

11 Partial derivatives and multivariable chain rule

11.1 Basic definitions and the Increment Theorem

One thing I would like to point out is that you've been taking partial derivatives all your calculus-life. When you compute df/dt for $f(t) = Ce^{-kt}$, you get $-Cke^{-kt}$ because C and k are constants. The notation df/dt tells you that t is the variables and everything else you see is a constant. If we use the notation f' instead, then we are relying on your knowing which is the independent variable. It's usually called something like "t", not "C" or "k", but every now and then we end up computing df/dk or df/dC , so watch out! The only rule is: everyone should understand which is the independent variable.

So now, studying partial derivatives, the only difference is that the other variables aren't constants – they vary – but you treat them as constants anyway. It's not a big difference because really, what is a constant? It's always possible to imagine some quantity changing. Mathematically we just need to be precise about what is holding steady and what is changing. In this section, only one variable at a time will change. Then in the next section (chain rule), we'll change more than one independent variable at a time and keep track of the total effect on the independent variable.

We assigned plenty of MML problems on this section because the computations aren't much different than ones you are already very good at. You can read the basics in Section 14.3. I will include one example as a self-check; if you are not able to cover up the answer and figure it out pretty easily, then you need to go back and re-read Section 14.3.

EXAMPLE: Let $f(x, t, q) = \frac{e^q - 1}{1 + xtq}$. What is $\frac{\partial f}{\partial t}$ at the point $(3, 1, 1)$ and what does this quantity signify?

Answer: treating everything other than t as a constant, by either the chain rule or the quotient rule you get $-xq(e^q - 1)/(1 + xtq)^2$. Evaluating at the point $(3, 1, 1)$ gives $-3(e - 1)/16$.

This means that if t is changes by a small amount from 1 while x is held fixed at 3 and q at 1, the value of f would change by roughly $3(e - 1)/16$ times as much in the opposite direction.

The Increment Theorem

By now I'm sure you remember the linearization in one-variable. The value of $f(x)$ near the point $x = a$ is well approximated by $L(x) = f(a) + f'(a) \cdot (x - a)$. Suppose we now want to approximate $f(x, y)$ near a point (a, b) where we know the value. Suppose, in fact that we change only x but not y . Then we might as well treat y as a constant and write

$$f(x + \Delta x, y) = f(x, y) + (\Delta x) \cdot \frac{\partial f}{\partial x}(x, y).$$

It's a partial derivative, not a total derivative, because there is another variable y which is being held fixed. Similarly, if we moved only y we would have

$$f(x, y + \Delta y) = f(x, y) + (\Delta y) \cdot \frac{\partial f}{\partial y}(x, y).$$

I hope it doesn't seem like too much of a leap to say that if you move both x and y you'll get both of these effects:

$$f(x, y + \Delta y) = f(x, y) + (\Delta x) \cdot \frac{\partial f}{\partial x}(x, y) + (\Delta y) \cdot \frac{\partial f}{\partial y}(x, y). \quad (11.1)$$

Equation (11.1) is called the Increment Theorem in the textbook and appears as Theorem 3 on page 818 (Section 14.3). You might wonder whether it's OK to assume that you can just add the two effects from moving x and moving y . In fact, after you move x , you really should be computing the y increment according to the $\partial f / \partial y$ at the new location, $(x + \Delta x, y)$. However, it's only an approximation anyway, and the new partial derivative is close enough to the old that the computation with the new partial derivative matches the computation with the old partial derivative to within the error you already introduce by linearizing.

EXAMPLE: About how much does $x^2/(1 + y)$ change if (x, y) changes from $(10, 4)$ to $(11, 3)$? Here $\Delta x = 1$ and $\Delta y = -1$. We compute $f_x = 2x/(1 + y)$ and $f_y = -x^2/(1 + y)^2$ so so $f_x(10, 4) = 4$ and $f_y(10, 4) = -4$. Thus,

$$\Delta f \approx f_x \Delta x + f_y \Delta y = 4(1) + (-4)(-1) = 8.$$

In fact, f changes from 20 to 30.25 so the 8 was kind of a crude estimate, but that's because Δx and Δy were pretty big. If we choose 0.1 and -0.1 instead, we get a linear estimate of $\Delta f = 0.8$ which is very close to the actual 0.818...

Application: marginal rates

Suppose the cost of a proposed building is a function $f(A, q, \ell)$ where A is the area of usable space in square feet, q is an index of the quality (thickness of walls, gauge of wiring, level of insulation, quantity of lighting, etc.) and ℓ is a location parameter measuring, for example, the desirability of the location. The average cost per square foot for a given proposed building is, by definition, $f(A, q, \ell)/A$. However, this statistic is far less useful than the *marginal* cost per square foot, that is, $\partial f/\partial A$. That's because most decisions are about whether to put a few extra dollars into one of these categories or to trim a few bucks from another category. Therefore, it is most useful to know how many dollars more you will spend or save with each square foot, rather than what all the square footage costs that is already in all the proposals being compared.

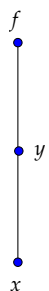
EXAMPLE: The total number P of people exposed to an recurring ad is a function of its market share, M , and the length of time, t , that stays in rotation⁵. The marginal increase in exposure per time run is $\partial f/\partial t$. The right time to yank the ad is when $v \cdot \partial f/\partial t$ drops below the cost per time to run the ad, where v is the value in dollars per unit of exposure. Note that the units match: v has units of dollars per exposure, $\partial f/\partial t$ has units of exposure per time and the cost to run the ad is priced in dollars per time: $(\$/\text{exp}) (\text{exp}/t) = \$/t$.

Note: the notion of marginal rates should already be familiar from univariate calculus. There isn't much added here, except to say that it makes sense to compute marginal rates when there are many quantities that could vary, by varying only one.

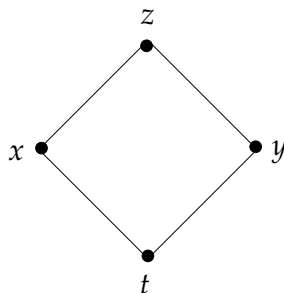
Branch diagrams

In applications, computing partial derivatives is often easier than knowing what partial derivatives to compute. With all these variables flying around, we need a way of writing down what depends on what. We do this by writing a branch diagram. Here are some common ones.

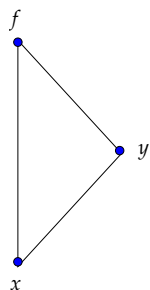
⁵It is not just the product of these because the longer it runs, the more redundancy there is in people seeing it multiple times.



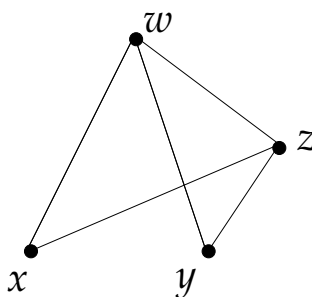
The branch diagram for the ordinary chain rule.



w is a function of x and y , both of which are functions of a single variable t (see page 823 of the textbook).



z depends on x and y but y is really a function of x



w is a function of x, y and z , but z is really a function of the other two.

Any variable at the top is an dependent variable. Any variable at the bottom is an independent variable; these drive the other variables and are the only ones we tweak directly. The variables in the middle are called *intermediate* variables. The independent variables drive them and they drive the dependent variables.

11.2 Chain rule

Think about the ordinary chain rule. A useful metaphor is that it is like a gear assembly⁶: y depends on u , which in turn depends on x . Each unit increase of x increases u by $u'(x)$ many units. Each unit increase of u increases y by $y'(u)$ units. Therefore each unit increase in x produces $u'(x) \cdot y'(u)$ units increase in y . That's what's going on in the first branch diagram.

In the second diagram, there is a single independent independent variable t , which we think of as a gear driving both x and y , while both x and y drive z . I am going to try now to explain why

$$\frac{dy}{dt} = \frac{\partial y}{\partial u} \frac{du}{dt} + \frac{\partial y}{\partial v} \frac{dv}{dt}. \quad (11.2)$$

⁶OK, you got me, that's a simile not a metaphor.

When t increases by Δt , both u and v increase. The increases are roughly $(\Delta t)(du/dt)$ and $(\Delta t)(dv/dt)$ respectively. As we just saw at the end of the previous section (with the function $z(x, y)$) each increase in u produces an increase in y that is $\partial y/\partial u$ times as great. So the increase in u of $\Delta t \frac{du}{dt}$ gives an increase in y of roughly $\Delta t \frac{du}{dt} \frac{\partial y}{\partial u}$. Simultaneously, the increase in t has produced an increase in v which produces another increase in y of roughly $\Delta t \frac{dv}{dt} \frac{\partial y}{\partial v}$. Thus the total increase in y is roughly

$$\Delta t \left[\frac{\partial y}{\partial u} \frac{du}{dt} + \frac{\partial y}{\partial v} \frac{dv}{dt} \right].$$

This means that the rate of change of y per change in t is given by equation (11.2). Note that we use partial derivative notation for derivatives of y with respect to u and v , as both u and v vary, but we use total derivative notation for derivatives of u and v with respect to t because each is a function of only the one variable; we also use total derivative notation dy/dt rather than $\partial y/\partial t$. Do you see why? Partial derivative notation would mean that t was changing while something else was being held fixed, which is not the case. Rather, all variables are functions of the single variable t .

That's the basic story. There are lots of variations, depending on how many independent variables there are (up till now there has been only one, all the others ultimately being functions of the one), how many intermediate variables and how they are related.

Where to evaluate?

The one thing you need to be careful about is evaluating all derivatives in the right place. It's just like the ordinary chain rule. For example, in (11.2), the derivatives du/dt and dv/dt are evaluated at some time t_0 . The partial derivative $\partial y/\partial u$ is evaluated at $u(t_0)$ and the partial derivative $\partial y/\partial v$ is evaluated at $v(t_0)$.

Example: Chain rule for $f(x, y)$ when y is a function of x

The heading says it all: we want to know how $f(x, y)$ changes when x and y change but there is really only one independent variable, say x , and y is a function of x . This

is captured by the third of the four branch diagrams on the previous page. Applying the chain rule gives

$$\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \cdot y'. \quad (11.3)$$

The notation really makes a difference here. Both df/dx and $\partial f/\partial x$ appear in the equation and they are not the same thing!

Derivative along an explicitly parametrized curve

One common application of the multivariate chain rule is when a point varies along a curve or surface and you need to figure the rate of change of some function of the moving point. The classical economics application is that price and quantity are moving together along the *demand curve* and we want to figure out how revenue changes along this curve (and in particular, we want to find where the revenue is maximized). In this section we solve the problem when the curve is known explicitly, saving the case of implicitly defined curves until we have discussed implicit differentiation.

Suppose a point varies along a curve as a function of time, and its coordinates are explicitly known: the coordinates at time t are $(x(t), y(t))$. The rate of change of the function $g(x, y)$ with respect to *time* along the curve is given by the formula we just computed: x and y are functions of t and g is a function of x and y , so

$$\frac{dg}{dt} = \frac{\partial g}{\partial x} \frac{dx}{dt} + \frac{\partial g}{\partial y} \frac{dy}{dt}. \quad (11.4)$$

I hope you realize this is the exact same equation as (11.2) but with the letter g in place of y , and x and y in place of u and v .

11.3 Implicit differentiation

The chain rule helps us to understand ordinary implicit differentiation. In Section 14.4 on page 826 the textbook re-explains finding the slope of an implicitly defined curve (first discussed in the textbook in Section 3.7). Here follows a quick recap of this.

Slope of an implicitly defined curve

Suppose a curve is defined by $F(x, y) = 0$. What is the slope of its tangent line? That's the same as asking, if we treat y as a function of x along the curve, what is dy/dx ? This is just (11.3) run backwards – we know that $df/dx = 0$ and want to solve for y' . Differentiating the relation $F(x, y) = 0$ with respect to x , where y is an intermediate variable that is a function of x , the chain rule gives $0 = F_x + F_y dy/dx$. Solving for dy/dx gives (see page 826 of the textbook):

$$\frac{dy}{dx} = -\frac{F_x}{F_y}. \quad (11.5)$$

Derivative along an implicitly parametrized curve

Now suppose a curve is defined implicitly by $F(x, y) = 0$. How fast does the function $g(x, y)$ change along the curve? We had better decide: how fast does $g(x, y)$ change with respect to *what*? Suppose we treat y as a function of x along the curve and ask for dg/dx . Using the chain rule for this case (11.3)

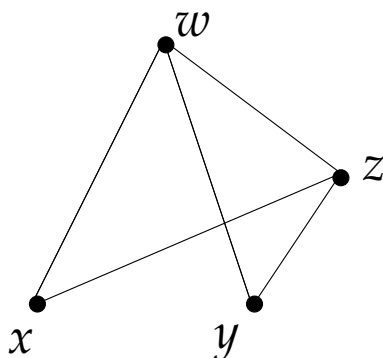
$$\begin{aligned} \frac{dg}{dx} &= \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} \frac{dy}{dx} \\ &= \frac{\partial g}{\partial x} - \frac{\partial g}{\partial y} \frac{\partial F/\partial x}{\partial F/\partial y}. \end{aligned}$$

In the last line, we used the expression for dy/dx given by implicit differentiation (11.5).

Implicitly defined surfaces

This is just like curves defined by an equation, only now there are three variables. Any equation $F(x, y, z) = 0$ defines a surface. If any two vary freely, the third changes as a function of the other two. When this happens, we can ask for the rate of change of one with respect to another. What should $\partial z/\partial x$ mean in this context? It means: consider z as a function of x and y , then find out the rate of change in z when x varies, y is held constant, and z changes in order still to satisfy the equation. Please take a moment to think this through now.

Computationally, how do we find $\partial z/\partial x$ when $F(x, y, z) = 0$? We differentiate, keeping in mind the branch diagram. Letting w denote $F(x, y, z)$, it is the same as one we have seen before:



The variables vary in such a way that w remains at zero. Taking the partial derivative with respect to x of the equation $w = 0$ gives

$$0 = \frac{\partial w}{\partial x}(x, y, z) = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial z} \frac{\partial z}{\partial x}.$$

Solving for $\partial z/\partial x$ we see that

$$\frac{\partial z}{\partial x} = \frac{-F_x}{F_z}.$$

This looks exactly the same as for two variables, x and z only; compare to equation (11.5). This is not a coincidence. If z is a function of x and y and we hold y constant, then y is playing a similar role to the constant k in the function e^{kx} . The problem really does reduce to the two variable problem. Let's try it on Example 4 from Section 14.3 of the textbook.

EXAMPLE: Find $\partial z/\partial x$ when the equation $F(x, y, z) = x + y + \ln z - yz = 0$ defines z as a function of x and y . We compute $F_x = 1$ and $F_z = 1/z - y$ therefore

$$\frac{\partial z}{\partial x} = \frac{-1}{1/z - y} = \frac{z}{yz - 1}.$$

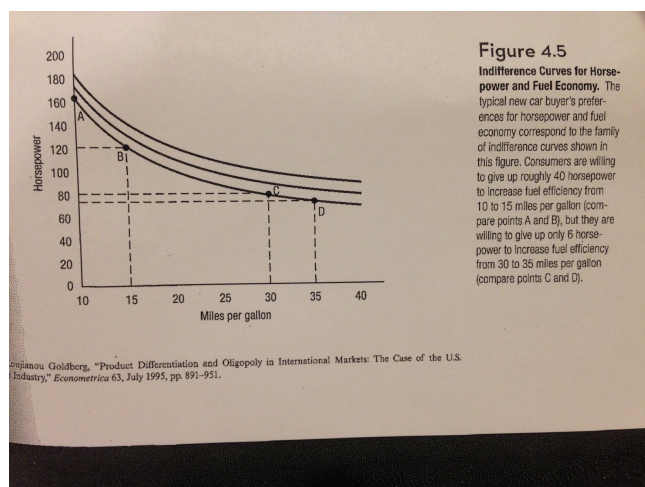
You should compare this to how the book does it (page 813); I think this way is simpler than the book's but either is OK.

11.4 Featured application: indifference curves

Remember level curves from our first day of multivariate calculus? They're back, in an economic application, under the name of "indifference curves". Suppose that the independent variables x and y represent quantities of two different things that will rival each other for importance in a single scenario.

Example 1: x is the horsepower of a car and y is its MPG.

Example 2: x is ounces of pizza at a meal and y is pints of FroYo.

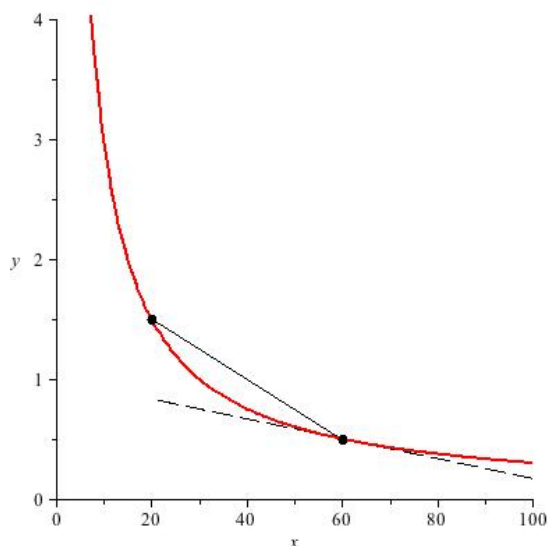


An indifference curve is a set of points in the x - y plane corresponding to bundles that the agent (often a consumer) likes equally well. The two examples above are taken from Berheim and Whinston (current textbook for BEPP 250). The indifference curve for horsepower versus fuel economy is taken from actual data. The indifference curve for pizza versus FroYo is a made up model. In either case, the indifference curves are just level contours for a utility function $u(x, y)$. The food example uses the utility function $u(x, y) = xy$ and shows indifference curves of $xy = 10$, $xy = 20$ and $xy = 30$.

Indifference curves are important for several reasons, one of which is that they describe incentives and reactions to changes in the quantities x and y . The *marginal rate of substitution* is the amount of x an agent would be willing to give up in order to increase y by one unit. This is not a static quantity, rather it depends on the present levels of x and y . If a group of diners has 10 pints of FroYo and only three ounces of pizza, they will not be willing to give up much pizza for one more pint of Froyo,

whereas a group with 60 ounces of pizza and half a pint of FroYo might well give up a lot of pizza for a pint of FroYo.

Two points on the same indifference curve, such as $(60, 1/2)$ and $(20, 3/2)$, determine an equivalence of utility. The slope of the line segment between these two points is a ratio for a trade the agent is willing to make in either direction (see the straight line in the figure). But the point $(20, 3/2)$ is quite far from $(60, 1/2)$ and does not represent the rate of substitution if the consumer is able to make continuous small adjustments. As the point (x, y) on the curve $u(x, y) = 30$ approaches $(60, 1/2)$, the slope of the line segment approaches the slope of the tangent line to the curve $u(x, y) = 30$ at $(60, 1/2)$ (dashed line in the figure).



Mathematically, the marginal rate of substitution is defined to be the negative of the slope of this tangent line (negative because the slope represents one quantity going down while the other goes up). This slope is just dy/dx , which we know how to compute via implicit differentiation. In the pizza and FroYo example, the level curve is $xy = 30$ and implicit differentiation gives $y + x(dy/dx) = 0$. Thus $dy/dx = -y/x$. At the point $(60, 1/2)$, this gives a marginal rate of substitution of $1/120$ pint of FroYo per ounce of pizza. On the other hand, at the point $(3, 10)$, the marginal rate of substitution is $10/3$ pints of FroYo per ounce of pizza. Whether or not you think xy is a reasonable utility function for this scenario, this model sheds light on consumer behavior and how to model it.