Multivariable Calculus

Geometric Representation of Functions.

- When we study functions from $\mathbb{R}$ to $\mathbb{R}$, we find it useful to visualize functions by drawing their graphs.
- When we are concerned with functions of two variables, i.e. from $\mathbb{R}^2$ to $\mathbb{R}$ this technique is even more useful.
- As we needed two dimensions to draw the graph of a function from $\mathbb{R}$ to $\mathbb{R}$, we need three dimensions to draw the graph of a function from $\mathbb{R}^2$ to $\mathbb{R}$.
- When drawing graphs of unfamiliar functions from $\mathbb{R}$ to $\mathbb{R}$ in two dimensions, we could mark out a number of points to get an idea of what the function looks like.
- When we have a function from $\mathbb{R}^2$ to $\mathbb{R}$, we can do the same, but because of the extra dimension, we need a lot more points to build up a graph of the function.
- One more systematic way to proceed is to draw the usual one dimensional graphs on various two-dimensional slices or *cross-sections* and then put these together to draw the graph.

Example 1. Say we are trying to graph the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $z = f(x, y) = x^2 - y^2$ for all $(x, y) \in \mathbb{R}^2$.

- We can take a slice at in the $y = b$ planes. In this plane $z = x^2 - b^2$ which is a parabola shifted down by b^2 units. We draw these cross-sections for several values of b.
- Next consider a cross-section in the plane $x = 0$. The restriction to this plane is the upside-down parabola $z = -y^2$.
- Finally attach each of your cross-sections in the $y = b$ plane to your cross-section in the $x = 0$ plane and sketch the rest.

Cross-Section Applet

- Another way to visualize a function from $\mathbb{R}^2$ to $\mathbb{R}$ is to use level curves. This technique only requires two-dimensional sketching.
- For each (x, y) we evaluate $f(x, y)$ to obtain, say b. Then we draw the locus of points (x, y) in the xy-plane for which f has the same value b.

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and let $b \in \mathbb{R}$. A *level set* is a set of the form
$$\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid f(x_1, \dots, x_n) = b\}.$$

- When $n = 2$,we call a level set a *level curve*, when $n = 3$ we call it a *level surface* and when $n \geq 3$. In general we may also call a level set a *level hypersurface*. ▲

Example 2. Consider the function given by $f(x, y) = x^2 - y^2$.

- Start with the point $(0, 0)$ at which f takes value 0. Now find all points where f is 0. This is the set $\{(x, y) \mid x^2 - y^2 = 0\}$, which is simply the inverse image of the set $\{0\}$ i.e. $f^*(0)$. This level curve is the set of all points for which $x^2 = y^2$ i.e. such that $x = \pm y$.
- Now consider the point $(2, 1)$ at which f takes a value of 3. The level curve in this case is
$$f^*(3) = \{(x, y) \mid x^2 - y^2 = 3\},$$
which is a hyperbola centred on the origin with asymptotes $y = x$ and $y = -x$ and focal points $\sqrt{3}$ and $-\sqrt{3}$.
- Compute more level sets for different b, say -9, -5, -3, 5 and 9. Draw them in the xy-plane and then think of "pulling" each level curve up into the $z = b$ plane.

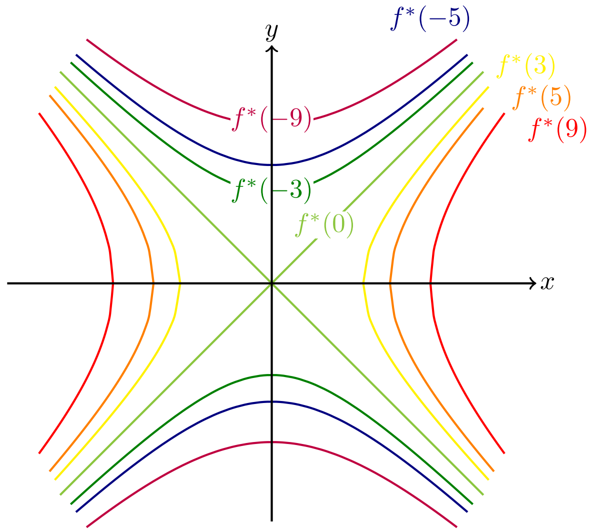

Figure 1: Level curves of the function given by $f(x) = x^2 - y^2$.

Level Curve Applet

Example 3. Level sets are common in economics.

- An indifference curve is just a level curve of the utility function. Consider a utility function $u : \mathbb{R}^2 \to \mathbb{R}$ given by $u(x, y)$. Then for any two bundles (x_1, y_1) and (x_2, y_2) on the same level curve the consumer is indifferent, because $u(x_1, y_1) = u(x_2, y_2)$.
- An isoquant is a level curve of a production function. Take a simple Cobb-Douglas production function given by $Q = KL$, where K and L are quantities of inputs, say captial and labour and Q is the amount output produced from the inputs.
  - To draw an isoquant for $Q = 10$, solve the equation $KL = 10$ for L in terms of K and then graph the result. Solving we find that the isoquant for $Q = 10$ is a branch of a hyperbola $L = 10/K$.
  - Do this for several values of Q, to build up a picture of the production function.
- So far we have looked at drawing graphs for functions from $\mathbb{R}$ or $\mathbb{R}^2$ to $\mathbb{R}$.
- We can also draw picture of functions from $\mathbb{R}$ to $\mathbb{R}^2$ or $\mathbb{R}^3$.
- A typical function from $\mathbb{R}$ to $\mathbb{R}^2$ would be written as $x(t) = (x_1(t), x_2(t))$, where are x_1 and x_2 are the coordinate functions of x.
- For each t, $x(t)$ is a point in $\mathbb{R}^2$. By marking each such point $x(t)$ in the x_1, x_2 plane, we trace out a curve in the plane. This curve is the image of x.

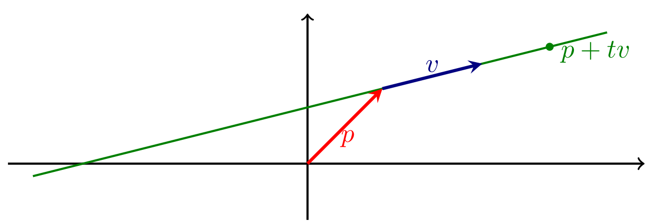

Figure 2: The parameterized line $x(t) = (p_1 + tv_1, p_2 + tv_2)$.

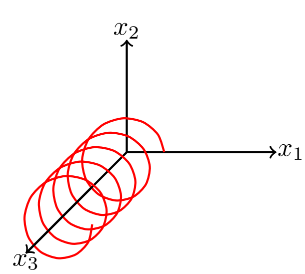

Figure 3: The parameterized curve $x(t) = (\cos t, \sin t, t)$.

Special Kinds of Functions.

Definition. A *linear function* from $\mathbb{R}^k$ to $\mathbb{R}^m$ is a function f that preserves the vector space structure, i.e.

$$f(x+y) = f(x) + f(y) \quad \text{and} \quad f(\lambda x) = \lambda f(x)$$

for all $x, y \in \mathbb{R}^k$ and all $\lambda \in \mathbb{R}$. Linear functions are sometimes called *linear transformations*. ▲

Example 4. An example of a linear function is the function $f : \mathbb{R}^k \to \mathbb{R}$ given by

$$f(x) = a \cdot x = a_1 x_1 + \cdots + a_k x_k,$$

for some $a \in \mathbb{R}^k$. ♦

- It turns out that every linear real-valued function is of the form given in the example above.

Theorem 1. *Let $f : \mathbb{R}^k \to \mathbb{R}$ be a linear function. Then, there exists a vector $a \in \mathbb{R}^k$ such that $f(x) = a \cdot x$ for all $x \in \mathbb{R}^k$.*

- Thus every real-valued function on $\mathbb{R}^k$ can be written as

$$f(x) = a \cdot x = \begin{pmatrix} a_1 & \cdots & a_k \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}.$$

- The level sets of a linear function into $\mathbb{R}$ are the sets $a \cdot x = b$ which we called hyperplanes when we were looking at vectors.

Theorem 2. *Let $f : \mathbb{R}^k \to \mathbb{R}^m$ be a linear function. Then, there exists an $m \times k$ matrix A such that $f(x) = Ax$ for all $x \in \mathbb{R}^k$.*

- This says there that every linear function from $\mathbb{R}^k$ to $\mathbb{R}^m$ can be associated with an $m \times k$ matrix A.

Definition. A *quadratic form* on $\mathbb{R}^n$ is a real-valued function of the form

$$Q(x) = \sum_{i,j=1}^{n} a_{ij} x_i x_j = x^T A x$$

where A is any symmetric $n \times n$ matrix. ▲

Example 5. The general two-dimensional quadratic form

$$a_{11} x_1^2 + 2 a_{12} x_1 x_2 + a_{22} x_2^2$$

can be written as

$$x^T A x = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Note the symmetry of A. ♦

- Linear functions and quadratic forms are special forms of class of functions called polynomials which are functions made up of the sum of monomials.

Definition. A function $f : \mathbb{R}^k \to \mathbb{R}$ is called a *monomial* if it is of the form

$$f(x) = cx_1^{a_1} x_2^{a_2} \cdots x_k^{a_k}$$

where $c \in \mathbb{R}$ and $a_1, \ldots, a_k$ are nonnegative integers. The sum of the exponents $a_1 + \cdots + a_k$ is called the *degree* of the monomial. ▲

Example 6.

1. $f(x_1, x_2) = -6x_1^3 x_2$ is a monomial of degree four.
2. $g(x_1, x_2, x_3) = 2x_1 x_2 x_3^3$ is a monomial of degree five.
3. A constant function is a monomial of degree zero. ♦

Definition. A function $f : \mathbb{R}^k \to \mathbb{R}$ is called a *polynomial* if it is a finite sum of monomials on $\mathbb{R}^k$. The highest degree of these monomials is called the *degree* of the polynomial. A function $f : \mathbb{R}^k \to \mathbb{R}^m$ is called a *polynomial* if each of its coordinate functions is a real-valued polynomial. ▲

Example 7.

1. $f(x_1, x_2, x_3) = 2x_1^2 x_3^3 - 5x_1 x_2 x_3^4 + 7x_2^8 x_3$ is a polynomial of degree nine.
2. A linear real-valued function is a polynomial of degree one.
3. A quadratic form is a polynomial of degree two. ♦

Definition. A function $f : \mathbb{R}^k \to \mathbb{R}^m$ is called an *affine function* if it is of the form

$$f(x) = Ax + b,$$

where A is an $m \times k$ matrix and $b \in \mathbb{R}^m$. ▲

- So an affine function is a polyonomial of degree one, and each component of f has the form

$$f_i(x) = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ik}x_k + b_i = a_i \cdot x + b_i.$$

Example 8.

1. $f(x) = 2x + 1$ is an affine function.
2. $g(x_1, x_2) = \begin{pmatrix} 1 & 4 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 2 \\ -6 \end{pmatrix} = \begin{pmatrix} x_1 + 4x_2 + 2 \\ -2x_1 + 3x_2 \end{pmatrix}$ is an affine function. ♦

Continuous Functions.

- Just as we defined continuity for functions from subsets of $\mathbb{R}$ to $\mathbb{R}$, we can define continuity for functions from subsets of $\mathbb{R}^k$ to $\mathbb{R}^m$.
- Again, the idea is that as the vector x gets near but not equal to x_0, the value of the function at x gets close to $f(x_0)$.

Definition. Let f be a function into $\mathbb{R}^m$ whose domain is a subset of $\mathbb{R}^k$.

- The function f is *continuous at* x_0 in $\operatorname{dom}(f)$ if, for every sequence (x_n) in $\operatorname{dom}(f)$ converging to x_0, we have $\lim f(x_n) = f(x_0)$.
- If f is continuous at each point of a set $S \subseteq \operatorname{dom}(f)$, then f is said to be *continuous on* S.
- The function f is said to be *continuous* if it is continuous on $\operatorname{dom}(f)$. ▲

- Again, we can formulate an ε-δ definition of continuity. Because of the higher dimension, we use ε-balls instead of ε-intervals in the definition.

Theorem 3 (ε-δ definition of continuity)**.** *Let f be a function into $\mathbb{R}^m$ whose domain is a subset of $\mathbb{R}^k$. Then f is continuous at $x_0 \in \operatorname{dom}(f)$ iff*

> *for each $\varepsilon > 0$ there exists $\delta > 0$ such that*
> *$x \in \operatorname{dom}(f)$ and $x \in B_\delta(x_0)$ imply $f(x) \in B_\varepsilon(f(x_0))$.*

- For a function into $\mathbb{R}$ in two variables x and y this says: if we draw two xy-planes, no matter how close together, we can always cut off a cylinder such that all that part of the surface which is contained in the cylinder lies between the planes.

Continuity Applet

Definition. Let $f, g : \mathbb{R}^k \to \mathbb{R}^m$ be functions and let $c \in \mathbb{R}$. We define new functions from $\mathbb{R}^k$ into $\mathbb{R}^m$ as follows.

- cf given by $(cf)(x) = cf(x) = (cf_1(x), \ldots, cf_m(x))$;
- $f + g$ given by $(f + g)(x) = (f_1(x) + g_1(x), \ldots, f_m(x) + g_m(x))$;
- fg given by $(fg)(x) = (f_1(x)g_1(x), \ldots, f_m(x)g_m(x))$;

for all $x \in \mathbb{R}^k$. ▲

Theorem 4. *Let $f, g : \mathbb{R}^k \to \mathbb{R}^m$ be functions that are continuous at $x_0 \in \mathbb{R}^k$ and let $c \in \mathbb{R}$. Then*

1. *cf is continuous at x_0;*

2. *$f + g$ is continuous at x_0;*

3. *fg is continuous at x_0.*

- The following theorem follows from the sequential definition of continuity and the fact that a sequence in $\mathbb{R}^m$ converges iff each of the m component sequences converges in $\mathbb{R}$.

- It can be used, together with our theorem about continuity of combinations of real-valued functions from $\mathbb{R}$, to prove the previous theorem.

Theorem 5. *Let $f = \mathbb{R}^k \to \mathbb{R}^m$ be a function. Then, f is continuous at $x_0 \in \mathbb{R}^k$ iff each of its coordinate functions $f_i : \mathbb{R}^k \to \mathbb{R}$ is continuous at x_0.*

Theorem 6. *If f is continuous at $x_0 \in \mathbb{R}^k$ and g is continuous at $f(x_0) \in \mathbb{R}^m$, then the composite function $g \circ f$ is continuous at x_0.*

- A partial derivative of a function of several variables is its derivative with respect to one of those variables with the others held constant.

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and let $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. We say that f has a *partial derivative* with respect to x_i at a if the limit

$$\lim_{h\to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h}$$

exists and is finite. We write $(\partial f/\partial x_i)(a)$ for the partial derivative of f at a with respect to x_i:

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h\to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h}. \qquad \blacktriangle$$

- Sometimes we write $f_i(a)$, $f_{x_i}(a)$ or $D_i f(a)$ to denote the partial derivative with respect to x_i at a.

Example 9. Consider the function f given by $f(x, y) = 3x^3 + 2x^2y - 2xy^2$.

Partial Derivatives Applet

The Total Derivative.

- Suppose we are interested in the behaviour of a function $f(x, y)$ of two variables in the neighbourhood of some point (x^*, y^*).

- If we hold y fixed at y^* and change x^* to $x^* + \Delta x$, then

$$f(x^* + \Delta x, y^*) - f(x^*, y^*) \approx \frac{\partial f}{\partial x}(x^*, y^*)\Delta x.$$

  Similarly, if we hold x fixed at x^* and change y^* to $y^* + \Delta y$, then

$$f(x^*, y^* + \Delta y) - f(x^*, y^*) \approx \frac{\partial f}{\partial y}(x^*, y^*)\Delta y.$$

- Since we are working with linear approximations, we can add the effects of the one-variable changes to find the approximate effect of a simultaneous change in x and y:

$$f(x^* + \Delta x, y^* + \Delta y) - f(x^*, y^*) \approx \frac{\partial f}{\partial x}(x^*, y^*)\Delta x + \frac{\partial f}{\partial y}(x^*, y^*)\Delta y.$$

- Often we write

$$f(x^* + \Delta x, y^* + \Delta y) \approx f(x^*, y^*) + \frac{\partial f}{\partial x}(x^*, y^*)\Delta x + \frac{\partial f}{\partial y}(x^*, y^*)\Delta y.$$

- To interpret this, consider the approximation for a one-variable function

$$f(x^* + h) \approx f(x^*) + f'(x^*)h.$$

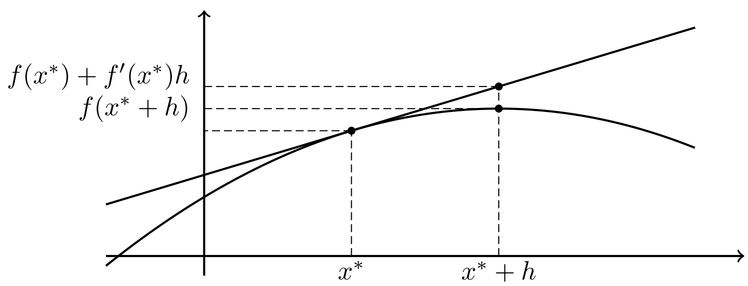

Figure 4: The tangent line to the graph of f at x^* is a good approximation of the graph in the vicinity of $(x^*, f(x^*)$.

- For a function $f(x, y)$ of two variables, the graph is a two-dimensional surface in $\mathbb{R}^3$ and the analogue of the tangent line is the tangent plane to the graph.

- We will show that

$$f(x^* + \Delta x, y^* + \Delta y) \approx f(x^*, y^*) + \frac{\partial f}{\partial x}(x^*, y^*)\Delta x + \frac{\partial f}{\partial y}(x^*, y^*)\Delta y,$$

states that the tangent plane to the graph at the point $p = (x^*, y^*, f(x^*, y^*))$ is a good approximation to the graph near p.

- Recall that to compute parameterized equation of the tangent plane $\mathcal{P}$ through the point p, we need two independent vectors u and v in the plane. In this case we parameterize the plane as

$$\{x \in \mathbb{R}^3 \mid x = p + su + tv \text{ for some } s, t \in \mathbb{R}\}.$$

- $u = (1, 0, \partial f/\partial x(x^*, y^*))$ and $v = (0, 1, \partial f/\partial x(x^*, y^*))$ are two independent vectors in the tangent plane.

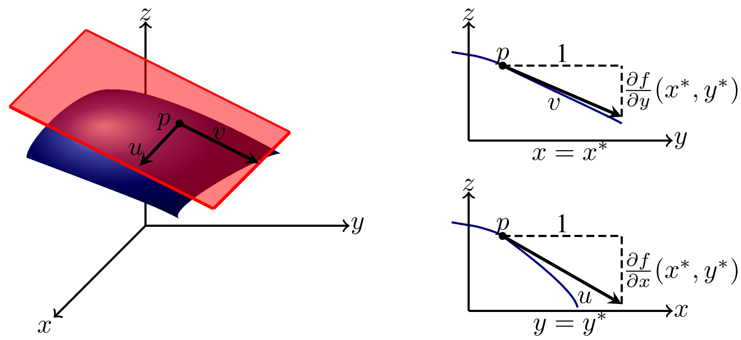

Figure 5: The tangent plane to the graph of f at $p = (x^*, y^*)$ is a good approximation of the graph in the vicinity of $(x^*, y^*, f(x^*))$.

Tangent Plane Applet

- Thus the tangent plane is given by

$$\begin{aligned}&(x^*, y^*, f(x^*, y^*)) + s(1, 0, \frac{\partial f}{\partial x}(x^*, y^*)) + t(0, 1, \frac{\partial f}{\partial y}(x^*, y^*)\\&= (x^* + s, y^* + t, f(x^*, y^*) + \frac{\partial f}{\partial x}(x^*, y^*)s + \frac{\partial f}{\partial y}(x^*, y^*)t)\end{aligned}$$

- If we replace s by Δx and t by Δy, we get our linear approximation of f about (x^*, y^*) i.e.

$$f(x^* + \Delta x, y^* + \Delta y) \approx f(x^*, y^*) + \frac{\partial f}{\partial x}(x^*, y^*)\Delta x + \frac{\partial f}{\partial y}(x^*, y^*)\Delta y.$$

  Therefore the above expression states that the tangent plane is a good approximation to the graph.

- When we are working on the tangent plane to the graph of f at (x^*, y^*), we use dx, dy and df. These variations on the tangent plane are called *differentials*.

- Rearranging the last expression, we get

$$\Delta f \approx \frac{\partial f}{\partial x}(x^*, y^*)\Delta x + \frac{\partial f}{\partial y}(x^*, y^*)\Delta y, \tag{1}$$

  which states that the change Δf on the graph of f is approximately the change df on the tangent plane. This equation in terms of df, dx and dy:

$$df = \frac{\partial f}{\partial x}(x^*, y^*)dx + \frac{\partial f}{\partial y}(x^*, y^*)dy,$$

  is called the *total differential* of f at (x^*, y^*).

- We saw that the tangent plane $\mathcal{P}$ can be thought of as the graph of the affine mapping
$$(s,t) \to f(x^*,y^*) + \frac{\partial f}{\partial x}(x^*,y^*)s + \frac{\partial f}{\partial y}(x^*,y^*)t$$

- Thus (1) says the change Δf can be approximated by the linear mapping
$$(s,t) \to \frac{\partial f}{\partial x}(x^*,y^*)s + \frac{\partial f}{\partial y}(x^*,y^*)t,$$
which we can write in matrix form as
$$\begin{pmatrix} \frac{\partial f}{\partial x}(x^*,y^*) & \frac{\partial f}{\partial y}(x^*,y^*) \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix}.$$

- Thus we consider the matrix
$$\begin{pmatrix} \frac{\partial f}{\partial x}(x^*,y^*) & \frac{\partial f}{\partial y}(x^*,y^*) \end{pmatrix}.$$
as representing the *linear approximation* of f around (x^*,y^*). We call this matrix, or the linear map it represents, the *(Jacobian) derivative* of f at (x^*,y^*) and denote it as $Df(x^*,y^*)$ or $Df_{(x^*,y^*)}$.

- We can generalize this to functions from $\mathbb{R}^n$ to $\mathbb{R}$.

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and let $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$.

- The *total differential* of f at point a is
$$df = \sum_{j=1}^{n} f_j(a)dx_j = \frac{\partial f}{\partial x_1}(a)dx_1 + \cdots + \frac{\partial f}{\partial x_n}(a)dx_n.$$

- The *(Jacobian) derivative* of f at a, denoted by $Df(a)$ or Df_a, is given by
$$Df(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) & \cdots & \frac{\partial f}{\partial x_n}(a) \end{pmatrix}.$$ ▲

- Now it is the tangent *hyperplane* to the n-dimensional graph of f in $\mathbb{R}^{n+1}$ which is a good approximation to the graph itself in the sense that the actual change Δf is well approximated by the total differential given above with $dx_i = \Delta x_i$.

- Sometimes we write the derivative of f at a as a column matrix:
$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{pmatrix}.$$
We denote this vector by $\nabla f(a)$ or grad$f(a)$, and call it the *gradient (vector)* of f at a.

The Chain Rule.

- Sometimes we are interested in how a function changes along a curve in its domain.
- For instance, if inputs are changing with time, we may want to know how the corresponding outputs are changing with time.

Definition. A *curve* in $\mathbb{R}^n$ is a function $x : \mathbb{R} \to \mathbb{R}^n$ given by

$$x(t) = (x_1(t), \ldots, x_n(t)),$$

for all $t \in \mathbb{R}$, where each $x_i : \mathbb{R} \to \mathbb{R}$ is a continuous function. The functions $x_i(t)$ are called *coordinate functions* and t is the *parameter* describing the curve. ▲

- The function $x(t)$ describes the coordinates of the curve at the point where the parameter is t.
- If we think of t as time, then $x(t)$ gives the position of a point on its trajectory in $\mathbb{R}^n$ at time t.
- If t is time, then $x_i'(t)$ is the instantaneous velocity of the ith coordinate along the curve at t.

Definition. Let $x(t) = (x_1(t), \ldots, x_n)$ be a paremeterized curve in $\mathbb{R}^n$. The vector

$$x(t) = (x_1'(t), \ldots, x_n'(t)),$$

is called the *velocity vector* or the *tangent vector* of the curve at t. ▲

Example 10. Consider the curve

$$x(t) = y^3, \quad y(t) = t^2.$$

When $t = 2$ we are at the point $(8, 4)$. The tangent vector there is $(3t^2, 2t)_{t=2} = (12, 4)$. Note $(x'(0), y'(0)) = (0, 0)$ – the curve has a cusp at the origin and the tangent vector there is not well-defined. ⧫

- We saw in our example that a curve can display irregular behaviour with the possibility of nonsmooth points such as cusps.
- To ensure the existence of a well defined tangent vector at all t, we impose a regularity condition on the curves.

Definition. A curve $x(t)$ is *regular* if each $x_i(t)$ is continuous in t and $x'(t) \neq 0$ for all t. ▲

- Often we want to know how a function f from $\mathbb{R}^n$ behaves along some regular curve $(x_1(t), \ldots, x_n(t))$, $a \leq t \leq b$.

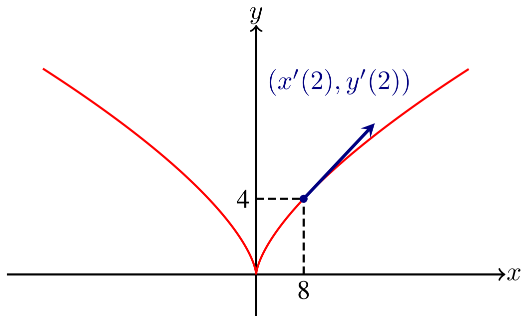

Figure 6: The parameterized curve $(x(t), y(t)) = (t^3, t^2)$.

- The value of the funtion at a point along the curve is given by

$$g(t) = f(x_1(t), \ldots, x_2(t)),\ a \le t \le b,$$

  where $g = f \circ x : \mathbb{R} \to \mathbb{R}$.

- The derivative $g'(t)$ gives the rate of change of f along the curve $x(t)$. Before we state the Chain Rule which tells us how to compute $g'(t)$, we need another definition.

Definition.

- Let U be an open subset of $\mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. We say f is *continuously differentiable* or C^1 on U if all its partial derivatives $(\partial f/\partial x_i)(a)$ exist and are continuous for all $a \in U$.

- A curve x from an open interval into $\mathbb{R}^n$ is continuously differentiable (or C^1) if each coordinate function x_i is continuously differentiable. ▲

Theorem 7 (Chain Rule I)**.** *Let $x(t) = (x_1(t), \ldots, x_n(t))$ be a C^1 curve on an open interval about a and let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function on an open ball about $x(a)$. Then $g = f \circ x$ is a C^1 function at a and*

$$\frac{dg}{dt}(a) = \frac{df}{dt}(x(a)) = \frac{\partial f}{\partial x_1}(x(a))x_1'(a) + \cdots + \frac{\partial f}{\partial x_n}(x(a))x_n'(a).$$

Example 11. Let $f(x, y) = 3x^2y$ and let $x(t) = 2t + 1$ and $y(t) = (t - 3)^3$. We will compute the total derivative of f with respect to t.

- We could do this directly. We have $f(x(t)) = 3(2t + 1)^2(t - 3)^3$, so that

$$\frac{df(x(t))}{dt} = 12(2t + 1)(t - 3)^3 + 9(2t + 1)^2(t - 3)^2.$$

- Using the chain rule, compute $\partial f/\partial x = 6xy$, $\partial f/\partial y = 3x^2$, $x'(t) = 2$ and $y'(t) = 3(t - 3)^2$. Thus

$$\begin{aligned}\frac{df(x(t))}{dt} &= (6xy)2 + (3x^2)3(t - 3)^2 \\ &= 12(2t + 1)(t - 3)^3 + 9(2t + 1)^2(t - 3)^3.\end{aligned}$$

♦

- It is important to distinguish between the total derivative and the partial derivative.
- Consider a funtion f of three variables x, y, and z.
- Usually we assume these variables are independent, but sometimes they may be dependent on each other – y and z, say, could be functions of x.
- In such cases the partial derivative of f with respect to x does not give the true rate of change of f with respect to x, as it does not take account of the dependency of y and z on x.
- The total derivative takes these dependencies into account.

Example 12.

1. Suppose $f(x, y, z) = xyz$.

    - The rate of change of f with respect to x is normally found by taking the partial derivative of f with respect to x. Here
    $$\frac{\partial f(x, y, z)}{\partial x} = yz.$$
    - However, if y and z are not truly independent but depend on x as well this does not give the right answer.
    - For a simple example, suppose $y = x$ and $z = x$.
    - Then $f(x, y(x), z(x)) = xy(x)z(x) = x^3$ and so the (total) derivative of f with respect to x is
    $$\frac{df(x, y(x), z(x))}{dx} = 3x^2.$$
    Notice that this is not equal to the partial derivative $yz = x^2$.

2. Consider the volume of a cone, which depends on the cone's height h and radius r according to the formula
$$V(r, h) = \frac{\pi r^2 h}{3}.$$
    - The partial derivative of V with respect to r is
    $$\frac{\partial V}{\partial r} = \frac{2\pi r h}{3}.$$
    It describes the rate with which the cone's volume changes if its radius is varied and its height is kept constant.

- The partial derivate with respect to h is

$$\frac{\partial V}{\partial h} = \frac{\pi r^2}{3},$$

and represents the rate at which the cone's volume changes if its height is changed and its radius kept constant.

2. - Now suppose that $r(h)$, or that $h(r)$. Then the total derivatives with respect to r or h are

$$\begin{aligned} \frac{dV}{dr} &= \frac{\partial V}{\partial r} + \frac{\partial V}{\partial h}\frac{dh}{dr} \\ &= \frac{2\pi rh}{3} + \frac{\pi r^2}{3}\frac{dh}{dr} \end{aligned} \qquad \begin{aligned} \frac{dV}{dh} &= \frac{\partial V}{\partial h} + \frac{\partial V}{\partial r}\frac{dr}{dh} \\ &= \frac{\pi r^2}{3} + \frac{2\pi rh}{3}\frac{dr}{dh}. \end{aligned}$$

- The difference between the total and partial derivatives is the ignorance of indirect dependencies in the latter.
- If, for some reason, the cone's proportions have to stay the same with height and radius in a fixed ratio k, we have

$$k = \frac{h}{r} = \frac{dh}{dr}$$

- Thus, the total derivative with respect to r is

$$\frac{dV}{dr} = \frac{2\pi rh}{3} + k\frac{\pi r^2}{3} = k\pi r^2.$$

♦

- Sometimes we write the chain rule as

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_1}{dt}.$$

Compare this with the total differential.

- We can generalize the chain rule to the case where the inside function depends on several variables.

Theorem 8 (Chain Rule II)**.** *Let $x : \mathbb{R}^s \to \mathbb{R}^n$, given by $x(t) = (x_1(t_1,\dots,t_s),\dots,x_n(t_1,\dots,t_s))$, and $f : \mathbb{R}^n \to \mathbb{R}$ be C^1 functions. Let $g = f \circ x$ be the composite function from $\mathbb{R}^s$ to $\mathbb{R}$. Then g is continuously differentiable and*

$$\frac{\partial g}{\partial t_i}(a) = \frac{\partial f}{\partial t_i}(x(a)) = \frac{\partial f}{\partial x_1}(x(a))\frac{\partial x_1}{\partial t_i}(a) + \cdots + \frac{\partial f}{\partial x_n}(x(a))\frac{\partial x_n}{\partial t_i}(a)$$

for all $a \in \mathbb{R}^s$.

- A diagrammatic way to remember the chain rule is given below for the example of a function $g : \mathbb{R}^2 \to \mathbb{R}$ given by $g(s,t) = f(p(s,t), q(t), r(s,t))$.

- To find $\partial f/\partial t$ ($\partial f/\partial s$) find the branches ending in t (s).

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial p}\frac{\partial p}{\partial t} + \frac{\partial f}{\partial q}\frac{dq}{dt} + \frac{\partial f}{\partial r}\frac{\partial r}{\partial t}$$
$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial p}\frac{\partial p}{\partial s} + \frac{\partial f}{\partial r}\frac{\partial r}{\partial s}$$

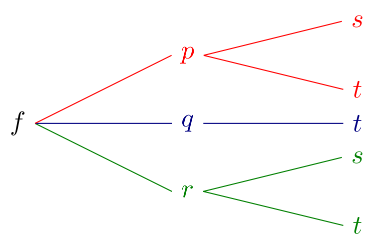

Figure 7: Chain Rule II.

Example 13. Suppose $u = x^2 + 2y$, where $x = r\sin(t)$ and $y = \sin^2(t)$.

- Note that $u = g(r,t) = f(x(r,t), y(r,t))$, where $f(x,y) = x^2 + 2y$.

$$\begin{aligned}\frac{\partial u}{\partial r} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial r}\left(= \frac{\partial g}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r}\right)\\ &= (2x)\sin(t) + 2(0) = 2r\sin^2(t)\end{aligned}$$

$$\begin{aligned}\frac{\partial u}{\partial t} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t}\left(= \frac{\partial g}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}\right)\\ &= (2x)r\cos(t) + 2(2\sin(t)\cos(t))\\ &= 2(r^2+2)\sin(t)\cos(t)\end{aligned}$$

Explicit Functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.

- Until now we have only looked at derivatives of functions with one endogenous variable.
- Often in economics we are interested in functions with several endogenous variables.
- For example, a firm producing m products using n inputs has a production funtion for each output:

$$\begin{aligned} q_1 &= f_1(x_1,\ldots,x_n)\\ q_2 &= f_2(x_1,\ldots,x_n)\\ &\vdots\\ q_m &= f_m(x_1,\ldots,x_n).\end{aligned}$$

- We can view the above collection of m functions in n variables as a single function f from $\mathbb{R}^n$ to $\mathbb{R}^m$:

$$f(x) = (f_1(x_1, \ldots, x_n), f_2(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)).$$

- Conversely, if we start with a single funtion $f : \mathbb{R}^n \to \mathbb{R}^n$ as above, we see that each component of f is a function from $\mathbb{R}^n$ to $\mathbb{R}$.
- Thus it is simple to apply our results for functions from $\mathbb{R}^n$ to $\mathbb{R}$, such as the chain rule, to the more general case of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.
- We just apply what we have learnt to each component function $f_i : \mathbb{R}^n \to \mathbb{R}$ and then put it all together in a matrix.
- If, for example, we want to approximate a function $f : \mathbb{R}^n \to \mathbb{R}^m$ (with component functions $f_1, \ldots, f_n$) using differentials, we apply our results to each component f_i (see p 324 S&B).
- We again obtain a matrix of partial derivatives which represents a linear map giving the linear approximation of f about a point a.

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function. The *(Jacobian) derivative* of f at a, denoted by $Df(a)$ or Df_a, is given by

$$Df(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}. \qquad \blacktriangle$$

- This is sometimes called the *Jacobian (matrix)*. An alternative notation is

$$\frac{\partial(f_1, \ldots, f_m)}{\partial(x_1, \ldots, x_n)}.$$

- When $m = n = 1$, we simply have the derivative of a function $f : \mathbb{R} \to \mathbb{R}$ and denote it as usual by $f'(a)$.

Example 14. Suppose, there are two commodities with constant elasticity demand functions

$$q_1(p_1, p_2, m) = 2\frac{p_2^3 m^2}{p_1} \quad \text{and} \quad q_2(p_1, p_2, m) = 3\frac{p_1^2 m}{p_2^2}$$

in the vicinity of current prices and income $(p_1^*, p_2^*, m) = (2, 4, 1)$. We want to find out the approximate change in demand for the two goods as a result of a simultaneous change in prices and income.

- We totally differentiate each component function q_i.

$$\begin{aligned} dq_1 &= \frac{\partial q_1}{\partial p_1}dp_1 + \frac{\partial q_1}{\partial p_2}dp_2 + \frac{\partial q_1}{\partial m}dm \\ &= (-2p_1^{-2}p_2^3m^2)dp_1 + (6p_1^{-1}p_2^2m^2)dp_2 + (4p_1^{-1}p_2^3m)dm \\ &= -32dp_1 + 48dp_2 + 32dm \qquad \text{at (2,4,1),} \end{aligned}$$

$$\begin{aligned} dq_2 &= \frac{\partial q_2}{\partial p_1}dp_1 + \frac{\partial q_2}{\partial p_2}dp_2 + \frac{\partial q_2}{\partial m}dm \\ &= (6p_1p_2^{-2}m)dp_1 + (-6p_1^2p_2^{-3}m)dp_2 + (4p_1^2p_2^{-2})dm \\ &= (3/4)dp_1 - (3/8)dp_2 + dm \qquad \text{at (2,4,1),} \end{aligned}$$

- Suppose the price of good 1 rises by 0.1 and the price of good 2 falls by 0.1 ($dp_1 = 0.1$, $dp_2 = -0.1$) and that income rises by 0.1 ($dm = 0.1$). Then $dq_1 = -11.2$ and $dq_2 = 0.2125$.

- In matrix notation

$$\begin{pmatrix} dq_1 \\ dq_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial q_1}{\partial p_1} & \frac{\partial q_1}{\partial p_2} & \frac{\partial q_1}{\partial m} \\ \frac{\partial q_2}{\partial p_1} & \frac{\partial q_2}{\partial p_2} & \frac{\partial q_2}{\partial m} \end{pmatrix} \begin{pmatrix} dp_1 \\ dp_2 \\ dm \end{pmatrix}$$

- So the changes in q_1 and q_2 in the tangent hyperplane at the point $(2,4,1)$ are

$$\begin{aligned} \begin{pmatrix} dq_1 \\ dq_2 \end{pmatrix} &= \begin{pmatrix} -32 & 48 & 32 \\ \frac{3}{4} & -\frac{3}{8} & 1 \end{pmatrix} \begin{pmatrix} 0.1 \\ -0.1 \\ 0.1 \end{pmatrix} \\ &= \begin{pmatrix} -11.2 \\ 0.2125 \end{pmatrix} \end{aligned}$$

- We can compare this linear approximation to the actual change in the function $q = (q_1, q_2)$ which can be calculated by substitution. The actual change is

$$\Delta q = (\Delta q_1, \Delta q_2) = (-7.506, 0.120)$$

  to three decimal places. ♦

Theorem 9 (Chain Rule III)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R} \to \mathbb{R}^n$be continuously differentiable functions. Let $h = f \circ g$ be the composite function from $\mathbb{R}$ to $\mathbb{R}^m$. Then h is continuously differentiable, and*

$$h'(a) = D(f \circ g)(a) = Df(g(a))g'(a).$$

for all $a \in \mathbb{R}$. That is

$$\begin{pmatrix} h_1'(a) \\ h_2'(a) \\ \vdots \\ h_m'(a) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(g(a)) & \frac{\partial f_1}{\partial x_2}(g(a)) & \cdots & \frac{\partial f_1}{\partial x_n}(g(a)) \\ \frac{\partial f_2}{\partial x_1}(g(a)) & \frac{\partial f_2}{\partial x_2}(g(a)) & \cdots & \frac{\partial f_2}{\partial x_n}(g(a)) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(g(a)) & \frac{\partial f_m}{\partial x_2}(g(a)) & \cdots & \frac{\partial f_m}{\partial x_n}(g(a)) \end{pmatrix} \begin{pmatrix} g_1'(a) \\ g_2'(a) \\ \vdots \\ g_n'(a) \end{pmatrix}.$$

- The ith component of the above derivative is

$$\begin{aligned} h_i'(a) = Df_i(g(a)) \cdot g'(a) &= \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(g_1(a), \ldots, g_n(a)) g_j'(a) \\ &= \frac{\partial f_i}{\partial x_1}(g(a)) g_1'(a) + \cdots + \frac{\partial f_i}{\partial x_n}(g(a)) g_n'(a) \end{aligned}$$

Example 15. Consider the demand functions from the previous example, and suppose now that p_1, p_2 and m vary over time according to the equations

$$p_1(t) = t^2 + 1, \quad p_2(t) = 4t, \text{ and } m(t) = \sqrt{t}.$$

We want to know the rate of change of demand with respect to time at $t = 1$.

- First note that $(p_1(1), p_2(1), m(1)) = (2, 4, 1)$. Therefore

$$\begin{aligned} \begin{pmatrix} \frac{dq_1}{dt}(1) \\ \frac{dq_2}{dt}(1) \end{pmatrix} &= \begin{pmatrix} \frac{\partial q_1}{\partial p_1}(p(1)) & \frac{\partial q_1}{\partial p_2}(p(1)) & \frac{\partial q_1}{\partial m}(p(1)) \\ \frac{\partial q_2}{\partial p_1}(p(1)) & \frac{\partial q_2}{\partial p_2}(p(1)) & \frac{\partial q_2}{\partial m}(p(1)) \end{pmatrix} \begin{pmatrix} p_1'(1) \\ p_2'(1) \\ m'(1) \end{pmatrix} \\ &= \begin{pmatrix} -32 & 48 & 32 \\ \frac{3}{4} & -\frac{3}{8} & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ \frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} 144 \\ \frac{1}{2} \end{pmatrix} \end{aligned}$$

  gives the rate of change of demand over time at $t = 1$. ♦

Theorem 10 (Chain Rule IV)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^s \to \mathbb{R}^n$ be continuously differentiable functions. Let $h = f \circ g$ be the composite function from $\mathbb{R}^s$ to $\mathbb{R}^m$. Then h is continuosly differentiable, and*

$$Dh(a) = D(f \circ g)(a) = Df(g(a))Dg(a).$$

for all $a \in \mathbb{R}^s$.

- Here $Df(g(a))$ is an $m \times n$ Jacobian matrix and $Dg(a)$ is an $n \times s$ Jacobian matrix.
- The product of these matrices is an $m \times s$ Jacobian matrix.
- Note that this chain rule is the most general and nests all the other three.
- Writing out the matrices explicitly, the chain rule is:

$$\begin{pmatrix} \frac{\partial h_1}{\partial x_1}(a) & \cdots & \frac{\partial h_1}{\partial x_s}(a) \\ \frac{\partial h_2}{\partial x_1}(a) & \cdots & \frac{\partial h_2}{\partial x_s}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1}(a) & \cdots & \frac{\partial h_m}{\partial x_s}(a) \end{pmatrix} =$$

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(g(a)) & \cdots & \frac{\partial f_1}{\partial x_n}(g(a)) \\ \frac{\partial f_2}{\partial x_1}(g(a)) & \cdots & \frac{\partial f_2}{\partial x_n}(g(a)) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(g(a)) & \cdots & \frac{\partial f_m}{\partial x_n}(g(a)) \end{pmatrix} \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(a) & \cdots & \frac{\partial g_1}{\partial x_s}(a) \\ \frac{\partial g_2}{\partial x_1}(a) & \cdots & \frac{\partial g_2}{\partial x_s}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1}(a) & \cdots & \frac{\partial g_n}{\partial x_s}(a) \end{pmatrix}.$$

Higher Order Derivatives.

- The partial derivative $\partial f/\partial x_i$ of a function given by $f(x_1, \ldots, x_n)$is itself a function of n variables. We can continue taking partial derivatives of these partial derivatives.
- Sometimes it is not possible to partially differentiate a function with respect to some variable. So we need some terminology describing how "smooth" functions are.

Definition. Let U be an open subset of $\mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be a function.

- We say f is *k-times differentiable* at $a \in U$ if all its partial derivatives of order less than k exist. If this is true for all $a \in U$, we say f is *k-times differentiable on U*.
- We say f is *k-times continuously differentiable* or C^k at a if all its partial derivatives exist and are continuous at a. If this is true for all $a \in U$, we say f is *k-times continuously differentiable* or C^k *on U*. ▲

- There are several types of notation you might see. Consider the function $y = f(x_1, \ldots, x_n)$.
- For the first order partial derivative, we had the notation

$$\frac{\partial f}{\partial x_i} = f_i = f_{x_i} = D_i f.$$

- For second order *own partial derivatives* we have

$$\frac{\partial^2 f}{\partial x_i^2} = f_{ii} = f_{x_i x_i} D_{ii} f.$$

- For second order *cross partial* or *mixed derivatives* we have

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = f_{ij} = f_{x_i x_j} = D_{ij} f.$$

- For higher order partial and mixed derivatives we have

$$\frac{\partial^{r+s+t} f}{\partial x_i^r \partial x_j^s \partial x_k^t}.$$

Example 16. Consider the Cobb-Douglas utility function $u : \mathbb{R}^2 \to \mathbb{R}$ given by $u(x,y) = 5x^{\frac{1}{5}} y^{\frac{4}{5}}$. We will find the second-order derivatives of u.

- First find the first order partial derivatives:

$$\frac{\partial u}{\partial x} = x^{-\frac{4}{5}} y^{\frac{4}{5}} \quad \text{and} \quad \frac{\partial u}{\partial y} = 4x^{\frac{1}{5}} y^{-\frac{1}{5}}.$$

- Now find the second order own partial derivatives:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial x}\left(x^{-\frac{4}{5}} y^{\frac{4}{5}}\right) = -\frac{4}{5} x^{-\frac{9}{5}} y^{\frac{4}{5}},$$

and

$$\frac{\partial^2 u}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\right) = \frac{\partial}{\partial y}\left(4x^{\frac{1}{5}} y^{-\frac{1}{5}}\right) = -\frac{4}{5} x^{\frac{1}{5}} y^{-\frac{6}{5}},$$

- Now find the second order cross partial derivatives:

$$\frac{\partial^2 u}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial y}\left(x^{-\frac{4}{5}} y^{\frac{4}{5}}\right) = \frac{4}{5} x^{-\frac{4}{5}} y^{-\frac{1}{5}},$$

and

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}\right) = \frac{\partial}{\partial x}\left(4x^{\frac{1}{5}} y^{-\frac{1}{5}}\right) = \frac{4}{5} x^{\frac{4}{5}} y^{-\frac{1}{5}},$$

♦

- Notice that the function above of two variables has four second order partial derivatives. In general, a real-valued function of n variables will have n^2 second order partial derivatives. We can array these in a matrix.

Definition. The *Hessian (matrix)* of a function $f : \mathbb{R}^n \to \mathbb{R}$ at a point a, denoted by $D^2 f(a)$ or $D^2 f_a$, is given by

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

It is the $n \times n$ matrix of cross-partial derivatives. ▲

- Note that the Hessian matrix is the derivative matrix of the vector-valued gradient function $\nabla f(x)$, i.e. $D^2 f(a) = D[\nabla f(x)]$.

- In our utility function example we had

$$\frac{\partial^2 u}{\partial y \partial x} = \frac{\partial^2}{\partial x \partial y},$$

  so that the order of differentiation did not matter.

- It turns out that for functions with continuous second order derivatives, this is always the case.

Theorem 11 (Young's Theorem)**.** *Let U be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}$ be a C^2 function. Then $D^2 f$ is a symmetric matrix, i.e. we have*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a)$$

for all $i, j = 1, \ldots, n$ and for all $a \in U$.

- This means the Hessian is a symmetric matrix, a result you will use when studying demand functions in economics.

- It means that for C^2 utility functions the substitution matrix is symmetric implying that the effect on compensated demand for good j of a rise in the price of good i is the same as the effect on compensated demand for good i of a rise in the price of good j.

- Young's theorem generalizes to the case of taking kth order partial derivatives of C^k functions.

- For example, if we take the $x_1 x_2 x_4$ derivative of order three, then

$$\begin{aligned} \frac{\partial^3 f}{\partial x_1 \partial x_2 \partial x_4} &= \frac{\partial^3 f}{\partial x_1 \partial x_4 \partial x_2} = \frac{\partial^3 f}{\partial x_2 \partial x_1 \partial x_4} \\ &= \frac{\partial^3 f}{\partial x_2 \partial x_4 \partial x_1} = \frac{\partial^3 f}{\partial x_4 \partial x_1 \partial x_2} = \frac{\partial^3 f}{\partial x_4 \partial x_2 \partial x_1}. \end{aligned}$$