ECONOMICS 200A MATHEMATICAL HANDOUT

Fall 2020 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:

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

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:

$$\frac{\partial f(x_1,x_2)}{\partial x_1} \;=\; f_1(x_1,x_2) \;=\; 2x_1 \cdot x_2 + 3 \qquad \text{and} \qquad \frac{\partial f(x_1,x_2)}{\partial x_2} \;=\; f_2(x_1,x_2) \;=\; x_1^2$$

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 $f(x) \equiv g(h(x))$ we have

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

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:

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

[1] If the material in this section is not *already* familiar to you, you probably won't be able to pass this course.

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

A **second partial derivative** of a function of several variables is analogous, i.e., we write:

$$f_{ii}(x_1,\ldots,x_n) \quad \text{or} \quad \frac{\partial^2 f(x_1,\ldots,x_n)}{\partial x_i^2}$$

to denote differentiating twice with respect to x_i.

We get a **cross partial derivative** when we differentiate first with respect to x_i and then with respect to some other variable x_j. We will denote this with the expressions:

$$f_{ij}(x_1,\ldots,x_n) \quad \text{or} \quad \frac{\partial^2 f(x_1,\ldots,x_n)}{\partial x_i \partial x_j}$$

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

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

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

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

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

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:

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

or in equivalent notation: $df = f_1(x_1,\ldots,x_n)\cdot dx_1 + \cdots + f_n(x_1,\ldots,x_n)\cdot dx_n$

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:

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

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:

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

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.

It is crucial to note that while 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 is exactly proportional to x, i.e., if we have $y = c\cdot x$, the *slope* of this curve would be c, but its *elasticity* is given by:

$$E_{y,x} = \left(\frac{dy}{dx}\right)\cdot\left(\frac{x}{y}\right) = c\cdot\left(\frac{x}{c\cdot x}\right) = 1$$

In other words, whenever y is exactly proportional to x, the elasticity of y with respect to x will be one, *regardless* of the value of the coefficient c.

Here's another example to show that elasticity is not the same as slope. The function $y = 3 + 4x$ obviously has a constant slope (namely 4). But it does *not* have a constant elasticity. To see this, use the formula again:

$$E_{y,x} \;=\; \left(\frac{dy}{dx}\right)\cdot\left(\frac{x}{y}\right) \;=\; 4\cdot\left(\frac{x}{3+4x}\right) \;=\; \left(\frac{4x}{3+4x}\right)$$

which is obviously not constant as x varies.

Finally, we note that if a function has a **constant elasticity**, it must take the form $f(x) \equiv c{\cdot}x^{\beta}$ for some constants $c > 0$ and β. We prove this by the calculation:

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

Note that β will be positive if the function is increasing, and negative if it is decreasing.

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 given by $f(x_1,x_2) = k$. 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:

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

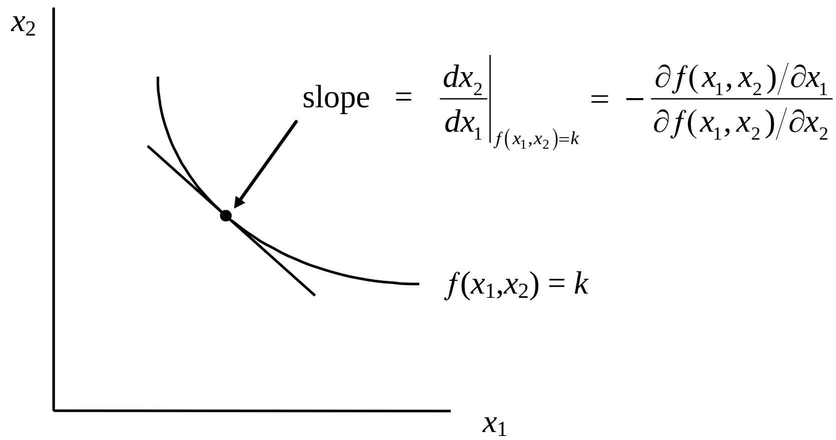
which implies that Δx_1 and Δx_2 will accordingly satisfy:

$$\left.\frac{\Delta x_2}{\Delta x_1}\right|_{f(x_1,x_2)=k} \;\approx\; -\frac{f_1(x_1,x_2)}{f_2(x_1,x_2)}$$

so that in the limit we have:

$$\left.\frac{dx_2}{dx_1}\right|_{f(x_1,x_2)=k} \;=\; -\frac{f_1(x_1,x_2)}{f_2(x_1,x_2)}$$

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:

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

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

D. POSSIBLE PROPERTIES OF FUNCTIONS

Cardinal vs. Ordinal Properties of a Function

Consider some domain $\mathcal{X} \subseteq \Re^n$. For convenience, we can express each point in $\mathcal{X}$ by the boldface symbol $\boldsymbol{x} = (x_1,...,x_n)$, $\boldsymbol{x}^* = (x_1^*,...,x_n^*)$, etc. *Every* real-valued function $f(\boldsymbol{x}) = f(x_1,...,x_n)$ over $\mathcal{X}$ implies (among other things) the following two types of "information:"

ordinal information: how the function "ranks" different points in its domain, i.e., the ranking $\succcurlyeq_{ord}$ over points in $\mathcal{X}$, as defined by

$$\boldsymbol{x}_A \succcurlyeq_{ord} \boldsymbol{x}_B \quad \Leftrightarrow \quad f(\boldsymbol{x}_A) \geq f(\boldsymbol{x}_B)$$

cardinal information: how the function ranks *differences* between pairs of points in the domain, the ranking $\succcurlyeq_{crd}$ over all *pairs of points* in $\mathcal{X}$, as defined by

$$\{\boldsymbol{x}_A,\boldsymbol{x}_B\} \succcurlyeq_{crd} \{\boldsymbol{x}_C,\boldsymbol{x}_D\} \quad \Leftrightarrow \quad f(\boldsymbol{x}_A) - f(\boldsymbol{x}_B) \geq f(\boldsymbol{x}_C) - f(\boldsymbol{x}_D)$$

It is easy to prove that the rankings $\succcurlyeq_{ord}$ and $\succcurlyeq_{crd}$ will always be transitive (can you do this?) and we define their associated relations $\preccurlyeq_{ord}$, $\sim_{ord}$, $\succ_{ord}$ and $\preccurlyeq_{crd}$, $\sim_{crd}$, $\succ_{crd}$ in the obvious manner.

Note that a pair of functions $f(\cdot)$ and $g(\cdot)$ over $\mathcal{X}$ will have the *same ordinal information as each other* if and only if they are linked by some **increasing transformation** of the form

$$g(\boldsymbol{x}) \equiv \phi(f(\boldsymbol{x})) \qquad \text{for some strictly increasing function } \phi(\cdot)$$

Similarly, a pair of functions $f(\cdot)$ and $g(\cdot)$ will have the *same cardinal information as each other* if and only if they are linked by some **increasing affine transformation** of the form

$$g(\boldsymbol{x}) \equiv a \cdot f(\boldsymbol{x}) + b \qquad \text{for some constants } a, b \text{ with } a > 0$$

Although it is important to recall that *every* function has both cardinal and ordinal information, we typically refer to a function as **ordinal** if only its ordinal information is relevant. An example is a standard utility function $U(x_1,...,x_n)$ over nonstochastic commodity bundles $(x_1,...,x_n)$, whose ordinal information is precisely the consumer's preference relation over such bundles. We refer to a function as **cardinal** if its cardinal information is relevant. An example of a cardinal function is a consumer's von Neumann-Morgenstern utility function which (under the expected utility hypothesis) represents their attitudes toward risk (if you have not yet seen this concept, don't worry.) The cardinal information of a function *includes* its ordinal information, since

$$\boldsymbol{x}^* \succcurlyeq_{ord} \boldsymbol{x} \;\Leftrightarrow\; f(\boldsymbol{x}^*) \geq f(\boldsymbol{x}) \;\Leftrightarrow\; f(\boldsymbol{x}^*) - f(\boldsymbol{x}) \geq f(\boldsymbol{x}) - f(\boldsymbol{x}) \;\Leftrightarrow\; \{\boldsymbol{x}^*, \boldsymbol{x}\} \succcurlyeq_{crd} \{\boldsymbol{x}, \boldsymbol{x}\}$$

But don't think that a function's cardinal information (the ranking $\succcurlyeq_{crd}$) exhausts the information of the function. For example, knowing all the cardinal information of a *production function* still leaves you with no idea of how much output it yields from any input bundle! (I.e., production functions are *more* than just cardinal functions.)

Scale Properties of a Function

The function $f(x_1,...,x_n)$ exhibits **constant returns to scale** or is **homogeneous of degree one**, if

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

i.e. if multiplying all arguments by λ implies the value of the function is also multiplied by λ.

A function $f(x_1,...,x_n)$ is said to exhibit **scale invariance** or is **homogeneous of degree zero** if

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

i.e., if multiplying all arguments by λ leads to *no change* in the value of the function.

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

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

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 $f(\lambda x_1,...,\lambda x_n) \equiv \lambda \cdot f(x_1,...,x_n)$ and this time differentiate with respect to x_i, to get:

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

or equivalently:

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

which establishes our result. In other words, if a production function exhibits constant returns to scale, the marginal products of all the factors will be scale invariant.

More generally, a function $f(x_1,...,x_n)$ is **homogeneous of degree *n*** if

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

i.e., if multiplying all arguments by λ implies that the value of the function is multiplied by λ^n. Another way to think of this property is that $f(x_1,...,x_n)$ has a constant "scale elasticity" of n. To check your understanding of arguments involving homogeneous functions, see if you can prove the following generalization of the previous paragraph's result:

"If $f(x_1,...,x_n)$ is homogeneous of degree k, then its partials are all homogeneous of degree $k-1$."

Concave and Convex Functions

A real-valued function $f(\boldsymbol{x}) = f(x_1,...,x_n)$ over a convex domain $\mathcal{X} \subseteq \Re^n$ is said to be **concave** if

$$f(\lambda\cdot\boldsymbol{x} + (1-\lambda)\cdot\boldsymbol{x}^*) \;\geq\; \lambda\cdot f(\boldsymbol{x}) + (1-\lambda)\cdot f(\boldsymbol{x}^*) \qquad \text{for all } \boldsymbol{x}, \boldsymbol{x}^*, \text{ and all } \lambda \in (0,1)$$

and **convex** if

$$f(\lambda\cdot\boldsymbol{x} + (1-\lambda)\cdot\boldsymbol{x}^*) \;\leq\; \lambda\cdot f(\boldsymbol{x}) + (1-\lambda)\cdot f(\boldsymbol{x}^*) \qquad \text{for all } \boldsymbol{x}, \boldsymbol{x}^*, \text{ and all } \lambda \in (0,1)$$

$f(\boldsymbol{x})$ is *strictly* concave/convex if the relevant inequality holds strictly for all $\boldsymbol{x}$, $\boldsymbol{x}^*$ and λ.[2]

If $f(x)$ is a concave/convex function of a single variable x, the above conditions imply that the chord linking any two points on its graph will lie everywhere on or below/above it. If $f(x)$ is twice differentiable, then it will be concave/convex if and only if its second derivative $f''(x)$ is everywhere nonpositive/nonnegative. However, the link with second derivatives is not so exact for *strict* concavity/convexity: the function $f(x) = x^4$ is twice differentiable and *strictly* convex over the entire real line, event though its second derivative $f''(x)$ is not positive at $x = 0$.

Finally, $f(\boldsymbol{x}) = f(x_1,...,x_n)$ is **affine** if it takes the form $f(x_1,...,x_n) \equiv \sum_{i=1}^{n} a_i \cdot x_i + b$ for some constants $a_1,...,a_n, b$. Affine functions are simultaneously weakly concave and weakly convex.

Quasiconcave and Quasiconvex Functions

A real-valued function $f(\boldsymbol{x}) = f(x_1,...,x_n)$ over a convex domain $\mathcal{X} \subseteq \Re^n$ is **quasiconcave** if

$$f(\lambda\cdot\boldsymbol{x} + (1-\lambda)\cdot\boldsymbol{x}^*) \;\geq\; f(\boldsymbol{x}) \qquad \text{for all } \boldsymbol{x}, \boldsymbol{x}^* \text{ such that } f(\boldsymbol{x}^*) = f(\boldsymbol{x}), \text{ and all } \lambda \in (0,1)$$

and **quasiconvex** if

$$f(\lambda\cdot\boldsymbol{x} + (1-\lambda)\cdot\boldsymbol{x}^*) \;\leq\; f(\boldsymbol{x}) \qquad \text{for all } \boldsymbol{x}, \boldsymbol{x}^* \text{ such that } f(\boldsymbol{x}^*) = f(\boldsymbol{x}), \text{ and all } \lambda \in (0,1)$$

We use the terms *strictly* quasiconcave/quasiconvex if the relevant inequality always holds strictly.

There are a couple other equivalent definitions of these concepts. Perhaps the most intuitive is

$f(\boldsymbol{x}) = f(x_1,...,x_n)$ is **quasiconcave** $\Leftrightarrow$ for every $\boldsymbol{x}^*$, the set $\{\boldsymbol{x} \,|\, f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ is a convex set $\}$

$f(\boldsymbol{x}) = f(x_1,...,x_n)$ is **quasiconvex** $\Leftrightarrow$ for every $\boldsymbol{x}^*$, the set $\{\boldsymbol{x} \,|\, f(\boldsymbol{x}) \leq f(\boldsymbol{x}^*)$ is a convex set $\}$

---

[2] Functions that are concave/convex, but not *strictly* so, are sometimes termed **weakly concave**/**weakly convex**.

Alternatively, $f(\boldsymbol{x}) = f(x_1,...,x_n)$ is

$$\textbf{quasiconcave} \;\Leftrightarrow\; \text{for all } \boldsymbol{x}, \boldsymbol{x}^*,\; f(\lambda\cdot\boldsymbol{x}+(1-\lambda)\cdot\boldsymbol{x}^*) \geq \min[f(\boldsymbol{x}^*),f(\boldsymbol{x})] \text{ for all } \lambda\in(0,1)$$

$$\textbf{quasiconvex} \;\Leftrightarrow\; \text{for all } \boldsymbol{x}, \boldsymbol{x}^*,\; f(\lambda\cdot\boldsymbol{x}+(1-\lambda)\cdot\boldsymbol{x}^*) \leq \max[f(\boldsymbol{x}^*),f(\boldsymbol{x})] \text{ for all } \lambda\in(0,1)$$

The fundamental difference between concavity/convexity and quasiconcavity/quasiconvexity is:

Quasiconcavity/quasiconvexity are *ordinal* concepts, in that they solely pertain to properties of a function's implied ranking $\succcurlyeq_{ord}$:

$$\textbf{quasiconcave} \;\Leftrightarrow\; \text{for all } \boldsymbol{x}^*, \text{ the set } \{\boldsymbol{x} \mid \boldsymbol{x} \succcurlyeq_{ord} \boldsymbol{x}^* \text{ is a convex set }\}$$

$$\textbf{quasiconvex} \;\Leftrightarrow\; \text{for all } \boldsymbol{x}^*, \text{ the set } \{\boldsymbol{x} \mid \boldsymbol{x} \preccurlyeq_{ord} \boldsymbol{x}^* \text{ is a convex set }\}$$

Thus if a function is quasiconcave/quasiconvex, so is *every increasing transformation* of it.

Concavity/convexity are *cardinal* concepts, in that they solely pertain to properties of the implied ranking $\succcurlyeq_{crd}$. On the assumption that the function is continuous, they can be shown to be equivalent to the condition that whenever $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x})$, then $f(\boldsymbol{x}^*) - f(\tfrac{1}{2}\cdot\boldsymbol{x}^*+\tfrac{1}{2}\cdot\boldsymbol{x}) \leq/\geq f(\tfrac{1}{2}\cdot\boldsymbol{x}^*+\tfrac{1}{2}\cdot\boldsymbol{x}) - f(\boldsymbol{x})$. This property can be expressed in terms of the ranking $\succcurlyeq_{crd}$ as:

$$\textbf{concave} \;\Leftrightarrow\; \text{if } \{\boldsymbol{x}^*,\boldsymbol{x}\} \succcurlyeq_{crd} \{\boldsymbol{x},\boldsymbol{x}\} \text{ then } \{\boldsymbol{x}^*, \tfrac{1}{2}\cdot\boldsymbol{x}^*+\tfrac{1}{2}\cdot\boldsymbol{x}\} \succcurlyeq_{crd} \{\tfrac{1}{2}\cdot\boldsymbol{x}^*+\tfrac{1}{2}\cdot\boldsymbol{x}, \boldsymbol{x}\}$$

$$\textbf{convex} \;\Leftrightarrow\; \text{if } \{\boldsymbol{x}^*,\boldsymbol{x}\} \succcurlyeq_{crd} \{\boldsymbol{x},\boldsymbol{x}\} \text{ then } \{\boldsymbol{x}^*, \tfrac{1}{2}\cdot\boldsymbol{x}^*+\tfrac{1}{2}\cdot\boldsymbol{x}\} \preccurlyeq_{crd} \{\tfrac{1}{2}\cdot\boldsymbol{x}^*+\tfrac{1}{2}\cdot\boldsymbol{x}, \boldsymbol{x}\}$$

Thus, if a function is concave/convex, so is every *increasing affine transformation* of it.

Note that every concave/convex function is necessarily quasiconcave/quasiconvex, but not vice versa: think about the function $f(x_1,x_2) \equiv x_1\cdot x_2$ over $\mathcal{X} = \{(x_1,x_2) \mid x_1 \geq 0, x_2 \geq 0\}$.

E. 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:

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

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 present **Cramer's Rule** for solving systems of linear equations. The solutions x_1^* and x_2^* 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:

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

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

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.

F. 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 will be a lot easier if we consider 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:

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

Although we will often have many parameters $\alpha_1,...,\alpha_m$ in a given problem, for simplicity we shall assume from now on that there is only a single parameter α. All of our results will apply to the many-parameter case, however.

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 one or more parameters, our problem becomes:

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

$$\text{subject to} \quad g(x_1,...,x_n;\beta) = 0$$

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:

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

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

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

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 $f(x_1,\ldots,x_n)$ with respect to each of the x_i's 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 $f(x_1, \ldots,x_n)$ be a weakly concave function of $(x_1,\ldots,x_n)$ (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 $f(x_1,\ldots,x_n)$ be a weakly convex function of $(x_1,\ldots,x_n)$ (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*).

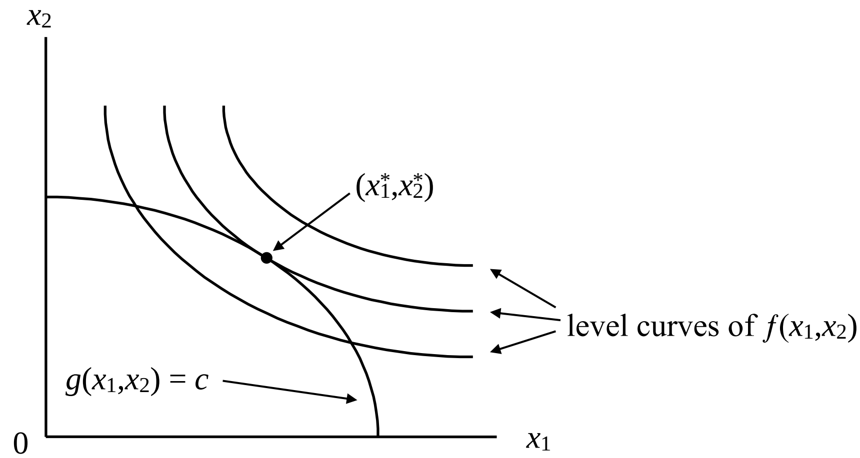
First Order Conditions for Constrained Optimization Problems (VERY important)

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

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

$$\text{subject to} \quad g(x_1,x_2) = c$$

are easy to see from the following diagram



The point (x_1^*,x_2^*) in the diagram 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:

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

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 $(x_1^*,\ldots,x_n^*)$ to be a solution to the constrained maximization problem:

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

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

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 $g(x_1,\ldots,x_n)$ must be the same as the tradeoff rate between x_i and x_j that preserves the value of $f(x_1,\ldots,x_n)$*. We thus have the condition:

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

or in other words, that:

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

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

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

To summarize: the first order conditions for the many-variable 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 $(x_1^*,...,x_n^*)$ satisfy the equations:

$$f_1(x_1^*,...,x_n^*) \;=\; \lambda\cdot g_1(x_1^*,...,x_n^*)$$
$$f_2(x_1^*,...,x_n^*) \;=\; \lambda\cdot g_2(x_1^*,...,x_n^*)$$
$$\vdots$$
$$f_n(x_1^*,...,x_n^*) \;=\; \lambda\cdot g_n(x_1^*,...,x_n^*)$$

and the constraint:

$$g(x_1^*,...,x_n^*) \;=\; c$$

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

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

and also that the constraint $g(x_1^*,...,x_n^*) = c$ 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 $x_1,...,x_n$ and λ. Personally, I suggest that you get these first order conditions the direct way by simply setting the normal vector of $f(x_1,...,x_n)$ to equal a scalar multiple of the normal vector of $g(x_1,...,x_n)$ (with the scale factor λ). However, another way to obtain these equations is to construct the **Lagrangian function**:

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

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

$$\partial\mathcal{L}(x_1^*,...,x_n^*,\lambda)/\partial x_1 \;=\; f_1(x_1^*,...,x_n^*) - \lambda\cdot g_1(x_1^*,...,x_n^*) \;=\; 0$$
$$\vdots$$
$$\partial\mathcal{L}(x_1^*,...,x_n^*,\lambda)/\partial x_n \;=\; f_n(x_1^*,...,x_n^*) - \lambda\cdot g_n(x_1^*,...,x_n^*) \;=\; 0$$
$$\partial\mathcal{L}(x_1^*,...,x_n^*,\lambda)/\partial\lambda \;=\; c - g(x_1^*,...,x_n^*) \;=\; 0$$

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 $f(x_1,...,x_n)$ be λ times the normal vector of $g(x_1,...,x_n)$, and the constraint $g(x_1,...,x_n) = c$ be satisfied.

See the standard texts for second order conditions for constrained optimization.

G. COMPARATIVE STATICS OF SOLUTION FUNCTIONS: IMPLICIT DIFFERENTIATION

Having arrived at 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). The easiest way to do this would be to actually solve the first order conditions for the solutions, that is, solve for the functions:

$$x_1^* = x_1^*(\alpha) \qquad x_2^* = x_2^*(\alpha) \qquad . . . \qquad x_n^* = x_n^*(\alpha)$$

then just differentiate these functions with respect to the parameter α to obtain the derivatives:

$$\frac{dx_1^*(\alpha)}{d\alpha} \quad \frac{dx_2^*(\alpha)}{d\alpha} \quad \cdots \quad \frac{dx_n^*(\alpha)}{d\alpha}$$

However, first order conditions are usually much too complicated to solve. Are we up a creek? No: there is another approach known as **implicit differentiation**, which always works.

The idea behind implicit differentiation is very simple. Since the optimal values:

$$x_1^* = x_1^*(\alpha) \qquad x_2^* = x_2^*(\alpha) \qquad . . . \qquad x_n^* = x_n^*(\alpha)$$

come from the first order conditions, it stands to reason that the *derivatives* of these optimal values will come from the *derivatives* of the first order conditions, where in each case, we are talking about derivatives with respect to the parameter α.

Let's do it for the simplest of cases. Consider the unconstrained maximization problem:

$$\max_x \ \phi(x,\alpha)$$

with the single control variable x, the single parameter α, and where $\partial^2\phi(x,\alpha)/\partial x^2 < 0$. We know that the solution $x^* = x^*(\alpha)$ solves the first order condition:

$$\phi_x(x^*,\alpha) \ \equiv \ 0$$

If we want to find $\partial x^*/\partial\alpha$, we would simply differentiate this first order condition with respect to the parameter α. Before doing this, it's useful to remind ourselves of exactly which variables will be affected by a change in α, or in other words, those variables which depend upon α. This can be done by drawing arrows over these variables, or alternatively, by representing all functional dependence explicitly, that is:

$$\phi_x(\overset{\downarrow}{x^*},\overset{\downarrow}{\alpha}) \ = \ 0 \qquad \text{or alternatively} \qquad \phi_x(x^*(\alpha),\alpha) \ = \ 0$$

Differentiating with respect to α yields:

$$\phi_{xx}(x^*(\alpha),\alpha)\cdot\frac{d\,x^*(\alpha)}{d\alpha} \ + \ \phi_{x\alpha}(x^*(\alpha),\alpha) \ \underset{\alpha}{\equiv} \ 0$$

Note that we get one term for each arrow we have drawn on the left (or each appearance of α on the right). Rearranging and solving for the derivative $\partial x^*(\alpha)/\partial\alpha$, we obtain:

$$\frac{d\,x^*(\alpha)}{d\alpha} \underset{\alpha}{\equiv} -\frac{\phi_{x\alpha}(x^*(\alpha),\alpha)}{\phi_{xx}(x^*(\alpha),\alpha)} \underset{\alpha}{\equiv} -\frac{\partial^2\phi(x^*(\alpha),\alpha)/\partial x\,\partial\alpha}{\partial^2\phi(x^*(\alpha),\alpha)/\partial x^2}$$

Let's take another example, this time with two control variables and two parameters. For the general unconstrained maximization problem:

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

the solutions $x_1^* = x_1^*(\alpha,\beta)$ and $x_2^* = x_2^*(\alpha,\beta)$ solve the first order conditions:

$$f_1(x_1^*,x_2^*;\alpha,\beta) = 0$$

$$f_2(x_1^*,x_2^*;\alpha,\beta) = 0$$

If, say, we wanted to find the derivative $\partial x_1^*/\partial\alpha$, we would differentiate these first order conditions with respect to the parameter α. Before doing this, use arrows to identify the variables that will be affected by a change in α. In this example, they are the solutions x_1^* and x_2^*, as well as α itself (the parameter β again remains fixed). We thus get:

$$f_1(\overset{\downarrow}{x_1^*},\overset{\downarrow}{x_2^*};\overset{\downarrow}{\alpha};\beta) \;=\; 0$$

$$f_2(\overset{\downarrow}{x_1^*},\overset{\downarrow}{x_2^*};\overset{\downarrow}{\alpha};\beta) \;=\; 0$$

Differentiating these two equations with respect to α gives us:

$$f_{11}(x_1^*,x_2^*;\alpha,\beta)\cdot\frac{\partial x_1^*}{\partial\alpha} \;+\; f_{12}(x_1^*,x_2^*;\alpha,\beta)\cdot\frac{\partial x_2^*}{\partial\alpha} \;+\; f_{13}(x_1^*,x_2^*;\alpha,\beta) \;=\; 0$$

$$f_{21}(x_1^*,x_2^*;\alpha,\beta)\cdot\frac{\partial x_1^*}{\partial\alpha} \;+\; f_{22}(x_1^*,x_2^*;\alpha,\beta)\cdot\frac{\partial x_2^*}{\partial\alpha} \;+\; f_{23}(x_1^*,x_2^*;\alpha,\beta) \;=\; 0$$

or equivalently:

$$f_{11}(x_1^*,x_2^*;\alpha,\beta)\cdot\frac{\partial x_1^*}{\partial\alpha} \;+\; f_{12}(x_1^*,x_2^*;\alpha,\beta)\cdot\frac{\partial x_2^*}{\partial\alpha} \;=\; -f_{13}(x_1^*,x_2^*;\alpha,\beta)$$

$$f_{21}(x_1^*,x_2^*;\alpha,\beta)\cdot\frac{\partial x_1^*}{\partial\alpha} \;+\; f_{22}(x_1^*,x_2^*;\alpha,\beta)\cdot\frac{\partial x_2^*}{\partial\alpha} \;=\; -f_{23}(x_1^*,x_2^*;\alpha,\beta)$$

This is a set of two linear equations in the two terms $\partial x_1^*/\partial\alpha$ and $\partial x_2^*/\partial\alpha$, and hence can be solved by substitution, or by Cramer's Rule (see below), to obtain:

$$\frac{\partial x_1^*}{\partial\alpha} = \frac{f_{23}(x_1^*,x_2^*;\alpha,\beta)\cdot f_{12}(x_1^*,x_2^*;\alpha,\beta) \;-\; f_{13}(x_1^*,x_2^*;\alpha,\beta)\cdot f_{22}(x_1^*,x_2^*;\alpha,\beta)}{f_{11}(x_1^*,x_2^*;\alpha,\beta)\cdot f_{22}(x_1^*,x_2^*;\alpha,\beta) \;-\; f_{12}(x_1^*,x_2^*;\alpha,\beta)\cdot f_{21}(x_1^*,x_2^*;\alpha,\beta)}$$

and

$$\frac{\partial x_2^*}{\partial\alpha} = \frac{f_{21}(x_1^*,x_2^*;\alpha,\beta)\cdot f_{13}(x_1^*,x_2^*;\alpha,\beta) \;-\; f_{11}(x_1^*,x_2^*;\alpha,\beta)\cdot f_{23}(x_1^*,x_2^*;\alpha,\beta)}{f_{11}(x_1^*,x_2^*;\alpha,\beta)\cdot f_{22}(x_1^*,x_2^*;\alpha,\beta) \;-\; f_{12}(x_1^*,x_2^*;\alpha,\beta)\cdot f_{21}(x_1^*,x_2^*;\alpha,\beta)}$$

We could do this same procedure in a constrained maximization problem as well, provided we remember to differentiate all n+1 of the first order conditions, including the last one (the constraint equation).

This procedure will *always* work in getting the derivatives of the optimal values of the control variables with respect to any parameter. To use it, just remember that it consists of three steps:

> **STEP 1**: *Derive the first order condition(s) for the maximization or minimization problem.* As usual, the exact form of the first order conditions depends on the number of control variables in the problem, and on whether it is an unconstrained or constrained optimization problem.
>
> **STEP 2**: *Differentiate the first order condition(s) with respect to the parameter that is changing.* Before doing this, it's always useful to draw arrows over those variables that will change as a result of the change in the parameter. These will generally consist of all of the control variables, the Lagrangian multiplier λ if we are working with a constrained optimization problem, and the parameter itself.
>
> **STEP 3**: *Solve for the desired derivative(s).* If there is a single control variable, this consists of solving for the single derivative (say) $\partial x^*/\partial\alpha$. If there is more than one control variable, this will involve solving a system of linear equations for the derivatives $\partial x_1^*/\partial\alpha$, $\partial x_2^*/\partial\alpha$, etc.

H. COMPARATIVE STATICS OF OPTIMAL VALUE FUNCTIONS: THE ENVELOPE THEOREM

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:

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

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

$$\phi(\alpha) \;\equiv\; f(x_1(\alpha),\dots,x_n(\alpha);\alpha)$$

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

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

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 $\partial f/\partial x_1 = \dots = \partial f/\partial x_n = 0$ at the optimum, *all of these first n terms are zero*, so that we just get:

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

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 effect 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.

The envelope theorem also applies to constrained maximization problems. Consider the problem

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

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

$$\phi(\alpha) \equiv f(x_1(\alpha),\ldots,x_n(\alpha);\alpha)$$

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

$$c - g(x_1(\alpha),\ldots,x_n(\alpha);\alpha) \equiv 0$$

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

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

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

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

Remember, even though it involves λ and the constraint, this equation is still the formula for the optimal value function (i.e. the highest attainable value of $f(x_1,\ldots,x_n;\alpha)$ subject to the constraint).

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

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

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

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

That is, 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.