{x^*(\alpha)}\cdot\frac{dx^*(\alpha)}{d\alpha} \;\underset{\alpha}{\equiv}\; 0$$

and solve, to get
$$\frac{dx^*(\alpha)}{d\alpha} \;\underset{\alpha}{\equiv}\; -\frac{2\cdot x^*(\alpha)}{2\cdot\alpha - e^{x^*(\alpha)}}$$

Comparative Statics when there are Several Parameters

Implicit differentiation also works when there is more than one parameter. Consider the problem

$$\max_x \; f(x;\alpha,\beta)$$

with first order condition

$$\frac{\partial f(x^*;\alpha,\beta)}{\partial x} \;=\; 0$$

Since the solution function $x^* = x^*(\alpha,\beta)$ satisfies this first order condition for all values of α and β, we have the identity

$$\frac{\partial f(x^*(\alpha,\beta);\alpha,\beta)}{\partial x} \;\underset{\alpha,\beta}{\equiv}\; 0$$

Note that the optimal value $x^* = x^*(\alpha,\beta)$ is affected by both changes in α as well as changes in β. To derive $\partial x^*(\alpha,\beta)/\partial\alpha$, we totally differentiate the above identity with respect to α, and then solve. If we want $\partial x^*(\alpha,\beta)/\partial\beta$, we totally differentiate the identity with respect to β, then solve.

For example, consider the maximization problem

$$\max_x \; a\cdot\ln(x)-\beta\cdot x^2$$

Since the first order condition is

$$\alpha\cdot[x^*]^{-1} \;-\; 2\cdot\beta\cdot x^* \;=\; 0$$

its solution function $x^*(\alpha,\beta)$ will satisfy the identity

$$\alpha\cdot[x^*(\alpha,\beta)]^{-1} \;-\; 2\cdot\beta\cdot x^*(\alpha,\beta) \;\underset{\alpha,\beta}{\equiv}\; 0$$

To get $\partial x^*(\alpha,\beta)/\partial\alpha$, totally differentiate this identity with respect to α:

$$[x^*(\alpha,\beta)]^{-1} \;-\; \alpha\cdot[x^*(\alpha,\beta)]^{-2}\cdot\frac{\partial x^*(\alpha,\beta)}{\partial\alpha} \;-\; 2\cdot\beta\cdot\frac{\partial x^*(\alpha,\beta)}{\partial\alpha} \;\underset{\alpha,\beta}{\equiv}\; 0$$

and solve to get:

$$\frac{\partial x^*(\alpha,\beta)}{\partial\alpha} \;=\; \frac{[x^*(\alpha,\beta)]^{-1}}{\alpha\cdot[x^*(\alpha,\beta)]^{-2} \;+\; 2\cdot\beta}$$

On the other hand, to get $\partial x^*(\alpha,\beta)/\partial\beta$, totally differentiate the identity with respect to β:

$$-\alpha\cdot[x^*(\alpha,\beta)]^{-2}\cdot\frac{\partial x^*(\alpha,\beta)}{\partial\beta} \;-\; 2\cdot x^*(\alpha,\beta) \;-\; 2\cdot\beta\cdot\frac{\partial x^*(\alpha,\beta)}{\partial\beta} \;=\; 0$$

and solve to get:

$$\frac{\partial x^*(\alpha,\beta)}{\partial\beta} \;=\; \frac{-\,2\cdot x^*(\alpha,\beta)}{\alpha\cdot[x^*(\alpha,\beta)]^{-2} \;+\; 2\cdot\beta}$$

Comparative Statics when there are Several Control Variables

Implicit differentiation also works when there is more than one control variable, and hence more than one equation in the first order condition. Consider the example

$$\max_{x_1,x_2} f(x_1, x_2;\alpha)$$

The first order conditions are that x_1^* and x_2^* solve the *pair* of equations

$$\frac{\partial f(x_1^*, x_2^*;\alpha)}{\partial x_1} = 0 \qquad \text{and} \qquad \frac{\partial f(x_1^*, x_2^*;\alpha)}{\partial x_2} = 0$$

so the solution functions $x_1^* = x_1^*(\alpha)$ and $x_2^* = x_2^*(\alpha)$ satisfy the pair of identities

$$\frac{\partial f(x_1^*(\alpha), x_2^*(\alpha);\alpha)}{\partial x_1} \underset{\alpha}{\equiv} 0 \qquad \text{and} \qquad \frac{\partial f(x_1^*(\alpha), x_2^*(\alpha);\alpha)}{\partial x_2} \underset{\alpha}{\equiv} 0$$

To get $\partial x_1^*(\alpha)/\partial\alpha$ and $\partial x_2^*(\alpha)/\partial\alpha$, totally differentiate *both* of these identities respect to α, to get

$$\frac{\partial^2 f\left(x_1^*(\alpha), x_2^*(\alpha);\alpha\right)}{\partial x_1^2}\cdot\frac{\partial x_1^*(\alpha)}{\partial\alpha} + \frac{\partial^2 f\left(x_1^*(\alpha), x_2^*(\alpha);\alpha\right)}{\partial x_1\,\partial x_2}\cdot\frac{\partial x_2^*(\alpha)}{\partial\alpha} + \frac{\partial^2 f\left(x_1^*(\alpha), x_2^*(\alpha);\alpha\right)}{\partial x_1\,\partial\alpha} \underset{\alpha}{\equiv} 0$$

$$\frac{\partial^2 f\left(x_1^*(\alpha), x_2^*(\alpha);\alpha\right)}{\partial x_1\,\partial x_2}\cdot\frac{\partial x_1^*(\alpha)}{\partial\alpha} + \frac{\partial^2 f\left(x_1^*(\alpha), x_2^*(\alpha);\alpha\right)}{\partial x_2^2}\cdot\frac{\partial x_2^*(\alpha)}{\partial\alpha} + \frac{\partial^2 f\left(x_1^*(\alpha), x_2^*(\alpha);\alpha\right)}{\partial x_2\,\partial\alpha} \underset{\alpha}{\equiv} 0$$

This is a set of two linear equations in the two derivatives $\partial x_1^*(\alpha)/\partial\alpha$ and $\partial x_2^*(\alpha)/\partial\alpha$, and we can solve for $\partial x_1^*(\alpha)/\partial\alpha$ and $\partial x_2^*(\alpha)/\partial\alpha$ by substitution, or by Cramer's Rule, or however.

Comparative Statics of Equilibria

Implicit differentiation isn't restricted to optimization problems. It also allows us to derive how changes in the parameters affect the *equilibrium values* in an economic system.

Consider a simple market system, with supply and demand and supply functions

$$Q^D = D(P,I) \qquad \text{and} \qquad Q^S = S(P,w)$$

where P is market price, and the parameters are income I and the wage rate w. Naturally, the *equilibrium price* is the value P^e solves the equilibrium condition

$$D(P^e,I) = S(P^e,w)$$

It is clear that the *equilibrium price function*, namely $P^e = P^e(I,w)$, must satisfy the identity

$$D(\,P^e(I,w)\,,I\,) \underset{I,w}{\equiv} S(\,P^e(I,w)\,,w\,)$$

So if we want to determine how a rise in income affects equilibrium price, totally differentiate the above identity with respect to I, to get

$$\frac{\partial D\left(P^e(I,w),I\right)}{\partial P^e}\cdot\frac{\partial P^e(I,w)}{\partial I} + \frac{\partial D\left(P^e(I,w),I\right)}{\partial I} \underset{w,I}{\equiv} \frac{\partial S\left(P^e(I,w),w\right)}{\partial P^e}\cdot\frac{\partial P^e(I,w)}{\partial I}$$

then solve to get

$$\frac{\partial P^e(I,w)}{\partial I} = \frac{\frac{\partial D(P^e(I,w),I)}{\partial I}}{\frac{\partial S(P^e(I,w),w)}{\partial P^e} - \frac{\partial D(P^e(I,w),I)}{\partial P^e}}$$

In class, we'll analyze this formula to see what it implies about the effect of changes in income upon equilibrium price in a market. For practice, see if you can derive the formula for the effect of changes in the wage rate upon the equilibrium price.

Summary of the Use of Implicit Differentiation to Obtain Comparative Statics Results

The approach of implicit differentiation is straightforward, yet robust and powerful. It is used extensively in economic analysis, and always consists of the following four steps:

STEP 1: *Obtain the first order conditions for the optimization problem, or the equilibrium conditions of the system.*

STEP 2: *Convert these conditions to identities in the parameters, by substituting in the solution functions.*

STEP 3: *Totally differentiate these identities with respect to the parameter that is changing.*

STEP 4: *Solve for the derivatives with respect to that parameter.*

G. OPTIMIZATION #3: COMPARATIVE STATICS OF OPTIMAL VALUES

The final question we can ask is how the optimal attainable value of the objective function varies when we change the parameters. This has a surprising aspect to it. In the unconstrained maximization problem:

$$\max_{x_1,\ldots,x_n} f(x_1,\ldots,x_n;\alpha)$$

recall that we get the optimal value function $\phi(\alpha)$ by substituting the solutions $x_1^*(\alpha),\ldots,x_n^*(\alpha)$ back into the objective function, i.e.:

$$\phi(\alpha) \underset{\alpha}{\equiv} f(x_1^*(\alpha),\ldots,x_n^*(\alpha);\alpha)$$

Thus, we could simply differentiate with respect to α to get:

$$\begin{aligned}\frac{d\phi(\alpha)}{d\alpha} &= \frac{\partial f(x_1^*(\alpha),\ldots,x_n^*(\alpha);\alpha)}{\partial x_1}\cdot\frac{dx_1^*(\alpha)}{d\alpha}\\ &\vdots\\ &+ \frac{\partial f(x_1^*(\alpha),\ldots,x_n^*(\alpha);\alpha)}{\partial x_n}\cdot\frac{dx_n^*(\alpha)}{d\alpha}\\ &+ \frac{\partial f(x_1^*(\alpha),\ldots,x_n^*(\alpha);\alpha)}{\partial \alpha}\end{aligned}$$

where the last term is obviously the direct effect of α upon the objective function. The first n terms are there because a change in α affects the optimal x_i values, which in turn affect the objective function. All in all, this derivative is a big mess.

However, if we recall the *first order conditions* to this problem, we see that since $\partial f/\partial x_1 = ... = \partial f/\partial x_n = 0$ at the optimum, *all of these first n terms are zero*, so that we just get:

$$\frac{d\phi(\alpha)}{d\alpha} = \frac{\partial f(x_1^*(\alpha),...,x_n^*(\alpha);\alpha)}{\partial\alpha}$$

This means that when we evaluate how the optimal value function is affected when we change a parameter, we only have to consider that parameter's direct affect on the objective function, and can *ignore* the indirect effects caused by the resulting changes in the optimal values of the control variables. If we keep this in mind, we can save a lot of time.

This also works for constrained maximization problems. Consider the problem

$$\max_{x_1,...,x_n} f(x_1,...,x_n;\alpha) \quad \text{subject to} \quad g(x_1,...,x_n;\alpha) = c$$

Once again, we get the optimal value function by plugging the optimal values of the control variables (namely $x_1^*(\alpha),...,x_n^*(\alpha)$) into the objective function:

$$\phi(\alpha) \underset{\alpha}{\equiv} f(x_1^*(\alpha),...,x_n^*(\alpha);\alpha)$$

Note that since these values must also satisfy the constraint, we also have:

$$c - g(x_1^*(\alpha),...,x_n^*(\alpha);\alpha) \underset{\alpha}{\equiv} 0$$

so we can multiply by $\lambda(\alpha)$ and add to the previous equation to get:

$$\phi(\alpha) \equiv f(x_1^*(\alpha),...,x_n^*(\alpha);\alpha) + \lambda(\alpha)\cdot[c - g(x_1^*(\alpha),...,x_n^*(\alpha);\alpha)]$$

which is the same as if we had plugged the optimal values $x_1^*(\alpha),...,x_n^*(\alpha)$ and $\lambda^*(\alpha)$ directly into the Lagrangian formula, or in other words:

$$\phi(\alpha) \equiv \mathcal{L}(x_1^*(\alpha),...,x_n^*(\alpha),\lambda^*(\alpha);\alpha) \equiv f(x_1^*(\alpha),...,x_n^*(\alpha);\alpha) + \lambda(\alpha)\cdot[c - g(x_1^*(\alpha),...,x_n^*(\alpha);\alpha)]$$

Now if we differentiate the above identity with respect to α, we get:

$$\begin{aligned}\frac{d\phi(\alpha)}{d\alpha} &= \frac{\partial\mathcal{L}(x_1^*(\alpha),...,x_n^*(\alpha),\lambda^*(\alpha);\alpha)}{\partial x_1}\cdot\frac{dx_1^*(\alpha)}{d\alpha} \\ &\vdots \\ &+ \frac{\partial\mathcal{L}(x_1^*(\alpha),...,x_n^*(\alpha),\lambda^*(\alpha);\alpha)}{\partial x_n}\cdot\frac{dx_n^*(\alpha)}{d\alpha} \\ &+ \frac{\partial\mathcal{L}(x_1^*(\alpha),...,x_n^*(\alpha),\lambda^*(\alpha);\alpha)}{\partial\lambda}\cdot\frac{d\lambda^*(\alpha)}{d\alpha} \\ &+ \frac{\partial\mathcal{L}(x_1^*(\alpha),...,x_n^*(\alpha),\lambda^*(\alpha);\alpha)}{\partial\alpha}\end{aligned}$$

But once again, since the first order conditions for the constrained maximization problem are $\partial\mathcal{L}/\partial x_1 = \cdots = \partial\mathcal{L}/\partial x_n = \partial\mathcal{L}/\partial\lambda = 0$, all but the last of these right hand terms are zero, so we get:

$$\frac{d\phi(\alpha)}{d\alpha} \quad = \quad \frac{\partial\mathcal{L}(x_1^*(\alpha),\ldots,x_n^*(\alpha),\lambda^*(\alpha);\alpha)}{\partial\alpha}$$

In other words, we only have to take into account the direct effect of α on the Lagrangian function, and can ignore the indirect effects due to changes in the optimal values of the x_i's and λ. A very helpful thing to know.

H. DETERMINANTS, SYSTEMS OF LINEAR EQUATIONS & CRAMER'S RULE

The Determinant of a Matrix

In order to solve systems of linear equations we need to define the **determinant** $|\mathbf{A}|$ of a square matrix $\mathbf{A}$. If $\mathbf{A}$ is a 1 × 1 matrix, that is, if $\mathbf{A} = [a_{11}]$, we define $|\mathbf{A}| = a_{11}$.

In the 2 × 2 case: if $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ we define $|\mathbf{A}| \;=\; a_{11}\cdot a_{22} - a_{12}\cdot a_{21}$

that is, the product along the downward sloping diagonal ($a_{11}\cdot a_{22}$), minus the product along the upward sloping diagonal ($a_{12}\cdot a_{21}$).

In the 3 × 3 case: if $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ then first form $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix}$

(i.e., recopy the first two columns). Then we define:

$$|\mathbf{A}| \;=\; a_{11}\cdot a_{22}\cdot a_{33} + a_{12}\cdot a_{23}\cdot a_{31} + a_{13}\cdot a_{21}\cdot a_{32} - a_{13}\cdot a_{22}\cdot a_{31} - a_{11}\cdot a_{23}\cdot a_{32} - a_{12}\cdot a_{21}\cdot a_{33}$$

in other words, add the products of all three downward sloping diagonals and subtract the products of all three upward sloping diagonals.

Unfortunately, this technique doesn't work for 4×4 or bigger matrices, so to hell with them.

Systems of Linear Equations and Cramer's Rule

The general form of a **system of *n* linear equations** in the n unknown variables $x_1,\ldots,x_n$ is:

$$\begin{matrix} a_{11}\cdot x_1 & + & a_{12}\cdot x_2 & + & \cdots & + & a_{1n}\cdot x_n & = & c_1 \\ a_{21}\cdot x_1 & + & a_{22}\cdot x_2 & + & \cdots & + & a_{2n}\cdot x_n & = & c_2 \\ & & & & \vdots & & & & \\ a_{n1}\cdot x_1 & + & a_{n2}\cdot x_2 & + & \cdots & + & a_{nn}\cdot x_n & = & c_n \end{matrix}$$

for some matrix of coefficients $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$ and vector of constants $\mathbf{C} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$

Note that the first subscript in the coefficient a_{ij} refers to its row and the second subscript refers to its column (thus, a_{ij} is the coefficient of x_j in the i'th equation).

We now give **Cramer's Rule** for solving linear systems. The solutions to the 2 × 2 linear system:

$$\begin{aligned} a_{11}\cdot x_1 + a_{12}\cdot x_2 &= c_1 \\ a_{21}\cdot x_1 + a_{22}\cdot x_2 &= c_2 \end{aligned}$$

are simply:

$$x_1^* = \frac{\begin{vmatrix} c_1 & a_{12} \\ c_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \quad \text{and} \quad x_2^* = \frac{\begin{vmatrix} a_{11} & c_1 \\ a_{21} & c_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}$$

The solutions to the 3 × 3 system:

$$\begin{aligned} a_{11}\cdot x_1 + a_{12}\cdot x_2 + a_{13}\cdot x_3 &= c_1 \\ a_{21}\cdot x_1 + a_{22}\cdot x_2 + a_{23}\cdot x_3 &= c_2 \\ a_{31}\cdot x_1 + a_{32}\cdot x_2 + a_{33}\cdot x_3 &= c_3 \end{aligned}$$

are simply:

$$x_1^* = \frac{\begin{vmatrix} c_1 & a_{12} & a_{13} \\ c_2 & a_{22} & a_{23} \\ c_3 & a_{32} & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}} \qquad x_2^* = \frac{\begin{vmatrix} a_{11} & c_1 & a_{13} \\ a_{21} & c_2 & a_{23} \\ a_{31} & c_3 & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}} \qquad x_3^* = \frac{\begin{vmatrix} a_{11} & a_{12} & c_1 \\ a_{21} & a_{22} & c_2 \\ a_{31} & a_{32} & c_3 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}$$

Note that in both the 2 × 2 and the 3 × 3 case we have that x_i^* is obtained as the ratio of two determinants. The denominator is always the determinant of the coefficient matrix **A**. The numerator is the determinant of a matrix which is just like the coefficient matrix, except that the j'th column has been replaced by the vector of right hand side constants.