Math Camp Notes: Differentiation

Definition of Derivative

We say that a function f is differentiable at x if

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

If f is differentiable for all x in its domain, then we say f is a differentiable function.

Example:

Find the derivative of $f(x) = x^2$.

$$f'(x) = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} f'(x) = \lim_{h \to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} = \lim_{h \to 0} \frac{2xh}{h} = 2x$$

Derivative Rules

Assume k is an arbitrary constant, and that the two functions $u(x)$ and $v(x)$ are differentiable. Then the following rules hold:

1. Addition Rule
$$\frac{d}{dx}\left(u(x) + v(x)\right) = \frac{du}{dx} + \frac{dv}{dx}$$
2. Multiplicative Constant Rule
$$\frac{d}{dx}\left(k \cdot u(x)\right) = k \cdot \frac{du}{dx}$$
3. Product Rule
$$\frac{d}{dx}\left(u(x) \cdot v(x)\right) = \frac{du}{dx} \cdot v(x) + \frac{dv}{dx} \cdot u(x)$$
4. Quotient Rule
$$\frac{d}{dx}\left(\frac{u(x)}{v(x)}\right) = \frac{\frac{du}{dx} \cdot v(x) - \frac{dv}{dx} \cdot u(x)}{v(x)^2}$$
5. Power Rule
$$\frac{d}{dx}\left(x^k\right) = k \cdot x^{k-1}$$

Leibnitz's Rule

In order to differentiate under a definite integral, we use Leibniz Rule,

$$\frac{\partial}{\partial z} \int_{a(z)}^{b(z)} f(x,z)dx = \int_{a(z)}^{b(z)} \frac{\partial f}{\partial z} dx + f[b(z), z]\frac{\partial b}{\partial z} - f[a(z), z]\frac{\partial a}{\partial z}$$

Example:

Find the derivative of $\int_0^1 x^2 y^2 dx$ with respect to y.

$$\frac{\partial}{\partial y} \int_0^1 x^2 y^2 dx = \int_0^1 2x^2 y dx$$

Example:
Find the derivative of $\int_{2y}^{3y} x^2 y^2 dx$ with respect to y.

$$\frac{\partial}{\partial y} \int_{2y}^{3y} x^2 y^2 dx = \int_{2y}^{3y} 2x^2 y dx + (3y)^2 y^2 \cdot 3 - (2y)^2 y^2 \cdot 2$$

Derivatives of a Single Function of Several Variables

Partial Derivative

Let f be a function of many variables. The partial derivative of f with respect to x, denoted $\frac{\partial f}{\partial x}$ or f_x, is the function obtained by differentiating f with repect to x and treating all other variables as constants. Rigorously, this is

$$f_x(x,y,z) = \lim_{h \to 0} \frac{f(x+h,y,z) - f(x,y,z)}{h}$$

Example:

Let $f(x,y) = x^2y^3$. Then

$$\frac{\partial f}{\partial x} = 2xy^3$$

Second order partial derivative

The second order partial derivative is just a partial derivative of a partial derivative. The derivative with respect to x_i of the derivative with respect to x_j of the function f, or $\frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_j})$, is denoted by $\frac{\partial^2 f}{\partial x_i \partial x_j}$.

Example:
Let $f(x,y) = x^2y^3$. Then

$$\frac{\partial^2 f}{\partial x^2} = 2y^3$$
$$\frac{\partial^2 f}{\partial y^2} = 6x^2y$$
$$\frac{\partial^2 f}{\partial x \partial y} = 6xy^2$$
$$\frac{\partial^2 f}{\partial y \partial x} = 6xy^2$$

Notice that $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$.

The Hessian Matrix

The Hessian is merely a matrix of the second order partial derivatives of a function. Consider a function $f(x_1, \ldots, f_n)$. Its Hessian is given by.

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

Note that a Hessian matrix is always a square symmetric matrix.

Example:

Let $f(x,y) = x^2y^3$. Then the Hessian is

$$\begin{pmatrix} 2y^3 & 6xy^2 \\ 6xy^2 & 6x^2y \end{pmatrix}.$$

The Total Derivative or Gradient

The total derivative is merely a vector containing all of the partial derivatives of a function,

$$DF(x,y,z) = \left(\lim_{h\to 0}\frac{F(x+h,y,z)-f(x,y,z)}{h}, \lim_{j\to 0}\frac{F(x,y+j,z)-f(x,y,z)}{h}, \lim_{l\to 0}\frac{F(x,y,z+l)-f(x,y,z)}{h}\right)$$

Or, in different notation,

$$DF(x,y,z) = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z}\right)$$

The transpose of this vector, evaluated at a specific point $a = (x^*, y^*, z^*)$ is called the gradient of F at a, or $\nabla F(a)$.

Example:

Let $f(x,y) = x^2y^3$, and $a = (1,2)$. Then

$$\nabla f = \begin{pmatrix} 2xy^3 \\ 3x^2y^2 \end{pmatrix}$$

and

$$\nabla f(a) = \begin{pmatrix} 16 \\ 12 \end{pmatrix}$$

The Total Differential

The total differential of a function $F(x,y)$ at a point $a = (x^*, y^*)$ is

$$dF = \frac{\partial F}{\partial x}(x^*)dx + \frac{\partial F}{\partial y}(y^*)dy.$$

The interpretation of the total differential is easy to see by looking at a tangent plane to the function F. Let dx and dy be thought of as small deviations from the points x^* and y^* respectively. The total differential states that the change in the value of the function F from this small deviation from (x^*, y^*) to $(x^* + dx, y^* + dy)$ is the slope of the plane at (x^*, y^*) in the x direction times the deviation dx, plus the slope of the plane at (x^*, y^*) in the y direction times the deviation dy.

Example:

Given a utility function of the form $U(C,N) = \ln(C) + \ln(N)$, find the approximate change in utility by consuming an additional 0.1 units of C and less 0.1 units of N. Assume the individual originally was consuming 0.5 units of each.

$$dU \approx \frac{1}{C}dC + \frac{1}{N}dN = \frac{1}{0.5}0.1 + \frac{1}{0.5}(-0.1) = 0$$

Note that the total differential is only valid with equality as $(dx, dy) \to (0,0)$. Otherwise it is only an approximation.

Chain Rule

Let $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))$. Also, let f be some function of $\mathbf{x}(t)$. Then

$$\frac{df}{dt}(\mathbf{x}(t)) = \frac{\partial f}{\partial x_1}\cdot\frac{dx_1}{dt} + \ldots + \frac{\partial f}{\partial x_n}\cdot\frac{dx_n}{dt}.$$

Example:

Let K (capital stock) and L (labor stock) be functions of time, such that $\frac{dK}{dt} = sY - \delta K$ and $\frac{dL}{dt} = nL$. If $Y = K^\alpha L^{1-\alpha}$, what is $\frac{dY}{dt}$?

$$\frac{dY}{dt} = \frac{\partial Y}{\partial K}\cdot\frac{dK}{dt} + \frac{\partial Y}{\partial L}\cdot\frac{dL}{dt}$$

$$\frac{dY}{dt} = \alpha\left(\frac{L}{K}\right)^{1-\alpha} \cdot (sY - \delta K) + (1-\alpha)\left(\frac{K}{L}\right)^{\alpha} \cdot nL$$
$$\frac{dY}{dt} = \left\{\alpha\left[s\left(\frac{L}{K}\right)^{1-\alpha} - \delta\right] + (1-\alpha)\,n\right\} \cdot Y$$

Directional Derivatives

Assume a function $f(\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_n)$ is a point in $\mathbb{R}^n$. Now let there be a vector $\mathbf{v} = (v_1, \ldots, v_n)$, and a parameter t. The point of a directional derivative is to see how the value of the function $f(\mathbf{x})$ changes as we move along the vector $\mathbf{v}$ from the original point $\mathbf{x}$. A line can be drawn from $\mathbf{x}$ along $\mathbf{v}$ in $\mathbb{R}^n$ by constructing a point $\mathbf{x} + t \cdot \mathbf{v}$, and allowing t to vary. Define a new function $g(t) = f(\mathbf{x} + t \cdot \mathbf{v}) = f(x_1 + t \cdot v_1, \ldots, x_n + t \cdot v_n)$, and derivate it with respect to t,

$$\frac{dg}{dt} = \frac{\partial f}{\partial x_1}(x_1 + t \cdot v_1, \ldots, x_n + t \cdot v_n) \cdot v_1 + \ldots + \frac{\partial f}{\partial x_n}(x_1 + t \cdot v_1, \ldots, x_n + t \cdot v_n) \cdot v_n.$$

Since we are interested in the change of f at the original point $\mathbf{x}$, we let $t = 0$ to get

$$\frac{dg}{dt}(0) = \frac{\partial f}{\partial x_1}(x_1, \ldots, x_n) \cdot v_1 + \ldots + \frac{\partial f}{\partial x_n}(x_1, \ldots, x_n) \cdot v_n = [\nabla f(\mathbf{x})]' \cdot \mathbf{v}.$$

This is the directional derivative. It is merely the dot product of the total derivative at $\mathbf{x}$ with the vector $\mathbf{v}$. Other notiations are $DF_{\mathbf{x}} \cdot \mathbf{v}$ or $\frac{\partial F}{\partial \mathbf{v}}(\mathbf{x})$.

Level Sets

Consider a function $F(\mathbf{x})$ with $\mathbf{x} = (x_1, \ldots, x_n)$. A level set is the set of all $\mathbf{x}$'s such that $F(\mathbf{x}) = c$, where c is some specified constant. For example, indifference curves are level sets, as are isoquant curves. We will talk more about level sets when we discuss the Implicit Function Theorem, quasi-concavity, and upper contour sets.

Derivatives of Multiple Functions of Several Variables

Instead of assuming we have one function F, say we now have m equations which are each a function of n variables:

$$y_1 = f_1(x_1, \ldots, x_n)$$
$$\vdots$$
$$y_m = f_m(x_1, \ldots, x_n)$$

We can denote this as a single function F which is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Why? Because we can think of the system of equations as a box into which we throw n inputs (the x's) and get m outputs (the y's). Recall that the total derivative of any function F is

$$DF = \begin{pmatrix} \frac{\partial F}{\partial x_1} & \cdots & \frac{\partial F}{\partial x_n} \end{pmatrix}$$

How do we interpret $\frac{\partial F}{\partial x_1}$? Since F is actually m functions, we say that

$$\frac{\partial F}{\partial x_1} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} \\ \vdots \\ \frac{\partial f_m}{\partial x_1} \end{pmatrix}$$

Therefore, DF is now an $n \times m$ matrix called the Jacobian matrix

$$DF(\mathbf{x}^*) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}^*) \end{pmatrix}$$

where $\mathbf{x}^*$ is a particular value of $\mathbf{x} = (x_1, \ldots, x_n)$.

Example:

Find the Jacobian of the following system:

$$W = \frac{X}{Y}$$

$$V = XY$$

Notice that

$$DW = (\frac{1}{Y}, -\frac{X}{Y^2})$$

and

$$DV = (Y, X),$$

so the Jacobian is

$$\begin{pmatrix} \frac{1}{Y} & -\frac{X}{Y^2} \\ Y & X \end{pmatrix}.$$

Homework

Differentiate the following:

1. $f(x) = x^2 + 3x - 4$
2. $y = x^{-\frac{2}{3}}$
3. $g(x) = x^2 + x^{-3}$
4. $y = \frac{x^2+4x+3}{\sqrt{x}}$
5. $f(x) = x^2 e^x$
6. $V(x) = (2x+3)\left(x^4 - 2x\right)$
7. $f(x) = \frac{x}{x+\frac{c}{x}}$
8. $F(y) = \left(\frac{1}{y^2} - \frac{3}{y^4}\right)\left(y + 5y^3\right)$
9. $y = e^{\sqrt{x}}$
10. $y = e^{ax^2+bx+c}$
11. $y = \ln(t+9)$
12. $y = \ln(x) - \ln(1+x)$
13. Differentiate $\int_x^{e^x} 2xe^{xy} dy$ with respect to x
14. Differentiate $\int_x^1 \lambda e^{-\lambda xy} dy$ with respect to x

In 1st semester micro, you will solve general equilibrium models. Sometimes when solving these models it is useful to see if utility functions are concave. One way of testing for concavity involves calculating the function's Hessian. Find the Hessian matrices of the following utility functions (these functions were used in last year's homeworks and tests):

1. (Core Exam) $U(x_1, x_2) = \frac{2}{3}\sqrt{x_1} + \frac{1}{3}\sqrt{x_2}$
2. (Final Exam) $U(x_1, x_2) = x_1 + \frac{\delta}{\alpha} x_2^\alpha$

3. (Second Midterm Exam) $U(x_1, x_2) = \frac{1}{3}\ln(x_1) + \frac{2}{3}\ln(x_2)$

4. (Homework problem) $U(x_1, x_2, x_3) = -\frac{1}{x_1} + x_2 - \delta\frac{1}{x_3}$

The Slutsky Equation breaks changes in demand into income effects and substitution effects. Last semester for one of the homework problems, we were asked to calculate the Slutsky equation to find the substitution effect and the income effect. In the problem, the utility function was the same as in question 3 above, and it can be shown that the direct demand functions are

$$x_1(p_1, p_2, w) = \frac{w}{3p_1}$$

$$x_2(p_1, p_2, w) = \frac{2w}{3p_2}$$

Let the function $\mathbf{x}(p, w) : \mathbb{R}^3 \to \mathbb{R}^2$ be the combined demand function, where $p = (p_1, p_2)$. The Slutsky equation is as follows:

$$D_p\mathbf{x} + D_w\mathbf{x} \cdot \mathbf{x}' = D_p\mathbf{v},$$

where the subscripts denote which derivative it is respect to (Note: these are total derivatives, not directional derivatives). Find $D_p\mathbf{x}$ (the total effect), $D_w\mathbf{x}\cdot\mathbf{x}'$(the wealth effect), and $D_p\mathbf{v}$ (the substitution effect). Hint: $D_p\mathbf{x}$ is a 2×2 matrix, $D_w\mathbf{x}$ is 2×1, and $\mathbf{x}$ is 2×1. This involves matrix multiplication, something we haven't covered yet, but that you should already know.

If the graph of a function $F : \mathbb{R}^3 \to \mathbb{R}^1$ lives in $\mathbb{R}^4$, in which space does the gradient ∇F live?

In 1st semester econometrics, you will be asked to integrate probability density functions (or p.d.f.s) in order to find the probability that certain events will occur. Sometimes it is necessary to transform the random variables. In order to do so, the Jacobian must be computed.

Assume and X_1 and X_2 are random variables with a known p.d.f. However, we are interested in finding the distributions of

$$Y_1 = \frac{X_1}{X_2}$$

$$Y_2 = X_1X_2.$$

We can transform this system so that we solve for the X's as a function of the Y's, and then use the known p.d.f. of the X's to calculate the densities of the Y's. Confirm that the new system can be written as

$$X_1 = \sqrt{Y_1Y_2}$$

$$X_2 = \sqrt{\frac{Y_1}{Y_2}},$$

and then calculate the Jacobian matrix of this system.