Note: Since many elementary MVC textbooks cover the relation between partial derivatives and total derivative in less depth than I believe is optimal for this course (or they do it without reference to matrices), I am providing the key issues here as special notes.

Partial derivatives

Definition of partial derivatives

Everybody who can do single-variable derivatives from calculus 1&2, can already do partial derivatives: E.g., for a 2-variable function f given by the formula $f(x,y) = (x^3 + 3x^2y^2 + y^3)\sin(x^2 + 3y)$, we can choose to treat y like a (fixed) parameter and view the expression as a function of the variable x alone, and we can take the derivative with respect to x. Alternatively, we can view this expression as a function of y alone, treating x like a parameter, and taking the derivative with respect to y. These derivatives are called *partial derivatives*, and the adjective 'partial' is fitting because they provide only part of the information that should be contained in the derivative.

The prime notation f' from single variable calculus won't serve us here, because the prime doesn't tell us with which variable (x or y) we are dealing. Instead, we use the Leibniz notation; in single variable calculus that would be $df(x)/dx$. — However, in order to provide a visual sign that these are *partial* derivatives, we replace the standard d with 'curly d's' ∂. The significance in this notation will become more transparent soon; for the moment it's just a reminder that we are dealing with one variable at a time in a situation where several variables are present.

So we can write an example for partial derivatives, with the f given above:

$$
\begin{aligned}
f(x,y) &= (x^3 + 3x^2y^2 + y^3)\sin(x^2 + 3y) \\
\frac{\partial f(x,y)}{\partial x} &= (3x^2 + 6xy^2)\sin(x^2 + 3y) + (x^3 + 3x^2y^2 + y^3)\cos(x^2 + 3y)\,2x \\
\frac{\partial f(x,y)}{\partial y} &= (6x^2y + 3y^2)\sin(x^2 + 3y) + (x^3 + 3x^2y^2 + y^3)\,3\cos(x^2 + 3y)
\end{aligned}
$$

How do the curly d's read aloud? For instance 'partial df over dx' or 'dell f over dell x'.

Clarification of the function concept and appropriate notation

To explain and interpret partial derivatives, we need to work a bit on a clean language about functions. Mind the mantra that a function is not the same as a formula. Think of a function as a slot machine, that takes certain inputs and assigns a specific output to each legitimate input. The output *may* be obtained from the input by means of a formula (or by means of several formulas, in the case of piecewise defined functions). But even if one formula provides the output, the function *is* not the formula, but rather the function is the whole input-output device.

In the example of partial derivatives given above, we are actually talking about three different functions (each of which has its output given by the same formula), but they are distinguished by different input slots. Namely we are talking about one two-variable function, and two single-variable functions.

Firstly, we have the two-variable function which we assigned the name f. A more elaborate (and maybe weird-looking) name would be $f(\cdot,\cdot)$. The dots represent the input slots. We customarily name the quantity that goes in the first slot as x, and the quantity that goes in the second slot as y. (Not always; in $f(1,2)$, the quantities are explicit numbers, and they don't need to be given another name.) The full picture of what the function does is given in the following notation (which pure math folks love for its conceptual clarity and unambiguity, but which is not so often used in calculus because it is lengthy):

$$f : (x,y) \mapsto f(x,y) = (x^3 + 3x^2y^2 + y^3)\sin(x^2+3y) \quad .$$

Before the colon, you have the name of the function (f), after the colon, the explanation what the function does. You see its input slots, with (convential, but arbitrary) names (x and y) given to the variables that go in these slots. Then the assignment arrow $\mapsto$, which symbolizes the function's operation of taking the input and converting it into an output. Finally the output, with its generic conventional symbol $f(x,y)$ (function with variables filled in the slots), and the actual formula that tells how the function calculates the input.

You'll save yourself a lot of heartache with partial derivatives if you make sure that your notion of function in your brain represents this pattern and is NOT *reduced to the formula at the end. And yes, I know, the 'function=formula' misconception has served you well so far, so it's hard to get rid of it, but now you'd better dump the misconception anyway, lest it become the source of mysterious confusions down the road. Unlike the neatly groomed textbooks, I will not provide an artificially screened and protected environment designed to cater for survival with the beloved misconception.*

In a rough metaphor, the function f is like a coke machine, with the two inputs being a coin and a button-push. The output $f(x,y)$ is the coke.

Now if somebody rigs up the input-output device and keeps the button for diet cherry coke permanently selected, leaving you only one input slot, than this device is a *new* function. We can give it a completely new name like g, or we can call it $f(\cdot,y)$ with the second slot already filled with some constant y, and only the first slot ready to take an input (called x). This function is now

$$f(\cdot,y) : x \mapsto f(x,y) = (x^3 + 3x^2y^2 + y^3)\sin(x^2+3y)$$

It is a single variable function, and its derivative at x is what we called $\partial f(x,y)/\partial x$ above. So you could write $[f'(\cdot,y)](x) = \partial f(x,y)/\partial x$. (But writing it this way serves no other purpose than to illustrate a notation that may seem weird to you yet.)

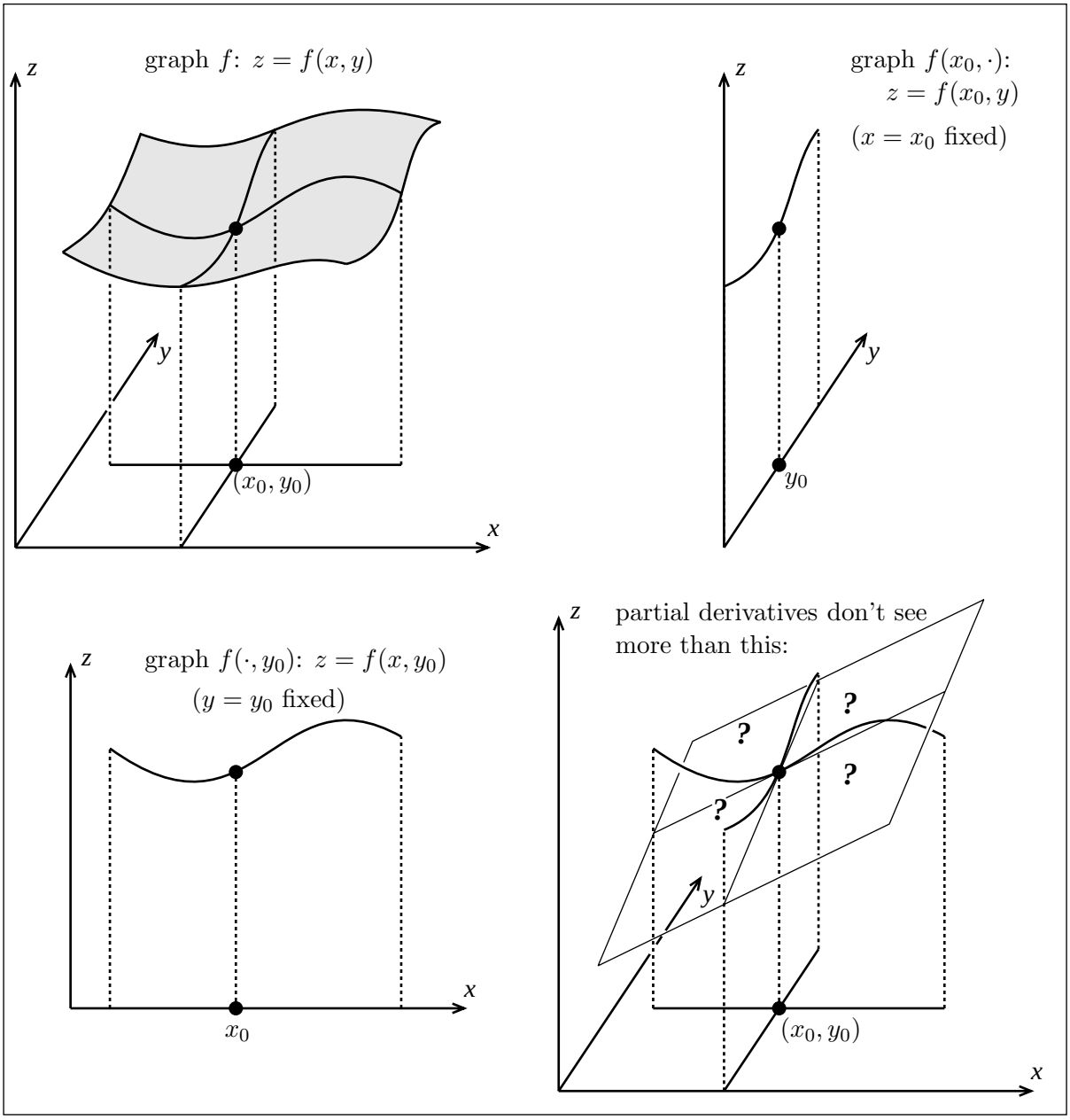
The third function we were considering in our example is

$$f(x,\cdot) : y \mapsto f(x,y) = (x^3 + 3x^2y^2 + y^3)\sin(x^2+3y)$$

Three different functions, all from the same formula!

You may not see the dot-slot notation too often, and it may be abhorrent to physicists, who might rather call the 2-variable function $f(x,y)$, and the 1-variable functions $f(x)$ and $f(y)$ respectively, using the variables to distinguish which function we are talking about, and not bothering to give a name to the function itself.

Interpretation of partial derivatives in the graph of a function



The upper left corner of this figure represents the graph of a 2-variable function f. Below, and to its right, you see the single variable functions $f(\cdot, y_0)$ and $f(x_0, \cdot)$ graphed, respectively. The graph of $f(x_0, \cdot)$ still appears tilted with respect to the paper plane, as in its orignal position. The lower right corner of the figure combines the two single-variable graphs again, and adds their tangent lines.

The partial derivatives $\frac{\partial f}{\partial x}(x_0, y_0)$ and $\frac{\partial f}{\partial y}(x_0, y_0)$ are precisely the slopes of tangent lines in the graphs of the single variable functions. If you think of these tangent lines drawn in the 3 dimensional figure, you can put a plane Π passing through these lines. This plane has also been plotted. We'd like to consider it as the tangent plane to the graph of the 2-variable function f in the point $(x_0, y_0, f(x_0, y_0))$.

In the example drawn here, the plane Π does indeed qualify as a tangent plane, and this

observation will justify calling the function f from the graph (totally) differentiable – in contradistinction to *partial* differentiability, which only refers to the existence of the partial derivatives. Look at the lower right figure: The information that enters into the partial derivatives does not 'see' anything the function does in the place where the question marks are. The 2-variable function f could be modified wildly in these quadrants without affecting the single variable functions from which the partial derivatives are calculated. In particular, it could be modified so wildly that neither Π nor any other plane could reasonably be considered to be tangent to the graph. And this brings us to the limitations of the partial derivative:

Limitations of the partial derivative, and what we will do about them

As I have just pointed out: If we were to call a multi-variable function differentiable if merely all partial derivatives exist, we would end up calling some rather wild functions 'differentiable' that do not deserve to be called differentiable.

Take for instance our old friend

$$(x,y) \mapsto f(x,y) := \begin{cases} \dfrac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

This function is not even continuous at $(0,0)$; but still the single-variable functions $f(\cdot,0) : x \mapsto 0$ and $f(0,\cdot) : y \mapsto 0$ are constant and have trivially the derivative 0.

We will define a notion of 'differentiable' that still gives us the theorem that we had in single variables, namely: "If f is differentiable, then it is continuous". In contrast, existence of the partial derivatives does not even guarantee continuity of a function, as we have just seen.

There is a practical limitation as well, and it is related to the theoretical limitation just mentioned. If the variables x and y are indeed cartesian coordinates of a point P, with the function depending on the point (geometrically), then the function $P \mapsto f(P)$ has a meaning independent of the coordinate directions we choose. We could construct a continuous function like, e.g. (note the square root this time),

$$(x,y) \mapsto g(x,y) := \begin{cases} \dfrac{xy}{\sqrt{x^2+y^2}} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

In polar coordinates, $g(x,y) = g(r\cos\varphi, r\sin\varphi) = r\sin\varphi\,\cos\varphi$. We can slice the graph of g in many directions, not only the two directions that were arbitrarily singled out as coordinate directions, and we get infinitely many single variable functions, just from slices through the origin; for instance $t \mapsto g(t,t) = \frac{1}{2}t$, if we slice along the diagonal of the x-y-plane. This time, the function is continuous in the origin, and we also have tangent lines in all coordinate directions, but these tangent lines still do not assemble into a plane.

If the input of the function has a geometric meaning, then partial derivatives single out certain directions arbitrarily as 'coordinate directions', at the neglect of other directions. (Think about the other meaning of 'partial', whose opposite is not 'total' but 'impartial'). We will also construct a notion of directional derivative, which generalizes partial derivatives in the sense that any direction can be used for getting a single-variable slice of the graph. With this notion, partial derivatives will simply be directional derivatives in coordinate directions. But again, even the existence of *all* directional derivatives in a point is not sufficient for the existence of a tangent plane, as the example g shows.

Worse even: Even if all directional derivatives in one point exist and are 0 (so that the tangent lines neatly fit together into a plane, namely a plane $z = const$, this still does not guarantee the continuity of the function in this point. The function from Homeworks #8,13 $h(x, y) := x^2y^4/(x^4+y^8)$ for $(x, y) \neq (0, 0)$ and $h(0, 0) = 0$ is an example for this phenomenon. (Details later.)

Planes, linear maps, derivatives, and matrices

Graphs that are planes; and linear maps

How does a 2-variable function T look like whose graph is a plane? – Well, all the single variable functions $T(\cdot, y) : x \mapsto T(x, y)$ should have graphs that are straight lines, and all these lines should have the same slopes: So $T(x, y) = g(y) + mx$. Since the single variable function $T(x, \cdot) : y \mapsto T(x_0, y)$ would also have to graph as a line, we need $g(y) = a + ny$ for some constants a and n. In other words, it is just the linear functions $T(x, y) = a + mx + ny$ whose graphs are planes.

In natural generalization, we call an ℓ-variable function T *linear inhomogeneous* (or: *affine*) if it is of the form $T(x_1, \ldots, x_\ell) = a + m_1x_1 + \ldots m_\ell x_\ell$ with constants a and m_j $(j = 1, \ldots, \ell)$. We call a vector-valued function $\vec{T}$ linear inhomogeneous (or: affine), if each component function T_i is a linear inhomogeneous function, i.e., if we can write, with constants a_i and m_{ij}:

$$\begin{aligned} T_1(x_1, \ldots, x_\ell) &= a_1 + m_{11}x_1 + \ldots + m_{1\ell}x_\ell \\ T_2(x_1, \ldots, x_\ell) &= a_2 + m_{21}x_1 + \ldots + m_{2\ell}x_\ell \\ &\vdots \\ T_k(x_1, \ldots, x_\ell) &= a_k + m_{k1}x_1 + \ldots + m_{k\ell}x_\ell \end{aligned} \tag{LF}$$

In the case $k = 1, \ell = 2$, which we can neatly graph in $\mathbb{R}^3$, the graph is a plane.

Note: The words 'linear inhomogeneous' or 'affine' are more popular in Linear Algebra. In Calculus, we often say simply 'linear' instead of linear inhomogeneous / affine. In Linear Algebra, the word 'linear' alone refers to the special case of (LF) where all a_i are zero.

Matrices

Matrices have been invented as a more concise notation for situations like (LF). This notation will actually condense (LF) to a form that makes it look very similar to the scalar valued single variable case $T(x) = a + mx$.

Earlier we had combined the components $x_1, \ldots, x_\ell$ into a vector, abbreviated as $\vec{x}$, which even turns out to have a geometric interpretation, so we are not merely talking about an abbreviation, but rather $\vec{x}$ is the 'real thing' and its components x_j are merely pieces of $\vec{x}$ defined in terms of an arbitrarily chosen cartesian coordinate system. In a similar spirit, we now arrange the coefficients m_{ij} into a rectangular array and call this whole array a *matrix*.

$$M := \begin{bmatrix} m_{11} & \cdots & m_{1\ell} \\ m_{21} & \cdots & m_{2\ell} \\ \vdots & \ddots & \vdots \\ m_{k1} & \cdots & m_{k\ell} \end{bmatrix}$$

More specifically we say M is a $k \times \ell$ matrix, namely a matrix with k rows and ℓ columns. We will strictly abide by the following convention: the fi**R**st index for entries of a matrix indicates the **R**ow, whereas the **second** index represents the **co**lumn in which that entry may be found.

For a matrix, an equally convenient geometric interpretation as in the case of vectors (viewed as arrows) is not visible at this moment, nevertheless you should still think of the matrix M as 'the real thing', and of its entres m_{ij} merely pieces determined from an arbitrarily chosen coordinate system.

Vectors are special cases of matrices, namely they are matrices with but one column (remember the 'vertical convention' for vectors).

> *Refer to the Linear Algebra Glossary (page 3) for the definition of multiplication of matrices ad the basic rules of matrix arithmetic; They belong here logically, but I won't repeat them here.*

Total Differentiability

Let us consider a function $\vec{f}$. Let it depend on several variables $x_1, \ldots, x_\ell$. So we can write

$$\vec{f} : (x_1, \ldots, x_\ell) \mapsto \vec{f}(x_1, \ldots, x_\ell) = \begin{bmatrix} f_1(x_1, \ldots, x_\ell) \\ \vdots \\ f_k(x_1, \ldots, x_\ell) \end{bmatrix}$$

We have written a vector valued function for generality, but the scalar valued case is included if the number k of components of $\vec{f}$ is 1. Moreover, the case $\ell = 1$ (the single variable case) is also included as a special case. For ease of notation, we treat the variables $(x_1, \ldots, x_\ell)$ as components of a vector $\vec{x} = [x_1, \ldots, x_\ell]^T$ and write $\vec{f}(\vec{x})$. Note however that the vectors $\vec{f}$ and $\vec{x}$ in general are of different nature, they may even have a different number of components.

We now select a particular input point $\vec{x}_*$. (For the moment, I don't want to call it $\vec{x}_0$ to avoid confusion between the 'index 0 for some particular point' and the vector component index.) We ask the question whether, for $\vec{x}$ close to $\vec{x}_*$, $\vec{f}(\vec{x})$ can be approximated well by some linear inhomogeneous function. Specifically, we want $\vec{f}(\vec{x})$ to be well approximated by $\vec{f}(\vec{x}_*) + T(\vec{x} - \vec{x}_*)$ for some matrix T. It does not suffice that this quantity goes to $\vec{0}$ as $\vec{x} \to \vec{x}_*$ (which would only mean continuity of $\vec{f}$). Rather it has to go to $\vec{0}$ 'faster' than $\vec{x} - \vec{x}_*$. Here is the formal definition:

Definition: *Let U be an open subset of $\mathbb{R}^\ell$ and $\vec{f}$ be a function from U into $\mathbb{R}^k$. Let $\vec{x}_* \in U$. We call $\vec{f}$ differentiable at $\vec{x}_*$ if there exists a $k \times \ell$ matrix T such that*

$$\lim_{\vec{h} \to \vec{0}} \frac{\|\vec{f}(\vec{x}_* + \vec{h}) - \vec{f}(\vec{x}_*) - T\vec{h}\|}{\|\vec{h}\|} = 0 \,. \qquad (TD)$$

A synonym, used to emphasize the distinction to partial derivatives is totally differentiable. We'll see in a moment that there can be at most one such matrix T; if T exists, we call it the (total) derivative of $\vec{f}$ in $\vec{x}_*$, and write it as $D\vec{f}(\vec{x}_*)$. Synonyms for 'total derivative' are 'functional matrix' or 'Jacobi matrix'.

With $\vec{h} = \vec{x} - \vec{x}_*$, the function $\vec{x} \mapsto \vec{f}(\vec{x_*}) + T(\vec{x} - \vec{x}_*)$ will have as its graph a tangent plane to the graph of $\vec{f}$ in $\vec{x}_*$. At least this holds in the case $k = 1$, $\ell = 2$ in which we can geometrically

draw such a plane in $\mathbb{R}^3$. In other cases, this sentence is merely a way of speaking, which, by metaphor carries over our geometric intuition into 'dimensions never beheld by human eyes'. More formally, consider this statement a definition of 'tangent plane' in these cases of higher dimension.

In the case $k = 1$, $\ell = 1$ of single variable calculus, our definition of total differentiability reduces to the old definition of differentiability for single variable functions. $Df(\vec{x}_*)$ would have to be a 1×1 matrix, which we usually identify with the number that is the one and only entry of this matrix; and this number is what was called $f'(x_*)$ in single variable calculus. Indeed, in the SV case, the norms $\|\cdot\|$ become absolute values $|\cdot|$, and our definition asserts that the limit

$$\lim_{h\to 0} \frac{f(x_* + h) - f(x_*) - Th}{h}$$

vanishes for some *number* ('1×1 matrix') T. But this means that $\lim_{h\to 0} \frac{f(x_*+h)-f(x_*)}{h}$ exists and is T. So T is the derivative $f'(x_*)$.

Let's drop the $*$ from the notation now. We study differentiability at a point $\vec{x} = [x_1\,,\ x_2\,,\ \ldots\,,\ x_\ell]^T$.

Next, we'll see that total differentiability of $\vec{f}$ implies the existence of the partial derivatives, and that the entries of T are precisely these partial derivatives. We do this by choosing special vectors $\vec{h}$, namely those that point in coordinate directions. For simplicity, assume that f is scalar valued. Let's choose $\vec{h} = [t\,,\ 0\,,\ 0\,,\ \ldots]^T$. We are looking for a (row) matrix $T = [T_1\,,\ T_2\,,\ \ldots\,,\ T_\ell]$ that satifies the definition. Note that $T\vec{h} = tT_1 + 0T_2 + \ldots + 0T_\ell = tT_1$. Differentiability requires in particular (for our chosen vector $\vec{h}$) that that

$$\lim_{t\to 0} \frac{f(x_1 + t, x_2, \ldots, x_\ell) - f(x_1, x_2, \ldots, x_\ell) - tT_1}{t} = 0\,.$$

But this identifies T_1 as the partial derivative $\partial f(\vec{x})/\partial x_1$. If we had chosen $\vec{h}$ to be $[0\,,\ t\,,\ 0\,,\ \ldots]^T$ instead, we would have selected the second entry T_2 of T and identified it with the partial derivative $\partial f(\vec{x})/\partial x_2$, and so on. It is clear from this deliberation that partial derivatives arise from the total derivative by choosing specific vectors $\vec{h}$ in the definition of differentiability. Total differentiability requires that the limit in the definition exists even without any restriction on *how* $\vec{h}$ goes to 0.

The same considerations carry over to the vector valued case. For a function $\vec{f}$ with component functions $f_1, \ldots, f_k$, the limit (TD) in the above definition will be zero if and only if the corresponding limit for each component function is 0. Our conclusion is:

If $\vec{f}$ is differentiable, then

$$D\vec{f}(\vec{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} & \cdots & \frac{\partial f_1}{\partial x_\ell} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} & \cdots & \frac{\partial f_2}{\partial x_\ell} \\ \vdots & & & & \vdots \\ \frac{\partial f_k}{\partial x_1} & \frac{\partial f_k}{\partial x_2} & \frac{\partial f_k}{\partial x_3} & \cdots & \frac{\partial f_k}{\partial x_\ell} \end{bmatrix}$$

So different rows of $D\vec{f}(\vec{x})$ correspond to different components of the function $\vec{f}$. For scalar valued functions f, the matrix $Df(\vec{x})$ is made up of only one row. — Different columns of $D\vec{f}(\vec{x})$ correspond to the different variables.

While the $k \times \ell$ matrix in this formula can always be constructed when the partial derivatives exist, this matrix only deserves the name $D\vec{f}(\vec{x})$ if $\vec{f}$ is totally differentiable at $\vec{x}$.

Only if $\vec{f}$ is totally differentiable does this matrix give an appropriate linear approximation to the function near $\vec{x}$.

From now on, I'll omit the vector arrow from $\vec{f}$, regardless of whether f is scalar valued or vector valued. I will retain the vector arrow on $\vec{x}$.

Proving Total Differentiability

Before entering into the task outlined in the headline, let's note a very easy consequence of differentiablilty:

Theorem: *If f is totally differentiable at $\vec{x}$, then it is continuous there.*

The proof is easy. If $\lim_{\vec{h}\to\vec{0}} \dfrac{\|f(\vec{x}_*+\vec{h})-f(\vec{x}_*)-T\vec{h}\|}{\|\vec{h}\|} = 0$, then in particular the numerator must go to 0. So we get $\lim_{\vec{h}\to\vec{0}} \left(f(\vec{x}_*+\vec{h})-f(\vec{x}_*)-T\vec{h}\right) = \vec{0}$ or 0 (as the case may be). Since $T\vec{h} \to \vec{0}$ or 0 automatically as $\vec{h}\to\vec{0}$, we conclude $f(\vec{x}+\vec{h}) \to f(\vec{x})$, i.e., f is continuous at $\vec{x}$.

Pedestrian Differentiability Proofs:

In principle, to prove that a function is totally differentiable, you first need to find an appropriate matrix T to be used in definition (TD), then you have to check the limit property that is required in the definition. Finding T is easy, because the matrix formed from the partial derivatives is the only possible candidate, and partial derivatives are easy to calculate. The labor then consists of checking the limit property. We'll see an example below, and anther one is in Hwk. #18.

Easy Differentiability Proofs:

Easy proofs are available if the partial derivatives you have computed as the only possible entries of the matrix T turn out to be *continuous* functions in a neighborhod of a point $\vec{x}$. (That means, of course, that they have to be continuous in the multi-variable sense; continuity of the single variable functions obtained by freezing all but one variable will not suffice.) In that case, there is a theorem that guarantees that f is differentiable at $\vec{x}$, and we save a lot of work.

Note: when I say 'continuous in a *neighborhood* of $\vec{x}$', I mean: there is a little ball around $\vec{x}$ in which the functions in question are continuous.

A proof of the theorem in question is a very useful exercise to begin understanding the notion of total differentiability; so you do not want to skip over this proof (below).

Example of a pedestrian differentiability proof:

We consider the 2-variable function $f(x,y) = \dfrac{x^2y^2}{x^2+y^2}$ for $(x,y) \neq (0,0)$ and $f(0,0) = 0$. For sake of comparison, we will also study the function $g(x,y) = \dfrac{xy}{x^2+y^2}$ for $(x,y) \neq (0,0)$ and $g(0,0) = 0$.

We prove that f is differentiable in the origin, but g is not.

First we note that $f(x,0) = 0x^2/(x^2+0) = 0$ for $x \neq 0$, and of course $f(0,0) = 0$ also. So the single variable function $x \mapsto f(x,0)$ is the constant 0. Its derivative at $x = 0$ is 0. (It's derivative is 0 everywhere, but it is $x = 0$ we are interested in.) We have concluded $\frac{\partial f}{\partial x}(0,0) = 0$. The very same argument applies to show $\frac{\partial f}{\partial y}(0,0) = 0$. The only matrix T that *could be* $Df(0,0)$ is $[0\,,\ 0]$.

So far, the very same could be said for g, with the same calculations except for the trivial change in the formula. The only matrix T that *could be* $Dg(0,0)$ is $[0\,,\ 0]$.

Next we show that T is indeed $Df(0,0)$. Thereafter, we will see that T does not qualify for $Dg(0,0)$. We choose the letters h and k for the components of the vector $\vec{h}$.

We want to show that

$$\lim_{(h,k)\to(0,0)} \frac{\left| f(0+h,0+k) - f(0,0) - [0\,,\ 0] \begin{bmatrix} h \\ k \end{bmatrix} \right|}{\sqrt{h^2+k^2}} = 0\ .$$

Since $(h,k) = (0,0)$ is *not* considered in the limit $(h,k) \to (0,0)$, we can use the $x^2y^2/(x^2+y^2)$ formula for f, and, simplifying, we have to show that

$$\lim_{(h,k)\to(0,0)} \frac{h^2k^2}{(h^2+k^2)^{3/2}} = 0\ .$$

Here is an easy trick how to do that. Note that by the agm inequality, $|hk| \le \frac{1}{2}(h^2+k^2)$ for all real numbers h,k. Therefore $|(hk)^2/(h^2+k^2)^{3/2}| \le \frac{1}{4}(h^2+k^2)^{1/2}$, and this goes to 0 trivially as $(h,k) \to (0,0)$. This proves that f is totally differentiable at the origin.

Now if we try to do the same on g, we would have to show that $|hk|/(h^2+k^2)^{3/2} \to 0$ as $(h,k) \to 0$. This however is not true. For instance, if we approach the origin along the diagonal $k = h$, we get $h^2/(2h^2)^{3/2} = 2^{-3/2}|h|^{-1}$, which does not go to 0. So g is not differentiable at the origin. – Of course we knew this from the onset, because differentiability implies continuity, and we had seen that g isn't even continuous at the origin.

Proof that continuous partials imply total differentiability:

We do this for a 3-variable function $(x,y,z) \mapsto f(x,y,z)$, using $\vec{h} = [h\,,\ k\,,\ l]^T$. The general case can be worked out completely analogously, except for bulkier writeup. In the numerator, we have to consider the difference $f(\vec{x}+\vec{h}) - f(\vec{x}) - T\vec{h}$, where T is the matrix of partial derivatives. Specifically, we have to deal with the difference

$$\text{Num} := f(x+h,y+k,z+l) - f(x,y,z) - \frac{\partial f(x,y,z)}{\partial x}h - \frac{\partial f(x,y,z)}{\partial y}k - \frac{\partial f(x,y,z)}{\partial z}l$$

In line with the fact that we have knowledge about partial derivatives, we write this as a sum of differences in such a way that in each of several terms, only one variable changes:

$$\begin{aligned}
&\text{Num} = \\
&\Big(f(x+h,y+k,z+l) - f(x,y+k,z+l)\Big) + \\
&\qquad +\Big(f(x,y+k,z+l) - f(x,y,z+l)\Big) + \\
&\qquad\qquad \Big(f(x,y,z+l) - f(x,y,z)\Big) \\
&\qquad - \frac{\partial f(x,y+k,z+l)}{\partial x}h \\
&\qquad\qquad - \frac{\partial f(x,y,z+l)}{\partial y}k \\
&\qquad\qquad\qquad - \frac{\partial f(x,y,z)}{\partial z}l \\
&+\Big(\frac{\partial f(x,y+k,z+l)}{\partial x} - \frac{\partial f(x,y,z)}{\partial x}\Big)h + \Big(\frac{\partial f(x,y,z+l)}{\partial y} - \frac{\partial f(x,y,z)}{\partial y}\Big)k
\end{aligned}$$

In this layout, the first 'staircase' just rewrites the two f terms as a 'telescoping sum'; the second 'staircase' models the partials we have in the formula for Num, but we have changed the arguments to match the ones in the first 'staircase'. The last line merely corrects for the modifications made in the second 'staircase'.

Let's begin with what this last line contributes to the fraction in (TD); to this end we throw in the denominator $\sqrt{h^2+k^2+l^2}$ again:

$$\Big(\frac{\partial f(x,y+k,z+l)}{\partial x}-\frac{\partial f(x,y,z)}{\partial x}\Big)\frac{h}{\sqrt{h^2+k^2+l^2}}+\Big(\frac{\partial f(x,y,z+l)}{\partial y}-\frac{\partial f(x,y,z)}{\partial y}\Big)\frac{k}{\sqrt{h^2+k^2+l^2}}$$

The fractions have absolute value ≤ 1, and the differences in the parentheses go to 0, because the partials are continuous.

Now we combine matching steps in the two staircases. The first of them contributes

$$\frac{f(x+h,y+k,z+l)-f(x,y+k,z+l)-\frac{\partial f(x,y+k,z+l)}{\partial x}h}{h}\,\frac{h}{\sqrt{h^2+k^2+l^2}}$$

to the fraction in (TD). Here we notice that $f(x+h,y+k,z+l)-f(x,y+k,z+l)=\frac{\partial f(*,y+k,z+l)}{\partial x}\,h$ by the mean value theorem for the single variable function $f(\cdot,y+k,z+l)$, where $*$ is some number between x and $x+h$. Again we have exhibited a contribution that goes to 0 as $h\to 0$ by the continuity of the partials.[1] The same reasoning applies for the other two steps of the staircases. So if we put the quantity Num into formula (TD), we obtain a sum of terms, each of which goes to 0 as $(h,k,l)\to(0,0,0)$. And this proves total differentiability of f at (x,y,z).

Directional Derivative, and Geometric Interpretation of $Df(\vec{x})$ as 'Vector Eater'

We have seen that total differentiability implies the existence of partial derivatives. To see this, we merely had to choose for the vector $\vec{h}$ vectors $t[1,0,0,\ldots]^T$, $t[0,1,0,\ldots]^T$ etc: vectors pointing in coordinate directions. Let us instead use vectors $\vec{h}:=t\vec{v}$ with $\vec{v}$ a fixed vector pointing in any direction, coordinate or not.

We then get a single variable function $t\mapsto f(\vec{x}+t\vec{v})$, which is obtained by restricting the multi-variable function f to inputs on the line $\{\vec{x}+t\vec{v}\mid t\in\mathbb{R}\}$. If the derivative of this single variable function at $t=0$ exists, we call this quantity the *directional derivative* of f at $\vec{x}$ in direction $\vec{v}$, and denote it as $\partial_{\vec{v}}f(\vec{x})$. In formulas

$$\partial_{\vec{v}}f(\vec{x}):=\frac{d}{dt}f(\vec{x}+t\vec{v})\Big|_{t=0}$$

Some authors use the word 'directional' derivative only if $\vec{v}$ has length 1, because the quantity in question depends both on the direction and the length of $\vec{v}$. Only by normalizing (fixing) the length a-priori do we get a quantity that depends only on the direction. In this class however, I will not restrict the length of $\vec{v}$ and accept the drawback that the word 'directional derivative' could then be slightly misleading.

[1] If you have a really excellent Hons Calc 2 vision, you'll see that we actually use that the partials are *uniformly* continuous, which is implied by continuity on a bounded and closed domain. If you don't see this sublety, ignore it in peace for now and try again seeing it after the course Math 341.

It is a healthy (and hopefully simple) exercise for you to prove the following
Theorem: *If f is totally differentiable at $\vec{x}$, then it has a directional derivative in each direction $\vec{v}$, and this derivative equals $Df(\vec{x})\vec{v}$.*

As with partial derivatives, even the existence of all directional derivatives in a point does not guarantee total differentiability, as is seen in Homework #13.

I used the symbol $\vec{v}$ for the direction vector and refrained from enforcing length 1 on it. The idea I have in mind is that $\vec{v}$ may be a velocity. Think of a function $f : \vec{x} \mapsto f(\vec{x})$ as a temperature function, depending on a location $\vec{x} \in \mathbb{R}^3$. Now if I start out at $\vec{x}$, thermometer in hand, and move with velocity $\vec{v}$, I'll be at location $\vec{x} + t\vec{v}$ at time t. My thermometer records the temperature at each time t. That is, it records the temperature at the location where I am at time t. The rate of change of this temperature with respect to time is what we called directional derivative. Of course if I move faster, I'll experience faster temperature changes: this accounts for the dependence on the length of $\vec{v}$ that is being hidden by the name 'directional derivative'. But more significantly, the rate of change of the temperature will in general depend on the direction in which I am moving. This issue is absent in single variable calculus, because there is only one direction on the real line. ('Negative direction' is nothing but (-1) times positive direction, so it contributes no independent information about rates of change.[2])

The notion of directional derivative is useful to understand why the total derivative has to be such a 'bulky' object like a matrix: It needs to have many pieces of information incorporated in it. In single variable calculus, every change dx in the input x is a multiple of one standard change $+1$. To tell how the output changes, all that is needed is one number $f'(x)$ that gives the amplification of the input change dx into an output change $dy = f'(x)dx$ (of course in linear approximation only). In multivariable calculus, if there is a notion of derivative that is to tell you the rate of change of the output $f(\vec{x})$ as you change the input $\vec{x}$, this thing 'derivative' must ask back: 'In which direction do you change the input?' So it asks for a vector $\vec{v}$, and in response it gives you a rate of change. Seen from this vantage point, it is clear that the derivative $Df(\vec{x})$ is *not* a vector, even though it has as many entries as a vector. Rather it is a 'vector eater': You must feed it a vector and it produces for you a rate of change (which is a number or a vector, depending on whether f is scalar valued or vector valued).

This distinction is reflected in the row vs column distinction: columns represent vectors, rows represent 'vector eaters'. They are called 'forms' in more advanced mathematical contexts, but let's keep the more descriptive word 'vector eater' just for fun for the purposes of this class.

In some MVC textbooks, you will see this distinction omitted 'for simplicity'. Such simplification is perfectly good for crunching calculational problems, but it comes at the expense of disconnecting the geometric intuition from the calculational formalism.

> An outlook far ahead: There are two 'upgrades' of MVC that you may encounter in more advanced courses:
>
> You may study 'infinitely many variables' (called functional analysis). In that context, the distinction between vector eaters and vectors becomes much more substantial and cannot be covered up by simply converting a row into a column.

[2] In linear algebra language, if you know it, I'd say that there is only one *linearly independent* direction on the real line

You may study multi-variable problems in which the variables are coordinates describing a curved surface (like longitude and latitude on a sphere). We can do this already now. But in a more advanced setting (called differential geometry), you may want to have the language reflect geometric issues precisely, in particular you may want to be able to write objects that should be independent of coordinates in a way that doesn't make formulas *look* as if things did depend on our choice of a coordinate system. This idea is foundational for modern developments of physics, and in particular, general relativity theory relies on the formalism developed in differential geometry. Here the conversion between vector eaters and vectors is explicitly dependent on core geometric information (physically measurable information) and cannot be made without reference to such geometric information. (Reference that is taken for granted without any discussion in the case of $\mathbb{R}^3$.)

In both of these situations, you'd create a lot of confusion if you trashed the distinction between vectors and vector eaters. The conviction underlying these course notes is that only a formalism that is ready to accomodate these generalizations naturally at a later time, will be a formalism that is genuinely intuitive for the purposes here and now. (As a matter of fact, the big book on Gravitation by Misner, Thorne and Wheeler contributed a good deal to my own understanding of how intuition and formalism connect in MVC.)

The Gradient

In this section, we consider only scalar valued functions f. And we assume that f is totally differentiable. Much of what we will do can be done already if only the partial derivatives exist (and some books define the gradient only in terms of partial derivatives, regardless of total differentiability or not). But such generality will serve no purpose for us at the moment; rather it would make the language clumsy when explaining some geometry highlights.

You will often see the partial derivatives being considered as components of a vector, called the gradient, written as $\nabla f(\vec{x})$. The symbol ∇ is called 'nabla'. With the vertical convention for vectors, $\nabla f(\vec{x}) = Df(\vec{x})^T$. The gradient is the transpose of the functional matrix of a scalar valued function.

Doing this is NOT in defiance of the geometric distinction I have stressed so far. Rather, the gradient has a geometric meaning of its own, which we will explore. The geometric distinction stressed before only amounts to insisting that the gradient is *not the same* as the derivative, but rather *is the transpose* of the derivative. We will explore the geometric meaning of the gradient here. By implication, if there is geometric significance in distinguishing $Df(\vec{x})$ from $\nabla f(\vec{x})$, there must be some hidden geometric meaning in the innocent-looking formal operation of transposing a row into a column. It is not so easy to tickle this geometric contents out at the level of a MVC course. The difficulty is of the same nature as the difficulty of explaining water to a fish. The fish will understand better when he gets out of the water. But I'll try it anyways, just for the heck of it, and for reference if you want to have another look later.

For reference, let me quote the familiar here:

$$Df(\vec{x}) = \begin{bmatrix} \frac{\partial f(\vec{x})}{\partial x_1} & \frac{\partial f(\vec{x})}{\partial x_2} & \cdots & \frac{\partial f(\vec{x})}{\partial x_\ell} \end{bmatrix}, \quad \nabla f(\vec{x}) = Df(\vec{x})^T = \begin{bmatrix} \frac{\partial f(\vec{x})}{\partial x_1} \\ \frac{\partial f(\vec{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\vec{x})}{\partial x_\ell} \end{bmatrix}$$

The directional derivative is

$$\partial_{\vec{v}} f(\vec{x}) = Df(\vec{x})\vec{v} = \nabla f(\vec{x}) \cdot \vec{v}$$

The product in $Df(\vec{x})\vec{v}$ is a matrix product, the product in $\nabla f(\vec{x}) \cdot \vec{v}$ is the dot product of vectors.

For the moment, we now do fix the length of $\vec{v}$ to be 1, since we will now be interested in effects of the direction of $\vec{v}$ only; we ask the question: In which direction $\vec{v}$ is the rate of change of f largest? You may be inclined to use calculus to answer this question, since it is a maximum problem after all. But algebra does it much more easily: We note, from the Cauchy-Schwarz inequality, that $\nabla f(\vec{x}) \cdot \vec{v} \leq \|\nabla f(\vec{x})\| \, \|\vec{v}\| = \|\nabla f(\vec{x})\|$. If $\vec{v}$ actually has the same direction as $\nabla f(\vec{x})$, then the dot product is equal to $\|\nabla f(\vec{x})\|$ by the geometric definition of the dot product, or by direct calculation with $\vec{v} := \nabla f(\vec{x})/\|\nabla f(\vec{x})\|$.

For all other directions, the directional derivative is strictly less than $\|\nabla f(\vec{x})\|$. Geometrically this is because then the $\cos\varphi$ in the definition of the dot product is stritly < 1. Algebraically speaking, we can see the same thing from a second look into the proof of Cauchy Schwarz. If we do this, we see that $\vec{a} \cdot \vec{b} = \|\vec{a}\| \, \|\vec{b}\|$ only if $\vec{a} = t\vec{b}$. (Assuming $\vec{b} \neq \vec{0}$.)

So here is what we conclude:

> The direction of $\nabla f(\vec{x})$ is the direction in which we have to go from $\vec{x}$ in order to experience the greatest rate of change. The rate of change we experience in this direction is the length (norm) of $\nabla f(\vec{x})$. If we move at right angle to $\nabla f(\vec{x})$, then the rate of change experienced is 0 (because in the dot product, the cosine of the angle is 0).

The following discussion is a tad informal and will become more rigorous after we have covered the multi-variable version of the chain rule: Assume we move not along a line $\vec{x}+t\vec{v}$ but along a level set, on which f is constant by definition of level set. The derivative (with respect to time t as we are moving) is therefore 0. At any moment, the velocity vector will be tangent to the level set, because we are moving within the level set. If the fact that we are not actually exploring f along a straight line but along a bent path doesn't cause trouble (and the chain rule will tell us it doesn't), we should still observe the directional derivative in direction $\vec{v}$, which is tangential to the level set. Since this directional derivative is 0, we would have to be moving orthogonal to the gradient (unless the gradient vanishes, in which case it does not specify a direction at all).

> This means that the gradient will always be orthogonal to the level sets of a function.

The following facts can be proved rigorously with more advanced methods, but can and should be appreciated at this stage: We consider a *continuously differentiable function* of two or three variables. (Could be more variables also, but I want to refer to your geometric intuition). Continuously differentiable means (a) differentiable and (b) the partial derivatives are continuous functions. In this case, the matrix-valued function $\vec{x} \mapsto Df(\vec{x})$ is continuous automatically. Then the following facts hold for level sets of f:

For two variables $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$: At any point $\vec{x}$ where $\nabla f(\vec{x})$ is NOT the zero vector, the level set that passes through $f(\vec{x})$ looks like a smooth curve (graph of a continuously differentiable function $y = g(x)$, or $x = h(y)$) in some ball around that point $\vec{x}$. (Look in particular at the level sets that were the solution of Hwk #11.)

For three variables $\vec{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$: At any point $\vec{x}$ where $\nabla f(\vec{x})$ is NOT the zero vector, the level set that passes through $f(\vec{x})$ looks like a smooth surfacee (graph of a continuously differentiable function $z = g(x, y)$, or $y = h(x, z)$, or $x = k(y, z)$) in some ball around that point $\vec{x}$.

At points where $\nabla f(\vec{x}) = \vec{0}$, the level sets may look weird or 'untypical': The following list of 'building blocks' for level sets in two variables is not exhaustive, but features the most common examples: At points where $\nabla f(\vec{x}) = \vec{0}$, the level set may consist of a single isolated point, or it could feature two (or sometimes more) smooth curves that are crossing each other. The level set *might* also look like a smooth piece of curve, giving no indication of the vanishing gradient.

We call any point where the gradient of f vanishes a *critical point* of f. The relevance of this notion is the following: If f has a local minimum or a local maximum at an interior point $\vec{x}_*$ of the domain of f, then the gradient of f vanishes there. (Can you see why? This can be seen using the single variable slice functions only.) Conversely, the vanishing of $\nabla f(\vec{x}_*)$ is no guarantee that f has a minimimum or a maximum at $\vec{x}_*$. (As in single variables, where the vanishing of the derivative doesn't guarantee a minimum or a maximum either.) A new alternative to minimum and maximum that occurs with several variables is the possibility of saddle points. A saddle point is one that looks like a single variable maximum in some directions and like a single variable minimum in some other directions. The origin is a saddle point in Hwk. #11. A level line that goes through a saddle point will typically have a crossing there. We will require second derivatives to distinguish minima, maxima, and saddle points, and this will be studied later.

Rules for differentiation; in particular the chain rule

The following simple differentiation rules carry over from single variable calculus and are easy to prove.

• The sum of differentiable functions is differentiable. If $h = f + g$, then $Dh(\vec{x}) = Df(\vec{x}) + Dg(\vec{x})$. (Similarly for differences.)

• The product of scalar valued differentiable functions is differentiable. If $h = fg$, then $Dh(\vec{x}) = f(\vec{x})Dg(\vec{x}) + Df(\vec{x})g(\vec{x})$. The products on the right hand side are of course 'scalar times matrix'.

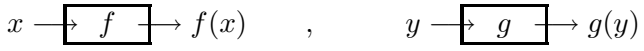
• The ratio of scalar valued differentiable functions is differentiable where the denominator doesn't vanish. If $h = f/g$, then $Dh(\vec{x}) = -\frac{f(\vec{x})}{g(\vec{x})^2}Dg(\vec{x}) + Df(\vec{x})\frac{1}{g(\vec{x})}$.

• The single-variable product rule carries over to the dot product of vector valued functions. If $h = \vec{f} \cdot \vec{g}$, then $h'(t) = \vec{f}'(t) \cdot \vec{g}(t) + \vec{f}(t) \cdot \vec{g}'(t)$.

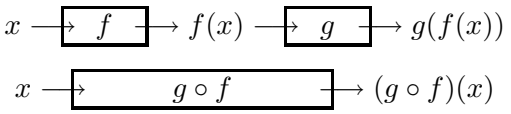
The one rule that requires discussion and training is the chain rule. Actually it also carries over without modification from single variable calculus if you rely on the total derivative and matrix multiplication consistently. However, most of the time you will use it in a form that

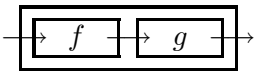
involves partial derivatives, and then it *looks* different from the single variable version. Our approach here will state the chain rule in matrix form first, then explore what it means to get some understanding for its inner workings (which includes an informal proof), and finally provide a formal proof.

Let's first review the chain rule from single variables. Remember to distinguish the name f of a function from its output (value) $f(x)$. So here is a graphical representation of the functions $f : x \mapsto f(x)$ and $g : y \mapsto g(y)$, where you should remember that the names x or y chosen for the variables are arbitrary, albeit common in the context in which we are using the functions here.

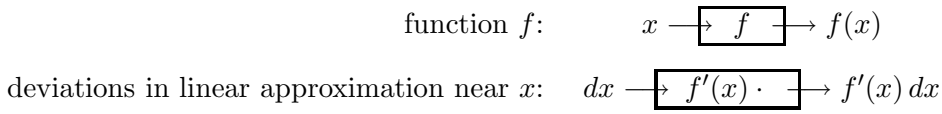


We can concatenate these two functions by feeding the output of the first function f as input into the second function g. The concatenated function bears the name $g \circ f$ (NOT $f \circ g$), because its value for input x is $g(f(x))$. Compositions are to be read 'from right to left' because in our notation the input value x stands to the right of the function symbol.

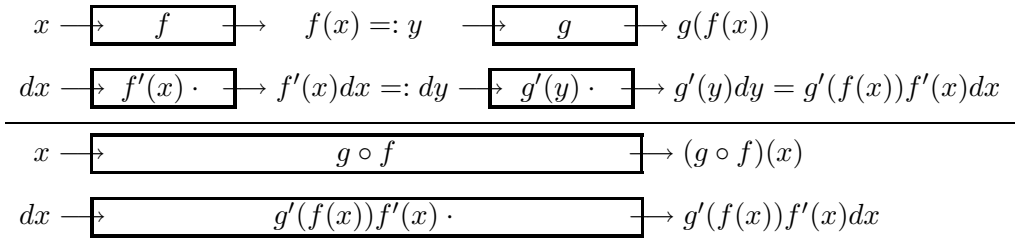


So $g \circ f$ is the name for the whole assembly  , and $(g \circ f)(x) = g(f(x))$.

Now if you change the input x to the function f by a small amount dx, the output $f(x)$ will be changed by an amount that is approximately $f'(x)dx$. (This linear approximation is only useful if dx is sufficiently small.) The derivative $f'(x)$ gives the amplification factor for small input errors dx, using linear approximation. Rather than viewing $f'(x)$ as a *number* that *gives an amplification factor*, we may view the derivative at x as a *linear function* itself, that *does the amplification*. The natural name for this linear function would be "$f'(x)$ times", because that's what it does: it multiplies an input dx by $f'(x)$:



Introducing the multiplication point after $f'(x)$ and viewing $f'(x)\cdot$ as a 'linear error amplification function' is the key to understanding the chain rule intuitively. This is true for single variables already, but it becomes particularly useful for multi-variable. So let's understand the chain rule in these terms:



This is NOT a proof of the chain rule, because we use as input into the second error amplification function not the actual deviation Δy, but instead the approximate deviation dy; an actual proof would need to give an account of how this error influences the outcome. The answer would be: In *linear* approximation, the effect cannot be seen; it only shows up if we study better than linear approximations (like 2nd order Taylor approximation).

But apart from this proof detail, this picture makes us *understand* why $(g \circ f)'(x) = g'(f(x))f'(x)$.

Now the punchline is that the very same argument carries over almost literally to the multi-variable setting: This is a benefit of working with the total derivative as the primary object and viewing the partial derivatives as 'parts' of the total derivative, rather than viewing the partial derivatives as the primary pieces of information that need to be 'somehow organized into a matrix or vector or whatever'.

The only changes that we need to make are: f and g may be vector valued, x and y may be vectors now, and instead of f', we have chosen to call the derivative Df. The input error 'amplification' is not merely achieved by multiplying with a number, but rather by multiplying with a matrix. This distinction is very natural, because deviation in different input variables may have different effects on the output; and matrix multiplication can achieve this effect, whereas multiplication by mere numbers cannot. So let's redo the previous picture in the new notation:

$$\begin{array}{l}
\vec{x} \longrightarrow \boxed{\quad \vec{f} \quad} \longrightarrow \quad \vec{f}(\vec{x}) =: \vec{y} \quad \longrightarrow \boxed{\quad \vec{g} \quad} \longrightarrow \vec{g}(\vec{f}(\vec{x})) \\
d\vec{x} \longrightarrow \boxed{D\vec{f}(\vec{x})\cdot} \longrightarrow D\vec{f}(\vec{x})d\vec{x} =: d\vec{y} \longrightarrow \boxed{D\vec{g}(\vec{y})\cdot} \longrightarrow D\vec{g}(\vec{y})d\vec{y} = D\vec{g}(\vec{f}(\vec{x}))D\vec{f}(\vec{x})d\vec{x} \\
\hline
\vec{x} \longrightarrow \boxed{\qquad \vec{g} \circ \vec{f} \qquad} \longrightarrow (\vec{g} \circ \vec{f})(x) \\
d\vec{x} \longrightarrow \boxed{\qquad D\vec{g}(\vec{f}(\vec{x}))D\vec{f}(\vec{x})\cdot \qquad} \longrightarrow D\vec{g}(\vec{f}(\vec{x}))D\vec{f}(\vec{x})d\vec{x}
\end{array}$$

When I put vector symbols over 'everything', I do not mean to say that all these quantities *must* be vectors. The scalar case is included as special case with 1-component vectors. The different vectors may have differently many components, if only the 'chain' fits together: For instance, $\vec{x}$ may have 3 components, and $\vec{f}(\vec{x})$ may have 2 components. Then the input variable $\vec{y}$ for $\vec{g}$ must also have two components, else the chain doesn't fit together; but then the output $\vec{g}(\vec{y})$ may have any number of components. The sizes of the matrices $D\vec{f}(\vec{x})$ and $D\vec{g}(\vec{y})$ are accordingly, and the size restriction on matrix multiplication is automatically satisfied!

Now with the theory all neat and slick, all we need to understand is what this matrix form of the chain rule means in practice for the crummy partial derivatives with which we do all the practical calculations. Let's do this in an example: We take a 3-variable function g (scalar valued), and assume its arguments x, y, z are themselves dependent on parameters s and t: Let's say $x = f_1(s,t)$, $y = f_2(s,t)$ and $z = f_3(s,t)$. If we insert these into g, we get $g(x,y,z) = g(f_1(s,t), f_2(s,t), f_3(s,t)) =: h(s,t)$ So now $h = g \circ f$. f is a 2-variable function whose values are 3-vectors (but I will omit the arrow on top of the f), and they fit into the 3-variable function g, which in turn has numbers as values. Now we want to calculate $\partial h/\partial s$ and $\partial h/\partial t$ in terms of the partials of the f_i and g. The chain rule says: $Dh(s,t) = Dg(f(s,t))Df(s,t)$, which written out in detail, means

$$\begin{bmatrix} \frac{\partial h}{\partial s}(s,t) & \frac{\partial h}{\partial t}(s,t) \end{bmatrix} = \begin{bmatrix} \frac{\partial g}{\partial x}(f(s,t)) & \frac{\partial g}{\partial y}(f(s,t)) & \frac{\partial g}{\partial z}(f(s,t)) \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial s}(s,t) & \frac{\partial f_1}{\partial t}(s,t) \\ \frac{\partial f_2}{\partial s}(s,t) & \frac{\partial f_2}{\partial t}(s,t) \\ \frac{\partial f_3}{\partial s}(s,t) & \frac{\partial f_3}{\partial t}(s,t) \end{bmatrix}$$

This can be written out as two equations:

$$\frac{\partial h}{\partial s}(s,t) = \frac{\partial g}{\partial x}(f(s,t))\frac{\partial f_1}{\partial s}(s,t) + \frac{\partial g}{\partial y}(f(s,t))\frac{\partial f_2}{\partial s}(s,t) + \frac{\partial g}{\partial z}(f(s,t))\frac{\partial f_3}{\partial s}(s,t)$$

and a similar equation for the partial with respect to t. Remember that the $f(s,t)$ inside g actually stands for three variables $(f_1(s,t), f_2(s,t), f_3(s,t))$.

With the identification $x = f_1$, $y = f_2$, $z = f_3$ (that is usually done with the physicist's convention about functions) and a common name like u for the output varaible of both g and $h = g \circ f$, this is often abbreviated as

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial s}$$

This is how you will find the chain rule in many books and many contexts. I have deliberately started with an involved and detailed notation, and then moved to this succinct and easy-to-remember version. The reason is that this 'easy' notation is ambiguous, and it is only the context that resolves the ambiguity. If you come to love the easy notation before having worked through the complicated one, you will find the issue of ambiguity in the curly ∂ notation rather difficult to stomach; and in situations where a hidden ambiguitiy does cause errors, it will then be very difficult to clear up the confusion. For the moment, let me make one simple comment about this issue: When we write $\partial u/\partial x$, our notation expresses which quantity varies (namely x), but it does not tell us which variables remain fixed (namely y and z). If the 'duh' answer "all other variables other than x remain fixed" really is clear enough to tell you that the other variables are y and z, then the context has resolved the ambiguity of the notation; and this hapens in many cases (but not in all). In thermodynamics, you can study the pressure of a gas as a function of volume and temperature; or you can study it as a function of volume and energy content. And then, if you take a partial with respect to volume, it is no longer clear whether the temperature or the energy content are to remain fixed. And this might make a difference.

Here is one obvious thing that can be seen from the above chain rule: curly ∂ terms cannot just be 'canceled' as you would do with the dx's and dy's in single variable calculus. And this is a very good reason why we use curly ∂'s for partial derivatives: as a reminder that formal cancellation yields WRONG results; not just sometimes, but nearly every time!

Applications of the chain rule

(1) The statement that the gradient is orthogonal to level lines (which we had discussed heuristically above) follows rigorously from the chain rule. Suppose $t \mapsto \vec{f}(t)$ describes a curve within a level set of a function g: then $g(\vec{f}(t)) = c$ for all t. The derivative of this (constant) single-variable function is therefore 0. By the chain rule,

$$0 = \frac{d}{dt} g(\vec{f}(t)) = Dg(\vec{f}(t))\vec{f}'(t) = \nabla g(\vec{f}(t)) \cdot \vec{f}'(t)$$

Now, $\vec{f}'(t)$ is tangent to the curve described by $\vec{f}(t)$. If we interprete t as a time, $\vec{f}'(t)$ is actually the velocity vector. For a 2-variable function g, the level set is typically a curve, and so $\vec{f}(t)$ must describe (part of) this curve. For 3 or more variable functions, the level set is a surface (or higher dimensional), and the curve described by $t \mapsto \vec{f}(t)$ lies in this surface. But since this argument can be made for any curve within the level surface, we still conclude that

$\nabla g(\vec{x})$ is orthogonal to the tangent of any curve in that surface, and therefore is orthogonal to the entire tangent plane in $g(\vec{x})$. But this is exactly what we mean when we say a vector is orthogonal to a surface: it is orthogonal to the tangent plane to this surface.

(2) By taking the composition of the 2-variable function $\mathsf{product} : (u, v) \mapsto uv$ with the 2-vector valued function $x \mapsto \begin{bmatrix} f(x) \\ g(x) \end{bmatrix}$, one can obtain the single variable product rule as a consequence of the multi-variable chain rule. See homework. Similarly a power rule for $\frac{d}{dx} f(x)^{g(x)}$ can be obtained.

(3) This example relies on the 'theorem' $\frac{d}{dx} \int_a^b g(x,t)\, dt = \int_a^b \frac{\partial g(x,t)}{\partial x}\, dt$. If we remember that the integral is a limit of Riemann sums and that we can differentiate sums term by term, this 'theorem' becomes plausible, but is by no means proved. The issue is that this 'theorem' pretends that the derivative of a limit (of Riemann sums) is the limit of the derivatives (of Riemann sums).

In reality, this 'theorem' is only true under certain hypotheses. I deliberately do not want to specify these hypotheses here. They are more appropriately dealt with in advanced courses. 'Easy' versions give the result under restrictive hypotheses (which in particular exclude improper integrals $\int_0^\infty$), but many interesting applications want the result under weaker hypotheses that allow for $\int_0^\infty$. More useful variants of the theorem rely on a more sophisticated notion of integral. In the present context (and only here), we are focusing on the mechanics of calculation, with a pragmatic, applied-science perspective, but all the while being aware that hypotheses *are* needed and are assumed to hold in our calculation. In most contexts in which you will want to do these calculations, the hypotheses will be satisfied.

So, suppose we have $f(x) := \int_a^x g(x,t)\, dt$. To find $f'(x)$ we consider $F(x,y) := \int_a^y g(x,t)\, dt$. Then $\partial F(x,y)/\partial x$ can be handled by differentiation under the integral sign (when the hypotheses for validity of this procedure are verified). On the other hand, $\partial F(x,y)/\partial y = g(x,y)$ by the fundamental theorem of calculus.

The chain rule says $f'(x) = \frac{d}{dx} F(x,x) = \frac{\partial F(x,y)}{\partial x}|_{y=x} + \frac{\partial F(x,y)}{\partial y}|_{y=x} \frac{\partial y}{\partial x}$. The latter partial derivative is $\frac{\partial y}{\partial x} = 1$ because $y = x$. Conclusion:

$$\frac{d}{dx} \int_a^x g(x,t)\, dt = g(x,x) + \int_a^x \frac{\partial g(x,t)}{\partial x}\, dt$$

This (kind of) example is among the most frequent usages of the MV chain rule in *practical* calculations with explicit formulas.

Cleaning up Notation a Bit: Slots vs variable names

Near the end of the section explaining the chain rule, I referred to different cultures of notation: Mathematicians (in particular pure mathematicians) prefer to give names to functions, and these names differ from the names for the variables that represent the value of a function. $f(x,y)$ is the value of the function f for input variable(s) (x,y). As such it is an expression dependent on x and y, and we can write its partial derivative with respect to x as $\frac{\partial f(x,y)}{\partial x}$. Physicists may give a name to the output variable (say z), with $z = f(x,y)$, and they would write $\frac{\partial z}{\partial x}$ instead.

However, so far, we have no notation for the partial derivative of a function itself regardless of the name(s) or value(s) of the input variable(s). Suppose $f(x,y) = x^2 + 2xy^3$. How do I write the partial with respect to x at the point $(x,y) = (2,-3)$? Should I write $\frac{\partial f(2,-3)}{\partial x}$. I don't like

this, because there is no x left in the 'numerator' with respect to which I could differentiate. Should I write $\frac{\partial f}{\partial x}(2,-3)$? Better, because now at least the order of operations is clear: First I take a derivative, then I plug in $(2,-3)$. But still, f is the name of a function, and the generic names for its variables are arbitrary. I could have given the very same function by $f(u,v) = u^2 + 2uv^3$, and then you would have written the same thing as $\frac{\partial f}{\partial u}(2,-3)$. The best, I think, that I can do with the previous notation is to write $\frac{\partial f(x,y)}{\partial x}|_{(x,y)=(2,-3)}$, and this is clumsy.

While you will see $\frac{\partial f(x,y)}{\partial x}$ written as $\frac{\partial f}{\partial x}(x,y)$, this latter is a 'mixed' notation. While $\frac{\partial f}{\partial x}$ clearly conveys that we take a partial derivative of the function f, which we subsequently evaluate at (x,y), the function f itself does not stipulate that its input variables be given specific names. What we really mean with $\frac{\partial f}{\partial x}$ is that we take a partial with respect to the *first* variable. And it is only because it is customary to call the first variable by the name of x that the notation identifies this fact. There is a 'pure' notation to indicate this: we write $\partial_1 f$ for the derivative of f with respect to its first argument. This is analog to the notation Df for the total derivative and to the Newton notation f' for the single variable derivative. Each refers to a function with no regard to what its arguments may be called.

To illustrate this issue, let me give you an example where both notations are needed and where confusion would arise if we didn't have a clean notation: Some functions have the property that its arguments can be swapped with impunity. For instance $f : (x,y) \mapsto x + y$, and $g : (x,y) \mapsto xy$ are such functions. Let's call them symmetric for the moment. More precisely, a 2-variable function f is called symmetric, iff $f(x,y) = f(y,x)$ for all (x,y). For instance, $h(x,y) = xy^2 + yx^2$ is symmetric, but $p(x,y) = x^y$ is not symmetric. Now we want to show the following claim: If f is a symmetric function, then the function g defined by $g(x,y) := \partial f(x,y)\partial x + \partial f(x,y)\partial y$ is symmetric. You see, since the very hypothesis reads $f(x,y) = f(y,x)$, you'd be doomed if you tried to identify slots by variable names.

Here is a clean proof:

$$\frac{\partial f(x,y)}{\partial x} + \frac{\partial f(x,y)}{\partial y} = (\partial_1 f)(x,y) + (\partial_2 f)(x,y)$$

So $g = \partial_1 f + \partial_2 f$. We want to show that $g(x,y) = g(y,x)$. Now

$$g(y,x) = (\partial_1 f)(y,x) + (\partial_2 f)(y,x) = \frac{\partial f(y,x)}{\partial y} + \frac{\partial f(y,x)}{\partial x} =_*$$
$$\frac{\partial f(x,y)}{\partial y} + \frac{\partial f(x,y)}{\partial x} = (\partial_2 f)(x,y) + (\partial_1 f)(x,y) = g(x,y)$$

It is at the $=$ sign marked with $*$ that we used the hypothesis that f is symmetric.

There is one more notation you will encounter: Since the notation with 'fractions' of curly ∂'s is sometimes bulky, you may see the subscript notation: ∂_x is often used instead of $\frac{\partial}{\partial x}$. So I could have rewritten the above proof as follows:

$$g(y,x) = (\partial_1 f)(y,x) + (\partial_2 f)(y,x) = \partial_y f(y,x) + \partial_x f(y,x) =_*$$
$$\partial_y f(x,y) + \partial_x f(x,y) = (\partial_2 f)(x,y) + (\partial_1 f)(x,y) = g(x,y)$$

Similarly, in the physicist style variable notation, u_x stands for $\frac{\partial u}{\partial x}$. When $u = f(x,y)$, you will also see the 'mixed' notation f_x in analogy to $\frac{\partial f}{\partial x}$.

My best advice is that in your own usage you should avoid 'mixed' notation altogether, i.e., never identify slots by default variable names, but be tolerant to the frequent occurrences when others use such notation.

I may be uptight on the notation issue, but students do suffer in courses on partial differential equations when they have fuzzy ideas about multi-variable calculus.

Proof of the chain rule

In this proof, $x, h, k, g(x), f(g(x))$ are all vectors, even though I don't adorn them with arrows.

In a preliminary consideration, we prove that for a matrix T and a vector h, we have the estimate $\|Th\| \leq c\|h\|$, where the constant C depends on the entries of the matrix T. For instance we can take $C = \sqrt{\sum_{ij}(T_{ij})^2}$. This is a consequence of the Cauchy Schwarz inequality. The first entry of the vector Th is $T_{11}h_1 + T_{12}h_2 + \ldots T_{1n}h_n$, which can be written as a dot product of the vector $[T_{11}\,,\ T_{12}\,,\ \ldots\,,\ T_{1n}]^T$ with h. Therefore its absolute value is less than the product of the norms, or:

$$(Th)_1^2 \leq (\textstyle\sum_j T_{1j}^2)\|h\|^2$$

Similarly for the other components of Th. Adding up these, we get

$$\|Th\|^2 \leq (\textstyle\sum_{ij} T_{ij}^2)\|h\|^2$$

Next, we want to show that $Df(g(x))Dg(x)$ is the total derivative of $f \circ g$. In other words, we have to show that

$$\lim_{h\to 0} \frac{\|f(g(x+h)) - f(g(x)) - Df(g(x))Dg(x)h\|}{\|h\|} = 0$$

Rewriting this using the ε-δ definition of the limit, we have to show: For every $\varepsilon > 0$, there exists $\delta > 0$ such that $\|h\| < \delta$ implies

$$\|f(g(x+h)) - f(g(x)) - Df(g(x))Dg(x)h\| \leq \varepsilon\|h\| \qquad (G)$$

(eqn (G) for goal). Similarly, we rewrite the hypotheses that (H1) f is differentiable at $g(x)$ and (H2) g is differentiable at x as: For every $\varepsilon_1 > 0$, there exists $\delta_1 > 0$ such that $\|k\| < \delta_1$ implies

$$\begin{gathered}\|f(g(x)+k) - f(g(x)) - Df(g(x))k\| \leq \varepsilon_1\|k\| \\ \text{in particular for} \quad k = g(x+h) - g(x)\end{gathered} \qquad (H1)$$

For every $\varepsilon_2 > 0$, there exists $\delta_2 > 0$ such that $\|h\| < \delta_2$ implies

$$\|g(x+h) - g(x) - Dg(x)h\| \leq \varepsilon_2\|h\| \qquad (H2)$$

We now calculate

$$\begin{aligned}&\|f(g(x+h)) - f(g(x)) - Df(g(x))Dg(x)h\| \\ &\qquad \leq \|f(g(x+h)) - f(g(x)) - Df(g(x))(g(x+h) - g(x))\| \\ &\qquad\quad + \|Df(g(x))(g(x+h) - g(x) - Dg(x)h)\| \\ &\qquad \leq \|f(g(x+h)) - f(g(x)) - Df(g(x))(g(x+h) - g(x))\| \\ &\qquad\quad + M_f\|(g(x+h) - g(x) - Dg(x)h)\|\end{aligned}$$

where M_f is the constant that comes from the matrix $T = Df(g(x))$ in the estimate $\|Tk\| \leq C\|k\|$. We aim to show that each of the two terms in the sum on the right is $\leq \frac{1}{2}\varepsilon\|h\|$,

provided $\|h\|$ is sufficiently small. For this purpose, we choose $\varepsilon_2 := \varepsilon/(2M_f)$ in (H2) and require $\|h\| < \delta_2$. This takes care of the second term.

For the first term, we want to argue that $k := g(x+h) - g(x)$ becomes small if h is small, and then we use (H1). More specifically, since

$$\|g(x+h) - g(x)\| \le \|g(x+h) - g(x) - Dg(x)h\| + \|Dg(x)h\|$$

we use first (H2) with $\varepsilon_2 := 1$ (requiring $\|h\| < \delta_2'$ for the corresponding δ_2'), and we use that $\|Dg(x)h\| \le M_g\|h\|$. This guarantees $\|g(x+h) - g(x)\| \le (M_g+1)\|h\|$.

So far, we have achieved

$$\begin{aligned} &\|f(g(x+h)) - f(g(x)) - Df(g(x))Dg(x)h\| \\ &\qquad \le \|f(g(x+h)) - f(g(x)) - Df(g(x))(g(x+h) - g(x))\| + \tfrac{1}{2}\varepsilon\|h\| \qquad (G0) \\ &\|g(x+h) - g(x)\| \le (M_g+1)\|h\| \end{aligned}$$

provided $\|h\| \le \min\{\delta_2, \delta_2'\}$.

Now we ue hypothesis (H1) with $\varepsilon_1 := \varepsilon/(2M_g+2)$, and we get a corresponding quantity δ_1, such that $\|f(g(x+h)) - f(g(x)) - Df(g(x))(g(x+h)-g(x))\| \le \varepsilon\|g(x+h)-g(x)\|/(2M_g+2)$ provided $\|g(x+h) - g(x)\| < \delta_1$.

Now we strengthen our requirement on h to

$$\|h\| \le \min\left\{\delta_2, \delta_2', \frac{\delta_1}{M_g+1}\right\} =: \delta$$

This guarantees that $\|g(x+h) - g(x)\| \le (M_g+1)\|h\| < \delta_1$ and therefore we do get

$$\begin{aligned} &\|f(g(x+h)) - f(g(x)) - Df(g(x))(g(x+h) - g(x))\| \le \\ &\qquad \le \varepsilon\|g(x+h) - g(x)\|/(2M_g+2) \le \tfrac{1}{2}\varepsilon\|h\| \ . \end{aligned}$$

Merging this with the previous estimate (G0), we obtain (G).

Continuous differentiability, and higher derivatives

It is possible to say "A function f is twice differentiable, if it is differentiable and its total derivative Df, which is a matrix-valued function $x \mapsto Df(x)$, is differentiable again". But for the purposes of a first course in multi-variable calculus, this approach tends to lead to a somewhat bulky formalism. Fortunately, there is an easier way out, and it relies on the fact (already visible in single-variable calculus) that the class of differentiable functions *with continuous derivative* is much more useful than the class of (merely) differentiable functions.

Whereas we have seen that (total) differentiability as such cannot be described in terms of partial derivatives alone, continuous differentiability, i.e., total differentiability with continuous derivative, can very well be described in terms of partial derivatives alone. This is due to the fact that continuity of all partial derivatives implies total differentiability.

So we define: A function f defined on an open set of $\mathbb{R}^n$, scalar valued or vector valued, is *continuously differentiable* (also called 'once continuously differentiable' and abbreviated as C^1), iff all its partial derivatives exist and are continuous functions. — This is equivalent to saying that f is totally differentiable and the matrix valued function Df is continuous.

With this in mind, we can now go on to say: A function f is twice continuously differentiable (C^2), if all its partial derivatives are (once) continuously differentiable; similarly, we define k times continuously differentiable functions (C^k functions).

A fundamental theorem says: If f is C^2, then the order of partial derivatives doesn't matter, more precisely, for an n variable function f that is C^2, and any $i, j \in \{1, 2, \ldots, n\}$, it holds:

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f(x_1, \ldots, x_n)}{\partial x_j}\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial f(x_1, \ldots, x_n)}{\partial x_i}\right)$$

Similarly, for C^k functions, partial derivatives of order up to k may be carried out in any order. [We'll skip the proof, even though it's not difficult; refer to a textbook if needed.]

Just one simple example to illustrate the theorem. Take $f(x, y) = x^2 y e^x$. Then $\frac{\partial}{\partial x} f(x, y) = 2xye^x + x^2 ye^x$ and $\frac{\partial}{\partial y}\frac{\partial}{\partial x} f(x, y) = 2xe^x + x^2 e^x$. Calculating in the opposite order, we get $\frac{\partial}{\partial y} f(x, y) = x^2 e^x$ and $\frac{\partial}{\partial x}\frac{\partial}{\partial y} f(x, y) = 2xe^x + x^2 e^x$, the same result.

It must be pointed out that the C^2 hypothesis is crucial. Here is a counterexample when the C^2 hypothesis fails: Take

$$f(x, y) := \begin{cases} \frac{xy(x^2 - y^2)}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

It is easy to check that $f(x, 0) = 0$ and $f(0, y) = 0$. With this observation, and the quotient rule applied in points outside the origin, we get

$$(\partial_1 f)(x, y) = \partial_x f(x, y) = \begin{cases} \frac{x^4 y + 4x^2 y^3 - y^5}{(x^2 + y^2)^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

$$(\partial_2 f)(x, y) = \partial_y f(x, y) = \begin{cases} \frac{x^5 - 4x^3 y^2 - xy^4}{(x^2 + y^2)^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

A quick conversion into polar coordinates shows that these partials are still continuous in the origin (they are r times some trig expression in the angle φ). So f is a C^1 function. It turns out that none of the 2nd partial derivatives has a limit as $(x, y) \to (0, 0)$, so f is not C^2.

From $(\partial_1 f)(0, y) = -y$, we obtain $(\partial_2 \partial_1 f)(0, 0) = -1$. From $(\partial_2 f)(x, 0) = x$, we obtain $(\partial_1 \partial_2 f)(0, 0) = 1$. So in this case the order of partials does matter.

Outside the origin, where all 2nd order partial derivatives are continuous, we have

$$(\partial_2 \partial_1 f)(x, y) = (\partial_1 \partial_2 f)(x, y) = \frac{x^6 + 9x^4 y^2 - 9x^2 y^4 - y^6}{(x^2 + y^2)^3}$$

Such counterexamples play a surprisingly insignificant role outside calculus textbooks. The reason is two-fold: Either applications (like partial differential equations) work with *continuous* differentiability (and then there is no counterexample), or else, one is in a situation in which continuous differentiability is not a useful hypothesis. In such a situation, total differentiability is often not a useful hypothesis either. One then rather deals with a yet more general notion of differentiability that looks at the function as a whole and is not concerned with the discrepancy in a single point like $(0, 0)$ in our counterexample. Details on this matter must be reserved to graduate level classes in partial differential equations.

The Hessian

Suppose we have a scalar valued function f. Then ∇f is a vector valued function. Its derivative $D\nabla f$ is a matrix valued function, We call it Hf, the Hessian matrix. So, specifically (for a C^2 function f),

$$Hf(\vec{x}) = \begin{bmatrix} \frac{\partial^2 f(\vec{x})}{\partial x_1^2} & \frac{\partial^2 f(\vec{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\vec{x})}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\vec{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\vec{x})}{\partial x_1 \partial x_n} & \frac{\partial^2 f(\vec{x})}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_n^2} \end{bmatrix}$$

Combining the derivative and the gradient in this way allows us to use the two matrix indices for the two partial derivatives.

Note that our hypothesis that f is C^2 guarantees that $Hf(\vec{x})$ is a *symmetric* matrix, i.e., it is equal to its own transpose.

The Hessian will play a similar role in multi-variable minimax problems as the second derivative does in single-variable minimax problems. To this end, we note the following simple formula about a 'directional second derivative':

If f is a scalar valued C^2 function, then

$$\frac{d^2}{dt^2} f(\vec{x} + t\vec{v})\Big|_{t=0} = \vec{v}^T Hf(\vec{x})\vec{v}$$

Proof:

$$\frac{d}{dt} f(\vec{x} + t\vec{v}) = Df(\vec{x} + t\vec{v})\vec{v} = \vec{v} \cdot \nabla f(\vec{x} + t\vec{v}) = \vec{v}^T \nabla f(\vec{x} + t\vec{v})$$

$$\frac{d^2}{dt^2} f(\vec{x} + t\vec{v}) = \vec{v}^T D\nabla f(\vec{x} + t\vec{v})\vec{v} = \vec{v}^T Hf(\vec{x})\vec{v}$$

(In the second line, we have used that the derivative may be moved past the multiplication with a constant matrix; think why?)

Minimax problems

Suppose f is a scalar valued multi-variable function. Analogously to single-variable calculus, we say: f has a local maximum (synonym: relative maximum) at $\vec{x}_*$, if $f(\vec{x}_*) \geq f(\vec{y})$ for all $\vec{y}$ in a certain ball $B_r(\vec{x}_*)$ about $\vec{x}_*$ that are also in the domain of f. Likewise, we say: f has a local minimum (synonym: relative minimum) at $\vec{x}_*$, if $f(\vec{x}_*) \leq f(\vec{y})$ for all $\vec{y}$ in a certain ball $B_r(\vec{x}_*)$ about $\vec{x}_*$ that are also in the domain of f.

Reminder from single variable calculus: If $f : x \mapsto f(x)$ is a differentiable single variable function and has a local maximum or a local minimum at x_*, and x_* is in the interior of the domain (this domain used to be an interval), then $f'(x_*) = 0$. If moreover, the function is twice differentiable and has a local maximum (resp., minimum) at x_* in the interior of the domain, then $f''(x_*) \leq 0$ (resp., $f''(x_*) \geq 0$). — Conversely, if f is C^2 and x_* is in the interior of the domain of f and $f'(x_*) = 0$ and $f''(x_*) < 0$ (resp., $f''(x_*) > 0$), then f has a local maximum (resp., local minimum) at x_*.

Now the good news is that this result carries over to multi-variable calculus, if we study directional derivatives in all directions $\vec{v}$ going out from $\vec{x}_*$:

If $f : \vec{x} \mapsto f(\vec{x})$ is a differentiable multi-variable function (with scalar values) and has a local maximum or local minimum at $\vec{x}_*$, a point in the interior of the domain of f, then $\partial_{\vec{v}} f(\vec{x}_*) = 0$ for every direction vector $\vec{v}$. (Equivalently: $\nabla f(\vec{x}_*) = \vec{0}$.) If moreover, the function is twice continuously differentiable and has a local maximum (resp., minimum) at $\vec{x}_*$ in the interior of the domain, then $\vec{v}^T Hf(\vec{x}_*)\vec{v} \leq 0$ (resp., $\vec{v}^T Hf(\vec{x}_*)\vec{v} \geq 0$) for *all* direction vectors $\vec{v}$. — Conversely, if f is C^2 and $\vec{x}_*$ is in the interior of the domain of f and $\nabla f(\vec{x}_*) = 0$ and $\vec{v}^T Hf(\vec{x}_*)\vec{v} < 0$ (resp., $\vec{v}^T Hf(\vec{x}_*)\vec{v} > 0$) for *all* direction vectors $\vec{v} \neq \vec{0}$, then f has a local maximum (resp., local minimum) at $\vec{x}_*$.

The first part is nearly obvious: For if f has a local maximum at $\vec{x}_*$, then in particular each restriction of f onto a straight line through $\vec{x}_*$ has a local maximum there as well. This restriction is a function $t \mapsto f(\vec{x}_* + t\vec{v})$, and it has a local maximum at $t = 0$. Then the single variable result about local maxima, together with the formulas $\frac{d}{dt}|_{t=0} f(\vec{x}_* + t\vec{v}) = \partial_{\vec{v}} f(\vec{x}_*)$, and the similar formula for the 2nd derivative involving the Hessian, produces the statement about *necessary* conditions for a local maximum.

Not quite so obvious is the statement about the sufficient conditions: If $\vec{x}_*$ satisfies the single variable conditions for a local maximum in every direction, then f does indeed have a local maximum at $\vec{x}_*$. The proof of this statement would use that we assumed the function to be twice *continuously* differentiable, and some version of the mean value theorem (somewhat similar to how we concluded total differentiability from continuous partial derivatives). We won't bother with a formal proof here. Rather, we will study a bit how we use this minimum/maximum test in practice. This endeavour has a number of interesting and nontrivial quirks of its own.

Consider the function given by $f(x, y) = x^4 + 2x^2y^2 + y^4 + 2y^2 - 2x^2$. Does it have local minima and/or maxima?

There is no boundary to consider since the domain is $\mathbb{R}^2$. So all points are in the interior. *If* f has a local minimum or maximum at any point (x, y), then the derivative (or gradient) must vanish there, i.e., both partial derivatives must vanish:

$$\begin{aligned} \partial_x f(x, y) &= 4x^3 + 4xy^2 - 4x = 4x(x^2 + y^2 - 1) = 0 \\ \partial_y f(x, y) &= 4x^2y + 4y^3 + 4y = 4y(x^2 + y^2 + 1) = 0 \end{aligned}$$

The second equation is equivalent to $y = 0$ (since $x^2 + y^2 + 1$ never vanishes.) The first equation then says: $x \in \{-1, 0, 1\}$.

By definition, those points (x, y) where the derivative of f vanishes are called *critical points* of f. They are the only candidates where f could have an interior minimum or maximum. We investigate the three critical points $(-1, 0)$, $(0, 0)$ and $(1, 0)$.

Let's calculate the Hessian:

$$Hf(x, y) = \begin{bmatrix} 12x^2 + 4y^2 - 4 & 8xy \\ 8xy & 4x^2 + 12y^2 + 4 \end{bmatrix}$$

$$Hf(\pm 1, 0) = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix} \qquad Hf(0, 0) = \begin{bmatrix} -4 & 0 \\ 0 & 4 \end{bmatrix}$$

$$\vec{v}^T Hf(\pm 1, 0)\vec{v} = 8v_1^2 + 8v_2^2 \qquad \vec{v}^T Hf(0, 0)\vec{v} = -4v_1^2 + 4v_2^2$$

Now indeed, $8v_1^2 + 8v_2^2 > 0$ for all vectors $\vec{v} = [v_1\,,\ v_2]^T \neq \vec{0}$, and therefore f has local minima at $(\pm 1, 0)$. However, $-4v_1^2 + 4v_2^2 =: Q$ does not have a specific sign for all vectors $\vec{v}$. Since

this expression fails to be ≥ 0 for all $\vec{v}$ (e.g., $\vec{v} = [1,0]^T$ is a counterexample), f cannot have a local minimum at $(0,0)$. But Q also fails to be ≤ 0 for all $\vec{v}$ (now $\vec{v} = [0,1]^T$ is a counterexample). So f cannot have a local maximum there either.

Let's study a more complicated example: $f(x,y) := \frac{7}{2}x^4 + x^3 + 2xy^2 + y^4$. Again we look for relative maxima and minima. We have the conditions for the critical points

$$\frac{\partial f(x,y)}{\partial x} = 14x^3 + 3x^2 + 2y^2 = 0 \qquad \frac{\partial f(x,y)}{\partial y} = 4xy + 4y^3 = 0$$

The second equation is equivalent to $y = 0$ or $y^2 = -x$.

Case 1: $y = 0$. Then the first equation becomes $x = 0$ or $x = -\frac{3}{14}$.
Case 2: $y^2 = -x$. Then the first equation becomes $14x^3 + 3x^2 - 2x = 0$, i.e., $x = 0$ or $x = -\frac{1}{2}$ or $x = \frac{2}{7}$. The latter choice however can be discarded, because it doesn't correspond to any real y.

So we have four critical points: $P_0 = (0,0)$, $P_1 = (-\frac{3}{14}, 0)$, and $P_{2\pm} = (-\frac{1}{2}, \pm\frac{1}{2}\sqrt{2})$. We need the Hessian at each of these points.

$$Hf(x,y) = \begin{bmatrix} 42x^2 + 6x & 4y \\ 4y & 4x + 12y^2 \end{bmatrix}$$

So

$$Hf(P_0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad Hf(P_1) = \begin{bmatrix} \frac{9}{14} & 0 \\ 0 & -\frac{6}{7} \end{bmatrix} \qquad Hf(P_{2\pm}) = \begin{bmatrix} \frac{15}{2} & \pm 2\sqrt{2} \\ \pm 2\sqrt{2} & 4 \end{bmatrix}$$

So at P_0, the quadratic form $\vec{v}^T Hf(P_0)\vec{v}$ vanishes identically. The 2nd derivative test is inconclusive.

At P_1, the quadratic form $\vec{v}^T Hf(P_1)\vec{v}$ is $\frac{9}{14}v_1^2 - \frac{6}{7}v_2^2$. This is positive for some $\vec{v}$ (e.g., $\vec{v} = [1\,,\ 0]^T$) and negative for other $\vec{v}$. So P_1 can be neither a maximum nor a minimum. A point where the quadratic form is positive for some direction vectors $\vec{v}$, but negative for others, is a saddle point: In some directions, the critical point looks like a minimum, in other directions it looks like a maximum.

Finally, at $P_{2\pm}$, we claim that the quadratic form $\vec{v}^T Hf(P_{2\pm})\vec{v} = \frac{15}{2}v_1^2 \pm 4\sqrt{2}v_1v_2 + 4v_2^2$ will be positive for all non-zero vectors $\vec{v}$. This means that $P_{2\pm}$ are local minima. Now let's look how I would see this positivity:

First consider a general rule-of-thumb approach: The pure squares v_1^2 and v_2^2 have positive coefficients (the diagonal entries of the Hessian), so they tend to make the expression positive. The mixed terms could contribute negative terms if the signs of v_1 and v_2 are chosen inconveniently. If their coefficient is small, they may not be able to out-compete the pure squares, but if their coefficient is large, then they will win over the pure squares. How big is too big? Let's complete the squares:

$$\frac{15}{2}v_1^2 \pm 4\sqrt{2}v_1v_2 + 4v_2^2 = 4\Big(v_2 \pm \frac{1}{2}\sqrt{2}v_1\Big)^2 + \frac{11}{2}v_1^2$$

So clearly, as a sum of squares, this is non-negative; and it will be strictly positive unless $v_1 = 0$ and the parenthesis vanishes, too. And this latter happens only if $v_2 = 0$ as well. So, having shown that $\vec{v}^T Hf(P_{2\pm})\vec{v} > 0$ for all nonzero direction vectors $\vec{v}$, we have identified $P_{2\pm}$ as local minima.

Symmetric matrices, quadratic forms and definiteness properties

We study here the algebraic task of distinguishing local minima and maxima and saddle points by means of the Hessian, as encountered in the previous examples.

Let H be a symmetric $n \times n$ matrix. Remember that 'symmetric' means that $H = H^T$; With such a matrix there is associated a quadratic expression in n variables $v_1, \ldots, v_n$, namely the expression $\vec{v}^T H \vec{v}$. Such an expression is called a 'quadratic form'. (No, I don't know of a good motivation for this choice of name.) For instance, writing $[u\ v\ w]$ instead of $[v_1\ v_2\ v_3]$, we have

$$\begin{bmatrix} u & v & w \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{12} & h_{22} & h_{23} \\ h_{13} & h_{23} & h_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = h_{11}u^2 + h_{22}v^2 + h_{33}w^2 + 2h_{12}uv + 2h_{13}uw + 2h_{23}vw$$

You can see that the diagonal entries of H give the coefficients of pure quadratic terms, whereas the off-diagonal entries give coefficients of mixed terms.

Given H, our task is to find out whether the quadratic form is positive for *all* vectors $\vec{v}$, or negative for *all* vectors $\vec{v}$, or positive for some and negative for other vectrs $\vec{v}$. And yes, there are borderline cases, e.g., where the form doesn't take on negative values but can take on a 0 value.

The following definitions are common:

A symmetric matrix H is called POSITIVE DEFINITE, if $\vec{v}^T H \vec{v} > 0$ for all $\vec{v} \neq \vec{0}$. It is called POSITIVE SEMIDEFINITE if $\vec{v}^T H \vec{v} \geq 0$ for all $\vec{v}$.

A symmetric matrix H is called NEGATIVE DEFINITE, if $\vec{v}^T H \vec{v} < 0$ for all $\vec{v} \neq \vec{0}$. It is called NEGATIVE SEMIDEFINITE if $\vec{v}^T H \vec{v} \leq 0$ for all $\vec{v}$. Clearly, H is negative (semi-)definite if and only if $-H$ is positive (semi-)definite.

A symmetric matrix H is called INDEFINITE, if it is neither positve nor negative semidefinite, in other words, if $\vec{v}^T H \vec{v} > 0$ for some $\vec{v}$ and $\vec{v}^T H \vec{v} < 0$ for some other $\vec{v}$.

By considering in particular coordinate direction unit vectors $\vec{v}$ (that have a single 1 and otherwise only 0's as components), we make the following easy observations:

$$\text{If a matrix is } \left\{ \begin{array}{c} \text{positive definite} \\ \text{positive semidefinite} \\ \text{negative definite} \\ \text{negative semidefinite} \end{array} \right\}, \text{ then all its diagonal entries must be } \left\{ \begin{array}{c} > 0 \\ \geq 0 \\ < 0 \\ \leq 0 \end{array} \right\}.$$

None of the converses hold. The off-diagonal entries of H contribute mixed terms in the quadratic form, and they could change the sign of the quadratic form against what the diagonal entries would 'vote' for. Intuitively, the diagonal entries get the say about definiteness properties only if the off-diagonal entries are not too large. We'll quantify this in a moment.

Didactic decision: I will give you two easy-to-use tests for any size matrix without proof, and a proved and explained version for the case of 2×2 matrices. A thorough study of definiteness properties would require the full Linear Algebra course as a prerequisite, and then some work on top of it. My outline will be such that it is 'workable now', but becomes more coherent when you revisit it with full Linear Algebra wisdom; it is written in sufficient generality that you will not attempt your own generalization from simpler special cases (which would certainly lead to wrong guesses). But you may not have the calculational tools to exploit the full generality right now, and then such calculations will not be required from you in this class.

First let's study a 2×2 symmetric matrix $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ and see exactly when it is positive definite. We take the quadratic form $au^2+2buv+cv^2$. We know that a (and also c) must be positive, if the matrix is to be positive definite. Assuming a positive, we use completion of squares to control the mixed term:

$$au^2+2buv+cv^2 = a\left(u+\frac{bv}{a}\right)^2 - \frac{b^2}{a}v^2 + cv^2 = a\left(u+\frac{bv}{a}\right)^2 + \frac{ac-b^2}{a}v^2$$

So if a and $ac-b^2$ are both positive, then the quadratic form is positive definite: namely, it's clearly ≥ 0; but for it to vanish we need both $v=0$ and $u+bv/a=0$, i.e., we need $u=v=0$. Conversely, if $ac-b^2$ is *not* positive, then the quadratic form is not positive for $[u\ v]=[-b/a\ 1]$.

Conclusion: $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ is positive definite if and only if $a>0$ and $ac-b^2>0$. — Equivalently we could show: this matrix is positive definite if and only if $c>0$ and $ac-b^2>0$.

The quantity $ac-b^2$ is called the *determinant* of the 2×2 matrix $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$. In linear algebra, a certain number is assigned to every square matrix, and this number is called the determinant of that matrix. There is a neat geometric interpretation (as a signed area, or a signed volume, or higher dimensional signed volume) of the determinant, and there is a variety of effective ways of calculating the determinant of any square matrix of not too large size. But we will skip all these, and I'll just give you some basic facts:

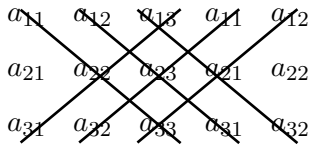
(1) The determinant of a 1×1 matrix $[a]$ is simply a. The only reason I am telling you this is that 1×1 matrices retrieve the 2nd derivative test for single variable minima as a special case of the 2nd derivative (Hessian) test for multi-variable minima.

(2) The determinant of a 2×2 matrix is the product of its diagonal entries minus the product of its off-diagonal entries: $\det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22}-a_{12}a_{21}$.

(3) The determinant of a 3×3 matrix is a sum/difference of six products, namely:

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{12}a_{21}a_{33}-a_{13}a_{22}a_{31}-a_{11}a_{23}a_{32}$$

The way to remember this mess is to copy the first two columns to the right and add the NW-SE products, and subtract the NE-SW products:



(4) In this course I will *not* teach you how to calculate determinants of an $n\times n$ matrix for $n\geq 4$; I only warn you that you should *not* attempt to guess-generalize the preceding formulas to larger matrices. It would almost certainly be a wrong guess. A linear algebra course will provide correct ways of getting larger determinants. I do want to tell you that $\det(-H)=\pm\det H$ for an $n\times n$ matrix H; here the $+$ applies if n is even, and the $-$ applies if n is odd. You may write this more concisely as $\det(-H)=(-1)^n\det H$.

Here is the correct generalization of the above test for positive definiteness of a 2×2 matrix:

Theorem (Hurwitz): *Given a symmetric matrix A of size $n \times n$, take the following sequence of determinants: start with the top-left corner, then in each step add the next row and column, and calculate the determinant in each step until you have calculated the determinant of the full matrix A. Now A is positive definite if and only if all of these determinants are positive.*

Example:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \text{ is positive definite}$$

if and only if

$$a_{11} > 0 \text{ and } \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} > 0 \text{ and } \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} > 0 \text{ and } \det \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} > 0$$

(and yes, remember, I haven't told you how to calc the last determinant, and will not ask you to).

In practice, you do not have to use the test in this order. You could start with any diagonal element (e.g., a_{33}) and then successively add one row&column at a time, in any order, but the same row and column number always together. For example, taking rows and columns in order 3,1,4,2, we get that the above matrix is positive definite if and only if

$$a_{33} > 0 \text{ and } \det \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix} > 0 \text{ and } \det \begin{bmatrix} a_{11} & a_{13} & a_{14} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{bmatrix} > 0 \text{ and } \det \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} > 0$$

Warning: Do not modify this test in any other way. In particular, do not swap the $>$ into $<$ and believe this tests for negative definite. It doesn't. Rather, to test A for negative definite, you test $-A$ for positive definite. You may use $\det(-A) = (-1)^n \det A$ to see what sign implications this has for the determinants obtained from A. — Also do not replace $>$ with $\geq$ hoping to test for positive semidefinite. This would still lead to a false 'theorem'.

There is another test that is much easier to use, but that may be inconclusive (whereas the Hurwitz test is never inconclusive).

Theorem (Gershgorin): *Given a symmetric matrix A, calculate the sum of absolute values of off-diagonal entries for each row and compare it with the diagonal entry in this row: If each diagonal entry is positive* and *is larger than the sum of the absolute values of the off diagonal entries in its row, then A is positive definite.*

Note: The converese does not hold: if the condition is violated, the matrix may or may not be positive definite.

Example:

$$\begin{bmatrix} 9 & -1 & 3 & -2 \\ -1 & 8 & 5 & 1 \\ 3 & 5 & 17 & 4 \\ -2 & 1 & 4 & 10 \end{bmatrix}$$

is positive definite because $9 > |-1| + |3| + |-2|$ and $8 > |-1| + |5| + |1|$ and $17 > |3| + |5| + |4|$ and $10 > |-2| + |1| + |4|$.

This note is for students who have studied and digested all of linear algebra, and it is provided for backwards reference at a later time. It is not part of the required material of the present course: A symmetric matrix is positive definite if and only if all its eigenvalues are positive. Every symmetric matrix A can be written in the form $A = QDQ^T$ where Q is an orthogonal matrix, i.e., $QQ^T = I$ and D is a diagonal matrix, whose diagonal entries are just the eigenvalues of A (and at the same time the eigenvalues of D). A is positive definite if and only if D is positive definite. These facts are key ingredients for a proof of the Hurwitz test; they are also behind a proof of Gershgorin's test.

Global (=absolute) extrema

The information that enters into the derivative tests for (local) extrema is of a local nature only: To calculate the gradient and the Hessian of a function at one point needs knowledge of that function in a neighborhood of that point only. It does not use info about the function far away from this point.

From this it is clear that the question of global maxima / minima (aka absolute maxima / minima) cannot be decided by means of derivative tests. Note that we say a function f has a global (= absolute) minimum at $\vec{x}$ if $f(\vec{x}) \leq f(\vec{y})$ for *all* $\vec{y}$ in the domain of f; unlike for a local (=relative) minimum, competitors $\vec{y}$ are allowed to come from anywhere in the domain, not only from a neighborhood of $\vec{x}$. A similar definition applies for global maxima.

Lower division calculus provides very few tools how to find global minima or maxima. However, there is one tool that is readily available, and it carries over almost verbatim from single to multi-variable:
Theorem: *If f is a* continuous *function defined on a* bounded *and* closed *subset of $\mathbb{R}^n$, then f does have a global minimum and a global maximum somewhere on this set.*

This theorem may look like a disappointment, because it merely asserts existence of a global min and a global max without giving any clue how to find them. Nevertheless, this abstract knowledge is in itself very valuable, and can serve as the basis to find them by means of other tools.

Note that all three hypotheses are needed: the function must be continuous (a requirement that is usually seen to be met obviously); the domain must be bounded (a requirement that sometimes causes us a headache); and it must be closed, i.e., it would include its boundary. This means we often have to split up our search into two parts: (1) an absolute minimum may be in the interior of the domain. As it would trivially be among the the local minima, we could retrieve it by the critical point test (gradient = 0). Or else, (2) an absolute minimum may be on the boundary. If we can describe the boundary points in terms of one variable less than the interior points (as we usually can), we may set up another minimum problem for the boundary alone. — Merging the two prongs of our search, we could finally argue that among the (usually finitely many) candidates that our search has netted, either in the interior or on the boundary, the one with the smallest value for f must be the absolute minimum. We may even omit the test with the Hessian if we are after the *absolute* minimum only.

But it is only by the a-priori knowledge that an absolute minimum *exists* that this procedure works. Short of such a-priori knowledge, you may well have filtered out many points, at which, for various reasons, f could not possibly have an absolute minimum, leaving over, say half a dozen points, which could not be ruled out (points where the gradient does vanish, and a few boundary points, and a few points where f isn't differentiable so the gradient test doesn't apply). If you'd now declare the one with the smallest value of f the absolute

minimum, you'd make a logical mistake. It would be like charging the only person without an alibi with having murdered a certain deceased person, but not establishing beforehand that the deceased person actually died as a result of foul play. (In one whodunnit, the person had actually died of natural causes, which meant the defendant was to be acquitted).

The homework gives an example where there is only one candidate for an absolute minimum, and a very promising candidate for that matter; but still there is no absolute minimum, and the sole promising candidate is only a relative minimum.

Example: We want to design a rectangular box, of sidelengths x, y, z, subject to the constraint (as might be imposed by the post office) that the sum $x + y + z$ is at most 3 feet. Within this constraint, we want to have a box of maximum volume xyz. In attempting to design this box, we believe (and will prove) that we cannot max out the volume except by maxing out the length constraint $x + y + z \leq 3$, so that we have $x + y + z = 3$. Indeed, for any box of dimensions x, y, z with $x + y + z < 3$, we can take a larger box with dimensions $(1 + \varepsilon)x, (1 + \varepsilon)y, (1 + \varepsilon)z$ of sum still ≤ 3 but with larger volume.

We could for instance solve for z and maximize $xy(3 - x - y)$ under the constraints $x \geq 0$, $y \geq 0$, $x + y \leq 3$. This domain is a triangle in the x, y plane. We have a continuous function on this triangle, a closed and bounded set. Therefore a maximum exists. On the boundary, $xy(3 - x - y)$ is zero, and this is certainly not the absolute maximum. So the absolute maximum is in the interior, and there it must obey the 'vanishing gradient' test: $\partial_x(xy(3 - x - y)) = 3y - 2xy - y^2 = 0$, $\partial_y(xy(3 - x - y)) = 3x - x^2 - 2xy = 0$. Since we have already disqualified any cases with $x = 0$ or $y = 0$ from being the location of the absolute maximum, we can simplify this to the system $3 = 2x + y$, $3 = x + 2y$, with the solution $x = y = 1$.

Constrained minima and maxima

The previous example had one aesthetic blemish: The role of x, y, z was entirely symmetric, but we arbitrarily selected z for elimination from the list of independent variables by means of the constraint $x + y + z = 3$. This is only an aesthetic issue, but in many more complicated examples, experience shows that messy calculations are avoided best by retaining any symmetry among variables that may exist in the problem.

Moreover, in some examples, using a constraint to eliminate a variable can make things really complicated calculationally, or may even be impossible practically.

Consider the following type of problem: Among all x, y, z satisfying a constraint $g(x, y, z) = 0$, we look for one that maximizes or minimizes the expression $f(x, y, z)$.

We say $f(x, y, z)$ has a constrained local (=relative) minimum at (x_0, y_0, z_0) (under the constraint $g(x, y, z) = 0$), if (x_0, y_0, z_0) satisfies this constraint: $g(x_0, y_0, z_0) = 0$, and $f(x_0, y_0, z_0) \leq f(x, y, z)$ for all (x, y, z) in some neighbourhood of (x_0, y_0, z_0) that also satisfy the constraint.

We assume that f and g are C^1 functions, and we assume that the gradient of g does not vanish on the level surface $g = 0$. (This technical assumption in particular guarantees that the level surface $g = 0$ is 'smooth', as we will discuss later). At a *constrained* local minimum of f, we cannot expect the gradient ∇f to vanish. The directional derivative in directions across the level set $g = 0$ may very well be non-zero. However, if we go in any direction $\vec{v}$ along the level set $g = 0$, i.e., tangentially to it, we will expect the directional derivative $\partial_{\vec{v}} f$ to vanish at the minimum.

More precisely: Let $t \mapsto \gamma(t)$ describe a curve within the level set $g = 0$ that passes through

(x_0, y_0, z_0) at $t = 0$; in formulas $\gamma(0) = [x_0, y_0, z_0]^T$ and $g(\gamma(t)) \equiv 0$ for all t. Then, since the composite function $t \mapsto f(\gamma(t))$ has a local minimum at $t = 0$, we must have $\frac{d}{dt}|_{t=0} f(\gamma(t)) = 0$. Evaluating this by the chain rule, we have $Df(\gamma(0))\gamma'(0) = 0$. Now we can do this reasoning for any curve γ passing through (x_0, y_0, z_0) within the level surface, and this way we can represent any direction vector $\vec{v}$ that is tangential to the level survace by such a curve: $\vec{v} = \gamma'(0)$. [3] Let's therefore sum up: $\partial_{\vec{v}} f(x_0, y_0, z_0) = \vec{v} \cdot \nabla f(x_0, y_0, z_0) = 0$ for any vector $\vec{v}$ that is tangential to the level surface $g(x, y, z) = 0$, or equivalently, for every vector $\vec{v}$ that is orthogonal to $\nabla g(x_0, y_0, z_0)$. Yet rewording this, we can say $\nabla f(x_0, y_0, z_0)$ is orthogonal to every vector $\vec{v}$ that is orthogonal to $\nabla g(x_0, y_0, z_0)$. A moments reflection may convince you that this simply means that $\nabla f(x_0, y_0, z_0)$ must be parallel to $\nabla g(x_0, y_0, z_0)$, i.e., there must be a number λ such that $\nabla f(x_0, y_0, z_0) = \lambda \nabla g(x_0, y_0, z_0)$.

We state this in full generality as a

Theorem: *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function whose gradient does not vanish on the level set $S_0 := \{\vec{x} | g(\vec{x}) = 0\}$. Let f be a C^1 function of n variables in a neighbourhood of this level set S. If f has a constrained relative minimum (or a constrained relative maximum) at $\vec{x}_0$, then there exists a real number λ (called a Lagrange multiplier) such that $\nabla f(\vec{x}_0) = \lambda \nabla g(\vec{x}_0)$.*

Note for those who know the notion of linear independence from linear algebra (it won't be required from you if you don't): If we have several constraint functions $g_1, \ldots g_k$, all C^1 and such that at each point on the joint level set $S_0 := \{\vec{x} | g_1(\vec{x}) = \ldots = g_k(\vec{x_0}) = 0\}$, the vectors $\nabla g_i(\vec{x})$ $(i = 1, \ldots, k)$ form a linearly independent set, and if a C^1 function f has a constrained local minimum or maximum at $\vec{x}_0$, then there exists Lagrange multipliers $\lambda_1, \ldots \lambda_k \in \mathbb{R}$ such that $\nabla f(\vec{x}_0) = \lambda_1 \nabla g_1(\vec{x}_0) + \ldots \lambda_k \nabla g_k(\vec{x}_0)$. – In short, if you have several constraints, you get a Lagrange multiplier for each constraint. The only subtlety to be observed is a hypothesis which is intended to guarantee that these several constraints are 'independent', namely the linear independence hypothesis.

Behind a rigorous proof of this method is the lemma (guaranteed by the theorem of implicit functions outlined in the next section) that in some neighbourhood of the presumed relative constrained minimum $\vec{x}_0$, one can indeed, in principle, eliminate one variable for each constraint. So while the proof is more akin to actually doing an elimination (in principle; practical elimination in formulas is not needed), like what we had done with the post office box example, the setup with Lagrange multipliers is designed to hide any elimination that was done in the proof that the method works in principle, and obtain a system of equations in which no elimination has been carried out explicitly.

The homework gives some examples how the method works in practice.

We are omitting any discussion of Hessians in connection with Lagrange multiplers. Some of this is discussed in the textbook by Marsden and Tromba. In many cases, the constraints restrict the domain to a bounded and closed set (hence guaranteeing existence of absolute maxima and minima a-priori), or to a domain that is at least closed, with a function to minimize that has a certain growth property so that all $\vec{x}$ that are outside a bounded set can be a-priori disqualified from consideration for an absolute minimum.

In such cases, the distinction into relative minima, relative maxima, and saddle points, is of

[3] We have neither proved here that the level surface $g = 0$ does have a tangent plane at (x_0, y_0, z_0), nor have we proved that an arbitrary tangential direction vector can be represented as the velocity vector of an appropriately chosen curve within the level surface. Rather we 'believe' this on intuitive grounds, having merely motivation in mind. Once the result we are aiming at is stated, based on our motivation, we would have to provide a rigorous proof yet, and it would be based on an advanced theorem called implicit function theorem, which we will briefly discuss afterwards. However, mathematically rigorous proofs in this matter are best left to a more advanced course.

less importance in practice.

Implicit function theorem

Example 1: we have a 2-variable function f, continuously differentiable, and wonder whether the equation $f(x,y)=0$ can in principle be solved for y to get $y=h(x)$. 'In principle' refers to the fact that we are not trying to do it in practice by means of explicit formula manipulation, but still that the equation $h(x,y)=0$ determines, for each x, a unique y solving the equation, and such that y can be viewed as a function of x, namely $y=h(x)$, where h is also a C^1 function. A specific example you might think of is the equation $x+e^x+y+e^y-2=0$.

This task is too general for us to say anything specific about the possibility of accomplishing it. The primary qualification to be added is that we are only interested in doing this job in the neighbourhood of some specific point. For instance, it is easy to see, that $(x,y)=(1,1)$ solves the equation $x+e^x+y+e^y-2=0$. So the unknown solution $y=h(x)$ would have $h(1)=1$. We now wonder how $h(x)$ might look for x near 1.

If we knew a-priori that there is a function h solving the equation, and that h has the number of derivatives we hope it to have, *then* we could use the chain rule to calculate more information about h. Indeed, the fact that $y=h(x)$ solves $f(x,y)=0$ means that $f(x,h(x))=0$ for every x. Taking a derivative with respect to x, we get $(\partial_1 f)(x,h(x))+(\partial_2 f)(x,h(x))h'(x)=0$. From this, with $x=1$ and $h(1)=1$, we can easily calculate $h'(1)$. All we need to be able to do is to divide by the value $(\partial_2 f)(1,1)$. So if this quantity vanishes, we may be in trouble, but if it does not vanish, then we can hope to be fine. The implicit function theorem will be the guy that tells us that we *are* fine, that a solution function h indeed exists, and that our calculation is therefore justified, provided only that $(\partial_2)f(1,1)$ doesn't vanish.

By taking further derivatives of $(\partial_1 f)(x,h(x))+(\partial_2 f)(x,h(x))h'(x)=0$, we could get higher derivatives of h, and in each step, the only solution condition is that $(\partial_2 f)(x,h(x))\neq 0$, since this is the quantity by which we must divide each time to get $h'(x)$, $h''(x)$, $h'''(x)$ successively.

Here is the precise wording of a simple version of the theorem on implicit functions:
Theorem: *Let $f:(x,y)\to f(x,y)$ be a C^1 function on some open subset of $\mathbb{R}^2$ and assume we have a point (x_0,y_0) solving the equation $f(x,y)=0$. If $(\partial_2 f)(x_0,y_0)\neq 0$, then there exists a C^1 function $h:x\mapsto h(x)$ in a neighbourhood of $x=x_0$ that solves the equation $f(x,y)=0$ for y, in the sense that $f(x,h(x))\equiv 0$. Moreover h is C^k if f is C^k. Similarly, if $(\partial_1 f)(x_0,y_0)\neq 0$, there is a C^1 function $g:y\mapsto g(y)$ defined in a neighbourhood of y_0 that solves $f(x,y)=0$ for x in the sense that $f(g(y),y)\equiv 0$. Moreover g is C^k if f is C^k. These solutions $y=h(x)$ or $x=g(y)$ are the only solutions within some small neighbourhood of (x_0,y_0). More precisely, there exists a small ball centered at (x_0,y_0) such that the intersection of the level set $f(x,y)=0$ with this ball is exactly the graph $y=h(x)$ (or $x=g(y)$), intersected with this ball.*

Note: It is NOT asserted that $y=h(x)$ is the only solution altogether. The implicit function theorem is local by its very nature.