This document is an introduction to the mathematical concepts and notation behind Einstein's general relativity. *It assumes the reader is already familiar with college-level math including vectors, derivatives, partial derivatives, and matrix math, as well as introductory physics.*

Numbers, Scalars, Vectors, Tensors

If I asked how many marbles you had in a bag, you might answer, "three." The number 3 by itself is enough information.

But if I were to ask how far is it to your house and you answered, "three," I would have to ask, "three *what*?". More information is required – a number with a unit, like "three km", is a **scalar**. Temperature is a scalar.

If I next asked how to get to your house from here and you said, "walk three km," again more information is needed – in what direction? So you might say, "three km *due north*." The scalar *3 km* now has *directional* information attached, which makes it a **vector**. If you were to tell me that to get to your house I must walk three blocks east, two blocks north, and go up three floors, the vector extending from "here" to "your house" would have three parts :

Three blocks east, Two blocks north, Three floors up

Each component of a vector **V** (**bold** means it's a vector) has scalar and vector parts, which can be represented as follows :

Scalar : a = 3 blocks, b = 2 blocks, and c = 3 floors are the vector's **components**

Vector : let **i** be a unit vector pointing east, **j** a unit vector pointing north, and **k** a unit vector pointing up ("unit vector" means it's one unit long, so in this case **i** and **j** are one block, **k** is one floor)

Then the total vector can be written as $\mathbf{V} = 3\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$, or more generally as $\mathbf{V} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$.

We usually drop the unit vectors, and just write **V** = (a, b, c). The parentheses indicate it is a vector.

Velocity (like wind speed) is a vector since it has a magnitude and direction, such as "10 miles per hour, heading east".

Why tensors are needed

Given a vector **V**, we can multiply **V** by a scalar (for example, *a*) to obtain *a***V**, a new vector with a different magnitude (length) but the same direction as **V**. But suppose we need to alter both the magnitude *and* the direction of a given vector – this is what tensors do.

For example, Fick's first law for diffusion in one dimension is :

$$J = -D\frac{\partial C}{\partial x}$$

This means that the *rate of flow J* (in number of molecules thru an area per second) moves *opposite* (minus sign) to the *gradient* (∂/∂x = rate of change in space) of the concentration *C*. The magnitude of *J* is proportional to the steepness of the gradient and a constant (scalar) *D* that depends on the situation. Imagine a drop of food coloring spreading out in a glass of water without stirring it. What this equation says is that the bigger the difference in concentration (like when you first drop the coloring in), the quicker it will diffuse. As it spreads out, the rate of spreading slows down. So the thinner it gets, the slower it spreads.

In three dimensions, this equation becomes :

$$\mathbf{J} = -D\,\nabla C$$

Here, ∇C and **J** are vectors and the ∇ ("del") operator is shorthand for the vector result $\nabla = (\partial/\partial x,\ \partial/\partial y,\ \partial/\partial z)$.

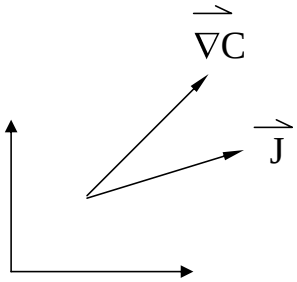
So this one equation really represents three equations, one for each component of the vector **J** :

$$J_x = -D\frac{\partial C}{\partial x},\quad J_y = -D\frac{\partial C}{\partial y}, \text{ and } J_z = -D\frac{\partial C}{\partial z}$$

This equation can be used as long as the material being studied acts the same in all directions, so that one value of the diffusion coefficient D works everywhere. Now suppose D is different in different directions – it has one value D_A in the *x* direction, another value D_B in the *y* direction, and D_C in the *z* direction. Then

J = – (D_A ∂C/∂x, D_B ∂C/∂y, D_C ∂C/∂z)

In this case **J** will <u>not</u> be parallel to the concentration gradient ∇C →



We can use matrix multiplication to represent this relationship as :

$$\begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix} = -\begin{bmatrix} D_A & 0 & 0 \\ 0 & D_B & 0 \\ 0 & 0 & D_C \end{bmatrix}\begin{bmatrix} \partial C/\partial x \\ \partial C/\partial y \\ \partial C/\partial z \end{bmatrix}$$

Symbolically, this can be written as :

J = – **<u>D</u>** ∇C

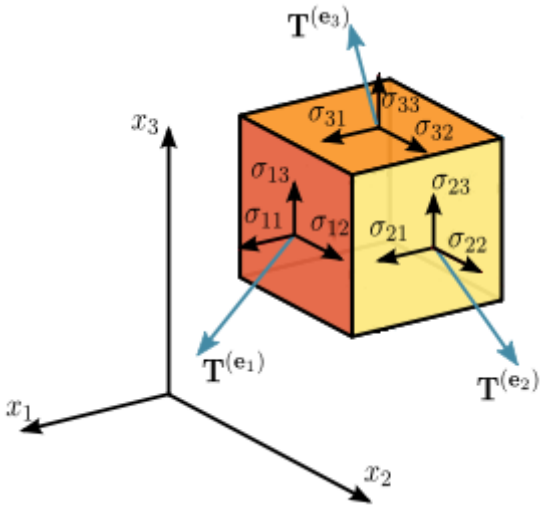
where the matrix **<u>D</u>** is a **tensor of rank 2** (the **<u>bold underline</u>** means it's a tensor) that represents the properties of the material in different directions. It is rank 2 because the matrix has two "directions" (rows and columns).

So the product of a vector and a tensor of rank 2 is another vector with both a new magnitude <u>and</u> a new direction. A tensor is like a function that "takes in" one vector and "produces" a new vector : **J** = **<u>D</u>**(∇**C**)

If something about the material causes <u>one</u> component of the gradient to affect a <u>different</u> component of J, then the tensor **<u>D</u>** is more generally written as

$$\begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix} = -\begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}\begin{bmatrix} \partial C/\partial x \\ \partial C/\partial y \\ \partial C/\partial z \end{bmatrix}$$

Where the vector (D_{xx} D_{xy} D_{xz}) is the direction of diffusion thru the surface pointing in the *x* direction, (D_{yx} D_{yy} D_{yz}) is the direction of diffusion thru the surface pointing in the *y* direction, and (D_{zx} D_{zy} D_{zz}) is the direction of diffusion thru the surface pointing in the *z* direction. → x1, x2, x3; all symbols wrong



"Tensor" is actually a general term – a tensor of rank 0 is a scalar, and a tensor of rank 1 is a vector. But unless indicated otherwise, from now on "tensor" will refer to a tensor of rank 2, as represented by a matrix. Vectors and tensors follow all the normal rules of matrix math.

Euclidean Spaces

Euclidean geometry is the study of <u>flat</u> spaces. Between every pair of points there is a unique *straight* line which is the shortest distance between those points. Furthermore, if you have a line and a point which isn't on the line, there is a unique second line running through the point which is *parallel* to the first line. And the

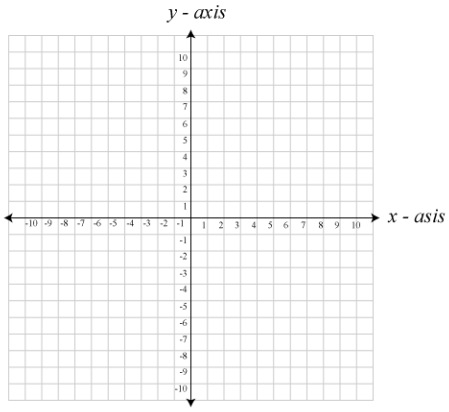
shortest line between two parallel lines is **orthogonal** (perpendicular) to both of the parallel lines. All of these ideas can be described by drawings on a *flat* piece of paper. From the laws of Euclidean geometry, we get Pythagoras' theorem and all the formulas you learned in geometry, like the distance between two points and how to find the circumference or area of a circle.

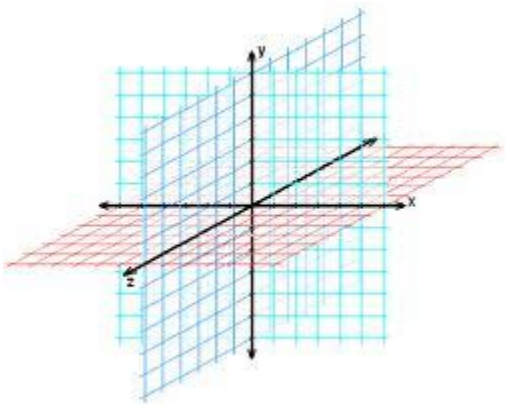
A two dimensional (or "2-D") Euclidean space is a plane (for example, a flat piece of paper); the ordinary world around us *appears* to be a 3-D Euclidean space.

Coordinate Systems

Coordinate systems are how we describe a space. For any given kind of space, there can be many coordinate systems that describe it. All of the following assumes a Euclidean space.

The most common (and intuitively obvious) coordinate system is the **Cartesian** system. In 2-D this is the typical graph with is horizontal *x*-axis and vertical *y*-axis. Every point on the plane is described by its distance along the *x* and *y* axis from the **origin** (where the axes intersect). In 3-D space, three values (*x*, *y*, *z*) are needed.





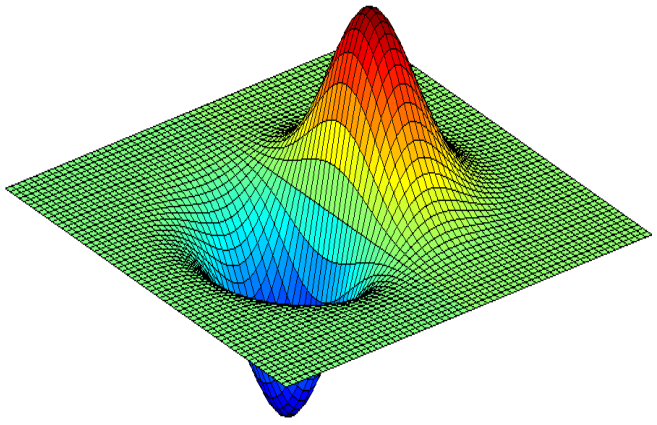
It is important to note that Cartesian axes are all at right angles to each other, or **orthogonal**.

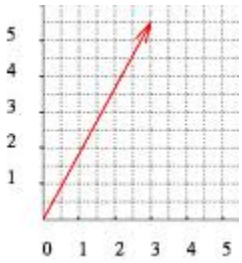
In 2-D, the distance between two points $u = (u_x, u_y)$ and $v = (v_x, v_y)$ is $[(u_x-v_x)^2 + (u_y-v_y)^2]^{1/2}$
In 3 dimensions, the distance between u and v is $[(u_x-v_x)^2 + (u_y-v_y)^2 + (u_z-v_z)^2]^{1/2}$

If T represents the temperature at a point, and using Cartesian coordinates to represent the space, we indicate that T depends on position by writing $T(x, y, z)$. Once we understand that a value depends on (x, y, z), we do not show it anymore, so we just write T.

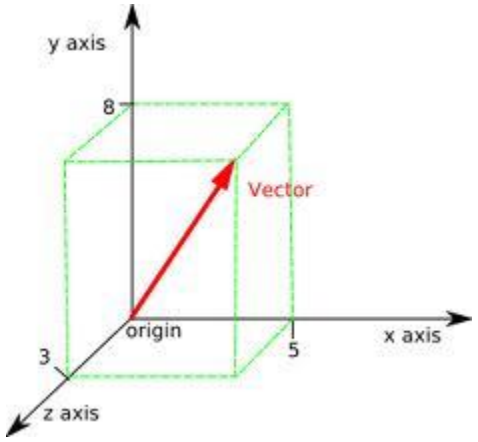
When a physical quantity (like temperature) is represented at every point in space by a scalar, we call it a **scalar field**. Scalar fields in 2-D space are often represented as surfaces, as shown below.

0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	-1	-2	-2	0	2	2	1	0	0
0	-1	-4	-10	-10	0	10	10	4	1	0
0	-3	-15	-34	-36	0	36	34	15	3	0
-1	-7	-31	-72	-76	0	76	72	31	7	1
-1	-9	-40	-92	-97	0	97	92	40	9	1
-1	-7	-31	-72	-76	0	76	72	31	7	1
0	-3	-15	-34	-36	0	36	34	15	3	0
0	-1	-4	-10	-10	0	10	10	4	1	0
0	0	-1	-2	-2	0	2	2	1	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0





To talk about vectors in a space, we need a way to describe the coordinates in terms of a **basis**, or <u>set of vectors</u> which together can represent every point in the space. For a 2-D Cartesian plane, the most natural basis vectors are <1,0> pointing in the *x* direction and <0,1> pointing in the *y* direction. There are many "standard" ways to represent basis vectors in formulas. Here, we will use $\mathbf{e}_x$ = <1,0> and $\mathbf{e}_y$ = <0,1>. Together, [$\mathbf{e}_x$, $\mathbf{e}_y$] is the basis. A basis vector will be represented by **e** or < > and a basis by [ ].

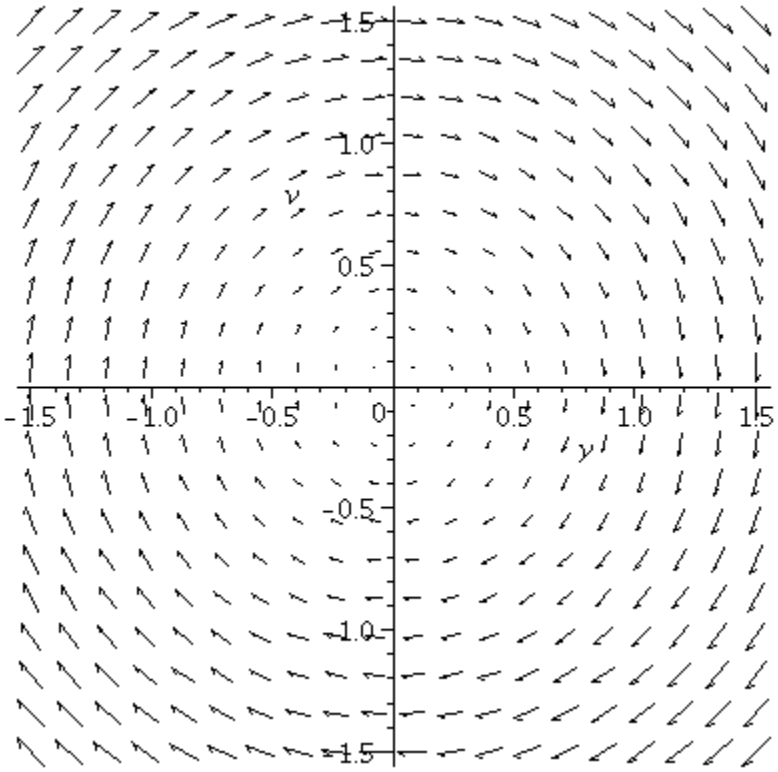


Once we have a basis, we can represent any vector in terms of its **components** in that basis. For example, in the 2-D figure to the upper right the vector **A** extends 3 units in the *x* or <1,0> direction, and 5.5 units in the *y* or <0,1> direction. So we can write $\mathbf{A} = 3\mathbf{e}_x + 5.5\mathbf{e}_y$, or just say the components of **A** are (3, 5.5). In the 3-D figure to the right, we would write **A** = (5, 8, 3). So when the axes or basis vectors are orthogonal, the way to find the i^{th} component of a vector is to project the vector perpendicularly onto the i^{th} axis.

When a physical quantity (such as wind velocity, diffusion, electric or magnetic field, etc.) is represented at every point in space by a vector, we call it a **vector field**. We might simply refer to the electric field as **E** but since it is a vector, it has three components (in 3-D Euclidean space) : $\mathbf{E} = (E_x, E_y, E_z)$. It is also understood that <u>each component</u> of **E** is a function of position :

$E_x = E_x(x, y, z)$ $E_y = E_y(x, y, z)$ $E_z = E_z(x, y, z)$

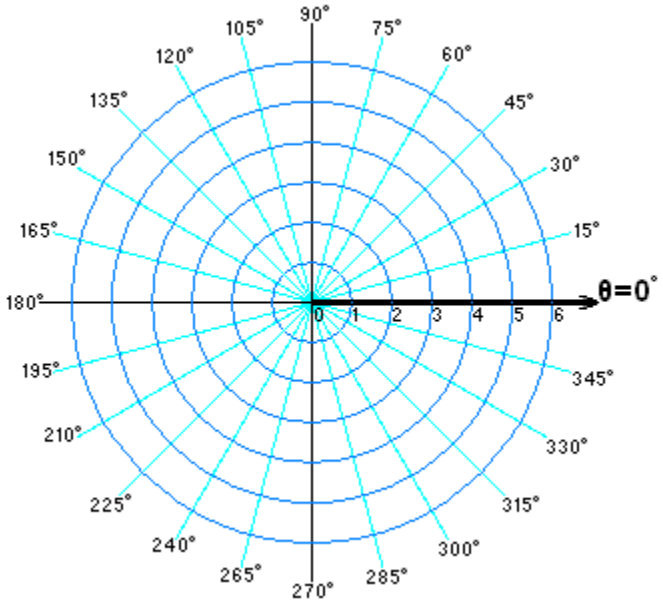
The image below is a visualization of some 2-D vector field. The arrows at each point show the direction the vector is pointing at that point. Since the direction of the arrows varies as *x* and *y* change, the vector components are functions of position. This image could represent (for example) the wind speed of a hurricane with the graph's origin at its center.

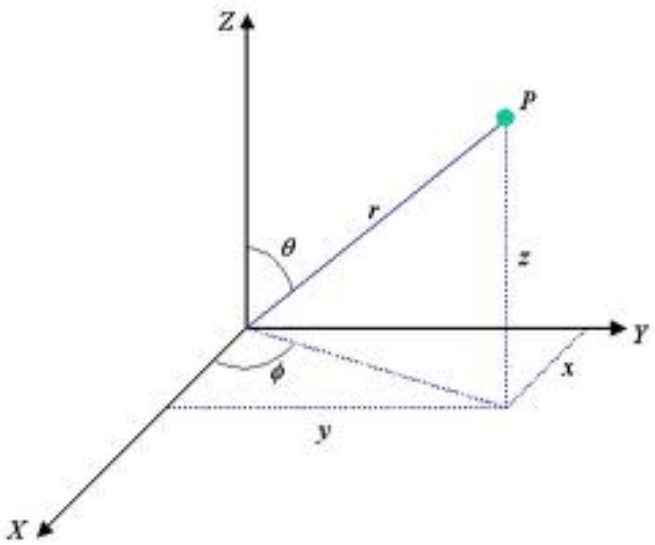


If the value of **<u>D</u>** (or any other tensor) changes with position in a material (due to defects, impurities) or with position in space, it is a **tensor field**.

Non-Cartesian Coordinate Systems

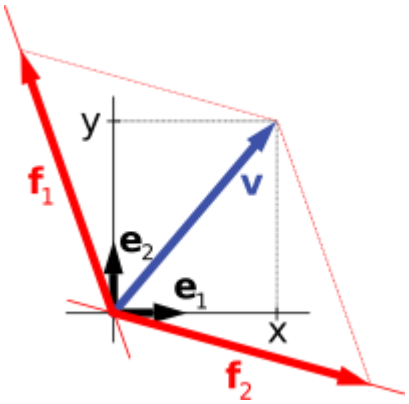
There are many different coordinate systems besides Cartesian that can be used to describe a space. A common one for 2-D Euclidean space is polar coordinates, where the two numbers to describe any point are its distance from the origin *r* and its angle θ to the *x*-axis (see below). In this case its basis is [$\mathbf{e}_R$, $\mathbf{e}_\theta$]. In 3-D Euclidean space two angles are needed, the E-W angle φ to the *x*-axis and θ, the N-S angle "down" from the *z*-axis (see below). So its basis is [$\mathbf{e}_R$, $\mathbf{e}_\varphi$, $\mathbf{e}_\theta$]. It is important to understand that both of these coordinate systems <u>are still orthogonal</u>, because as we vary the angle(s) we are always moving (locally) perpendicularly to the radial axis.

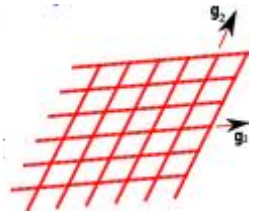




Non-orthogonal Coordinate Systems

There is nothing that says basis vectors (axes) have to be orthogonal to each other. It is perfectly valid to create a basis using [$\mathbf{f}_1$, $\mathbf{f}_2$]. → The vector **v** is not changing, just the way we describe it.





A non-orthogonal coordinate grid

Curvilinear Coordinate Systems

If we consider the Cartesian coordinate system, we find some specific characteristics :

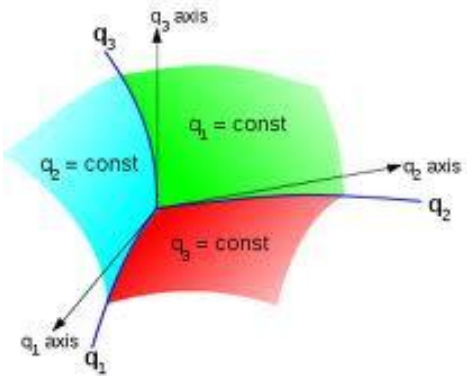
The coordinate axes are straight lines defined to intersect at a single point, the origin.
The coordinate axes are mutually orthogonal.
The coordinate planes are completely determined by the axes.

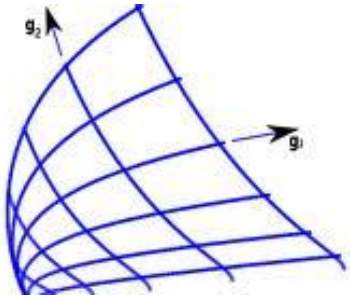
Suppose we relax these conditions :

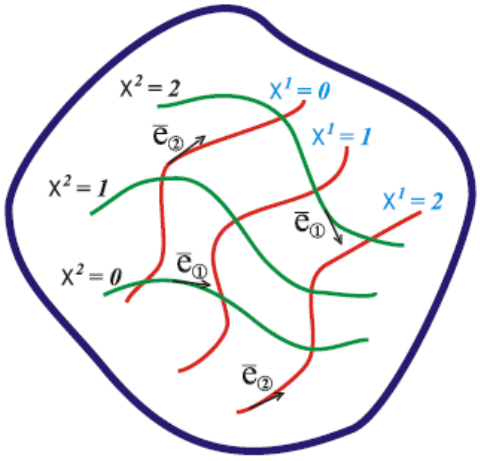
The coordinate axes are *general curves* defined to all intersect at least once. Any point of intersection can be chosen as the origin.
The coordinate axes are not *necessarily* mutually orthogonal.
Pairs of coordinate axes uniquely determine curvilinear coordinate surfaces as *product spaces.*

What is a **product space**? First, consider the case of Cartesian axes. If one axis is slid along another, a plane is created. The plane is said to be the product space of the two lines. The coordinates of all points in this plane are automatically determined in the process, for the moving axis has coordinates (…–2, –1, 0, 1, 2,…) marked on it, and as it slides across the other axis which also has coordinates (…–2, –1, 0, 1, 2,…) marked on it, each point P in the plane is uniquely marked by a pair of numbers : P = (x_0, y_0). Next, consider a straight line and a circle. Let the line extend perpendicularly out from the center of the circle. Now slide the circle along the line. The result is that a *cylinder* will be swept out. The cylinder is the product space of the circle and the line. If coordinates are marked on the line and the circle, then a unique pair of numbers will specify every point on the surface of the cylinder.

Finally, start with *any* two curves, and let the curves intersect at one point. Mark the curves with coordinates as before. Slide one curve along the other to produce a surface. Then a pair of numbers from the curves will specify any point on the surface in perfect analogy with the Cartesian plane. If the curves are called *u* and *v*, we then say that we have a *u*-axis and a *v*-axis. These axes together produce a *uv*-surface (a coordinate surface) as a product space. As before, any point P in this surface is also specified by a pair of numbers : P = (u_0, v_0). Such a system is called a **curvilinear** (or "generalized") coordinate system, in which the angles between axes can change from point-to-point, so the basis vectors are functions of position, and are usually non-orthogonal.







It is important to understand that (so far) the space itself is still flat, it is just the coordinate axes we are using to describe it that are curved. Polar and spherical coordinates are curvilinear (but orthogonal).

Coordinate Independence

Tensors (of any rank) are defined by how they change with respect to a "coordinate transformation". The physical quantity that a tensor represents must appear in certain ways to different observers with different points of view. Suppose I measure the temperature (°C) at a given point P at a given time. You also measure the temperature (°C) at P at the same time but from a different location that is in motion relative to my location. Would it make any sense if we measured different values; for example, my thermometer measured 25 °C and yours measured 125 °C? No – we must both obtain the same values from our measurements. Put another way, suppose that I call my point of view (my "coordinate system") K and yours K*. Let T be the temperature (°C) measured at P in K and T* be the temperature (°C) measured in K*. We then *require* T = T*. *Only scalars that remain* ***invariant*** *between coordinate systems like this can be called "tensors of rank 0".*

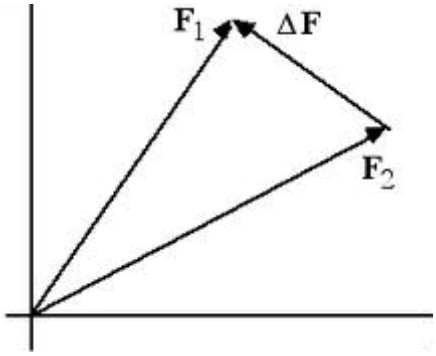
Now let *f* be the frequency of light coming from a laser at P. Again, let two observers, K and K*, measure the frequency of the light at P at the same time using the same units (Hz). If I am stationary relative to the source, the light will have a certain frequency, for example *f*. If, on the other hand, you are moving toward or away from the source, the light will be red or blue shifted to frequency f*. Obviously $f \neq f^*$ in this case, and although frequency is a scalar, it is *not* a tensor of rank 0.

A similar argument holds for vectors. Suppose a vector quantity **V** exists at a point P. Again, assume two reference frames, K and K*. Let **V** be the vector measured in K, and **V*** be the same vector measured in K* at the same time. As with the temperature example, we again require that **V** = **V*** since, after all, K and K* are both observing the same vector. Any other result would not make physical sense. Any vector that transforms according to the expression **V** = **V*** is *defined* to be a *"tensor of rank 1*". We say that the transformation laws T = T* or **V** = **V*** require the quantities represented by T or **V** to be *coordinate independent*.

As with scalars, not all vectors are tensors of rank 1. For example, let **V** be the *position vector* extending from the origin of K to a particular point P, and **V*** be the position vector from the origin of K* to that same point. If the origins of K and K* *do not coincide*, then **V** ≠ **V***. So position vectors are coordinate *dependent*, and are not used in general relativity.

But suppose that **V1** and **V2** were position vectors of points P1 and P2 in K (see the figure to the right), and that **V1*** and **V2*** were position vectors to the same points P1 and P2 in K*. Then the vector extending from P1 to P2 *must be the same vector in both systems*. This vector is **V2** – **V1** in K and **V2*** – **V1*** in K*. Thus we have **V2** – **V1** = **V2*** – **V1***. So while the position vector itself is not a tensor, the *difference* between any two position vectors is a tensor of rank 1! Similarly, dV = dV*, which means the differential of the position vector is a tensor of rank 1.

F1→V1, F2→V2, ΔF→V2-V1; add F*'s?



Index Notation

It is often useful to represent vectors and tensors using an index notation. Consider the 3-D vector

$\mathbf{v} = (v_x, v_y, v_z)$

We can also represent **v** by using the notation "v_i" where it is understood that the index *i* represents "*x*", "*y*", and "*z*" (all three at once, which means we are talking about the entire vector, not a specific single component).

So an equation like $a_i + S \cdot b_i = 0$ really means

$a_x + S \cdot b_x = 0$ $\qquad$ $a_y + S \cdot b_y = 0$ $\qquad$ $a_z + S \cdot b_z = 0$

In addition, we are usually not concerned with a particular coordinate system, so instead of *i* representing "*x*", "*y*", and "*z*" (which implies Cartesian coordinates), we let *i* represent generic coordinates "x_1", "x_2", and "x_3".

In Cartesian coordinates : $x_1 = x$ $\quad$ $x_2 = y$ $\quad$ $x_3 = z$
In spherical coordinates : $x_1 = r$ $\quad$ $x_2 = \varphi$ $\quad$ $x_3 = \theta$

Sometimes we even drop the "x" part altogether and just let *i* = 1, 2, 3 :

In Cartesian coordinates : $1 = x$ $\quad$ $2 = y$ $\quad$ $3 = z$
In spherical coordinates : $1 = r$ $\quad$ $2 = \varphi$ $\quad$ $3 = \theta$

So now an equation like $a_i + S \cdot b_i = 0$ means

$a_1 + S \cdot b_1 = 0$ $\qquad$ $a_2 + S \cdot b_2 = 0$ $\qquad$ $a_3 + S \cdot b_3 = 0$

Which in Cartesian coordinates would still be :

$a_x + S \cdot b_x = 0$ $\qquad$ $a_y + S \cdot b_y = 0$ $\qquad$ $a_z + S \cdot b_z = 0$

But in spherical coordinates would be :

$a_r + S \cdot b_r = 0$ $\qquad$ $a_\theta + S \cdot b_\varphi = 0$ $\qquad$ $a_\varphi + S \cdot b_\theta = 0$

We can also represent a tensor **D** as

D_{ij}

where the index *i* corresponds to the row and *j* to the column of the tensor's matrix. For example, D_{12} would be the element in the first row and second column of **D**. If we were using the Cartesian coordinate system, this would be D_{xy}. As for vectors, writing D_{ij} means that we are talking about the entire tensor **D**, not a single component.

Note that when using index notation, the number of indexes automatically tells you whether the symbol is a scalar (no indexes), vector (one index), or a tensor (two indexes), so there is no need for bolding or underlines. But a bold letter with an index such as $\mathbf{e}_i$ represents a set of vectors such as a basis : $\mathbf{e}_i = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$.

Einstein (Summation) Notation

Einstein notation is a convention that says in terms with indexed vectors or tensors, if an index letter is *repeated as a subscript and a superscript*, then the equation is *summed* over all values of the index. The difference between a subscript ("lower") and superscript ("upper") index is explained in GR1b, just ignore it for now. The index may vary over (1,2,3) when representing just space, or (0,1,2,3) when representing spacetime (0 represents the time coordinate). NOTE : in some cases (usually non-general-relativity physics), the "as a subscript and a superscript" requirement is ignored.

In relativity, the Greek alphabet and the Roman alphabet are often used to distinguish whether the index is summed over 1,2,3 or 0,1,2,3. Often, Roman (i, j, k, ...) is used for 1,2,3 and Greek (α, β, μ, ν, ...) is used for 0,1,2,3. NOTE : some authors reverse the roles of Roman and Greek letters, or ignore this convention. Unless stated otherwise, Greek and Roman letters are used interchangeably here.

This allows an equation such as

$y = c_1x_1 + c_2x_2 + c_3x_3 + \ldots + c_nx_n$

Typically written as

$$y = \sum_{i=1}^{n} c_i x_i$$

To be replaced with something even simpler :

$y = c_i\, x^i$

For a familiar example, consider the dot product of two vectors **a** and **b** :

$a^\alpha b_\alpha = a^1b_1 + a^2b_2 + a^3b_3 = \mathbf{a} \cdot \mathbf{b}$

More examples :

$$A_{ik}B^{ks}C_{js} = \sum_{k=1}^{n}\sum_{s=1}^{n} A_{ik}B^{ks}C_{js}$$

$$T^{\alpha\beta}S_{\alpha\beta} = \sum_{\alpha,\beta=0}^{3} T^{\alpha\beta}S_{\alpha\beta} = T^{00}S_{00} + T^{10}S_{10} + T^{20}S_{20} + T^{30}S_{30} + T^{01}S_{01} + \ldots$$

$$R^{\rho}{}_{\mu\rho\nu} = \sum_{\rho=0}^{3} R^{\rho}{}_{\mu\rho\nu} = R^{0}{}_{\mu 0\nu} + R^{1}{}_{\mu 1\nu} + R^{2}{}_{\mu 2\nu} + R^{3}{}_{\mu 3\nu}$$

The index being summed over is called a **dummy**, **summation**, **repeated**, or **bound** index, must appear exactly twice in any term as an upper/lower pair, and should only appear on the right-hand side of the equation. An index that is not summed over is a **free index**, cannot appear more than once in any term, should appear on both sides of the equation, and must be consistently upper or lower.

Some examples of improper use include :

$A_{ij}B^{i}C^{i}$	more than two instances of i
$A_{ij}A_{jk}$	j is not both subscript and superscript
$A_{ij} = B_{ik}C^{k}$	free index j isn't on right-hand side
$A^{i} = B^{i}{}_{j}C^{ij}$	free index i appears twice in a term

Any index can be changed to another letter, as long as every instance of the letter is changed consistently, and does not conflict with existing letters. Dummy indexes may be changed term-by-term if needed.

For example : $A_s = F_{kms} + G_{mks}$ (*s* is the free index, *k* and *m* are dummy indexes)

Can be changed to :

$A_b = F_{kmb} + G_{mkb}$ or $A_s = F_{bns} + G_{nbs}$

Cannot be changed to :

$A_b = F_{kmc} + G_{mkc}$ or $A_s = F_{nns} + G_{nns}$

Another example : $\partial\mathbf{V}/\partial x^{\beta} = (\partial V^{\alpha}/\partial x^{\beta})\mathbf{e}_{\alpha} + V^{\alpha}\Gamma^{\mu}{}_{\alpha\beta}\mathbf{e}_{\mu}$

Becomes (with $\alpha \leftrightarrow \mu$ in the second term)

$\partial\mathbf{V}/\partial x^{\beta} = (\partial V^{\alpha}/\partial x^{\beta} + V^{\mu}\Gamma^{\alpha}{}_{\mu\beta})\mathbf{e}_{\alpha}$

Comma, Semi-colon, and Del Notation

Remember that ∇ ("del") is shorthand for the vector result ($\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$) in Cartesian coordinates.

A comma followed by an index (*i* for example) means to take the partial derivative with respect to the i^{th} coordinate.

So the gradient of a scalar function J is represented by : $J_{,i} = \nabla_i J = \partial J / \partial x_i$ for all i = (1,2,3)

Other examples :

$J_{ij,k} = \partial J_{ij} / \partial x_k$ for all i,j,k = (1,2,3)

$J^{i}{}_{,i} = \partial_i J^{i} = \partial J^{i} / \partial x_i = \partial J^{1}/\partial x_1 + \partial J^{2}/\partial x_2 + \partial J^{3}/\partial x_3$ (divergence)

Note : if there are multiple partial derivatives, they commute : $J_{i,jk} = J_{i,kj}$

A semi-colon means to take the "covariant derivative", which is a derivative of the vectors tangent to the surface (see the "Manifolds" section below). The equation it represents is too complicated to explain briefly, but is explained fully in GR2c. For example, in general relativity the "continuity equation" is written as :

$$J^{i}{}_{;i} = \nabla_i J^i = 0$$

Note that $\nabla_\alpha J^\beta$ is a *covariant derivative* (because J^β implies a vector), while $\nabla_\alpha J$ is the *gradient* of the scalar J (no indexes).

Putting It All Together

Earlier, Fick's law in three dimensions was given as

$$\mathbf{J} = -\mathbf{D}\,\nabla C$$

Using index and comma notation this can be written as

$$J_i = -\sum_j D_{ij} C_{,j}$$

Or in Einstein notation (don't worry about the lower i/upper j right now) as

$$J_i = -D_i^j\, C_{,j}$$

It is important to understand that this equation actually represents 3 equations, one *each* for $i = 1, 2, 3$, and that each of these equations contains 3 terms, *summed* over $j = 1, 2, 3$. The expanded form using Cartesian coordinates is

$$J_x = -D_{xx}\frac{\partial C}{\partial x} - D_{xy}\frac{\partial C}{\partial y} - D_{xz}\frac{\partial C}{\partial z}$$

$$J_y = -D_{yx}\frac{\partial C}{\partial x} - D_{yy}\frac{\partial C}{\partial y} - D_{yz}\frac{\partial C}{\partial z}$$

$$J_z = -D_{zx}\frac{\partial C}{\partial x} - D_{zy}\frac{\partial C}{\partial y} - D_{zz}\frac{\partial C}{\partial z}$$

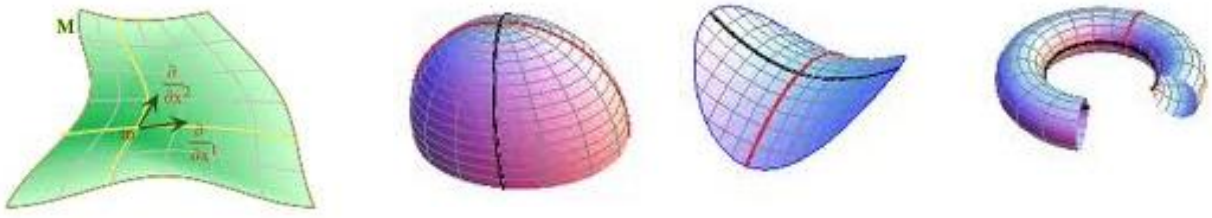
Manifolds

Remember the description of Euclidean spaces from before. Now, suppose instead of having a flat piece of paper, you have a curved piece of paper, such as a cylinder or a sphere. The shortest curved line between any pair of points on a *curved* surface is called a **geodesic** (such as along the equator of the Earth). You can find a geodesic between two points by stretching a rubber band between them. Depending on the surface, there may be more than one geodesic between two points (for example, there are many geodesics between the north and south poles of a globe).

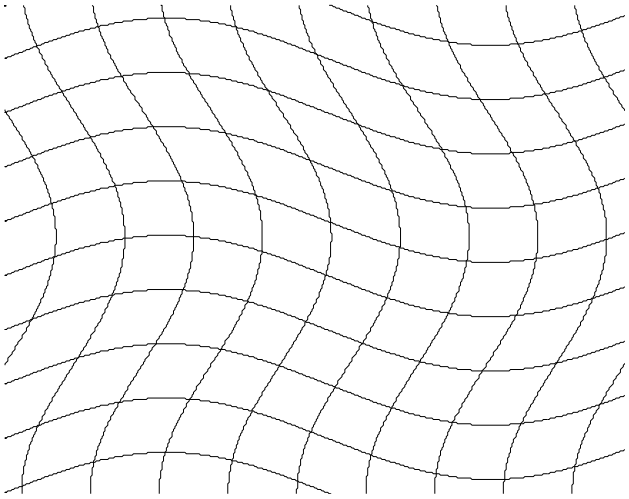
Surfaces like these are harder to study than flat surfaces, but there are still formulas which can be used to calculate the length of the hypotenuse of a triangle, the area inside the circle, etc.

A **manifold** is a space that may be curved but which is *locally* flat. "Locally flat" means that you can cover the entire space with "patches" (squares, cubes, etc.) that are small enough so that they are approximately Euclidean. So a line and a circle are 1-D manifolds, and a plane and the surface of a sphere are 2-D manifolds. Consider the surface of the earth – ignoring hills and such, locally it looks quite flat.

In general relativity, spacetime is a 4-D manifold. While the figures below can give some idea of what a curved manifold is, keep in mind that they are 2-D manifolds, and the only reason we can see their curvature is because they are shown "embedded" in a 3-D space. So to correctly *visualize* our 4-D spacetime's curvature, we would need to place it in an even higher-dimensional space! Good luck visualizing that…



But this does not mean that our spacetime is actually embedded in some higher-dimensional space – when our spacetime stretches, it is really stretching "within" itself. The following image is perhaps a better example – imagine a flat rubber sheet with a Cartesian grid on it, which is then pulled up on the top left edge and to the right on the right edge, all the time staying flat :



Creatures living in these 2-D spaces would be stuck "in" the surface, and would not necessarily realize that their space is curved because they are curving with it! **A creature in the above picture would think that their coordinate system is Cartesian.** Similarly, we cannot tell "just by looking" when our 4-D spacetime is curved (but it can be determined by making certain measurements).

From now on the word "surface" will be used instead of "manifold" to help visualize what's going on. But don't be fooled – in reality, the entire universe is the "surface" in general relativity, and it doesn't "bend" into a larger space!

Tangent Vectors and Tangent Planes

General relativity often refers to the "**tangent vectors**" at a point, or "vectors tangent to the surface". What this means is that the vector touches the surface at that point, and (assuming the surface is curved) the vector is parallel to the local surface at that point. The pictures below should give some idea of this concept. Note that in 2-D (or higher) surfaces, there can be more than one vector tangent to the same point.

For an N-dimensional surface, N tangent vectors are needed to form a *basis* for any **tangent plane**. In the 2-D surface on the right, the green plane is tangent to the purple surface at one point, and all the tangent vectors lie within the plane. Any two non co-linear vectors could be a (possibly non-orthogonal) basis for the plane. Again, the phrase "tangent plane" will be used to aid with visualizing what's going on, but in a 4-D spacetime this would really be a 4-D tangent "volume" (with time as one of its sides!).