DS 5:
Partial Derivatives and the Wave Equation

Philip Cherian

February 18, 2022

1 The idea of the derivative

Imagine for a moment that you're travelling on some sort of steep track (say, you're at a theme park and you're riding a roller coaster, or perhaps you're going to Shimla and you've got on the toy train from Kalka). If someone were to ask you how steep the track was, the question would have a clear answer: all you would have to do is find a nearby point, and calculate the *slope*. In other words, you would need to find the change in the height with horizontal distance. Mathematically, if you wanted to represent such a function, you would do so in the following way: suppose the height of the track can be described by a continuous function $f(x)$,[1] so that we can write

$$m_h = \frac{f(x+h) - f(x)}{h}. \tag{1}$$

Of course, the above quantity would give different numerical results, depending on what value of h was being chosen, as has been indicated by the subscript h. So at first glance, you could have a problem with this: the slope, you might say, should not depend on some arbitrary constant h, the answer should be unambiguous. So does the quantity m even make sense physically?

To answer this question, it is more instructive to look at a different problem: that of a falling object. In this case, we know that the object has both a position $y(t)$ as well as a velocity $v(t)$. The relation between the position and the velocity is mathematically exactly the same as the relation between the height of the track and its slope from our previous example. However, we know that the object must have a well-defined velocity at every instant of time, because it moves in a clear, reproducible, way. However, if you wished to calculate this velocity, you would use a relation like the one before:

$$v_{\Delta t} = \frac{y(t+\Delta t) - y(t)}{\Delta t}. \tag{2}$$

Here too the "velocity" that we have measured seems to depend on an ad-hoc quantity (Δt), which means that if you and a friend were trying to calculate the velocity of this falling object, you would get different results depending on what value of Δt you each chose!

Exercise: Using the fact that – for an object starting from rest – the displacement as a function of time is given by the formula below, calculate the velocity at (say) $t = 3.5$ using lower and lower values of Δt. (Take $g = 9.8\ \mathrm{ms}^{-1}$.)

$$y = \frac{1}{2} g t^2.$$

[1] If it wasn't, it wouldn't be a very pleasant trip!

If you do the above exercise, you should see that the calculated velocities are as follows:

Δt (s)	$v_{\Delta t}$ (m s^{-1})
1	39.2
0.1	34.79
0.01	34.349
0.001	34.3049
0.0001	34.3005

Table 1: For any desired number of significant digits (or "precision"), there is always a critical Δt_c, such that for any choice of $\Delta t < \Delta t_c$, the calculated velocity is independent of Δt.

Now, as we discussed earlier, the problem with this Δt prescription is that the velocity seems like it changes with Δt. However, if you stare at the above table for long enough, you will arrive at an interesting conclusion: as you reduce Δt, the velocity does indeed change, but if you fix a precision (i.e. a number of significant digits you are interested in looking at), then for every precision there is a "critical" Δt below which any Δt would give you the same answer, up to that precision.

Let's be very specific here: suppose we were using a terrible measuring device for velocity, and we were only able to measure velocities of the order of 1 ms^{-1}. In other words, any variation in the velocity of the order of 0.1 ms^{-1} would not be detectable by our device. What would our "threshold" Δt be? Well, if you and your friend used any $\Delta t < 0.1$, you could always end up agreeing on the first two digits of the velocity! Similarly, if you had access to a device with a much higher precision, (say, you could measure variations up to 0.01 ms^{-1}, then you would need to use a Δt of at least 0.001 s, but any Δt less than that would give the same value of the velocity (up to that precision).

So we have had to give up our "common-sense" result a little: the calculated velocity *does* depend on Δt, but at least there exists some critical Δt below which it doesn't. If you wish to express this mathematically, what you're saying is essentially that for any precision ϵ, there exists a critical Δt_c, such that

$$\forall \quad \Delta t < \Delta t_c \qquad \Longrightarrow \qquad \left| v_{\Delta t} - v_{\Delta t_c} \right| < \epsilon.$$

Of course, if you want *infinite* precision, then you would need to keep making a Δt smaller and smaller. This is called taking a *limit*. As a result, you can define the *instantaneous* velocity as:

$$v(t) = \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} = \lim_{\Delta t \to 0} \left(\frac{y(t+\Delta t) - y(t)}{\Delta t} \right) \equiv \frac{\mathrm{d}y}{\mathrm{d}t}. \tag{3}$$

The notation on the right (using $\mathrm{d}t$) indicates that we are taking an infinitesimal value for Δt. This idea doesn't just work for position and velocity, but also works for the height and slope problem we initially considered. There too, we can define the slope as being the *derivative* of the height function $f(x)$:

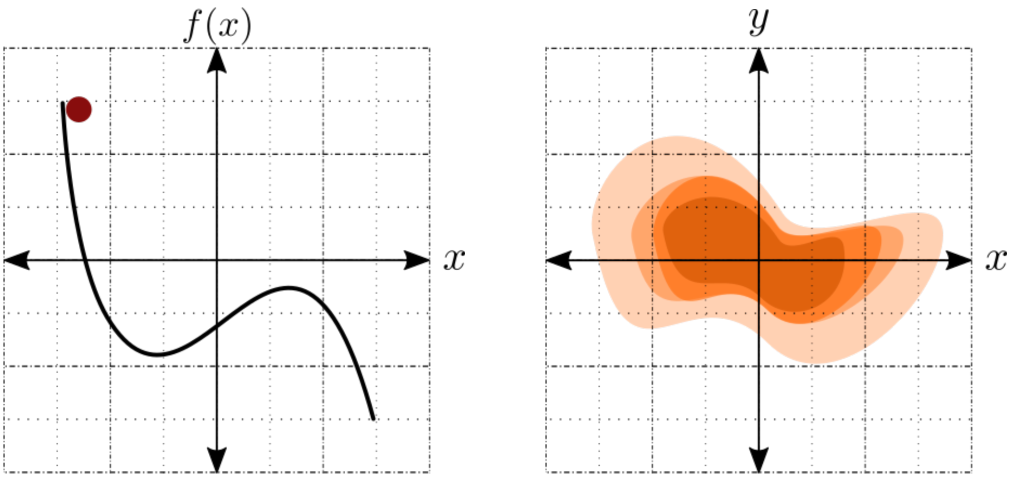
$$m(x) = \frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}. \tag{4}$$

Notice how in all of the above cases we took the value of a function at two different points, separated by some small interval (h, Δt, Δx, etc.) and divided it by the interval. In one dimension, this is quite intuitive, as you have only one of two choices: either choose a positive, or a negative, interval, and the former measures the downward slope at a point, and the latter the "upward" slope at the same point. However, when we move to two-dimensions, the situation gets considerably more complicated.

2 Derivatives in two dimensions

The derivative involves looking at two points that are very close to each other, which should already allow you to see why things become more complicated in the two-dimensional case. Unlike the one-dimensional case where the only liberty one had was to choose a point either to the “left” or to the “right”, and where both of these quantities measured essentially the same thing (the “rise” or “fall” at a specific point), in two-dimensions you have an infinite number of directions that you can choose your “nearby” point to be in.

Intuitively, imagine your two-dimensional function to be represented by a hill. The height at any point is the value of the function, and it is a function of both x and y. Generally, such two-dimensional functions are represented by a *three*-dimensional graph, with the $z-$axis being the value of the function (here, the height). However, we could equivalently represent the height by a contour diagram, as shown in Figure (1b), where the different colours each represent one value of height. The rate at which the shades of colour change in some direction therefore represents the slope in that direction. Contrast this with the case in Figure (1a). While in the case of the roller-coaster or the train, you could easily answer the question “how steep is the track?”, if someone asked you how steep the hill is, they would also have to tell you in which *direction* they want you to measure the steepness!



(a) A function of one variable $f(x)$

(b) A function of two variables, $f(x, y)$

Figure 1: Examples of one and two-dimensional functions. In the one-dimensional case, only one unique derivative exists, measuring either the rise or the fall of the function at a point, depending on the sign you choose. In the two-dimensional case, an infinite number of directions exist, representing the fall (or rise) along any intermediate direction.

2.1 The “partial” derivative

We would nevertheless like to work with derivatives in two-dimensions in as we would in one dimension. In the one dimensional case, we could say that for a function $f(x)$, the change in the function when we move by a small distance Δx is given by

$$\mathrm{d}f \approx \left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)\Delta x + \mathcal{O}(\Delta x^2), \tag{5}$$

where the "correction" is of order Δx^2, and thus becomes smaller and smaller as $\Delta x \to 0$. Using the same limiting process that we defined earlier, we could always choose a Δx sufficiently small to allow us to ignore these "second-order" corrections. Let us now generalise this to a function of two dimensions, $f(x, y)$ describing, say, the height of the hill we were talking about earlier. Imagine that you moved from some point (x, y) to some other point $(x + \mathrm{d}x, y + \mathrm{d}y)$. In this case, how much would the height of the hill have changed? You can answer this question by first moving along x, i.e., going from $(x, y) \to (x+\mathrm{d}x, y)$, an then going from $(x + \mathrm{d}x, y) \to (x + \mathrm{d}x, y + \mathrm{d}y)$. Now, we need to have a new type of notation that distinguishes these motions. In the first case, the motion is purely along x, keeping y constant, and in the second case it's purely along y, keeping x constant. In order to differentiate between these two actions, we define the following "partial" derivatives:

$$\begin{aligned}\left(\frac{\partial f}{\partial x}\right)_y &= \lim_{h\to 0}\left(\frac{f(x+h,y)-f(x,y)}{h}\right) \quad \text{Derivative with respect to } x \text{ at constant } y,\\ \left(\frac{\partial f}{\partial y}\right)_x &= \lim_{h\to 0}\left(\frac{f(x,y+h)-f(x,y)}{h}\right) \quad \text{Derivative with respect to } y \text{ at constant } x.\end{aligned} \tag{6}$$

Now, following this process, what is the change in the height when we move from (x, y) to $(x+\mathrm{d}x, y+\mathrm{d}y)$? It is the sum of the changes in the heights along x, and then along y:

$$\mathrm{d}f \approx \left(\frac{\partial f}{\partial x}\right)_y \Bigg|_{(x,y)} \mathrm{d}x \quad + \quad \left(\frac{\partial f}{\partial y}\right)_x \Bigg|_{(x+\mathrm{d}x,y)} \mathrm{d}y \quad + \quad \mathcal{O}(\mathrm{d}x^2) + \mathcal{O}(\mathrm{d}y^2), \tag{7}$$

where the vertical bar represents the point at which these derivatives are evaluated. Now, we would like to have the right-hand side evaluated at the same point (x, y), and we realise that the difference between $\left(\partial f/\partial y\right)_{(x+\mathrm{d}x,y)}$ and $\left(\partial f/\partial y\right)_{(x,y)}$ is a quantity that depends on the rate of change of $\left(\partial f/\partial y\right)$ along x, i.e. the difference between the two quantities is $\left(\partial^2 f/\partial x \partial y\right)\mathrm{d}x$. But this term adds a term of the order of $\mathcal{O}(\mathrm{d}x\mathrm{d}y)$ in the definition of $\mathrm{d}f$, and since we're already ignoring quadratic terms, we can ignore this one too. Therefore, we can rewrite the above equation as

$$\mathrm{d}f \approx \left(\frac{\partial f}{\partial x}\right)_y \Bigg|_{(x,y)} \mathrm{d}x \quad + \quad \left(\frac{\partial f}{\partial y}\right)_x \Bigg|_{(x,y)} \mathrm{d}y \quad + \quad \mathcal{O}(\mathrm{d}x^2) + \mathcal{O}(\mathrm{d}y^2) + \mathcal{O}(\mathrm{d}x\mathrm{d}y). \tag{8}$$

Now, this notation is a little cumbersome, and since we know that we will always be evaluating small quantities like $\mathrm{d}f$ in the limit that all other small quantities like $\mathrm{d}x$ and $\mathrm{d}y$ tend to zero, we can omit the higher-order terms, and also assume that all the partial derivatives are to be evaluated at (x, y). Thus, we can write:

$$\mathrm{d}f = \left(\frac{\partial f}{\partial x}\right)_y \mathrm{d}x + \left(\frac{\partial f}{\partial y}\right)_x \mathrm{d}y. \tag{9}$$

We still keep the parentheses to remind ourselves which quantity is being held constant. It might seem pointless now, but – as we shall soon see – it can give rise to a lot of confusion when the same function f can be described using different coordinate systems.

Through all of our analysis so far, we have been using an "obvious" coordinate system, x and y, to describe the system. However, you might object to the importance that we are giving these two coordinates. After all, the hill exists independently of our choice to describe it, and so why should the choice of x or y be so important when we're determining the height of a point or its steepness? Indeed, what would happen if we

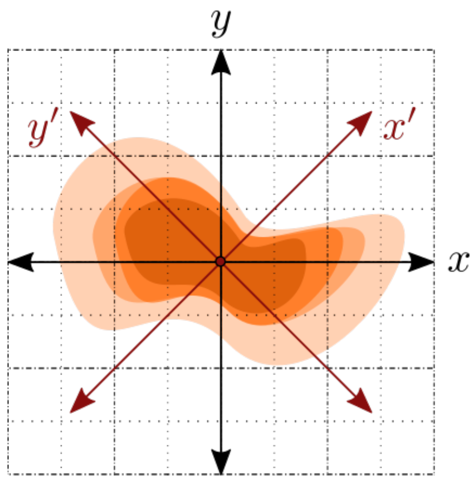


Figure 2: Two different coordinate systems. The primed coordinate system is rotated by some angle θ with respect to the un-primed coordinate system. The height of the hill, represented by the function f, is independent of our choice of coordinates, and therefore we should be able to represent it by either system. How do the partial derivatives in the un-primed system relate to those in the primed system?

were to use a different set of coordinate axes instead of (x, y)? There are, of course, many, many, possibilities: you could translate the original system so that the origin was not at the top of the hill. Alternatively, one could shift from Cartesian to polar coordinates, or even cylindrical coordinates. For simplicity, we will only consider a new coordinate system that keeps the origin fixed, but has its coordinate axes *rotated* by some angle (say, $\pi/4$).

2.2 Transformations of partial derivatives

One of the reasons that we use the rotated coordinate system is because we know quite well how the coordinates change in this case: the transformation is *linear*, and it is described using a rotation matrix R_θ, i.e. any point (x, y) is moved to another point (x', y') using the relations:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}. \tag{10}$$

Let's assume that you and a friend are trying to measure the steepness of a hill. You each choose a different coordinate system, and – from the same point – determine two different sets of partial derivatives. Let's say that you would like to know how these derivatives are related, so that you can compare the two sets. In order to do this, first keep in mind that the hill has an independent existence from the coordinate system. In other words, the height at some (x, y) and at the corresponding (x', y') is obviously the same, i.e. $f(x, y) = f(x', y')$. Let's now imagine that your friend (in the "primed" coordinates) wants to relate your derivatives to hers. To do this, she would first start with the relation for $\mathrm{d}f$ that we obtained in the previous section (Equation (9)), and then divide $\mathrm{d}f$ by a small change in x', keeping y' constant. In other words:

$$\left(\frac{\mathrm{d}f}{\mathrm{d}x'}\right)_{y'} = \left(\frac{\partial f}{\partial x}\right)_y \left(\frac{\mathrm{d}x}{\mathrm{d}x'}\right)_{y'} + \left(\frac{\partial f}{\partial y}\right)_x \left(\frac{\mathrm{d}y}{\mathrm{d}x'}\right)_{y'}. \tag{11}$$

Two things have to be kept in mind: firstly, we *did not* differentiate Equation (9), we just divided it by $\mathrm{d}x'$, and kept y' constant. This is why the partial derivatives with respect to x and y are left untouched. And secondly, the left-hand side is now just the partial derivative of f with respect to x' (keeping y' constant). All we need to do to calculate this is to find the two terms marked in red above.

In order to compute these terms, we need to be careful. We need to find the how x changes with x', along the curve $y' =$ constant. This would be different in general, from the way that x changes with x' along the curve $y =$ constant. To eliminate any chance of error, let us start by writing out all of the equations that describe how x, y, x', and y' are related, from the matrices described above:

$$\begin{aligned} x' &= x\cos\theta - y\sin\theta \quad &\textbf{(A)} \qquad x &= x'\cos\theta + y'\sin\theta \quad &\textbf{(C)} \\ y' &= x\sin\theta + y\cos\theta \quad &\textbf{(B)} \qquad y &= -x'\sin\theta + y'\cos\theta \quad &\textbf{(D)} \end{aligned} \tag{12}$$

Two of these equations (**A**, and **B**) take you from (x, y) to (x', y'), and the other two (**C**, and **D**) take you from (x', y') to (x, y). Now, we need to use two of these four equations to compute the two quantities we need. However, the standard mistake that is made is that the wrong equations are used. Luckily, it is quite simple to know which equations to use. In order to calculate a quantity like

$$\left(\frac{\mathrm{d}a}{\mathrm{d}b}\right)_c,$$

we need to use an equation with contains a, b, and c. You should convince yourself that

$$\left(\frac{\mathrm{d}x}{\mathrm{d}x'}\right)_{y'} \longrightarrow \text{Use Equation } \textbf{(C)},$$

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x'}\right)_{y'} \longrightarrow \text{Use Equation } \textbf{(D)}.$$

Exercise: Use the above argument to show that

$$\left(\frac{\mathrm{d}x}{\mathrm{d}x'}\right)_{y'} = \cos\theta, \qquad \text{and} \qquad \left(\frac{\mathrm{d}y}{\mathrm{d}x'}\right)_{y'} = -\sin\theta. \tag{13}$$

Putting everything together, you should be able to see that

$$\left(\frac{\partial f}{\partial x'}\right)_{y'} = \cos\theta\left(\frac{\partial f}{\partial x}\right)_y - \sin\theta\left(\frac{\partial f}{\partial y}\right)_x. \tag{14}$$

The above relation should hold *independent* of f. Notice how we did not use anywhere the type of function that f is, just so long that it is well-behaved and differentiable, and so on. Therefore, the above relation can be expressed as an *operator* relation:

$$\frac{\partial}{\partial x'} = \cos\theta\frac{\partial}{\partial x} - \sin\theta\frac{\partial}{\partial y}. \tag{15}$$

Notice how the quantities held constant have now been omitted; they will be inferred from the context. This is quite obvious to do in the case of rotated coordinate systems, but not quite so obvious in other domains like in Thermodynamics where one can take a host of different paths between two states of a system by holding different quantities constant (like the pressure, or the temperature, or the volume, and so on). However, the underlying idea is the same, and the sort of methods we have used to establish relations here should carry over very well into your course on Thermodynamics.

Exercise: Use the same methods as before to show that

$$\frac{\partial}{\partial y'} = \sin\theta \frac{\partial}{\partial x} + \cos\theta \frac{\partial}{\partial y}. \tag{16}$$

Using the above relations, it is easy to show that (in the case of two-dimensional rotations) the gradient transforms as a vector, meaning that if you define a column vector with the components of the partial derivatives in it, then you can write

$$\begin{pmatrix} \frac{\partial}{\partial x'} \\ \frac{\partial}{\partial y'} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix} \tag{17}$$

In fact, this is precisely why we often write the “gradient” operator as a vector $\vec{\nabla}$, because its components transform as a vector’s does under rotations.

3 Applications of the transformations of partial derivatives

3.1 Cartesian to polar coordinates

In the above case, we have dealt with linear transformations. However, our method generalises very easily to non-linear transformations. For example, consider the transformations that take us from Cartesian to polar coordinates given below (as before, we write both the forward and inverse transformations):

$$\begin{aligned} x &= r\cos\theta \quad \textbf{(A)} & r &= \sqrt{x^2+y^2} \quad \textbf{(C)} \\ y &= r\sin\theta \quad \textbf{(B)} & \theta &= \arctan\left(\frac{y}{x}\right) \quad \textbf{(D)} \end{aligned} \tag{18}$$

Just as before, we can relate the partial derivatives with respect to r and θ in terms of the partial derivatives with respect to x and y, using

$$\begin{aligned} \left(\frac{\partial}{\partial r}\right)_\theta &= \left(\frac{\partial x}{\partial r}\right)_\theta \left(\frac{\partial}{\partial x}\right)_y + \left(\frac{\partial y}{\partial r}\right)_\theta \left(\frac{\partial}{\partial y}\right)_x, \\ \left(\frac{\partial}{\partial \theta}\right)_r &= \left(\frac{\partial x}{\partial \theta}\right)_r \left(\frac{\partial}{\partial x}\right)_y + \left(\frac{\partial y}{\partial \theta}\right)_r \left(\frac{\partial}{\partial y}\right)_x. \end{aligned} \tag{19}$$

Exercise: Show that the partial derivative with respect to r can be written in terms of the partial derivatives of x and y as:

$$\begin{aligned} \frac{\partial}{\partial r} &= \frac{x}{\sqrt{x^2+y^2}}\frac{\partial}{\partial x} + \frac{y}{\sqrt{x^2+y^2}}\frac{\partial}{\partial y}, \\ \frac{\partial}{\partial \theta} &= -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}. \end{aligned} \tag{20}$$

Exercise: Now find the relations that get you from the partial derivatives with respect to r and θ to those with respect to x and y. Again, the method is identical.

3.2 Special Relativity

An interesting application of our analysis so far is to look at how the wave equation transforms in Special Relativity. From your earlier courses, you should be familiar with the Lorentz Transformations that take you between two inertial frames S and S', which moves with respect to S at some positive velocity v. These transformations (and their inverses) are given by

$$\begin{aligned} x' &= \gamma(x - vt) \quad \textbf{(A)} \qquad & x &= \gamma(x' + vt') \quad \textbf{(C)} \\ t' &= \gamma\left(t - \frac{v}{c^2}x\right) \quad \textbf{(B)} \qquad & t &= \gamma\left(t' + \frac{v}{c^2}x'\right) \quad \textbf{(D)} \end{aligned} \tag{21}$$

A more symmetric way of writing the Lorentz Transformations can be obtained if we use ct instead of just t. (Occasionally, the notation $x^0 = ct$ is used.) In this case, we can write them (in matrix form) as

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \underbrace{\gamma\begin{pmatrix} 1 & -v/c \\ -v/c & 1 \end{pmatrix}}_{\Lambda}\begin{pmatrix} ct \\ x \end{pmatrix} \tag{22}$$

The matrix above is often referred to as the Lorentz Boost matrix, since it takes you from one inertial frame (S) to another (S').[a] Quantities that transform using the Λ matrix are known as **four-vectors**, just as – in three-dimensional space – quantities that transform under a rotation using the rotation matrix R are known as vectors. As a result, we can find out if any quantity transforms as a four-vector by looking at how it changes under a boost.

[a] This action is called a "boost".

Using the same method outlined above, we can try to see how the partial derivatives transform.

Exercise: Show that the partial derivatives in the primed coordinates can be related to the partial derivatives in the un-primed coordinates using:

$$\begin{aligned} \frac{\partial}{\partial x'} &= \gamma\left(\frac{\partial}{\partial x} + \frac{v}{c} \times \frac{1}{c}\frac{\partial}{\partial t}\right), \\ \frac{1}{c}\frac{\partial}{\partial t'} &= \gamma\left(\frac{1}{c}\frac{\partial}{\partial t} + \frac{v}{c}\frac{\partial}{\partial x}\right). \end{aligned} \tag{23}$$

This can actually lead us to an interesting conclusion. If we write the above equations in matrix form, it shows us that

$$\begin{pmatrix} \frac{1}{c}\frac{\partial}{\partial t'} \\ \frac{\partial}{\partial x'} \end{pmatrix} = \gamma \begin{pmatrix} 1 & v/c \\ v/c & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{c}\frac{\partial}{\partial t} \\ \frac{\partial}{\partial x} \end{pmatrix} \tag{24}$$

You should be able to see quite clearly that the above matrix is *not* the Lorentz Transformation matrix we saw earlier. In the case of rotations, we saw that the three-gradient transformed as a three-vector (see Equation (17)). However, it looks like the four-gradient, i.e. the equivalent quantity with an additional partial derivative with respect to time, *does not* transform as a four-vector!

Exercise: Now, go back to Equation (23) and show the quantity formed by the $\left(-\partial/\partial(ct) \quad \partial/\partial x\right)^{\mathsf{T}}$ does indeed transform like a four-vector. In other words, show that:

$$\begin{pmatrix} -\frac{1}{c}\frac{\partial}{\partial t'} \\ \frac{\partial}{\partial x'} \end{pmatrix} = \gamma \begin{pmatrix} 1 & -v/c \\ -v/c & 1 \end{pmatrix} \begin{pmatrix} -\frac{1}{c}\frac{\partial}{\partial t} \\ \frac{\partial}{\partial x} \end{pmatrix} \tag{25}$$

This has some very important consequences, related deeply to the structure of space and time, which we will not go into at this moment.

3.3 The relativistic wave equation

Using the relations given in Equation (23), we can now try to see how the *wave-equation* transforms under a relativistic transformation. Let us consider a wave described by some function f, with some velocity u. Such an object satisfies the following wave equation:

$$\left(-\frac{1}{u^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2}\right) f = 0. \tag{26}$$

Exercise: Now, using the transformations of partial derivatives mentioned earlier, find out what this equation looks like in the inertial frame S'. In order to do this, you will need to find out how the second order partial derivatives transform. This is simple, but work through the calculations carefully.

Exercise: Now, show that if – and only if – $u = c$ does the wave equation conserve its form in both inertial frames. In other words,

$$\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2}\right) f = 0 \qquad \Longleftrightarrow \qquad \left(-\frac{1}{c^2}\frac{\partial^2}{\partial t'^2} + \frac{\partial^2}{\partial x'^2}\right) f = 0. \tag{27}$$

The above result has a very important implication: waves in one inertial frame don't transform into identical waves in another inertial frame, *unless* they are travelling at the speed of light. Which explains why the wave equation that arises naturally out of Maxwell's Equations – the wave equation that describes light – is frame independent. This is a very interesting result: it shows that the theory of classical Electromagnetism has, contained within it, a consistent logic of Special Relativity. In fact, one might go so far as to say that the discovery of Electromagnetism began not only a new field in physics, but also opened up a void that would soon be filled by theories described Nature much better than Newton's Laws.