1 Maps between real vector spaces

1.1 General notation

- $\mathbb{R}$: the set of real numbers, which we think of as points on a line.
- $\mathbb{R}^n$: the set of ordered n-tuples $(x_1, x_2, \ldots, x_n)$ where each x_i is real ($x_i \in \mathbb{R}$). Such an n-tuple can be thought of as the cartesian coordinates of a point in n-dimensional space. $\mathbb{R}^n$ is a real vector space (see Linear Algebra I for the full set of axioms of a real vector space), and so we refer to elements of $\mathbb{R}^n$ as n-dimensional vectors. You may be used to using column vector notation for such vectors, but here we will use row vector notation.
- Recall that the standard basis vector e_i is defined as the vector with a 1 in the i^{th} position, and a 0 in all other positions, e.g. $x_2 = (0, 1, 0, \ldots, 0)$.

  The standard basis vectors $\underline{e}_i$ are othonormal (orthogonal, and normalised to have length 1) with respect to the scalar (dot) product on $\mathbb{R}^n$:

  $$\underline{e}_i . \underline{e}_j \;=\; \left\{ \begin{array}{ll} 1, & i = j \\ 0, & i \neq j \end{array} \right\} \;\equiv\; \delta_{ij}.$$

  δ_{ij} as just defined is called the Kronecker delta, and we will revisit this in our later section on index notation 4.
- In terms of the standard basis of $\mathbb{R}^n$, $\{e_1, e_2, \ldots, e_n\}$, the position vector of a point in $\mathbb{R}^n$, $\underline{x}$ can be written as

  $$\underline{x} \;=\; x_1\underline{e}_1 + x_2\underline{e}_2 + \cdots + x_n\underline{e}_n \;=\; \sum_{i=1}^{n} x_i\underline{e}_i \;=\; x_i\underline{e}_i.$$

- Note: Index notation (as in the final expression in (1.1), $x_i\underline{e}_i$) is a very important convention: when we see an index (like i) repeated, we will assume that it is to be summed over from 1 to n and leave off the "$\sum_{i=1}^{n}$". This convention is known as the **Einstein Summation Convention** (ESC). Index notation is sufficiently important in vector calculus that we will have a whole section on it later in the course.
- For low values of n, we will often write $x_1, x_2, \ldots$ as $x, y, \ldots$

  $$\begin{array}{lrcl} \text{e.g.} & n = 2: & \underline{x} & = x\underline{e}_1 + y\underline{e}_2 \\ & n = 3: & \underline{x} & = x\underline{e}_1 + y\underline{e}_2 + z\underline{e}_3. \end{array}$$

- Important: **Don't forget the vector signs**, i.e. $\underline{x} \neq x$. xe_i has a quite different meaning to $\underline{x}.e_i$.
- Given two vectors $\underline{u}, \underline{v} \in \mathbb{R}^n$

  $$\underline{u} = u_1\underline{e}_1 + u_2\underline{e}_2 + \cdots + u_n\underline{e}_n \;=\; \sum_{i=1}^{n} u_i\underline{e}_i \;=\; u_i\underline{e}_i$$

  $$\underline{v} = v_1\underline{e}_1 + v_2\underline{e}_2 + \cdots + v_n\underline{e}_n \;=\; \sum_{i=1}^{n} v_i\underline{e}_i \;=\; v_i\underline{e}_i$$

  (note the use of the Einstein summation convention again here), the scalar (dot) product between these vectors is then

  $$\underline{u}.\underline{v} \;=\; u_1v_1 + u_2v_2 + \cdots + u_nv_n \;=\; \sum_{i=1}^{n} u_iv_i \;=\; u_iv_i.$$

  This can easily be proved by multiplying out $\underline{u}.\underline{v}$, and using the fact that the standard basis vectors are orthonormal.

- The <u>length</u> or <u>magnitude</u> of $\underline{u}$ is found using the dot product between $\underline{u}$ and itself,

$$|\underline{u}| \; = \; \sqrt{\underline{u}.\underline{u}},$$

and if θ is the <u>angle</u> between $\underline{u}$ and $\underline{v}$, then

$$\underline{u}.\underline{v} \; = \; |\underline{u}| \; |\underline{v}| \cos(\theta).$$

$|\underline{x}|$ is sometimes denoted r (note that this is a scalar quantity, and not a vector), since it is the radial coordinate of $\underline{x}$ in spherical polars. Similarly $\underline{x}$ is sometimes denoted $\underline{r}$.

- Draw vectors as arrows:

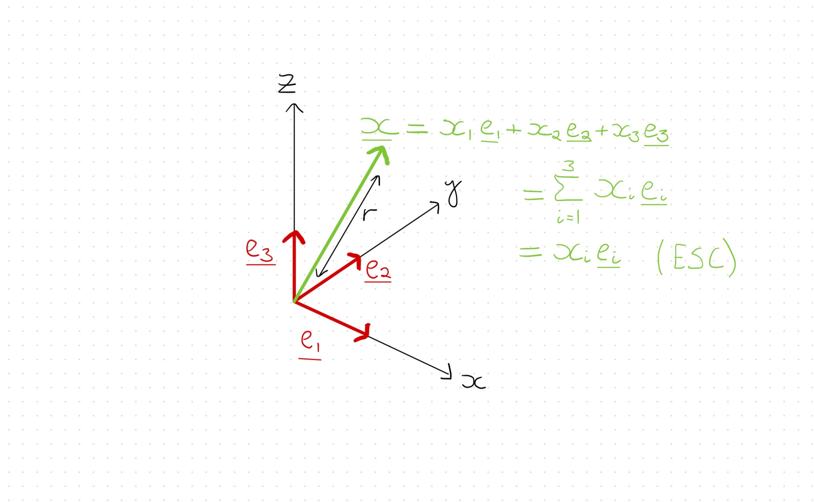


Figure 2: The vector $\underline{x}$ illustrated as the position vector of a point in $\mathbb{R}^3$. The length of $\underline{x}$, r, is shown, as are the standard basis vectors of $\mathbb{R}^3$. Using Einstein Summation Convention, we can write $\underline{x}$ in terms of the standard basis as $\underline{x} = x_i e_i$

1.2 Scalar fields, vector fields and curves

Now we can introduce the main objects of study in this course, scalar fields, vector fields and curves. These are all defined as maps from one real vector space to another, that is as maps from $\mathbb{R}^m \to \mathbb{R}^n$.

- **Scalar fields** are real-valued functions on $\mathbb{R}^n$, i.e. maps

$$\begin{aligned} \mathbb{R}^n &\to \mathbb{R} \\ \underline{x} &\mapsto f(\underline{x}) \end{aligned}$$

e.g. for $n = 3$, we could have the function f defined as

$$f(\underline{x}) \; = \; \frac{xy}{\tan z}.$$

Note that here the argument of the function is underlined to show that it's a vector quantity. $f(\underline{x})$ could also be written as $f(x, y, z)$.

The functions of two variables that you studied in the Epiphany term of Calculus I were all examples of scalar fields.

- **Vector fields** are vector-valued functions on $\mathbb{R}^n$, i.e. maps

$$\begin{aligned} \mathbb{R}^n &\to \mathbb{R}^n \\ \underline{x} &\mapsto \underline{f}(\underline{x}). \end{aligned}$$

Note that since the image of $\underline{x}$ is also a vector, we underline the function $\underline{f}$ to indicate this. In textbooks you may also see vector fields written in bold, rather than underlined.

Example 1. Given a constant vector $\underline{a} \in \mathbb{R}^n$, we might have the vector field $\underline{f}$ on $\mathbb{R}^n$, given by

$$\underline{f}(\underline{x}) = (\underline{a}.\underline{x})\underline{x}.$$

If we let $n = 2$ and take $\underline{a} = (1, 1)$, then the vector field near the origin looks as in Figure 3.

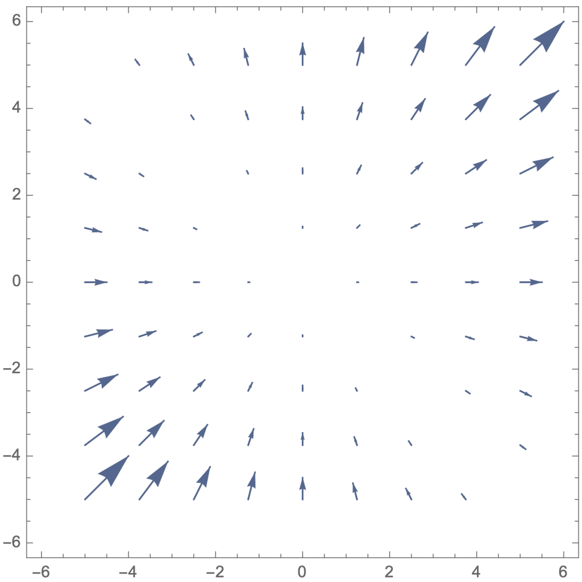


Figure 3: A plot of the vector field $\underline{f}(\underline{x}) = (a.x)x$ near the origin in 2 dimensions with $a = (1, 1)$. The vectors are drawn at a sample of points $\underline{\mathrm{x}}$, as arrows.

Pay careful attention to the two different types of multiplication being used in this example. We take the *scalar product* between $\underline{a}$ and $\underline{x}$, and we can then use *scalar multiplication* to multiply the vector $\underline{x}$ by the scalar $(\underline{a}.\underline{x})$.

Note: We can also write the formula "in components", i.e. using index notation - by giving a formula for the ith component of $\underline{f}$. So example 1 would be (using index notation and ESC):

$$f_i \quad = \quad (a_j x_j)x_i$$

- **Curves** in $\mathbb{R}^n$ are given parametrically by specifying $\underline{x}$ as a function $\underline{x}(t)$ of some parameter, t say. That is, a curve is a map

$$\begin{aligned} \mathbb{R} &\to \mathbb{R}^n \\ t &\mapsto \underline{x}(t). \end{aligned}$$

Since the image of t is a vector quantity, we underline the function $\underline{x}(t)$ to indicate this. t itself is a scalar quantity however, and so is not underlined.

Example 2. Given constant vectors $\underline{a}, \underline{b} \in \mathbb{R}^n$, the curve

$$\text{e.g.} \quad \underline{x}(t) = \underline{a} + t\underline{b},$$

which can be written in components as

$$x_i(t) = a_i + tb_i,$$

is a straight line in $\mathbb{R}^n$, which goes through the point $\underline{a}$ and is parallel to $\underline{b}$.

If $\underline{x}(t)$ is differentiable, then $\frac{d\underline{x}}{dt}$ is tangent to the curve (if non-zero). (If you studied Dynamics I, then you've already come across this idea in that course. There, the *trajectory* of a particle $\underline{r}(t)$ was a curve in space, parameterised by time t, the *velocity* was the derivative of the trajectory with respect to time, $\frac{d\underline{r}(t)}{dt}$, and the *acceleration* was the second derivative $\frac{d^2\underline{r}(t)}{dt^2}$.)

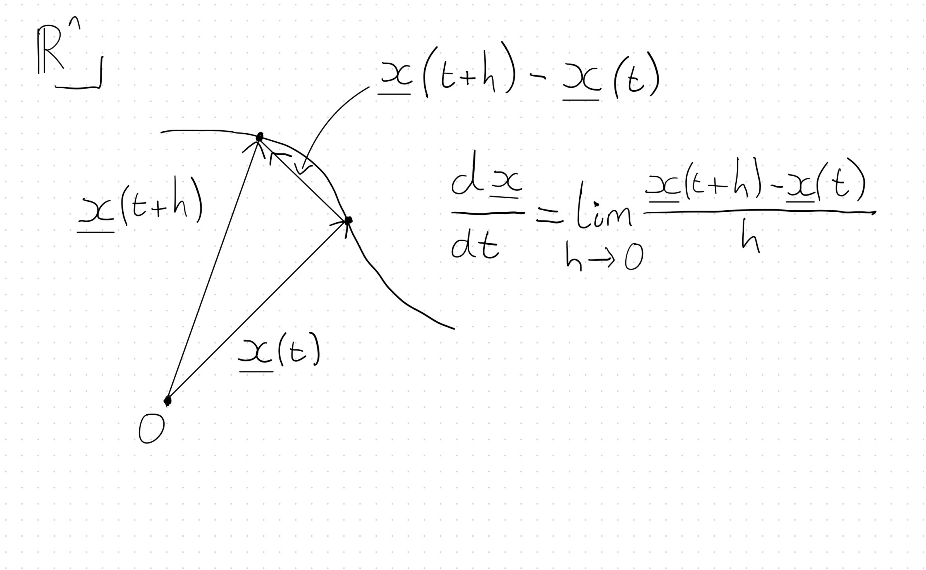


Figure 4: The tangent to a curve $\underline{x}(t)$ is found by differentiating the curve with respect to t.

Example 3. A helix in $\mathbb{R}^3$ can be parameterised as

$$\underline{x}(t) = \cos(t)\underline{e}_1 + \sin(t)\underline{e}_2 + t\underline{e}_3.$$

The tangent to the helix is therefore given by

$$\frac{d\underline{x}}{dt}(t) = -\sin(t)\underline{e}_1 + \cos(t)\underline{e}_2 + \underline{e}_3.$$

Note that the standard basis vectors are constant, and hence the components of the derivative of $\underline{x}(t)$ with respect to t, are just the derivatives of the components of $\underline{x}(t)$.

Example 4. If a curve is parameterised in terms of the so-called *arc-length* s (we won't define this precisely here) along the curve from a fixed point on it, then $|\frac{d\underline{x}}{ds}| = 1$.

$$\begin{aligned}
&\text{with} \quad t = s, \quad h = \delta s, \\
&|\underline{x}(s+\delta s) - \underline{x}(s)| \simeq \delta s \\
\Rightarrow \quad \lim_{\delta s \to 0} &\frac{|\underline{x}(s+\delta s) - \underline{x}(s)|}{\delta s} = 1 \\
&\Rightarrow \quad |\frac{d\underline{x}}{ds}| = 1 \qquad \text{a unit vector.}
\end{aligned}$$

1.3 Partial derivatives and the chain rule

In Calculus I, given a function of two variables $f(x,y) : \mathbb{R}^2 \to \mathbb{R}$, you learned that the partial derivatives

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \lim_{h\to 0} \frac{f(x+h,y) - f(x,y)}{h} \\
\frac{\partial f}{\partial y} &= \lim_{h\to 0} \frac{f(x,y+h) - f(x,y)}{h},
\end{aligned}$$

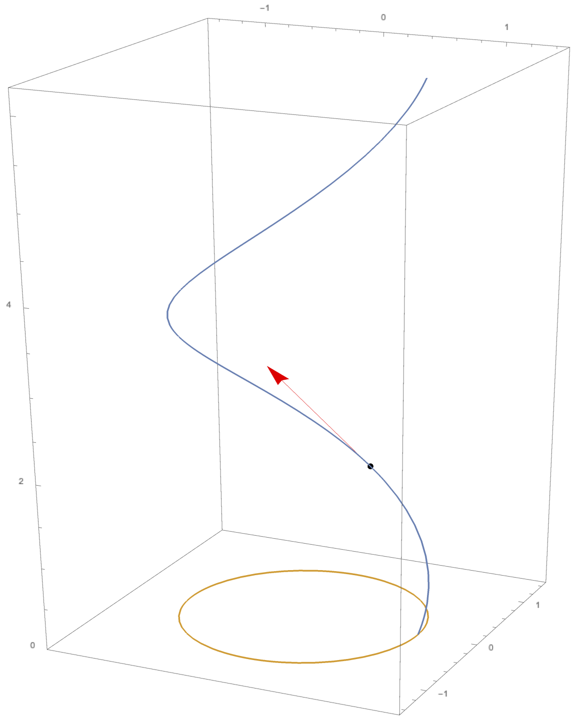


Figure 5: A helix in $\mathbb{R}^3$, with the tangent at a point shown.

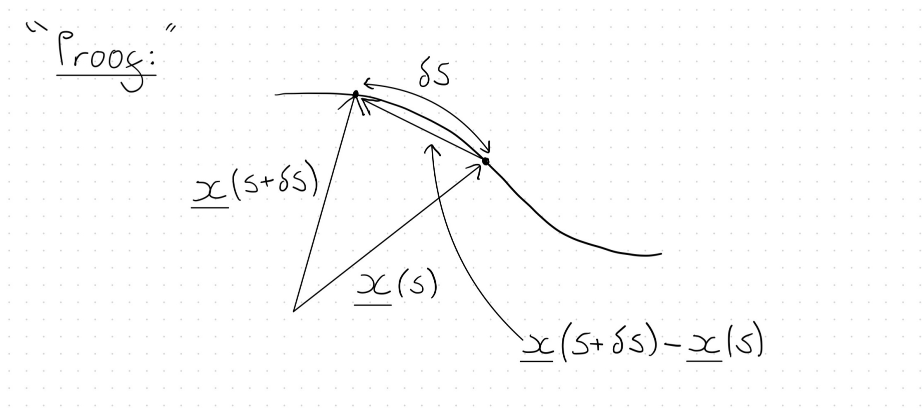


Figure 6: If the curve is parameterised by its arc-length (which we haven't rigorously defined), then for small δs, the length of the vector $\underline{x}(s+\delta s) - \underline{x}(s)$ becomes approximately the same as the length of the curve segment δs. In the limit that $\delta s \to 0$, these become equal, and hence the length of the derivative of the curve with respect to s (the tangent) becomes 1.

tell us the rate of change of the function f as we move parallel to the x- and y- axes respectively. If we write x, y as x_1, x_2 & $f(x, y)$ as $f(\underline{x})$, with $\underline{x} = x_1\underline{e}_1 + x_2\underline{e}_2$, then we can re-express the partial derivatives using vector notation as

$$\begin{aligned}\frac{\partial f}{\partial x} \equiv \frac{\partial f}{\partial x_1} &= \lim_{h \to 0} \frac{f(\underline{x} + h\underline{e}_1) - f(\underline{x})}{h} \\ \frac{\partial f}{\partial y} \equiv \frac{\partial f}{\partial x_2} &= \lim_{h \to 0} \frac{f(\underline{x} + h\underline{e}_2) - f(\underline{x})}{h}.\end{aligned}$$

This now suggests the obvious generalisation to scalar fields in n dimensions. If we let $f(\underline{x}) : \mathbb{R}^n \to \mathbb{R}$, then the function f has n (1^{st} order) partial derivatives given by

$$\frac{\partial f(\underline{x})}{\partial x_a} = \lim_{h \to 0} \frac{f(\underline{x} + h\underline{e}_a) - f(\underline{x})}{h},$$

for $a = 1, 2, \ldots, n$. These partial derivatives tell us about the rate of change of the function as we move parallel to any of the n coordinate axes in n dimensions.

You also learned that the rate of change of a function $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$ along a parametrically defined curve C given as $(x(t), y(t))$ can be found using the chain rule. Along this curve we have $F(t) \equiv f(x(t), y(t))$. Note that I don't write the function of t as $f(t)$, since f is a map from $\mathbb{R}^2$ and hence strictly speaking $f(t)$ is a different function, this time a map from $\mathbb{R}$. To avoid the confusion of two different functions with the same name, we will denote the restricted function as $F(t)$. The chain rule then tells us that

$$\frac{dF(t)}{dt} = \frac{dx}{dt}\frac{\partial f}{\partial x} + \frac{dy}{dt}\frac{\partial f}{\partial y}.$$

To extend this to the n-dimensional case using vector notation, we first note that in the two-dimensional case, the curve C is given by

$$\underline{x}(t) = x_1(t)\underline{e}_1 + x_2(t)\underline{e}_2,$$

with $x_1(t) = x(t)$ and $x_2(t) = y(t)$. Similarly, $f(x, y)$ is $f(\underline{x})$, so

$$\frac{dF(t)}{dt} = \frac{df(\underline{x}(t))}{dt} = \frac{dx_1}{dt}\frac{\partial f}{\partial x_1} + \frac{dx_2}{dt}\frac{\partial f}{\partial x_2}.$$

We can now see how this should be generalised to the case of a scalar field in n dimensions. The curve C can be given parametrically as

$$\underline{x}(t) = x_1(t)\underline{e}_1 + \ldots + x_n(t)\underline{e}_n.$$

Our scalar field is given as $f(\underline{x}) : \mathbb{R}^n \to \mathbb{R}$, and the restriction of the scalar field to the curve C can then be written as $F(t) = f(\underline{x}(t))$. The chain rule then tells us that

$$\frac{dF(t)}{dt} = \frac{d}{dt}f(\underline{x}(t)) = \frac{dx_1}{dt}\frac{\partial f}{\partial x_1} + \cdots + \frac{dx_n}{dt}\frac{\partial f}{\partial x_n}.$$

Note: The chain rule holds for differentiable functions of two variables. You defined what this meant in the context of functions of two variables in Calculus I, and we shall revisit precisely what it means for a scalar function in n dimensions to be differentiable in section 5. For now we simply assume that our scalar fields are indeed differentiable.

2 The gradient of a scalar field

2.1 Differential operators and $\underline{\nabla}$

In the previous section, we saw that when we can compute the rate of change of a scalar field $f(\underline{x})$ along a curve C given by $\underline{x}(t) = x_i(t)\underline{e}_i$ (ESC), using the chain rule as

$$\frac{dF(t)}{dt} = \frac{d}{dt}f(\underline{x}(t)) \;=\; \frac{dx_1}{dt}\frac{\partial f}{\partial x_1} + \cdots + \frac{dx_n}{dt}\frac{\partial f}{\partial x_n} = \frac{dx_i}{dt}\frac{\partial f}{\partial x_i},$$

where $F(t)$ is the restriction of $f(\underline{x})$ to the curve $\underline{x}(t)$, and where again we've used ESC in the final equality.

Since this is true for all (differentiable) scalar fields, it's often useful to write this rule in terms of the derivative operators themselves, separate from the field $F(t)$. In this form, the chain rule can be given as

$$\frac{d}{dt} \;=\; \frac{dx_1}{dt}\frac{\partial}{\partial x_1} + \cdots + \frac{dx_n}{dt}\frac{\partial}{\partial x_n}$$

This is known as a *differential operator*, which can be thought of as a map which takes functions to functions using derivatives.

When using operator notation, we need to be careful about exactly what the differential operator is acting on.

Example 5. Given two real functions $f(x), g(x) : \mathbb{R} \to \mathbb{R}$, then:

- $f(x)\frac{d}{dx}$ is a differential operator which can act on $g(x)$ to give $f(x)\frac{dg(x)}{dx}$.
- $\frac{d}{dx}f(x)$ is a differential operator which can on $g(x)$ to give $\frac{d}{dx}(f(x)g(x)) = \frac{df(x)}{dx}g(x) + f(x)\frac{dg(x)}{dx}$ by the product rule.
- If I want an operator which multiples $g(x)$ by $\frac{df(x)}{dx}$, then I should write the operator as $\left(\frac{d}{dx}f(x)\right)$, where the brackets make it clear the the derivative is "used up", only acting on $f(x)$.

The derivative with respect to t along the curve $C : \underline{x}(t) = x_i(t)\underline{e}_i$ as above, in operator $\frac{d}{dt}$ form, can then be rewritten using the scalar product as

$$\begin{aligned}\frac{d}{dt} &= \frac{dx_1}{dt}\frac{\partial}{\partial x_1} + \cdots + \frac{dx_n}{dt}\frac{\partial}{\partial x_n} \\ &= \left(\underline{e}_1\frac{dx_1}{dt} + \cdots + \underline{e}_n\frac{dx_n}{dt}\right)\cdot\left(\underline{e}_1\frac{\partial}{\partial x_1} + \cdots + \underline{e}_n\frac{\partial}{\partial x_n}\right) \\ &= \frac{d\underline{x}}{dt}\cdot\underline{\nabla},\end{aligned}$$

where

$$\underline{\nabla} = \underline{e}_1\frac{\partial}{\partial x_1} + \cdots + \underline{e}_n\frac{\partial}{\partial x_n} \;=\; \underline{e}_i\frac{\partial}{\partial x_i}. \quad \text{(ESC)}$$

This differential operator $\underline{\nabla}$ is called 'del', or 'nabla', and is one of the most important objects in this course. Note that since $\underline{\nabla}$ is a vector quantity, we always write it with an underline.

If $f(\underline{x}) : \mathbb{R}^n \to \mathbb{R}$ is a scalar field, then we define its gradient ("grad f") to be given by the action of $\underline{\nabla}$ on f:

$$\underline{\nabla} f \;\equiv\; \text{grad } f = \underline{e}_1\frac{\partial f}{\partial x_1} + \underline{e}_2\frac{\partial f}{\partial x_2} + \cdots + \underline{e}_n\frac{\partial f}{\partial x_n}.$$

The gradient of a scalar field is therefore a vector field, with components $\frac{\partial f}{\partial x_a}$.

Example 6. In two dimensions, with $\underline{x} = x\underline{e}_1 + y\underline{e}_2$, let $f(\underline{x}) = (x^2 + y^2)/4$. Then we have

$$\begin{aligned}\Rightarrow\quad \frac{\partial f}{\partial x} &= \frac{x}{2} \quad \frac{\partial f}{\partial y} = \frac{y}{2} \\ \Rightarrow\quad \underline{\nabla} f &= \frac{1}{2}x\underline{e}_1 + \frac{1}{2}y\underline{e}_2 \;=\; \frac{1}{2}\underline{x}.\end{aligned}$$

The vector field can be drawn by arrows of length $|\underline{\nabla} f|$ and direction parallel to $\underline{\nabla} f$ starting at a variety of sample points:

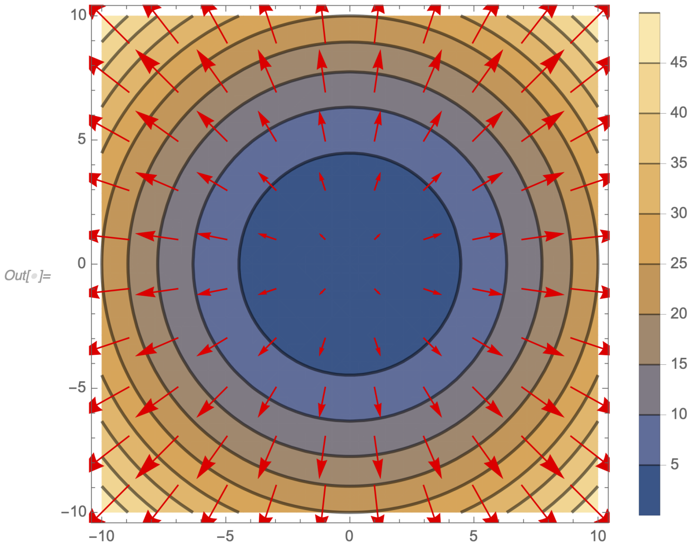


Figure 7: Level sets of the function $f(\underline{x}) = (x^2 + y^2)/4$ are shown in a contour plot. A plot of the vector field $\underline{\nabla} f(\underline{x})$ is overlaid on top of this, showing the vectors pointing away from the origin, parallel to the position vectors $\underline{x}$. The vectors in the vector field $\underline{\nabla} f(\underline{x})$ are perpendicular to the level sets of $f(\underline{x})$.

Example 7. In three dimensions, with $\underline{x} = x\underline{e}_1 + y\underline{e}_2 + z\underline{e}3$, let $\underline{a} = a_1\underline{e}_1 + a_2\underline{e}_2 + a_3\underline{e}_3$ be a constant vector, and let $f(\underline{x}) = \underline{a}.\underline{x} - \underline{x}.\underline{x}$. Then we have

$$\begin{aligned} f &= a_1 x + a_2 y + a_3 z - (x^2 + y^2 + z^2) \\ \Rightarrow \quad \frac{\partial f}{\partial x} &= a_1 - 2x, \qquad \frac{\partial f}{\partial y} = a_2 - 2y, \qquad \frac{\partial f}{\partial z} = a_3 - 2z \\ \Rightarrow \quad \underline{\nabla} f &= \underline{e}_1(a_1 - 2x) + \underline{e}_2(a_2 - 2y) + \underline{e}_3(a_3 - 2z) \\ &= \underline{a} - 2\underline{x}. \end{aligned}$$

Although the picture is less easy to interpret than the 2-dimensional example, for completeness this is included as Figure 8.

2.2 Directional derivatives

Let $C : \underline{x} = \underline{x}(t)$ be a curve in $\mathbb{R}^n$, and $f : \mathbb{R}^n \to \mathbb{R}$ a scalar field. Then $f(\underline{x}(t)) : \mathbb{R} \to \mathbb{R}$ is f restricted to C and

$$\frac{d}{dt} f(\underline{x}(t)) = \frac{d\underline{x}}{dt}.\underline{\nabla} f \qquad \text{by chain rule.}$$

As we saw in subsection 1.2, $\frac{d\underline{x}}{dt}$ is tangent to C at $\underline{x}(t)$. If we change to a parameterisation in terms of the arc-length s, such that the tangent $\frac{d\underline{x}}{ds} = \hat{\underline{n}}$ is a unit tangent (see example 4 for a justification as to why this is possible), then we have

$$\frac{df(\underline{x}(s))}{ds} = \hat{\underline{n}} \,.\, \underline{\nabla} f. \tag{2.1}$$

Now $\frac{df(\underline{x}(s))}{ds}$ is the rate of change of f with respect to distance (arc length) in the direction $\hat{\underline{n}}$. This is called the <u>directional derivative</u> of f in the direction $\hat{\underline{n}}$ (and is sometimes written $\frac{df}{d\hat{\underline{n}}}$).

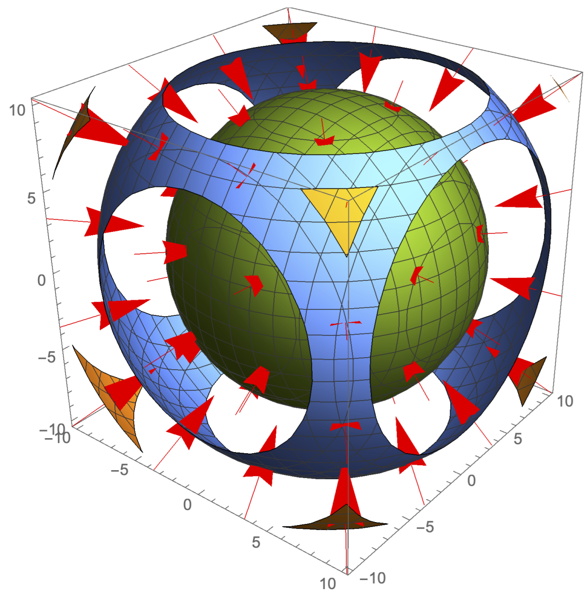


Figure 8: The contour plot of a scalar field overlaid with the gradient vector field in 3d. Parts of three level sets can be seen as spherical shells (with what center?) and representative vectors of $\underline{\nabla} f$ can be seen as red arrows. As in example 6, the vectors of $\underline{\nabla} f$ can be seen to be normal to the level sets of f. We return to this idea in the next section.

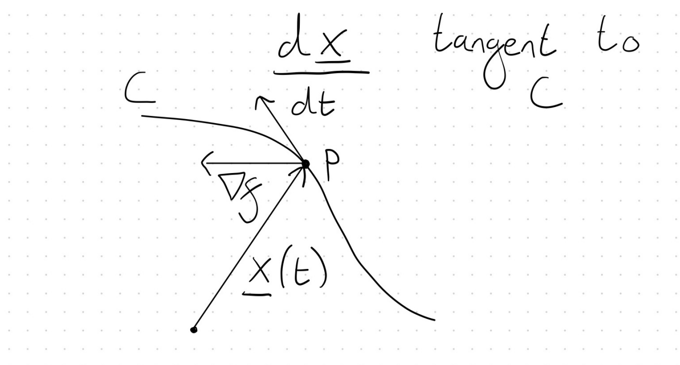


Figure 9: The tangent to a curve C at a point p is shown alongside an example of the gradient of a scalar field at the same point p.

Notice that:

$$\frac{df}{ds} = \underline{\hat{n}} \cdot \underline{\nabla} f \;=\; |\underline{\hat{n}}|\,|\underline{\nabla} f| \cos\theta$$
$$= |\underline{\nabla} f| \cos\theta \;\leq\; |\underline{\nabla} f|.$$

Therefore $|\underline{\nabla} f|$ is the greatest value of the directional derivative over all possible directions $\underline{\hat{n}}$. This value is achieved when $\theta = 0$, i.e. when $\underline{\hat{n}} \parallel \underline{\nabla} f$. Therefore $\underline{\nabla} f$ points in the direction where f increases fastest.

Note that in example 6, the vectors in the vector field $\underline{\nabla} f$ are normal to the curves of constant $|\underline{x}|$, which were the curves of constant f. This also holds more generally. In $\mathbb{R}^n$, suppose C lies entirely in the level set $f(\underline{x}) = k$ for k some constant. Call this whole level set S, so $C \subset S$.

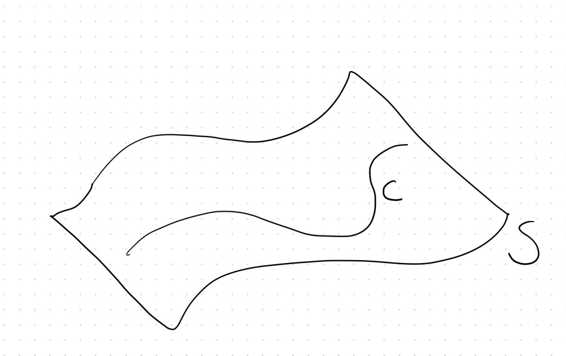


Figure 10: The curve C contained entirely within a level set of the function f. In two dimension the level sets are themselves curves, but in higher dimension spaces these level sets become surfaces, and hypersurfaces. You should think that the condition $f(\underline{x}) = k$ is a single constraint on a n-dimensional space, so the points which satisfy this constraint for an $(n-1)$-dimensional hypersurface.

So on this C, $f(\underline{x}(t)) = k$ and

$$0 = \frac{df}{dt} \;=\; \frac{d\underline{x}}{dt} \cdot \underline{\nabla} f$$
$$\Rightarrow \quad \frac{d\underline{x}}{dt} \perp \underline{\nabla} f,$$

i.e. At all points p in a level set of f, $\underline{\nabla} f$ is orthogonal to any curve through p contained in the level set.

In $\mathbb{R}^3$, the plane through P orthogonal to $\underline{\nabla} f$ is called the tangent plane to S at P. The tangent to any curve in S through P lies in this plane. In Calculus I you already used this to find the equation of the tangent plane to a surface at a point.

2.3 Some properties of the gradient

If $f, g : \mathbb{R}^n \to \mathbb{R}$ are scalar fields, ϕ is a function $\phi : \mathbb{R} \to \mathbb{R}$, and a & b are constants then,

$$\begin{aligned} (i) \quad & \underline{\nabla}(af + bg) &=& \; a\underline{\nabla} f + b\underline{\nabla} g \\ (ii) \quad & \underline{\nabla}(fg) &=& \; (\underline{\nabla} f)g + f(\underline{\nabla} g) \\ (iii) \quad & \underline{\nabla}\phi(f) &=& \; (\underline{\nabla} f)\frac{d\phi}{df}. \end{aligned}$$

Note that as always with differential operators, the brackets are very important here. This is because $\underline{\nabla} fg$ means $\underline{\nabla}(fg)$, since derivatives ($\frac{\partial}{\partial x}$ etc.) act on everything to the right.

Example 8. Let $n = 2$ and $\underline{x} = x\underline{e}_1 + y\underline{e}_2$. If we now take $\phi(f) = f^2$ and $f(x, y) = x \sin y$,

If we now take

$$\phi(f(x, y)) = x^2 \sin^2 y$$
$$\Rightarrow \quad \underline{\nabla}\phi(f) = \underline{\nabla}(x^2 \sin^2 y)$$
$$= 2x \sin^2 y \; \underline{e}_1 + 2x^2 \cos y \sin y \; \underline{e}_2,$$

by direct calculation. Or:

$$
\begin{aligned}
\underline{\nabla} f &= \sin y \ \underline{e}_1 + x \cos y \ \underline{e}_2 \\
\Rightarrow \quad \frac{d\phi}{df} &= 2f \\
\Rightarrow \quad (\underline{\nabla} f)\frac{d\phi}{df} &= (\sin y \ \underline{e}_1 + x \cos y \ \underline{e}_2) 2x \sin y \\
\Rightarrow \quad \underline{\nabla} \phi(f) &= 2x \sin^2 y \ \underline{e}_1 + 2x^2 \cos y \sin y \ \underline{e}_2,
\end{aligned}
$$

which shows that property iii holds in this case.

If we now let $\underline{x} = x_1\underline{e}_1 + \ldots + x_n\underline{e}_n$, then these properties can be shown to hold in the general case as follows:

$$
\begin{aligned}
(i) \quad \underline{\nabla}(af + bg) &= \underline{e}_1 \frac{\partial}{\partial x_1}(af + bg) + \ldots + \underline{e}_n \frac{\partial}{\partial x_n}(af + bg) \\
&= \underline{e}_1 \left(a\frac{\partial f}{\partial x_1} + b\frac{\partial g}{\partial x_1} \right) + \ldots + \underline{e}_n \left(a\frac{\partial f}{\partial x_n} + b\frac{\partial g}{\partial x_n} \right) \\
&= a \left(\underline{e}_1 \frac{\partial f}{\partial x_1} + \ldots + \underline{e}_n \frac{\partial f}{\partial y} \right) + b \left(\underline{e}_1 \frac{\partial g}{\partial x_1} + \underline{e}_2 \frac{\partial g}{\partial x_n} \right) \\
&= a\underline{\nabla} f + b\underline{\nabla} g.
\end{aligned}
$$

$$
\begin{aligned}
(ii) \quad \underline{\nabla}(fg) &= \underline{e}_1 \frac{\partial}{\partial x_1}(fg) + \ldots + \underline{e}_n \frac{\partial}{\partial x_n}(fg) \\
&= \underline{e}_1 \left(\left(\frac{\partial f}{\partial x_1} \right) g + f\frac{\partial g}{\partial x_1} \right) + \ldots + \underline{e}_n \left(\left(\frac{\partial f}{\partial x_n} \right) g + f\frac{\partial g}{\partial x_n} \right) \\
&= \left(\underline{e}_1 \frac{\partial f}{\partial x_1} + \ldots + \underline{e}_n \frac{\partial f}{\partial x_n} \right) g + f \left(\underline{e}_1 \frac{\partial g}{\partial x_1} + \ldots + \underline{e}_n \frac{\partial g}{\partial x_n} \right) \\
&= (\underline{\nabla} f)g + f\underline{\nabla} g.
\end{aligned}
$$

$$
\begin{aligned}
(iii) \quad \underline{\nabla} \phi &= \underline{e}_1 \frac{\partial \phi}{\partial x_1} + \cdots + \underline{e}_n \frac{\partial \phi}{\partial x_n} \\
&= \underline{e}_1 \frac{\partial f}{\partial x_1} \frac{d\phi}{df} + \cdots + \underline{e}_n \frac{\partial f}{\partial x_n} \frac{d\phi}{df} \\
&= (\underline{\nabla} f)\frac{d\phi}{df}.
\end{aligned}
$$

3 $\underline{\nabla}$ acting on vector fields

So far we've introduced scalar fields and vector fields, and seen how the differential operator $\underline{\nabla}$ can act on a scalar field $f(\underline{x}) : \mathbb{R}^n \to \mathbb{R}$ to give a vector field known as the gradient of the scalar field, $\underline{\nabla} f$. We also saw that the gradient tells us the direction in which the scalar field increases fastest. As the title of this section suggests, we're now going to see how we can define an action of $\underline{\nabla}$ on vector fields $\underline{v}(\underline{x}) : \mathbb{R}^n \to \mathbb{R}^n$, and what these quantities can tell us about the original vector field.

3.1 Divergence (div)

Since $\underline{\nabla}$ is a vector operator, and vector fields assign a vector $\underline{v}(\underline{x})$ for each point $\underline{x} \in \mathbb{R}^n$, we can take the dot product between $\underline{\nabla}$ and $\underline{v}(\underline{x})$. This quantity $\underline{\nabla}.\underline{v}$ is the known as the *divergence* of $\underline{v}$, and is also written as div $\underline{v}$.

In the standard cartesian basis for $\mathbb{R}^n$

$$\begin{aligned}
\underline{v}(\underline{x}) &= \underline{e}_1 v_1(\underline{x}) + \underline{e}_2 v_2(\underline{x}) + \cdots + \underline{e}_n v_n(\underline{x}) \\
\Rightarrow \qquad \underline{\nabla}.\underline{v} \ \equiv \ \text{div}\, \underline{v} &\equiv (\underline{e}_1 \frac{\partial}{\partial x_1} + \cdots + \underline{e}_n \frac{\partial}{\partial x_n}) \,.\, (\underline{e}_1 v_1(\underline{x}) + \cdots + \underline{e}_n v_n(\underline{x})) \\
&= \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \cdots + \frac{\partial v_n}{\partial x_n} \\
&= \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i} \qquad \text{(index notation)} \\
&= \frac{\partial v_i}{\partial x_i}, \qquad \text{(Einstein Summation Convention)}
\end{aligned}$$

since the $\underline{e}_a$ are orthonormal and constant.

Beware! In other coordinate systems the basis vectors $\underline{e}_a$ might vary with $\underline{x}$ and hence the formula needs more care. We may return to the divergence in other coordinate systems later in the term, depending on time.

Note: $\underline{\nabla}.\underline{v}(\underline{x})$ is a scalar field, as can be seen in the following example.

Example 9.

$$\begin{aligned}
\underline{v}(\underline{x}) = (v_1(\underline{x}), v_2(\underline{x}), v_3(\underline{x})) \ &= \ (x^2, y^2, z^2) \\
\underline{\nabla}.\underline{v} &= \frac{\partial x^2}{\partial x} + \frac{\partial y^2}{\partial y} + \frac{\partial z^2}{\partial z} \\
&= 2(x + y + z),
\end{aligned}$$

a number for each point $\underline{x} = (x, y, z)$ not a vector.

Although we treat $\underline{\nabla}$ like a vector, note that it is a vector differential operator, and therefore is not actually a vector in $\mathbb{R}^n$ like $\underline{x}$ is. Therefore although the inner product on $\mathbb{R}^n$ is symmetric, note that

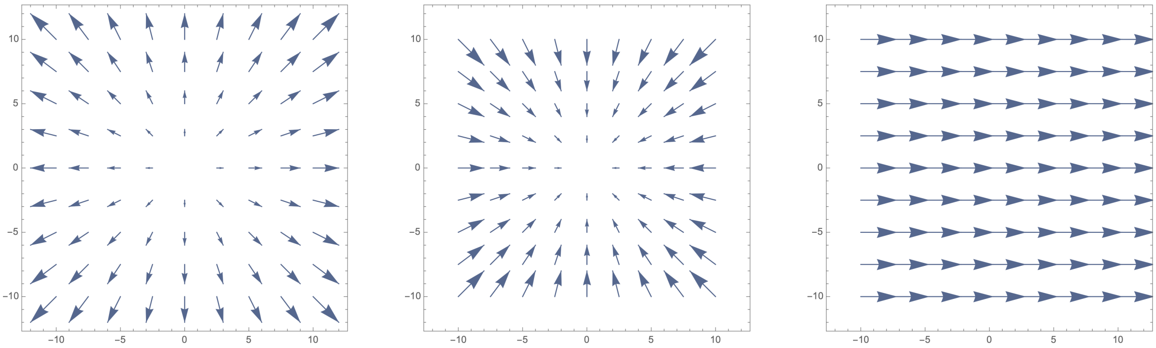
$$\underline{\nabla}.\underline{v} \ \neq \ \underline{v}.\underline{\nabla},$$

as the left-hand side of this is a scalar field, whereas the right-hand side is the (scalar) differential operator

$$\left(x^2 \frac{\partial}{\partial x} + y^2 \frac{\partial}{\partial y} + z^2 \frac{\partial}{\partial z} \right)$$

acting on scalar fields.

To get an intuition of what the divergence tells us about our vector field, we should think of our vector field as if it were a fluid, where the direction of a vector at a point tells us the direction of fluid flow at that point, and the magnitude of the vector tells us how fast the fluid is flowing at the point. The divergence of the vector field at a point then tells us whether the point is acting like a *source* (corresponding to positive divergence) or a *sink* (negative divergence) for the fluid. That is, whether more of the 'fluid' is entering than the point than is leaving. This is shown for a vector field $\underline{v}$ in Figure 11, where the divergence at the origin is positive, negative and zero respectively.



$\underline{\nabla}.\underline{v} > 0$ $\qquad$ $\underline{\nabla}.\underline{v} < 0$ $\qquad$ $\underline{\nabla}.\underline{v} = 0$

Figure 11: The divergence of a vector field at a point tells us whether the point acts like a source or a sink for the vector field, if we think of the vector field as describing the flow of a fluid.

Properties of div:

Let a, b be constants, f, g be scalar fields and $\underline{v}, \underline{w}$ vector fields, all in $\mathbb{R}^n$. Then

$$\begin{aligned}(i)\underline{\nabla}.(a\underline{v} + b\underline{w}) &= a\underline{\nabla}.\underline{v} + b\underline{\nabla}.\underline{w} \\ (ii)\underline{\nabla}.(f\underline{v}) &= (\underline{\nabla} f).\underline{v} + f\underline{\nabla}.\underline{v}\end{aligned}$$

<u>Proof:</u> (i) This follows from linearity of the partial derivative:

$$\begin{aligned}\underline{\nabla}.(a\underline{v} + b\underline{w}) &= \frac{\partial(av_1 + bw_1)}{\partial x_1} + \cdots + \frac{\partial(av_n + bw_n)}{\partial x_n} \\ &= a\frac{\partial v_1}{\partial x_1} + b\frac{\partial w_1}{\partial x_1} + \cdots + a\frac{\partial v_n}{\partial x_n} + b\frac{\partial w_n}{\partial x_n} \\ &= a\left(\frac{\partial v_1}{\partial x_1} + \cdots + \frac{\partial v_n}{\partial x_n}\right) + b\left(\frac{\partial w_1}{\partial x_1} + \cdots + \frac{\partial w_n}{\partial x_n}\right) \\ &= a\underline{\nabla}.\underline{v} + b\underline{\nabla}.\underline{w}.\end{aligned}$$

(ii) First note that $f\underline{v}$ is a vector field with components $(fv_1, fv_2, fv_3, \ldots, fv_n)$, so

$$\begin{aligned}\underline{\nabla}.(f\underline{v}) &= \frac{\partial fv_1}{\partial x_1} + \cdots + \frac{\partial fv_n}{\partial x_n} \\ &= \frac{\partial f}{\partial x_1}v_1 + f\frac{\partial v_1}{\partial x_1} + \cdots + \frac{\partial f}{\partial x_n}v_n + f\frac{\partial v_n}{\partial x_n} \\ &= \frac{\partial f}{\partial x_1}v_1 + \cdots + \frac{\partial f}{\partial x_n}v_n \ + \ f\frac{\partial v_1}{\partial x_1} + \cdots + f\frac{\partial v_n}{\partial x_n} \\ &= (\underline{\nabla} f).\underline{v} \ + \ f(\underline{\nabla}.\underline{v}).\end{aligned}$$

Once we finish section 4, we'll be able to derive this more effectively by using index notation.

Example 10. Suppose

$$\begin{aligned}
f(\underline{x}) &= \underline{a}.\underline{x} \\
\underline{v}(\underline{x}) &= \underline{a} \qquad \text{a constant} \\
\Rightarrow \qquad f\underline{v} &= (\underline{a}.\underline{x})\underline{a} \\
&= (a_1x + a_2y + a_3z)(a_1\underline{e}_1 + a_2\underline{e}_2 + a_3\underline{e}_3) \\
\Rightarrow \quad \underline{\nabla}.((\underline{a}.\underline{x})\underline{a}) &= \frac{\partial}{\partial x}(a_1(a_1x + a_2y + a_3z)) \\
&+ \frac{\partial}{\partial y}(a_2(a_1x + a_2y + a_3z)) \\
&+ \frac{\partial}{\partial z}(a_3(a_1x + a_2y + a_3z)) \\
&= a_1^2 + a_2^2 + a_3^2 \; = \; \|\underline{a}\|^2
\end{aligned}$$

by direct calculation. Or, using property (ii):

$$\underline{\nabla}.((\underline{a}.\underline{x})\underline{a}) \; = \; (\underline{\nabla}\underline{a}.\underline{x}).\underline{a} + (\underline{a}.\underline{x})\underline{\nabla}.\underline{a}.$$

But $\underline{\nabla}.\underline{a} = 0$, while

$$\begin{aligned}
\underline{\nabla}(\underline{a}.\underline{x}) &= (\underline{e}_1\frac{\partial}{\partial x} + \underline{e}_2\frac{\partial}{\partial y} + \underline{e}_3\frac{\partial}{\partial z})(a_1x + a_2y + a_3z) \\
&= \underline{e}_1a_1 + \underline{e}_2a_2 + \underline{e}_3a_3 \; = \; \underline{a} \\
\underline{\nabla}.((\underline{a}.\underline{x})\underline{a}) &= \underline{a}.\underline{a} \; = \; \|\underline{a}\|^2
\end{aligned}$$

agreeing with the direct calculation.

3.2 Curl

In 3 dimensions, there is a second type of product one can take between vectors, and this is the vector cross product. Recall that we define the vector product of two vectors $\underline{A}$ and $\underline{B}$ as

$$\begin{aligned}
\underline{A} \times \underline{B} &= \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix} \\
&= \underline{e}_1(A_2B_3 - A_3B_2) + \underline{e}_2(A_3B_1 - A_1B_3) + \underline{e}_3(A_1B_2 - A_2B_1).
\end{aligned}$$

Then, for a vector field $\underline{v}(\underline{x})$ in 3 dimensions, define the <u>curl</u> of $\underline{v}$ as

$$\underline{\nabla} \times \underline{v}(\underline{x}) \; \equiv \; \text{curl}\, \underline{v} \; \equiv \; \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ v_1 & v_2 & v_3 \end{vmatrix},$$

where we have to expand this *always making sure that the derivatives* $\frac{\partial}{\partial x_i}$ *are on the left of the* v_i. So therefore we have

$$\underline{\nabla} \times \underline{v}(\underline{x}) = \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ v_1 & v_2 & v_3 \end{vmatrix} = \underline{e}_1(\frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z}) + \underline{e}_2(\frac{\partial v_1}{\partial z} - \frac{\partial v_3}{\partial x}) + \underline{e}_3(\frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}).$$

Note that the curl of a vector field is therefore a *new* <u>vector</u> field.

Example 11. if $\underline{v} : \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field, and can be expressed in terms of its components as

$\underline{v} = (x^2 z, xyz, x)$, then

$$\begin{aligned}
\text{curl } \underline{v} = \underline{\nabla} \times \underline{v} \ &= \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ x^2 z & xyz & x \end{vmatrix} \\
&= \underline{e}_1 \left(\frac{\partial x}{\partial y} - \frac{\partial xyz}{\partial z}\right) + \underline{e}_2 \left(\frac{\partial x^2 z}{\partial z} - \frac{\partial x}{\partial x}\right) + \underline{e}_3 \left(\frac{\partial xyz}{\partial x} - \frac{\partial x^2 z}{\partial y}\right) \\
&= -xy\underline{e}_1 + (x^2 - 1)\underline{e}_2 + yz\underline{e}_3
\end{aligned}$$

Note that since $\underline{\nabla} \times \underline{v}$ is a vector field, we can calculate its divergence. In the case of $\underline{v}$ being the vector field from Example 11,

$$\begin{aligned}
\underline{\nabla}.(\underline{\nabla} \times \underline{v}) &= \frac{\partial}{\partial x}(-xy) + \frac{\partial}{\partial y}(x^2 - 1) + \frac{\partial}{\partial z}(yz) \\
&= -y + 0 + y \ = \ 0.
\end{aligned}$$

In turns out that this is always true so long as $\underline{v}$ has components with continuous second partial derivatives. We'll come back to this in subsection 3.3.

The curl of a vector field $\underline{v}$, tells us how much $\underline{v}$ is 'curling around' at a point. If we imagine our vector field $\underline{v}$ as a fluid, like when we thought about the meaning of the divergence (see subsection 3.1), the magnitude of the curl then tells us about the rotational speed of the fluid, and the direction of the curl then tells us which axis the fluid is rotating around. This axis is determined using the so-called 'right-hand rule': If you curl the fingers of your right hand, such that your fingers represent the rotation of the fluid, then your thumb points in the direction of the curl vector.

Example 12. Consider the vector field $\underline{v}$ with components $\underline{v} = (-y, x, 0)$. To imagine this vector field, realise that it is independent of z, and so you can imagine the vector field in the $z = 0$ plane, and the vector field in any other plane of z is then just a translation of the vector field at $z = 0$. This $z = 0$ plane is shown in Figure 12

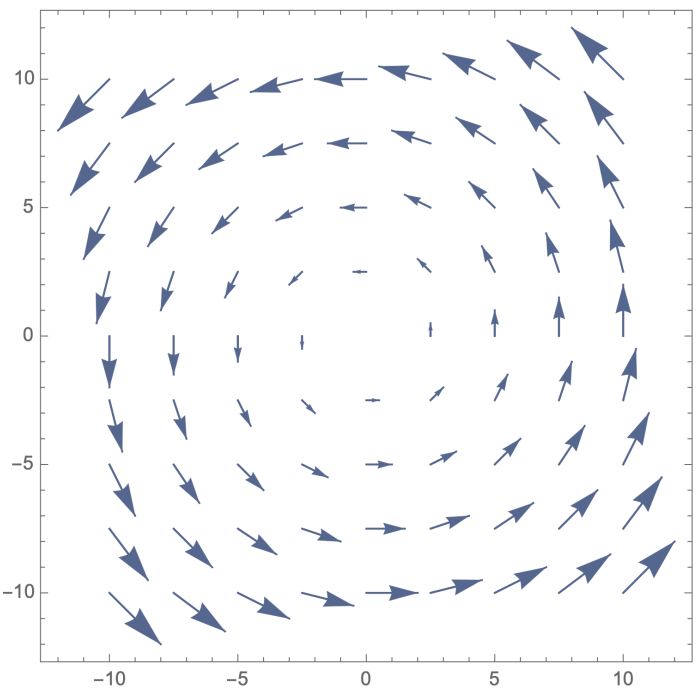


Figure 12: The plane $z = 0$ of the vector field $\underline{v} = (-y, x, 0)$, where the horizontal axis of the image is the x-axis, and the vertical axis is the y-axis. You should imagine the z-axis as coming straight out of the page.

The curl of this vector field can then easily be checked to be

$$\underline{\nabla} \times \underline{v} = (0, 0, 2),$$

so is a vector field of constant magnitude pointing in the positive z direction, as one would expect using the right-hand rule.

Properties of curl:

Let a, b be constants, $\underline{v}, \underline{w}$ be vector fields, and f be a scalar field, all in $\mathbb{R}^3$. Then:

$$\begin{aligned}
(i)\underline{\nabla} \times (a\underline{u} + b\underline{v}) &= a\underline{\nabla} \times \underline{u} + b\underline{\nabla} \times \underline{v} \\
(ii)\underline{\nabla} \times (f\underline{v}) &= (\underline{\nabla} f) \times \underline{v} + f\underline{\nabla} \times \underline{v}
\end{aligned}$$

Proof:

(i) follows from linearity of derivatives, as before.

Lets check (ii):

$$\begin{aligned}
\underline{\nabla} \times (f\underline{v}) =& \underline{e}_1 \left[\frac{\partial}{\partial y}(fv_3) - \frac{\partial}{\partial z}(fv_2)\right] + \underline{e}_2 \left[\frac{\partial}{\partial z}(fv_1) - \frac{\partial}{\partial x}(fv_3)\right] \\
&+ \underline{e}_3 \left[\frac{\partial}{\partial x}(fv_2) - \frac{\partial}{\partial y}(fv_1)\right] \\
&= \underline{e}_1 \left[\left(\frac{\partial f}{\partial y}\right) v_3 + f\frac{\partial v_3}{\partial y} - \left(\frac{\partial f}{\partial z}\right) v_2 - f\frac{\partial v_2}{\partial z}\right] + [\dots] + [\dots] \\
&= \underline{e}_1 \left[\left(\frac{\partial f}{\partial y}\right) v_3 - \left(\frac{\partial f}{\partial z}\right) v_2\right] \;+\; \underline{e}_1 f \left[\frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z}\right] + \dots + \dots \\
&= \underline{\nabla} f \times \underline{v} \;+\; f\underline{\nabla} \times \underline{v}
\end{aligned}$$

This is tedious. Once we develop our index notation in section 4 we will be able to do this in a slicker way.

3.3 Applying $\underline{\nabla}$ twice

If $f : \mathbb{R}^n \to \mathbb{R}$ is a scalar field then $\underline{\nabla} f$ is a vector field and we can take its divergence. This is

$$\begin{aligned}
\underline{\nabla}.(\underline{\nabla} f) &= \text{div grad } f \equiv \nabla^2 f \equiv \Delta f \\
&= \left(\underline{e}_1 \frac{\partial}{\partial x_1} + \dots + \underline{e}_n \frac{\partial}{\partial x_n}\right) \cdot \left(\underline{e}_1 \frac{\partial}{\partial x_1} + \dots + \underline{e}_n \frac{\partial}{\partial x_n}\right) f \\
&= \frac{\partial^2 f}{\partial x_1^2} + \dots + \frac{\partial^2 f}{\partial x_n^2}
\end{aligned}$$

in cartesian coordinates - note that it will look different in other coordinates if the basis vectors are not constant (as in the case of polar coordinates for example).

$\Delta f \equiv \nabla^2 f$ is called the Laplacian of f. Since the Laplacian is the divergence of a vector field, it is a scalar field.

Note that you actually met the Laplacian in Calculus I, when you considered *Linear Partial Differential Equations*. In particular, the Laplacian is a differential operator that appears in *Laplace's Equation*, the *Heat Equation* and the *Wave Equation*. More on this next term.

Example 13. For $n = 2$, let $f = log(x^2 + y^2) = log(\underline{x}.\underline{x})$. Then

$$\begin{aligned}
\Rightarrow \quad \frac{\partial f}{\partial x} = \frac{2x}{x^2+y^2}, \quad \frac{\partial^2 f}{\partial x^2} &= \frac{2}{x^2+y^2} - \frac{4x^2}{(x^2+y^2)^2} = \frac{2(y^2-x^2)}{(x^2+y^2)^2} \\
\Rightarrow \quad \frac{\partial f}{\partial y} = \frac{2y}{x^2+y^2}, \quad \frac{\partial^2 f}{\partial y^2} &= \frac{2(x^2-y^2)}{(x^2+y^2)^2} \qquad \text{(just swap } x \text{ and } y\text{)} \\
\Rightarrow \quad \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} &= \frac{2(y^2-x^2)}{(x^2+y^2)^2} + \frac{2(x^2-y^2)}{(x^2+y^2)^2} \\
= 0 \qquad \text{if } \underline{x} \neq \underline{0}
\end{aligned}$$

Example 14. Again, with $n = 2$ let $f = x^3 - 3xy^2$. Then

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x}(3x^2 - 3y^2) = 6x \\
\frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y}(-6xy) = -6x \\
\Rightarrow \quad \nabla^2 f &= 0
\end{aligned}$$

<u>Notice</u>: If we let $z = x + iy$, then $f = \text{Re}[(x + iy)^3] = \text{Re}[z^3]$. If you study Complex Analysis II, you will see that this is a differentiable complex function. As such, the real and imaginary parts of the function satisfy the Cauchy-Riemann equations. One can show (exercise) that in this case the Laplacian of both the real and imaginary parts of the function have Laplacian equal to 0.

For $n = 3$ (i.e. in $\mathbb{R}^3$) there are a couple of other natural combinations. Firstly, since the gradient of a scalar field is a vector field, we can take its curl:

$$\begin{aligned}
\underline{\nabla} \times \underline{\nabla} f &= \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{vmatrix} \\
&= \underline{e}_1\left(\frac{\partial}{\partial y}\frac{\partial f}{\partial z} - \frac{\partial}{\partial z}\frac{\partial f}{\partial y}\right) + \underline{e}_2\left(\frac{\partial}{\partial z}\frac{\partial f}{\partial x} - \frac{\partial}{\partial x}\frac{\partial f}{\partial z}\right) + \underline{e}_3\left(\frac{\partial}{\partial x}\frac{\partial f}{\partial y} - \frac{\partial}{\partial y}\frac{\partial f}{\partial x}\right) \\
&= \underline{0},
\end{aligned}$$

assuming the 2nd partial derivatives of f are continuous (so we can equate $\frac{\partial^2 f}{\partial y \partial z} = \frac{\partial^2 f}{\partial z \partial y}$ etc.). Also in $\mathbb{R}^3$, we can find the divergence of a curl. We did this once follow example 11, but in general

$$\begin{aligned}
\underline{\nabla}.(\underline{\nabla} \times \underline{v}) &= \underline{\nabla}. \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ v_1 & v_2 & v_3 \end{vmatrix} \\
&= \frac{\partial}{\partial x}\left(\frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial v_1}{\partial z} - \frac{\partial v_3}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}\right) \\
&= 0,
\end{aligned}$$

again, assuming the 2nd partial derivatives of f are continuous. Later we will redo these cases in a more elegant way using index notation.

4 Index notation

4.1 Einstein Summation Convention

Recall that in n dimensions, the indices i, j etc. labelling the components of vectors run from 1 to n, e.g. we write $\underline{v} = (v_1, v_2, \ldots, v_n)$.

Einstein spotted that in quantities like

$$\begin{aligned}
\underline{u}.\underline{v} &= \sum_{i=1}^{n} u_i v_i \\
\underline{\nabla}.\underline{u} &= \sum_{i=1}^{n} \frac{\partial u_i}{\partial x_i} \\
((\underline{a}.\underline{x})\underline{x})_i &= \sum_{j=1}^{n} a_j x_j x_i,
\end{aligned}$$

the index to be summed appears exactly twice in a term or product of terms, while all other indices appear only once (the reason for this is to do with invariance under rotations, or for those of you studying Special Relativity this year, Lorentz transformations). He suggested dropping the summation sign, with the convention that wherever an index is repeated you sum over it.

So, we write

$$\begin{aligned}
\underline{u}.\underline{v} &= \sum_{i} u_i v_i \;=\; u_i v_i \\
\underline{\nabla}.\underline{u} &= \sum_{i} \frac{\partial u_i}{\partial x_i} \;=\; \frac{\partial u_i}{\partial x_i} \\
((\underline{a}.\underline{x})\underline{x})_i &= a_j x_j x_i.
\end{aligned}$$

We call the repeated indices dummy indices, and those that are not repeated are called free indices. The dummy indices can be renamed without changing the expression, i.e.

$$a_j x_j x_i = a_k x_k x_i,$$

since clearly

$$\sum_{j=1}^{n} a_j x_j x_i = \sum_{k=1}^{n} a_k x_k x_i = (\underline{a}.\underline{x}) x_i.$$

However, the free indices must match on both sides of an equation.

We must also be careful never to repeat an index more than twice in any single term or product of terms in an expression. If we were to write $a_i x_i x_i$, we can't tell whether this is supposed to be the component form of $(\underline{a}.\underline{x})x_i$, or of $(\underline{x}.\underline{x})a_i$. So if we want to write $(\underline{u}.\underline{v})^2$ in index notation, we should write $u_i v_i u_j v_j$ and not $u_i v_i u_i v_i$.

Although writing an expression like this in index notation can sometimes looks messy, we'll see shortly that it can be incredibly efficient for calculations.

4.2 The Kronecker delta, δ_{ij}

A very useful object for manipulating expressions in terms of their components, is the Kronecker delta. This is an object with two indices, defined by

$$\delta_{ij} \;\equiv\; \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} .$$

We can think of the Kronecker delta as the components of the $n \times n$ identity matrix I, and in fact this was how you first met δ_{ij}. E.g. for $n = 3$

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \Rightarrow \quad I_{ij} = \delta_{ij},$$

where I_{ij} represents the element of I on the ith row and jth column.

The Kronecker delta appears naturally when we take partial derivatives, as we have

$$\frac{\partial x_i}{\partial x_j} = \delta_{ij}$$

Example 15. If δ_{ij} is the 3-dim Kronecker delta, and $\underline{A} = (A_1, A_2, A_3)$, simplify

1. $A_s\delta_{ts}$
2. $\delta_{rs}\delta_{st}$
3. $\delta_{rs}\delta_{sr}$

Answers:

1.
$$\begin{aligned} A_s\delta_{ts} &= A_1\delta_{t1} + A_2\delta_{t2} + A_3\delta_{t3} \quad (0 \text{ if } t \neq 3 \text{ etc.}) \\ &= A_t \end{aligned}$$

2.
$$\begin{aligned} \delta_{rs}\delta_{st} &= \delta_{r1}\delta_{1t} + \delta_{r2}\delta_{2t} + \delta_{r3}\delta_{3t} \quad (0 \text{ unless } r = 3, t = 3 \text{ etc.}) \\ &= \delta_{rt} \end{aligned}$$

   From these two examples, we can now see the formal rule:

   If a δ has a dummy index, then delete the δ and replace the dummy index in the rest of the expression by the other index on the deleted δ. E.g. $A_s\delta_{ts} = A_t$.

3.
$$\begin{aligned} \delta_{rs}\delta_{sr} &= \delta_{rr} \qquad (\text{by } (2.)) \\ &= \delta_{11} + \delta_{22} + \delta_{33} = 3 \end{aligned}$$

4.3 The Levi-Cevita symbol, ϵ_{ijk}

If we are working in 3 dimension, there is a another device which is useful for handling expressions involving the vector cross product, such as $\underline{A} \times (\underline{B} \times \underline{C})$ or $\underline{\nabla} \times (\underline{\nabla} \times \underline{v})$.

Consider $\underline{C} = \underline{A} \times \underline{B}$. If $\underline{A} = A_1\underline{e}_1 + A_2\underline{e}_2 + A_3\underline{e}_3$, and $\underline{B} = B_1\underline{e}_1 + B_2\underline{e}_2 + B_3\underline{e}_3$, then

$$\begin{aligned} \underline{C} = \underline{A} \times \underline{B} &= \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix} \\ &= \underline{e}_1(A_2B_3 - A_3B_2) + \underline{e}_2(A_3B_1 - A_1B_3) + \underline{e}_3(A_1B_2 - A_2B_1). \end{aligned}$$

The components of C are then given by

$$\begin{aligned} C_1 &= A_2B_3 - A_3B_2 \\ C_2 &= A_3B_1 - A_1B_3 \\ C_3 &= A_1B_2 - A_2B_1. \end{aligned}$$

We can write these equations as a single equation, by introducing ϵ_{ijk}, a set of numbers labelled by three indices i, j and k, each of which is equal to 1, 2 or 3.

This symbol is know as the <u>Levi-Civita</u> symbol, ϵ_{ijk} and is <u>defined</u> by:

$$\begin{array}{lll} I & \epsilon_{ijk} = -\epsilon_{jik} = -\epsilon_{ikj} & \text{(antisymmetric)} \\ II & \epsilon_{123} = 1 & \end{array}$$

These definitions imply the following properties:

1. $\epsilon_{ijk} = -\epsilon_{kji}$ (i.e. also antisymmetric when swapping 1st and 3rd index).

   Proof:
   $$\epsilon_{ijk} = -\epsilon_{jik} = +\epsilon_{jki} = -\epsilon_{kji}$$

2. $\epsilon_{ijk} = 0$ if any two indices have the same value.

   Proof: e.g.
   $$\epsilon_{112} = -\epsilon_{112} \;\Rightarrow\; 2\epsilon_{112} = 0 \;\Rightarrow\; \epsilon_{112} = 0$$

3. The only non-zero ϵ_{ijk} therefore have ijk all different (by property 2.), so (ijk) is some permutation of (123).

4. $\epsilon_{ijk} = +1$ if ijk is an <u>even</u> permutation of 123 ("even" = even # swaps)

   $\epsilon_{ijk} = -1$ if ijk is an <u>odd</u> permutation of 123 ("odd" = odd # swaps)

5. $\epsilon_{ijk} = \epsilon_{kij} = \epsilon_{jki}$ (cyclic permutations).

   So e.g. $1 = \epsilon_{123} = \epsilon_{312} = \epsilon_{231}$, and $-1 = \epsilon_{132} = \epsilon_{321} = \epsilon_{213}$

Most importantly, now the vector product $\underline{C} = \underline{A} \times \underline{B}$ can be written as:

$$C_i \;=\; \epsilon_{ijk} A_j B_k. \tag{4.1}$$

<u>Check</u>:

$$\begin{aligned} C_1 = \epsilon_{1jk} A_j B_k &= \sum_{j,k=1}^{3} \epsilon_{1jk} A_j B_k \\ &\quad \text{(must have } j,k \neq 1 \text{ and } j \neq k \text{ by property 2.,} \\ &\quad \text{so either } j=2, k=3 \text{ or } j=3, k=2) \\ &= \epsilon_{123} A_2 B_3 + \epsilon_{132} A_3 B_2 \\ &\quad \text{but } \epsilon_{123} = 1 \text{ by } (II) \text{ above, and } \epsilon_{132} = -1 \text{ by both } (I) \text{ and } (II) \\ &= A_2 B_3 - A_3 B_2 \end{aligned}$$

I leave it as an exercise for you to check other two components C_2 and C_3.

4.4 The very useful ϵ_{ijk} formula

If we want to write out the vector triple product $\underline{A} \times (\underline{B} \times \underline{C})$, for example, we'll need to be able to put these ϵs together. Luckily there is a neat way to do it:

$$\sum_{k=1}^{3} \epsilon_{ijk}\epsilon_{klm} \;=\; \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}. \qquad (\dagger) \tag{4.2}$$

<u>Best just to remember this formula!</u>

<u>Check:</u> Let's first consider the left-hand side (LHS) or Equation (4.2).

$$LHS \text{ of Equation (4.2)} \;=\; \epsilon_{ij1}\epsilon_{1lm} + \epsilon_{ij2}\epsilon_{2lm} + \epsilon_{ij3}\epsilon_{3lm} \qquad (*)$$

Now $\epsilon_{ij1}\epsilon_{1lm} = 0$ unless

$$
\begin{aligned}
(i,j) &= (2,3) \text{ or } (3,2) \\
\text{and} \quad (l,m) &= (2,3) \text{ or } (3,2) \\
\Rightarrow \quad \epsilon_{ij1}\epsilon_{1lm} &= \begin{cases} +1 & \text{if} \quad (l,m) = (i,j) \\ -1 & \text{if} \quad (l,m) = (j,i) \end{cases} ,
\end{aligned}
$$

in which case the second and third terms in (*) will be zero.

A similar argument holds for other terms in (*), $\epsilon_{ij2}\epsilon_{2lm}$ and $\epsilon_{ij3}\epsilon_{3lm}$. Therefore the sum in (*) will be zero, except when

$$
\begin{aligned}
i \neq j, l \neq m, \quad (i,j) = (l,m) \quad &\Rightarrow \quad (*) \;=\; +1 \\
i \neq j, l \neq m, \quad (i,j) = (m,l) \quad &\Rightarrow \quad (*) \;=\; -1.
\end{aligned}
$$

Let's now consider the right-hand side (RHS) of Equation (4.2).

$$
RHS \text{ of Equation (4.2)} \;=\; \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}
$$

We have $\delta_{il}\delta_{jm} = 1$ if $i = l$ and $j = m \Leftrightarrow (i,j) = (l,m)$, otherwise this term is zero.

Similarly, $-\delta_{im}\delta_{jl} = -1$ if $i = m$ and $j = l \Leftrightarrow (i,j) = (m,l)$, otherwise this term is zero.

The combination of both terms is zero if either $i = j$ or $l = m$.

This implies that LHS of Equation (4.2) $= RHS$ of Equation (4.2), and hence the formula holds.

Example 16. Show that $\underline{A} \times (\underline{B} \times \underline{C}) \;=\; \underline{B}(\underline{A}.\underline{C}) - \underline{C}(\underline{A}.\underline{B})$

To do this using index notation, we should compute the i^{th} component of $\underline{A} \times (\underline{B} \times \underline{C})$. We can write this as $[\underline{A} \times (\underline{B} \times \underline{C})]_i$. Going slowly, this gives

$$
\begin{aligned}
[\underline{A} \times (\underline{B} \times \underline{C})]_i &= \epsilon_{ijk} A_j \,(\underline{B} \times \underline{C})_k && j,k: \text{ dummy indices, } i: \text{ free index, 1,2,3} \\
&= \epsilon_{ijk} A_j \epsilon_{klm} B_l C_m && l,m: \text{ more dummy indices} \\
&= (\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}) A_j B_l C_m && \text{by useful formula } (\dagger) \\
&= \delta_{il}\delta_{jm} A_j B_l C_m - \delta_{im}\delta_{jl} A_j B_l C_m \\
&= \delta_{il} A_j B_l C_j - \delta_{im} A_j B_j C_m && \text{by the rule for } \delta \\
&= A_j B_i C_j - A_j B_j C_i && \text{by the rule for } \delta \\
&= B_i (A_j C_j) - C_i (A_j B_j) \\
&= B_i (\underline{A}.\underline{C}) - C_i (\underline{A}.\underline{B}) \\
&= [\underline{B}(\underline{A}.\underline{C}) - \underline{C}(\underline{A}.\underline{B})]_i \,,
\end{aligned}
$$

i.e. i^{th} component of $\underline{A} \times (\underline{B} \times \underline{C}) \;=\; i^{\text{th}}$ component of $\underline{B}(\underline{A}.\underline{C}) - \underline{C}(\underline{A}.\underline{B})$. Since this is true for all $i = 1, 2, 3$ the result follows.

4.5 Applications

We've already seen in example 16 that index notation can be used to prove the vector triple product identity, $\underline{A} \times (\underline{B} \times \underline{C}) \;=\; \underline{B}(\underline{A}.\underline{C}) - \underline{C}(\underline{A}.\underline{B})$.

For many vector calculus calculations we need $\underline{\nabla} f$, $\underline{\nabla}.\underline{v}$ and $\underline{\nabla} \times \underline{v}$ in index notation

- The ith component of $\underline{\nabla} f$ is simply

$$
(\underline{\nabla} f)_i \;=\; \frac{\partial f}{\partial x_i}. \tag{4.3}
$$

  A common notation used to simplify this further is to write $\frac{\partial}{\partial x_i} \equiv \partial_i$, so then we can write

$$
(\underline{\nabla} f)_i \;=\; \partial_i f. \tag{4.4}
$$

- $\underline{\nabla}.\underline{v}$ can be thought of simply as $\underline{A}.\underline{v}$ where $\underline{A}$ has "components" $\frac{\partial}{\partial x_i}$:

$$\underline{\nabla}.\underline{v} \;=\; \frac{\partial v_i}{\partial x_i}, \tag{4.5}$$

  and using the notation from above this can also be written as

$$\underline{\nabla}.\underline{v} \;=\; \partial_i v_i. \tag{4.6}$$

- Similarly, $\underline{\nabla} \times \underline{v}$ is like $\underline{A} \times \underline{v}$. So if $\underline{u} = \underline{\nabla} \times \underline{v}$, then its ith component is:

$$u_i \;=\; (\underline{\nabla} \times \underline{v})_i \;=\; \epsilon_{ijk}\frac{\partial}{\partial x_j}v_k \;=\; \epsilon_{ijk}\frac{\partial v_k}{\partial x_j}, \tag{4.7}$$

  and using the same notation as above this can also be written as

$$u_i \;=\; (\underline{\nabla} \times \underline{v})_i \;=\; \epsilon_{ijk}\partial_j v_k. \tag{4.8}$$

Here are some more examples that show how useful index notation can be for proving identities in vector calculus.

Example 17. Find the gradient of $f(\underline{x}) = |\underline{x}|^2$ in $\mathbb{R}^n$.

$$\begin{aligned}
& f(\underline{x}) = |\underline{x}|^2 \;=\; \underline{x}.\underline{x} \;=\; x_i x_i \;=\; x_j x_j \\
\Rightarrow \quad & (\underline{\nabla} f)_i = \frac{\partial}{\partial x_i}(x_j x_j) \qquad \text{important not to reuse the dummy index } j \\
& = \left(\frac{\partial x_j}{\partial x_i}\right) x_j + x_j \left(\frac{\partial x_j}{\partial x_i}\right) \qquad \text{product rule} \\
& = 2\delta_{ij}x_j \;=\; 2x_i \;=\; (2\underline{x})_i \\
\Rightarrow \quad & \underline{\nabla} f = 2\underline{x}
\end{aligned}$$

Example 18. Find the divergence of $\underline{v}(\underline{x}) = \underline{x}$ and $\underline{u}(\underline{x}) = (\underline{a}.\underline{x})\underline{x}$ in $\mathbb{R}^3$.

$$\begin{aligned}
& v_i = x_i \\
\Rightarrow \quad & \underline{\nabla}.\underline{v} = \frac{\partial x_i}{\partial x_i} \;=\; \delta_{ii} \;=\; 3 \\
\text{While} \quad & u_i = (\underline{a}.\underline{x})x_i \;=\; a_j x_j x_i \\
\Rightarrow \quad & \underline{\nabla}.\underline{u} = \frac{\partial}{\partial x_i}(a_j x_j x_i) \\
& = a_j \left(\frac{\partial x_j}{\partial x_i}x_i + x_j \frac{\partial x_i}{\partial x_i}\right) \\
& = a_j(\delta_{ij}x_i + x_j \delta_{ii}) \\
& = a_j(x_j + 3x_j) \;=\; 4a_j x_j \\
& = 4\,\underline{a}.\underline{x}
\end{aligned}$$

Example 19. Find the curl of $\underline{v}(\underline{x}) = \underline{x}$ and $\underline{u}(\underline{x}) = (\underline{a}.\underline{x})\underline{x}$ in $\mathbb{R}^3$.

$$\begin{aligned}
v_i &= x_i \\
\Rightarrow \qquad (\underline{\nabla} \times \underline{v})_i &= \epsilon_{ijk} \frac{\partial}{\partial x_j} x_k \\
&= \epsilon_{ijk} \delta_{jk} \\
&= \epsilon_{ijj} \\
&= \epsilon_{i11} + \epsilon_{i22} + \epsilon_{i33} \\
&= 0 \\
\text{While} \qquad u_k &= a_j x_j x_k \;=\; a_l x_l x_k \\
\Rightarrow \qquad (\underline{\nabla} \times \underline{u})_i &= \epsilon_{ijk} \frac{\partial}{\partial x_j} u_k \\
&= \epsilon_{ijk} \frac{\partial}{\partial x_j} (a_l x_l x_k) \\
&= a_l \epsilon_{ijk} \left(\frac{\partial x_l}{\partial x_j} x_k + x_l \frac{\partial x_k}{\partial x_j} \right) \\
&= a_l \epsilon_{ijk} \left(\delta_{lj} x_k + x_l \delta_{kj} \right) \\
&= a_l \epsilon_{ijk} \delta_{lj} x_k + a_l \epsilon_{ijk} x_l \delta_{kj} \\
&= a_j \epsilon_{ijk} x_k + a_l \epsilon_{ijj} x_l \\
&= \epsilon_{ijk} a_j x_k + 0 \\
&= (\underline{a} \times \underline{x})_i \\
\Rightarrow \qquad \underline{\nabla} \times \underline{u} &= \underline{a} \times \underline{x}
\end{aligned}$$

Example 20. Show that $\underline{\nabla}.(\underline{\nabla} \times \underline{v}) = 0$, (if v_k has continuous 2nd partial derivatives).

$$\begin{aligned}
\underline{\nabla}.(\underline{\nabla} \times \underline{v}) &= \frac{\partial}{\partial x_i} (\underline{\nabla} \times \underline{v})_i \\
&= \frac{\partial}{\partial x_i} (\epsilon_{ijk} \frac{\partial}{\partial x_j} v_k) \\
&= \epsilon_{ijk} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} v_k
\end{aligned}$$

Now ϵ_{ijk} is <u>anti</u>-symmetric when i and j swap, whereas $\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} v_k$ is symmetric, so answer is zero(!).

In more detail:

$$\begin{aligned}
\underline{\nabla}.(\underline{\nabla} \times \underline{v}) &= \epsilon_{ijk} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} v_k \\
&= \epsilon_{jik} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} v_k \qquad \text{swapping labelling of dummy indices } i \text{ and } j \\
&= -\epsilon_{ijk} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} v_k \qquad \text{anti-symmetric } \epsilon_{ijk} \\
&= -\epsilon_{ijk} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} v_k \qquad \text{symmetric derivatives, due to continuity} \\
&= -\underline{\nabla}.(\underline{\nabla} \times \underline{v}) \\
\Rightarrow \qquad \underline{\nabla}.(\underline{\nabla} \times \underline{v}) &= 0
\end{aligned}$$

Much more elegant (and quicker!) then writing it all out.