2 Regular Surfaces

In this half of the course we approach surfaces in $\mathbb{E}^3$ in a similar way to which we considered curves. A parameterized surface will be a function[1] $\mathbf{x} : U \to \mathbb{E}^3$ where U is some open subset of the plane $\mathbb{R}^2$. Our purpose is twofold:

1. To be able to measure quantities such as *length* (of curves), *angle* (between curves on a surface), *area* using the *parameterization space* U. This requires us to create some method of taking tangent vectors to the surface and 'pulling-back' to U where we will perform our calculations.[2]

2. We want to find ways of defining and measuring the *curvature* of a surface.

Before starting, we recall some of the important background terms and concepts from other classes.

Notation

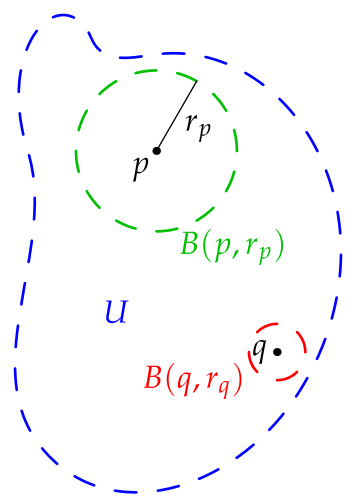
Surfaces being functions $\mathbf{x} : U \subseteq \mathbb{R}^2 \to \mathbb{E}^3$, we will preserve some of the notational differences between $\mathbb{R}^2$ and $\mathbb{E}^3$. Thus:

- *Co-ordinate points* in the parameterization space $U \subset \mathbb{R}^2$ will be written as lower case letters or, more commonly, *row vectors*: for example $p = (u, v) \in \mathbb{R}^2$.
- *Points* in $\mathbb{E}^3$ will be written using capital letters and row vectors: for example $P = (3, 4, 8) \in \mathbb{E}^3$.
- *Vectors* in $\mathbb{E}^3$ will be written bold-face or as *column-vectors*: for example $\mathbf{v} = \begin{pmatrix} 2 \\ -1 \\ \pi^2 \end{pmatrix}$.

Sometimes it will be convenient to abuse notation and add a vector $\mathbf{v}$ to a point P, the result will be a new point $P + \mathbf{v}$.

Open Sets in $\mathbb{R}^n$

The domains of our parameterized functions will always be open sets in $\mathbb{R}^2$. These are a little harder to describe than open intervals in $\mathbb{R}$: the definitions are here for reference.



Definition 2.1. Let $p \in \mathbb{R}^n$. The *open ball of radius r around p* is the set of all points whose distance to p is less than r: i.e.

$$B(p, r) = \{x \in \mathbb{R}^n : |x - p| < r\}$$

A set $U \subseteq \mathbb{R}^n$ is *open* if for all points $p \in U$ there is some radius $r_p > 0$ such that $B(p, r_p) \subseteq U$.

Otherwise said, around any point of an open set, there exists an open ball which is contained entirely in that set.

[1]For example $\mathbf{x}(u, v) = \begin{pmatrix} u \\ v \\ u^2+v^2 \end{pmatrix}$ describes a *paraboloid*.

[2]One of the great difficulties of differential geometry is that it is hard to see at this stage why this process is beneficial: why not measure all these things directly within $\mathbb{E}^3$? Part of the reason is that we want to be able to generalize: we want to be able to deal with surfaces (and eventually, though not in this class) with higher-dimensional objects (manifolds) without reference to some larger space in which they lie. As an example, suppose that a physicist says, 'space-time is curved.' They are not imagining an observer measuring its bendiness from the *outside*. On the contrary, the curvature is detected *within* spacetime: in general relativity gravity is precisely the curvature of space-time.

Examples

1. In $\mathbb{R}$, $B(p,r)$ is just the open interval $(p-r, p+r)$.
2. $(0,1) \subseteq \mathbb{R}$ is open: Let $p \in (0,1)$ and choose any r_p with $0 < r_p \leq \min(p, 1-p)$, then $B(p, r_p) \subseteq (0,1)$.
3. $[0,1) \subseteq \mathbb{R}$ is not open: for $p = 0$ it is impossible to find a radius $r > 0$ such that $B(p,r) = (-r,r)$ is contained entirely in $[0,1)$.
4. In $\mathbb{R}^2$, $B(p,r)$ is the interior (without edge) of the circle of radius r centered at p. In $\mathbb{R}^3$ it is the interior of the sphere of radius r centered at p.

We use the concept of an open set in $\mathbb{R}^n$ similarly to how we used open intervals when describing curves: a curve was usually defined as a function $\mathbf{x} : I \to \mathbb{E}^n$ where I was some open interval; surfaces will be defined on open subsets $U \subset \mathbb{R}^2$ similarly, e.g. $\mathbf{x} : U \to \mathbb{E}^3$. Openness is important because edges of sets cause difficulties for differentiation. Unless otherwise said, you should assume that the domain of any parameterized surface is open.

Partial Differentiation and the Gradient Vector

A multivariable scalar function is an assignment $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, i.e. $f(x_1, x_2, x_3) = x_1^2 - x_3 \cos x_2$ is a function $f : \mathbb{R}^3 \to \mathbb{R}$. Note that we could write $f(x) = x_1^2 - x_3 \cos x_2$, where $x = (x_1, x_2, x_3)$.
Supposing it exists, the *partial derivative* of f with respect to x_i is the function

$$\frac{\partial f}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}$$

The function f is *differentiable* if $\frac{\partial f}{\partial x_i}$ exists for all $i = 1, \ldots, n$. It is *smooth* if partial derivatives of all orders and all combinations exist: for example, if $f : \mathbb{R}^3 \to \mathbb{R}$ is smooth then $\frac{\partial^4 f}{\partial x_1 \partial x_2^2 \partial x_3}$ is defined and continuous. Essentially all our surfaces will be smooth. Some notations for functions:

- $C(U, V)$ is the set of *continuous* functions $f : U \to V$.
- $C^1(U, V)$ is the set of *continuously differentiable* functions $f : U \to V$ with (differentiable functions with continuous derivatives).
- $C^\infty(U, V)$ is the set of *smooth* functions.

It is common to say, for instance, that "f is C^1" rather than writing $f \in C^1(U, V)$. Sets C^2, C^3, etc., are defined similarly.

Example If $f(x) = x_1^2 - x_3 \cos x_2$, then it is a smooth function $f \in C^\infty(\mathbb{R}^3, \mathbb{R})$. Moreover,

$$\frac{\partial f}{\partial x_1} = 2x_1, \qquad \frac{\partial f}{\partial x_2} = x_3 \sin x_2, \qquad \frac{\partial f}{\partial x_3} = -\cos x_2$$

Definition 2.2. The *gradient* of a differentiable function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is the function $\nabla f : U \to \mathbb{R}^n$ defined by

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

If f is as above, we have $\nabla f = (2x_1, x_3 \sin x_2, -\cos x_3)$.

2.1 Vector Fields and 1-forms

We are almost ready to make the critical definition where we turn differential operators into vectors. The first step involves thinking about the gradient of a scalar function $f : \mathbb{E}^n \to \mathbb{R}$: in such a situation we would write ∇f as a column vector. Moreover, as the next definition facilitates, the gradient encodes both the *direction* in which the value of f changes maximally, and the *magnitude* of that rate of change.

Definition 2.3. If $\mathbf{v} = (v_1, \ldots, v_n)^T$ is a vector based at a point $P \in U \subseteq \mathbb{E}^n$ and $f : U \to \mathbb{R}$ is a function, then the *directional derivative* of f at P in the direction $\mathbf{v}$ is the *scalar*

$$D_{\mathbf{v}} f(P) := \mathbf{v} \cdot \nabla f(P) = \sum_{k=1}^{n} v_k \left.\frac{\partial f}{\partial x_k}\right|_P$$

The directional derivative allows one to compute the rate of change of f as one travels in any direction. Indeed:

Theorem 2.4. 1. *If h is a small number, then $f(P + h\mathbf{v}) \approx f(P) + D_{\mathbf{v}} f(P)h$.*

2. *If $\mathbf{x}(t)$ is a curve passing through P at $t = 0$ and with tangent vector $\mathbf{x}'(0) = \mathbf{v}$, then*

$$D_{\mathbf{v}} f(\mathbf{x}(0)) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(\mathbf{x}(t))$$

is the rate of change of f at P as you travel along the curve*.*

3. *If $\mathbf{v}$ is a unit vector, then $D_{\mathbf{v}} f(P)$ is maximal when $\mathbf{v}$ points in the same direction as the gradient vector.*

4. *∇f points in the direction of greatest increase of f at P. If $\mathbf{v}$ is the unit vector in that direction, then $|\nabla f(P)| = D_{\mathbf{v}} f(P)$.*

Example In $\mathbb{E}^3$, let $P = (x_1, x_2, x_3)$ and suppose that $f(P) = x_1^2 + x_2x_3 + x_3 - x_1$. Then

$$\nabla f = \begin{pmatrix} 2x_1 - 1 \\ x_3 \\ x_2 + 1 \end{pmatrix}$$

Now let $\mathbf{v} = \left(\begin{smallmatrix} 1 \\ 2 \\ -1 \end{smallmatrix}\right)$. Then the directional derivative in the direction of $\mathbf{v}$ is

$$D_{\mathbf{v}} f = \mathbf{v} \cdot \nabla f = \sum_{k=1}^{3} v_k \frac{\partial f}{\partial x_k} = 2x_1 - 1 + 2x_3 - (x_2 + 1)$$

Be careful with notation! $D_{\mathbf{v}} f$ is a *continuous function* $\mathbb{E}^3 \to \mathbb{R}$. If we evaluate at a specific point $P = (0, 1, 2)$, say, then $D_{\mathbf{v}} f(P) = -1 + 4 - 2 = 1$ is a *number.*

The Directional Derivative Operator

One of the reasons for putting the function f at the end of the directional derivative expression is that it tempts us to write the directional derivative at a point P as an *operator*:

$$D_{\mathbf{v}}|_P = \sum_{k=1}^{n} v_k \left.\frac{\partial}{\partial x_k}\right|_P$$

This is a map (function) from the space of continuously differentiable functions $U \subseteq \mathbb{E}^n \to \mathbb{R}$ to the real numbers. It is similarly tempting to drop any reference to the point P and define a more general operator

$$D_{\mathbf{v}} = \sum_{k=1}^{n} v_k \frac{\partial}{\partial x_k}$$

This is a map from the space of differentiable functions $U \subseteq \mathbb{E}^n \to \mathbb{R}$ to the space of continuous functions $U \to \mathbb{R}$: feed a C^1 function f to $D_{\mathbf{v}}$ and the result is a continuous function $D_{\mathbf{v}} f$.
It is worth noting that $\mathbf{v}$ determines the operator $D_{\mathbf{v}}$ and vice versa.[3]

One of the principal ideas of vector fields in differential geometry is to view them as directional derivative operators or, as we've seen above, *linear first-order partial differential operators*. This sounds artificially complicated but the rational is simple: we really only care about vectors in terms of how functions/surfaces change in said direction.

We now come to one of the fundamental definitions of differential geometry.

Definition 2.5. Let $U \subseteq \mathbb{R}^n$ be open and let $p \in U$. The *tangent space* to at p, denoted T_pU is the set of all directional derivative operators $D_v|_p$ at p. A *vector field* v on U is a (smooth) choice for each $p \in U$ of an element of T_pU.

By smooth, we mean that the entries of v are infinitely differentiable functions. Note that we currently use a lower-case $v \in \mathbb{R}^n$. We will often write $v|_p$ for the particular *tangent vector*[4] which is the value of the vector field at p.

Example $v = 3x_1 \frac{\partial}{\partial x_1} + 2x_1x_3 \frac{\partial}{\partial x_2} - x_1 \frac{\partial}{\partial x_3}$ is a vector field on $\mathbb{R}^3$. If $\mathbb{R}^3$ is Euclidean space $\mathbb{E}^3$, then this corresponds to the vector-valued function

$$\mathbf{v} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3x_1 \\ 2x_1x_3 \\ -x_1 \end{pmatrix}$$

$\mathbf{v}$ is an infinitely differentiable function $\mathbb{E}^3 \to \mathbb{E}^3$. It should be clear that the operators $\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}$ form a *basis* of the tangent space at each point.

[3]Given a linear partial differential operator $D_{\mathbf{v}}$, observe that the co-ordinate $x_k : \mathbb{E}^n \to \mathbb{R}$ is a C^1 function. Thus $v_k = D_{\mathbf{v}} x_k$, which recovers $\mathbf{v}$, entry by entry.

[4]The prefix *tangent* helps us discriminate between these and *position* vectors. To reiterate: in differential geometry almost every vector is a tangent vector: a directional derivative operator.

Definition 2.6. If v is a vector field on $U \subseteq \mathbb{R}^n$, and $f : U \to \mathbb{R}$ is a function, then we write $v[f]$ for the result of applying the vector field v to the function f. I.e.,

$$v = \sum_{k=1}^{n} v_k \frac{\partial}{\partial x_k} \implies v[f] = \sum_{k=1}^{n} v_k \frac{\partial f}{\partial x_k}$$

If f is also smooth, then $v[f]$ is itself a *smooth function* $U \to \mathbb{R}$.

Example Let $f(x) = x_1^2 x_2$ be a function on $\mathbb{R}^3$, and v be the vector field in the previous example. Then

$$v[f] = 3x_1 \frac{\partial f}{\partial x_1} + 2x_1 x_3 \frac{\partial f}{\partial x_2} - x_1 \frac{\partial f}{\partial x_3} = 6x_1^2 x_2 + 2x_1^3 x_3$$

Theorem 2.7. *Let v, w be vector fields on $U \subseteq \mathbb{R}^n$ and let $f, g : U \to \mathbb{R}$ be smooth. Then*

1. *Each tangent space is a vector space: indeed fv and $v + w$ are* vector fields, *where*

   $$(fv)[g] = fv[g], \qquad (v + w)[f] = v[f] + w[f]$$

2. *Vector fields act linearly on smooth functions: if $a, b \in \mathbb{R}$ are constant, then*

   $$v[af + bg] = av[f] + bv[g]$$

3. *Vector fields satisfy a product rule known as the* Leibniz rule*:*

   $$v[fg] = fv[g] + gv[f]$$

All these results are straightforward from the definition of a vector field as a linear first order partial differential operator.

Notation warning! Much of the difficulty of working with vector fields (and the associated 1-forms which are to come) is of becoming confused with notation. Remember that $v[f]$ is a function $\mathbb{R}^n \to \mathbb{R}$, whose value at $p \in \mathbb{R}^n$ is $v[f](p) = v|_p\,[f]$. The previous Theorem written at the point p, produces a total mess! In practice, extra brackets are very helpful...

$$\begin{aligned}
&(fv)[g](p) = f(p)v[g](p), \qquad (v + w)[f](p) = v[f](p) + w[f](p),\\
&v[af + bg](p) = av[f](p) + bv[g](p),\\
&v[fg](p) = f(p)v[g](p) + g(p)v[f](p)
\end{aligned}$$

Note in particular that fv denotes the vector field obtained by *multiplying* each vector in the vector field by the *value* of the function f at the corresponding point. It does *not* mean *apply the function f to the vector v*, which makes no sense.

For example, if $v = x_2^2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2}$ is a vector field on $\mathbb{R}^2$, $f(x_1, x_2) = x_1^2 x_2$ and $p = (2, -1)$, then

$fv = x_1^2 x_2^3 \frac{\partial}{\partial x_1} - x_1^3 x_2 \frac{\partial}{\partial x_2}$ is a *vector field.*

$v[f] = x_2^2 \frac{\partial}{\partial x_1}(x_1^2 x_2) - x_1 \frac{\partial}{\partial x_2}(x_1^2 x_2) = 2x_1 x_2^3 - x_1^3$ is a *function.*

$(fv)(p) = f(p)\, v|_p = -4\frac{\partial}{\partial x_1} + 8\frac{\partial}{\partial x_2}$ is a *tangent vector.*

$(v[f])(p) = -4 - 8 = -12$ is a *number.*

1-forms

Make sure you understand what a tangent vector is *before* moving on to this section, as 1-forms are defined in terms of tangent vectors.

Definition 2.8. A *1-form* at $p \in U \subseteq \mathbb{R}^n$ (sometimes called a *covector*), is a linear map

$$\alpha|_p : T_pU \to \mathbb{R}$$

The set of 1-forms at p is the *cotangent space* T_p^*U.
If $v|_p$ is a tangent vector at p, then we write $\alpha|_p\,(v|_p)$ for the value of this function.

Example Let $p = (1, 0, -2) \in \mathbb{R}^3$. Then $\alpha|_p$ defined by

$$\alpha|_p \left(a\, \frac{\partial}{\partial x_1}\bigg|_p + b\, \frac{\partial}{\partial x_2}\bigg|_p + c\, \frac{\partial}{\partial x_3}\bigg|_p \right) = 7a + 3b + c$$

is a 1-form. Indeed

$$\alpha|_p \left(-2\, \frac{\partial}{\partial x_1}\bigg|_p + 3\, \frac{\partial}{\partial x_2}\bigg|_p - 4\, \frac{\partial}{\partial x_3}\bigg|_p \right) = -2 \cdot 7 + 3 \cdot 3 - 4 \cdot 1 = -9$$

Definition 2.9. Write $\mathrm{d}x_k$ for the 1-form at p defined by

$$\mathrm{d}x_k \left(\frac{\partial}{\partial x_j}\bigg|_p \right) = \delta_{jk} = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases} \tag{†}$$

The 1-forms $\mathrm{d}x_k$ (where $k = 1, \ldots, n$) form a basis[5] of the cotangent space T_p^*U, hence the general 1-form at p has the form

$$\alpha|_p = \sum_{k=1}^{n} a_k\, \mathrm{d}x_k$$

for unique constants a_k.

Definition 2.10. A *1-form* on $U \subseteq \mathbb{R}^n$ is an assignment

$$\alpha = \sum_{k=1}^{n} a_k\, \mathrm{d}x_k$$

where each $a_k : U \to \mathbb{R}$ is a smooth function. We also write $\mathrm{d}x_k$ for the 1-form on U whose restriction to a point p is $\mathrm{d}x_k$.
If v is a vector field on U, we write $\alpha(v)$ for the *function* $U \to \mathbb{R}$ obtained by mapping $p \mapsto \alpha|_p\,(v|_p)$.

[5] The tangent space and the cotangent space at p both have the same dimension n. The * denotes 'dual space' in linear algebra: if V is a vector space over a field $\mathbb{F}$, then V^* is the vector space of linear maps $V \to \mathbb{F}$. Equation (†) says that the vectors $\frac{\partial}{\partial x_k}\big|_p$ and the covectors $\mathrm{d}x_k$ are dual bases to each other.

Example In $\mathbb{R}^2$, let $\alpha = 2x_2\,\mathrm{d}x_1 - 3\,\mathrm{d}x_2$ and $v = 3x_1^2\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2}$. Then $\alpha(v) = 6x_1^2x_2 - 3$ is a function $\mathbb{R}^2 \to \mathbb{R}$.

Theorem 2.11. *Let $f : U \to \mathbb{R}$ be smooth. Then there exists a unique 1-form denoted $\mathrm{d}f$ with the property that $\mathrm{d}f(v) = v[f]$ for all vector fields v on U.*

Definition 2.12. $\mathrm{d}f$ is the *exterior derivative* of f.
If a 1-form $\alpha = \mathrm{d}g$ is the exterior derivative of a function g, we say that it is *exact.*

Proof. A linear map is defined uniquely by what it does to a basis, thus, at each $p \in U$, it is enough to define $\mathrm{d}f|_p$ by

$$\mathrm{d}f|_p\left(\left.\frac{\partial}{\partial x_i}\right|_p\right) = \left.\frac{\partial f}{\partial x_i}\right|_p$$

$\mathrm{d}f$ is then the assignment $p \mapsto \mathrm{d}f|_p$. Otherwise said,

$$\mathrm{d}f = \frac{\partial f}{\partial x_1}\mathrm{d}x_1 + \cdots + \frac{\partial f}{\partial x_n}\mathrm{d}x_n \tag{$*$}$$

This is smooth since all derivatives of f are smooth. ∎

Example If $f(x_1, x_2) = 2x_1^2x_2 - 1$, then $\mathrm{d}f = 4x_1x_2\,\mathrm{d}x_1 + 2x_1^2\,\mathrm{d}x_2$.

Proposition 2.13. *If f, g are smooth functions, then*

1. $\mathrm{d}(f+g) = \mathrm{d}f + \mathrm{d}g$
2. $\mathrm{d}(fg) = f\,\mathrm{d}g + g\,\mathrm{d}f$
3. $\mathrm{d}f = 0 \iff f$ *is a constant function*

Proof. 1. and 2. follow immediately from the definition of $\mathrm{d}f$ and the relevant parts of Theorem 2.7. For 3.,

$$\begin{aligned}
\mathrm{d}f = 0 &\iff \mathrm{d}f[v] = 0 \text{ for all vector fields v} \\
&\iff \mathrm{d}f\left(\frac{\partial}{\partial x_i}\right) = 0 \text{ for all } i = 1, \ldots, n \\
&\iff \frac{\partial f}{\partial x_i} = 0 \text{ for all } i \\
&\iff f \text{ is constant}
\end{aligned}$$

∎

Example: Polar Co-ordinates

Co-ordinate functions $x_i : U \to \mathbb{R}$ have played a large role in our discussion of vector fields and 1-forms: at each point $p \in U$ the tangent and contangent spaces have bases generated by these functions:

$$T_pU = \operatorname{Span}\left\{\left.\frac{\partial}{\partial x_1}\right|_p, \ldots, \left.\frac{\partial}{\partial x_n}\right|_p\right\} \qquad T_p^*U = \operatorname{Span}\left\{\mathrm{d}x_1, \ldots, \mathrm{d}x_n\right\}$$

As an example of this, we consider the change of co-ordinates in $U = \mathbb{R}^2 \setminus \{(0,0)\}$ from rectangular to polar. Recall:

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \qquad \leftrightsquigarrow \qquad \begin{cases} r = \sqrt{x^2+y^2} \\ \tan\theta = \frac{y}{x} \end{cases}$$

The chain rule tells us:

$$\frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x}\frac{\partial}{\partial \theta} = \frac{x}{\sqrt{x^2+y^2}}\frac{\partial}{\partial r} - \frac{y}{x^2+y^2}\frac{\partial}{\partial \theta} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}$$

Similarly

$$\frac{\partial}{\partial y} = \frac{\partial r}{\partial y}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial y}\frac{\partial}{\partial \theta} = \frac{y}{\sqrt{x^2+y^2}}\frac{\partial}{\partial r} + \frac{x}{x^2+y^2}\frac{\partial}{\partial \theta} = \sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}$$

We can also compute the 1-forms: by $(*)$,

$$\mathrm{d}x = \mathrm{d}(r\cos\theta) = \cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta, \qquad \mathrm{d}y = \sin\theta\,\mathrm{d}r + r\cos\theta\,\mathrm{d}\theta$$

We can check that the dual basis relations hold:

$$\begin{aligned} \mathrm{d}x\left(\frac{\partial}{\partial x}\right) &= (\cos\theta\,\mathrm{d}r - r\sin\theta\,\mathrm{d}\theta)\left(\cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}\right) \\ &= \cos^2\theta\,\mathrm{d}r\left(\frac{\partial}{\partial r}\right) - \frac{\cos\theta\sin\theta}{r}\,\mathrm{d}r\left(\frac{\partial}{\partial \theta}\right) - r\sin\theta\cos\theta\,\mathrm{d}\theta\left(\frac{\partial}{\partial r}\right) + \sin^2\theta\,\mathrm{d}\theta\left(\frac{\partial}{\partial \theta}\right) = 1 \end{aligned}$$

Similarly

$$\mathrm{d}y\left(\frac{\partial}{\partial y}\right) = 1 \qquad \mathrm{d}x\left(\frac{\partial}{\partial y}\right) = \mathrm{d}y\left(\frac{\partial}{\partial x}\right) = 0$$

One can also compute $\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}, \mathrm{d}r, \mathrm{d}\theta$ in terms of $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \mathrm{d}x, \mathrm{d}y$, either by pure linear algebra from the above, or directly using the chain rule.

Relation to elementary calculus

When $n = 1$ we are used to writing $\frac{\mathrm{d}f}{\mathrm{d}x}$ for the derivative of a function $f : \mathbb{R} \to \mathbb{R}$. Contrary to what you might expect, $\frac{\mathrm{d}f}{\mathrm{d}x}$ is *not* the quotient of two 1-forms, rather it is the result of the application of the vector field $\frac{\mathrm{d}}{\mathrm{d}x}$ to the function f, or equivalently the application of the 1-form $\mathrm{d}f$ to the vector field $\frac{\mathrm{d}}{\mathrm{d}x}$:

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}[f] = \mathrm{d}f\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)$$

Vector fields in $\mathbb{R}$ are written with a straight d rather than partial ∂ because there is only one direction in which to differentiate. The notion of tangent vector is at best useless and at worst misleading in $\mathbb{R}$!

Line integrals

Where you have seen 1-forms before is in integration. $\int_0^1 g(x)\mathrm{d}x$ is the integral of the 1-form $g\,\mathrm{d}x$ over the interval $[0,1]$. Indeed we integrate 1-forms over oriented curves.

Definition 2.14. If $x : [a,b] \to U \subseteq \mathbb{R}^n$ is a curve, and α is a 1-form on U, we define

$$\int_x \alpha := \int_a^b \alpha\left(x'(t)\right)\mathrm{d}t$$

In order to feed the 1-form a tangent vector, we are obliged think of the derivative of the curve $x(t)$ as a family of tangent vectors along the curve,[6] i.e.

$$x(t) = \begin{pmatrix} f_1(t) \\ f_2(t) \end{pmatrix} \implies x'(t) = f_1'(t)\frac{\partial}{\partial x_1} + f_2'(t)\frac{\partial}{\partial x_2}$$

Examples

1. Integrate $\alpha = x_1\mathrm{d}x_2$ over the unit-circle counter-clockwise. Here $x(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$, so that $x'(t) = -\sin t\frac{\partial}{\partial x_1} + \cos t\frac{\partial}{\partial x_2}$. Then $\alpha(x'(t)) = x_1 \cos t$ which, restricted to the curve $x(t)$, is $\cos^2 t$. Around the unit circle is $t = 0$ to 2π, thus,

$$\int_x \alpha = \int_0^{2\pi} \cos^2 t\,\mathrm{d}t = \frac{1}{2}\int_0^{2\pi} 1 + \cos 2t\,\mathrm{d}t = \pi$$

2. Integrate $\alpha = x_2^2\mathrm{d}x_1 - x_1^2\mathrm{d}x_2$ over the curve $x(t) = (t,t^2)$ between $(0,0)$ and $(1,1)$.

$$\int_x \alpha = \int_0^1 \alpha(x'(t))\mathrm{d}t = \int_0^1 (t^2)^2 - 2t^3\mathrm{d}t = \frac{1}{5} - \frac{1}{2} = -\frac{3}{10}$$

Theorem 2.15. *The integral of a 1-form along a curve is independent of the choice of (orientation-preserving) parameterization.*

Proof. Suppose we are integrating α over a curve parameterized by $x(t)$ where t runs from a to b. Now suppose that t is a function of s ($t(p) = a,\ t(q) = b$) where $s'(t) > 0$ (orientation-preserving) and integrate using s as a parameter. The critical observation is that $\frac{\mathrm{d}}{\mathrm{d}s}x(t(s)) = x'(t(s))\frac{\mathrm{d}t}{\mathrm{d}s}$ by the chain rule.

$$\begin{aligned}
\int_x \alpha &= \int_p^q \alpha\left(\frac{\mathrm{d}}{\mathrm{d}s}x(t(s))\right)\mathrm{d}s = \int_p^q \alpha\left(\frac{\mathrm{d}t}{\mathrm{d}s}x'(t(s))\right)\mathrm{d}s \\
&= \int_p^q \alpha(x'(t(s)))\frac{\mathrm{d}t}{\mathrm{d}s}\mathrm{d}s = \int_a^b \alpha(x'(t))\mathrm{d}t \\
&= \int_x \alpha \qquad \text{(integrating with respect to } t\text{)}
\end{aligned}$$

∎

[6] This should make sense: the velocity x' indicates magnitude and direction or travel, *not* position, so it should be thought of as a tangent vector rather than a position vector.

If we change the orientation of the curve then the order of the limits is reversed and $\int \alpha$ becomes $-\int \alpha$.

As the next theorem shows, integration of 1-forms that are the exterior derivative of a function is independent of the path you choose. This is essentially the fundamental theorem of calculus for curves.

Theorem 2.16 (Fundamental Theorem of Line Integrals)**.** *Let $x(t)$ be a curve for $a \leq t \leq b$. If f is a function on $\mathbb{R}^n$, then,*

$$\int_x \mathrm{d}f = f(x(b)) - f(x(a))$$

The integral of $\mathrm{d}f$ *over any curve between two points depends only on the values of f at those points.*

Proof.

$$\int_x \mathrm{d}f = \int_a^b \mathrm{d}f(x')\,\mathrm{d}t = \int_a^b x'[f]\,\mathrm{d}t$$

However,

$$x'[f] = x_1' \frac{\partial f}{\partial x_1} + \cdots + x_n' \frac{\partial f}{\partial x_n} = \frac{\mathrm{d}x_1}{\mathrm{d}t}\frac{\partial f}{\partial x_1} + \cdots + \frac{\mathrm{d}x_n}{\mathrm{d}t}\frac{\partial f}{\partial x_n} = \frac{\mathrm{d}}{\mathrm{d}t}(f(x(t)))$$

Hence

$$\int_x \mathrm{d}f = \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}(f(x(t)))\mathrm{d}t = f(x(b)) - f(x(a))$$

■

Example If $\alpha = \cos(x_1 x_2)(x_2 \mathrm{d}x_1 + x_1 \mathrm{d}x_2)$, find the integral of α over any curve between the points $\left(\pi, \frac{1}{3}\right)$ and $\left(\frac{1}{2}, \pi\right)$.

Observe that $\alpha = \mathrm{d}\sin(x_1 x_2)$. Thus if γ is any curve between $(1,2)$ and $(3,4)$, we have

$$\int_\gamma \alpha = \sin(x_1 x_2)\Big|_{\left(\pi,\frac{1}{3}\right)}^{\left(\frac{1}{2},\pi\right)} = \sin\frac{\pi}{2} - \sin\frac{\pi}{3} = 1 - \frac{\sqrt{3}}{2}$$

Summary

There should be two take-aways from this discussion:

- Vector fields and tangent vectors allow one to express the notion that vectors encode a direction in which something can change (i.e. a derivative).
- Vector fields and 1-forms essentially break standard derivatives into two pieces: the result is an alternative, and ultimately more flexible, language for describing familiar results from multi-variable calculus.

The real pay-off is hard to see until we apply our new language to surfaces. Thus…

2.2 Surfaces

A surface can be thought of in two ways:

1. A two-dimensional subset of $\mathbb{E}^3$ with certain properties.

2. The image of a function $\mathbf{x} : U \to \mathbb{E}^3$ defined on an open set $U \subseteq \mathbb{R}^2$.

The second notion is easier to calculate with since we can describe points on the surface in terms of co-ordinates on U. It is, however, less general and arguably less geometric. In practice, a surface is typically defined as being the union of the images of several functions $\mathbf{x}_i$ with suitable rules governing what happens on overlaps.[7]

Definition 2.17. 1. A (local) parameterized surface S is the image of a map $\mathbf{x} : U \to \mathbb{E}^3$ where U is an open subset of $\mathbb{R}^2$. I.e. $S = \mathbf{x}(U)$. In what follows, we will often abuse language and refer to the parameterization $\mathbf{x}$ as the surface.

2. Let u, v be co-ordinates on U. The *co-ordinate tangent vector fields* to $\mathbf{x}$ are

$$\mathbf{x}_u = \frac{\partial \mathbf{x}}{\partial u} \qquad \mathbf{x}_v = \frac{\partial \mathbf{x}}{\partial v}$$

3. The surface $\mathbf{x}$ (with co-ordinates u, v) is *regular* at a point $p \in U$ if $\mathbf{x}_u(p), \mathbf{x}_v(p)$ are linearly independent.

4. If $\mathbf{x}$ is regular at $p \in U$, then the *tangent plane* to the surface at $\mathbf{x}(p)$ is $\operatorname{Span}\{\mathbf{x}_u(p), \mathbf{x}_v(p)\}$.

5. A surface is *regular* if it is regular at all points $p \in U$.

6. If $\mathbf{x}$ is regular, then the *unit normal vector field* to the surface is

$$\mathbf{U} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{|\mathbf{x}_u \times \mathbf{x}_v|}$$

   A choice of one of the two unit normal directions is called choosing an *orientation* of the surface.

Throughout we will assume that a surface is smooth and regular: i.e. all derivatives of $\mathbf{x}$ exist at all points of U and the tangent plane is well defined at each point of $\mathbf{x}(U)$.

We now make use of some of the language of the previous section and obtain an important result.

Definition 2.18. Suppose that $\mathbf{x} : U \to \mathbb{E}^3$ is a smooth surface. The *differential* of $\mathbf{x}$ is the vector-valued 1-form $\mathrm{d}\mathbf{x}$.

Suppose that u, v are co-ordinates on $U \subseteq \mathbb{R}^2$ and that $\mathbf{x}$ is a smooth surface. Then the differential of $\mathbf{x}$ can be written

$$\mathrm{d}\mathbf{x} = \frac{\partial \mathbf{x}}{\partial u}\mathrm{d}u + \frac{\partial \mathbf{x}}{\partial v}\mathrm{d}v = \mathbf{x}_u \mathrm{d}u + \mathbf{x}_v \mathrm{d}v$$

where $\mathrm{d}u, \mathrm{d}v$ are the co-ordinate 1-forms on U.

[7] See the aside below…

Theorem 2.19. *If* $\mathbf{x} : U \to \mathbb{E}^3$ *is a surface and* $y \in T_pU$ *is a tangent vector to* U *at* $p \in U$*, then* $\mathrm{d}\mathbf{x}(y)$ *is tangent to the surface at* $\mathbf{x}(p)$*. Indeed at each point* $p \in U$ *we have a* linear map

$$\mathrm{d}\mathbf{x}|_p : T_pU \to T_{\mathbf{x}(p)}S$$

Proof. Write $y = a\, \frac{\partial}{\partial u}\Big|_p + b\, \frac{\partial}{\partial v}\Big|_p$. Then $\mathrm{d}\mathbf{x}(y) = a\mathbf{x}_u + b\mathbf{x}_v$ is clearly tangent to the surface at $\mathbf{x}(p)$. ∎

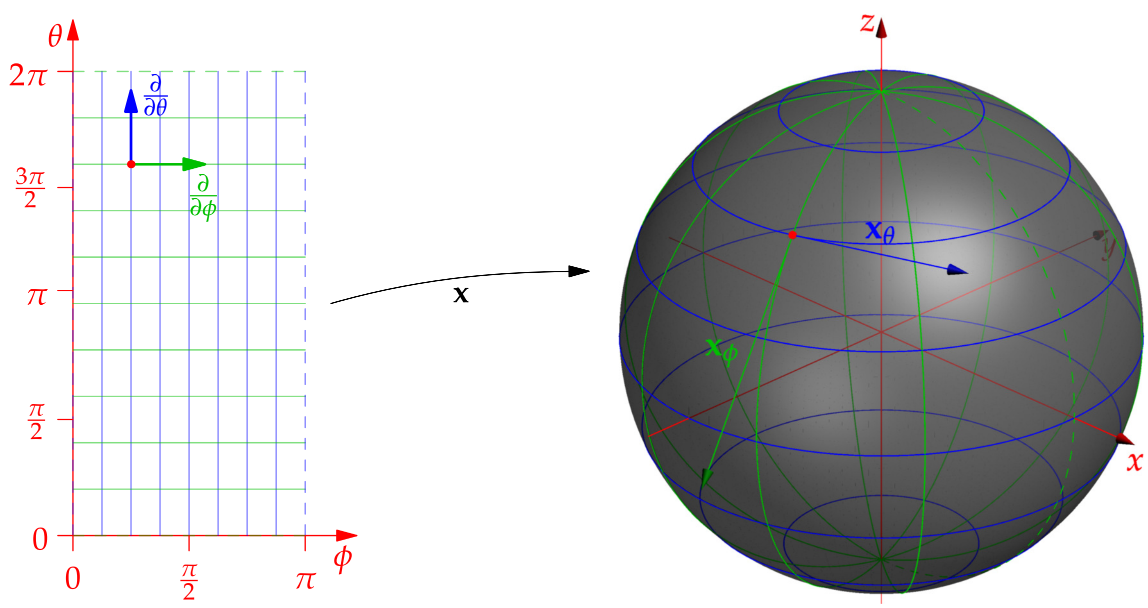
We say that $\mathrm{d}\mathbf{x}$ is a 1-form with values in $\mathbb{E}^3$: it maps tangent vectors to U to genuine tangent vectors to the surface. This indeed is one reason for calling linear 1st-order differential operators at $p \in U$ tangent vectors.

Examples

1. Sphere of radius a. Use spherical polar co-ordinates:

$$\mathbf{x}(\phi, \theta) = a \begin{pmatrix} \sin\phi\cos\theta \\ \sin\phi\sin\theta \\ \cos\phi \end{pmatrix}$$

   Here $\mathbf{x} : U \to \mathbb{E}^3$, where $U = (0, \pi) \times (0, 2\pi)$. Note that the image of $\mathbf{x}$ is the sphere minus the (dashed) semicircle defined by $\theta = 0$, so that our domain is open. We could argue for extending θ modulo 2π so that co-ordinates coontinue round the sphere. In contrast, however, the co-ordinates cannot be extended to the north or south poles without losing *regularity*: in this parameterization the sphere is not regular at the poles. This explains the term *local*: we usually need more than one parameterization to cover a complicated surface.



   Observe in the picture how the tangent vectors $\frac{\partial}{\partial\phi}, \frac{\partial}{\partial\theta}$ to U are mapped by the differential $\mathrm{d}\mathbf{x}$ to the tangent vectors

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\phi} = \mathrm{d}\mathbf{x}\left(\frac{\partial}{\partial\phi}\right), \qquad \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\theta} = \mathrm{d}\mathbf{x}\left(\frac{\partial}{\partial\theta}\right)$$

2. Graph of a smooth function $z = f(x,y)$. Parameterize by $\mathbf{x}(u,v) = \begin{pmatrix} u \\ v \\ f(u,v) \end{pmatrix}$. This has

$$\mathrm{d}\mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ f_u \end{pmatrix} \mathrm{d}u + \begin{pmatrix} 0 \\ 1 \\ f_v \end{pmatrix} \mathrm{d}v \qquad \mathbf{U} = \frac{1}{\sqrt{1 + f_u^2 + f_v^2}} \begin{pmatrix} -f_u \\ -f_v \\ 1 \end{pmatrix}$$

where $f_u = \frac{\partial f}{\partial u}$, etc. This is clearly regular at all points, regardless of the function f.

For example, the paraboloid $z = x^2 + y^2$ may be parameterized $\mathbf{x}(u,v) = \begin{pmatrix} u \\ v \\ u^2 + v^2 \end{pmatrix}$.

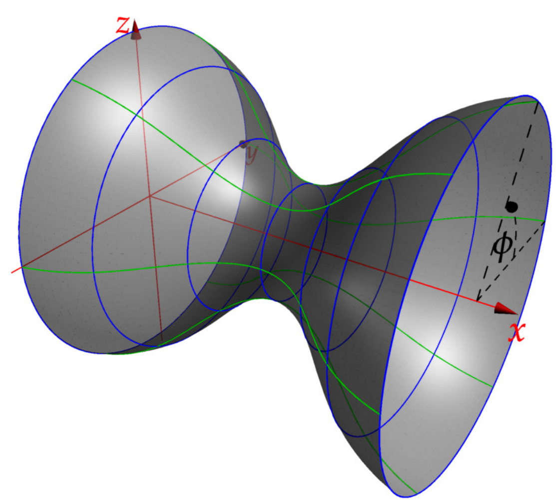
3. Surfaces of revolution. For instance, a smooth curve $z = f(x)$ can be rotated around the x-axis to obtain

$$\mathbf{x}(\phi, u) = \begin{pmatrix} u \\ f(u)\cos\phi \\ f(u)\sin\phi \end{pmatrix} \qquad (\phi, u) \in (0, 2\pi) \times \mathrm{dom}(f)$$

This has differential and unit normal field

$$\mathrm{d}\mathbf{x} = \begin{pmatrix} 0 \\ -f(u)\sin\phi \\ f(u)\cos\phi \end{pmatrix} \mathrm{d}\phi + \begin{pmatrix} 1 \\ f'(u)\cos\phi \\ f'(u)\sin\phi \end{pmatrix} \mathrm{d}u$$

$$\mathbf{U} = \frac{1}{\sqrt{1 + f'(u)^2}} \begin{pmatrix} -f'(u) \\ \cos\phi \\ \sin\phi \end{pmatrix}$$



The surface is regular everywhere.

Just as with the sphere, we should ignore the curve $\phi = 0$, but it is not crucial. The picture is the surface of revolution defined by the curve $z = f(x) = 2 + \cos x$ for $x \in (0, 2\pi)$. The co-ordinate ϕ essentially measures the angle around the axis of rotation.

4. Ruled surfaces. Given two curves $\mathbf{y}(u), \mathbf{z}(u)$, define

$$\mathbf{x}(u,v) = \mathbf{y}(u) + v\mathbf{z}(u)$$

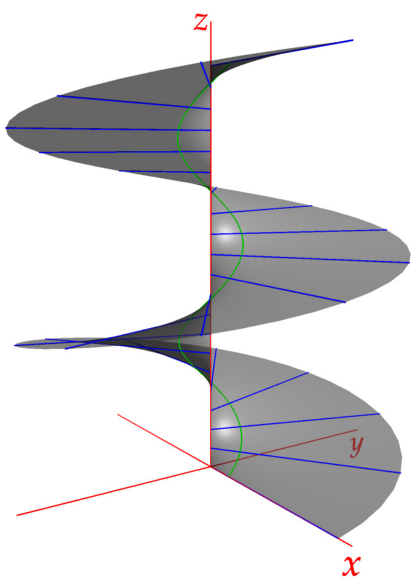
In general, the surface is ruled by the lines parameterized by v: that is, through each point on the surface, there exists a line parameterized which is contained entirely within the surface. The particular case where $\mathbf{z}(u) = \mathbf{y}'(u)$ is called the *tangent developable* of $\mathbf{y}(u)$.

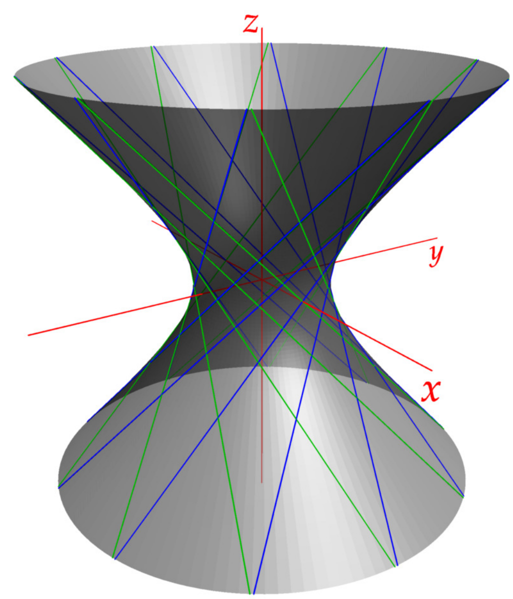
Ruled surfaces are very useful in engineering applications as they are very simple to construct: for example corrugated iron is a ruled surface built on a curve which is approximately a sine wave. Two famous examples of ruled surfaces are shown below:

- The *helicoid* is built from a helix by attaching to each point a horizontal line pointing to the z-axis. A particular example is given by $\mathbf{x}(u,v) = \begin{pmatrix} v\cos u \\ v\sin u \\ u \end{pmatrix}$ for $v > 0$.
- The *hyperboloid of one sheet* is in fact a *doubly ruled surface*: through each point there are *two* lines lying on the surface. It may be parameterized as a ruled surfaces by

$$\mathbf{x}(u,v) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + u \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + v \begin{pmatrix} 2u \\ u^2 - 1 \\ u^2 + 1 \end{pmatrix}$$

  though convincing yourself of the double ruling might take a little more work...



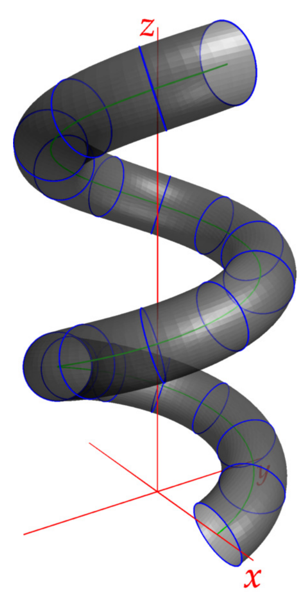


5. Tube of radius $a > 0$ about a curve $\mathbf{y}(u)$. Parameterize as

$$\mathbf{x}(u,\phi) = \mathbf{y}(u) + a\cos\phi\mathbf{N}(u) + a\sin\phi\mathbf{B}(u)$$

   where $\mathbf{N}, \mathbf{B}$ are the normal and binormal vectors of the Frenet frame of $\mathbf{y}$.

   A tube around a helix is drawn.



Implicitly Defined Surfaces

You are probably more used to surfaces being defined implicitly—as the level sets of some function—rather than being parameterized. This method of defining surfaces is extremely useful. Not having an explicit parameterization means that calculations are harder, however you gain by having a more global view of a surface.

Definition 2.20. An *implicitly defined surface* is the zero set of a smooth function $f : \mathbb{E}^3 \to \mathbb{R}$. We say that it is *regular* if ∇f (equivalently $\mathrm{d}f$) does not vanish on the surface.

Examples

1. The sphere of radius a is implicitly defined by $f(x, y, z) = x^2 + y^2 + z^2 - a^2$. Here

   $$\mathrm{d}f = 2(x\mathrm{d}x + y\mathrm{d}y + z\mathrm{d}z)$$

   is never zero since at least one of x, y, z is non-zero at all points on the surface $f = 0$. The sphere is thus regular. Note the contrast with our earlier example of the *parameterized* sphere which we could not make regular at the north and south pole. The lack of regularity in this case is a function of the parameterization, not the surface itself.

2. The function $f(x, y, z) = x^2 + y^2 - z^2 - a$ has

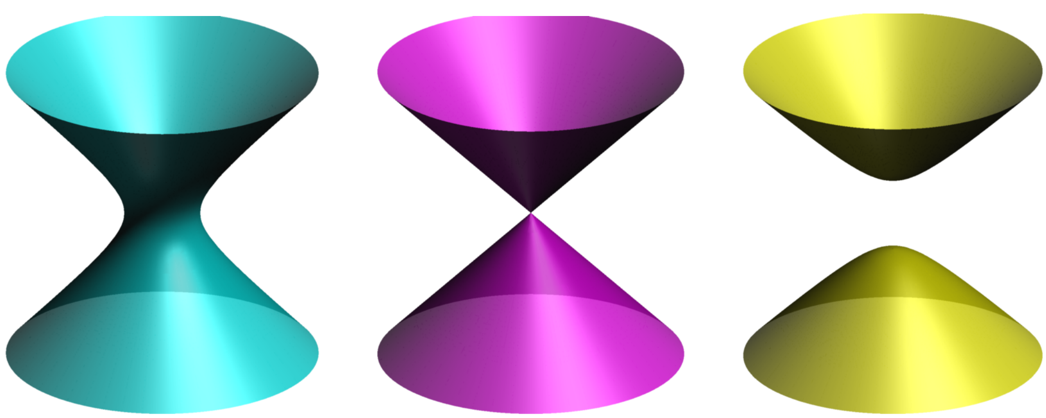
   $$\mathrm{d}f = 2(x\mathrm{d}x + y\mathrm{d}y - z\mathrm{d}z),$$

   which is non-zero away from $(x, y, z) = (0, 0, 0)$. Depending on the sign of a, the set of points with $f = 0$ is a hyperboloid or a cone.

   $a > 0$ Hyperboloid of 1-sheet: $x^2 + y^2 = z^2 + a > 0$ for all z

   $a = 0$ Cone: $(0, 0, 0)$ is not a regular point of the cone

   $a > 0$ Hyperboloid of 2-sheets: $x^2 + y^2 = z^2 - a \geq 0$ only when $|z| \geq \sqrt{a}$



The next theorem ties together the currently distinct notions of regular. More importantly, it says that we can always assume the existence of *local* co-ordinates.

Theorem 2.21. *A regular implicitly defined surface is (locally) the image of a regular local surface.*

We omit the proof, as it is essentially the famous (and difficult!) *Implicit Function Theorem*[8] from multi-variable calculus in disguise.
In order to calculate with implicitly defined surfaces, you need some of the ingredients of parameterized surfaces. In particular: if $f(x,y,z) = 0$ describes a regular surface S and $s \in S$, then $\nabla f(s)$ is orthogonal to the surface at s and so the tangent plane to S at s is

$$T_sS = \{\mathbf{x} \in \mathbb{E}^3 : (\mathbf{x} - s) \cdot \nabla f(s) = \mathbf{0}\}$$

Example The hyperboloid of one sheet defined implicitly by the equation $x^2 + y^2 - z^2 = 12$ has unit normal vector

$$\mathbf{U}(x,y,z) = \frac{1}{\sqrt{x^2+y^2+z^2}}\begin{pmatrix} x \\ y \\ -z \end{pmatrix} = \frac{1}{\sqrt{12+2z^2}}\begin{pmatrix} x \\ y \\ -z \end{pmatrix}$$

whenever (x,y,z) is a point on the hyperboloid. Thus at the point $(3,2,1)$, the unit normal is $\frac{1}{\sqrt{14}}(3,2,-1)$, and the tangent plane has equation

$$3x + 2y - z = 12$$

Of course, the hyperboloid could have been parameterized: it is a surface of revolution (around the z-axis) so we could choose

$$\mathbf{x}(u,\theta) = \begin{pmatrix} \sqrt{12+u^2}\cos\theta \\ \sqrt{12+u^2}\sin\theta \\ u \end{pmatrix}$$

Then the differential is

$$\mathrm{d}\mathbf{x} = \begin{pmatrix} u(12+u^2)^{-1/2}\cos\theta \\ u(12+u^2)^{-1/2}\sin\theta \\ 1 \end{pmatrix} \mathrm{d}u + \begin{pmatrix} -\sqrt{12+u^2}\sin\theta \\ \sqrt{12+u^2}\cos\theta \\ 0 \end{pmatrix} \mathrm{d}v$$

with resulting normal field

$$\mathbf{U} = \frac{\mathbf{x}_u \times \mathbf{x}_\theta}{|\mathbf{x}_u \times \mathbf{x}_\theta|} = \frac{1}{\sqrt{12+2u^2}}\begin{pmatrix} -\sqrt{12+u^2}\cos\theta \\ -\sqrt{12+u^2}\sin\theta \\ u \end{pmatrix}$$

Notice that this is precisely the negative of what we obtained earlier, but no matter. Yet another expression could be obtained using our parameterization as a ruled surface (page 14).

[8] If $f(x,y,z) = 0$ is the implicit surface and $\nabla f \neq \mathbf{0}$, then, at any given $P = (x_0, y_0, z_0)$, at least one of the partial derivatives of f is non-zero: suppose WLOG that $\frac{\partial f}{\partial z}\big|_P \neq 0$. The implicit function theorem says that there exists an open set $U \subseteq \mathbb{R}^2$ and a unique function $g : U \to \mathbb{R}$ such that $g(x_0, y_0) = z_0$ and $f\big(x, y, g(x,y)\big) = 0$. The surface may then (locally) be parameterized as a graph:

$$\mathbf{x} : U \to \mathbb{E}^3 : (u,v) \mapsto \begin{pmatrix} u \\ v \\ g(u,v) \end{pmatrix}$$

Aside: Global surfaces We have only rigorously defined a local parameterized surface. In general a surface S is any subset of $\mathbb{E}^3$ which locally 'looks like' a parameterized surface. That is, for any point $s \in S$, there exists a subset $V \subseteq S$ containing s, an open subset U of $\mathbb{R}^2$ and a smooth local surface $\mathbf{x} : U \to V$.
The upshot is that we may require more than one co-ordinate chart (subsets $\mathbf{x}(U) \subseteq S$) to cover an entire surface. For example, the sphere described using spherical polar co-ordinates covers the whole sphere in a regular fashion, except for a semi-circle (or, pushing things with periodicity in θ, everywhere except two points). To cover the entire sphere with co-ordinate charts we need to choose another point to be the center of a new chart. Indeed a famous topological result (the Hairy Dog Theorem) says that it is *impossible* to find global co-ordinates on the sphere: the best you can do is to have the entire sphere except for one point on a single chart (stereographic projection does the job, if you know what this is...).

When we start thinking about surfaces globally (such as the entire sphere at once), we see that it is possible to construct non-orientable surfaces such as the Möbius strip. A Möbius strip is certainly a surface: near any point it looks exactly like a bit of $\mathbb{R}^2$, and so we can find local co-ordinates. However if you try to look at the surface globally you find you cannot orient it. Consider walking around the strip once. An orientation at a point is a choice of unit normal vector, a choice of which way is 'up'. On circuiting the strip you find that 'up' is now 'down'. It is thus impossible to choose a global *regular* parameterization.

Change of co-ordinates Suppose that $\mathbf{y}(s,t) = \mathbf{x}\big(F(s,t)\big)$ where $F(s,t) = (u,v)$ is a change of co-ordinates on the parameterization space U. The differential takes care of everything. By the chain rule the differentials $\mathrm{d}\mathbf{x}$ and $\mathrm{d}\mathbf{y}$ may be thought of as related by a matrix multiplication:

$$\begin{pmatrix} \frac{\partial \mathbf{y}}{\partial s} \\ \frac{\partial \mathbf{y}}{\partial t} \end{pmatrix} = \begin{pmatrix} \frac{\partial u}{\partial s} & \frac{\partial v}{\partial s} \\ \frac{\partial u}{\partial t} & \frac{\partial v}{\partial t} \end{pmatrix} \begin{pmatrix} \frac{\partial \mathbf{x}}{\partial u} \\ \frac{\partial \mathbf{x}}{\partial v} \end{pmatrix} \qquad (\mathrm{d}u\ \mathrm{d}v) = (\mathrm{d}s\ \mathrm{d}t) \begin{pmatrix} \frac{\partial u}{\partial s} & \frac{\partial v}{\partial s} \\ \frac{\partial u}{\partial t} & \frac{\partial v}{\partial t} \end{pmatrix} \implies (\mathrm{d}s\ \mathrm{d}t) = (\mathrm{d}u\ \mathrm{d}v) \begin{pmatrix} \frac{\partial u}{\partial s} & \frac{\partial v}{\partial s} \\ \frac{\partial u}{\partial t} & \frac{\partial v}{\partial t} \end{pmatrix}^{-1}$$

from which

$$\begin{aligned} \mathrm{d}\mathbf{y} &= (\mathrm{d}s\ \mathrm{d}t) \begin{pmatrix} \frac{\partial \mathbf{y}}{\partial s} \\ \frac{\partial \mathbf{y}}{\partial t} \end{pmatrix} = (\mathrm{d}u\ \mathrm{d}v) \begin{pmatrix} \frac{\partial u}{\partial s} & \frac{\partial v}{\partial s} \\ \frac{\partial u}{\partial t} & \frac{\partial v}{\partial t} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial u}{\partial s} & \frac{\partial v}{\partial s} \\ \frac{\partial u}{\partial t} & \frac{\partial v}{\partial t} \end{pmatrix} \begin{pmatrix} \frac{\partial \mathbf{x}}{\partial u} \\ \frac{\partial \mathbf{x}}{\partial v} \end{pmatrix} \\ &= (\mathrm{d}u\ \mathrm{d}v) \begin{pmatrix} \frac{\partial \mathbf{x}}{\partial u} \\ \frac{\partial \mathbf{x}}{\partial v} \end{pmatrix} = \mathrm{d}\mathbf{x} \end{aligned}$$

The matrix of partial derivatives is the *Jacobian* of the co-ordinate change, as you should have met in multi-variable calculus. Strictly speaking, $\mathrm{d}\mathbf{x}$ and $\mathrm{d}\mathbf{y}$ are not identical as they feed on tangent vectors with respect to different co-ordinates. We should properly write $\mathrm{d}\mathbf{y} = \mathrm{d}\mathbf{x} \circ \mathrm{d}F$ as a composition of linear functions, where $\mathrm{d}F$ maps tangent vectors in $\mathrm{Span}\{\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\}$ to those in $\mathrm{Span}\{\frac{\partial}{\partial u}, \frac{\partial}{\partial v}\}$.

2.3 The Fundamental Forms

Now that we have differentials, we can start using them to describe surfaces. The required objects a based on a little more algebra.

Definition 2.22. Given 1-forms α, β at $p \in \mathbb{R}^n$, define a symmetric bilinear form $\alpha\beta$ of two vectors $X, Y \in T_p\mathbb{R}^n$ by,

$$\alpha\beta(X,Y) = \frac{1}{2}(\alpha(X)\beta(Y) + \alpha(Y)\beta(X))$$

Lemma 2.23. *$\alpha\beta$ is indeed a symmetric bilinear form on each tangent space $T_p\mathbb{R}^n$.*

Proof. $\alpha\beta(X,Y) = \alpha\beta(Y,X)$ is clear from the formula, hence the product is symmetric. Moreover, fixing X, we see that $\alpha\beta(X,-) : Y \to \alpha\beta(X,Y)$ is linear since both α, β are linear on $T_p\mathbb{R}^n$. By symmetry we have bilinearity. ■

Not only is $\alpha\beta$ a symmetric bilinear form on each tangent space, it is symmetric itself: $\alpha\beta = \beta\alpha$. We write $\alpha^2 = \alpha\alpha$. Clearly

$$\alpha^2(X,Y) = \alpha(X)\alpha(Y)$$

Examples

1. In $\mathbb{R}^2$, let $\alpha = x_1\,\mathrm{d}x_1 - \mathrm{d}x_2$, $\beta = x_1x_2\,\mathrm{d}x_2$, $X = \frac{\partial}{\partial x_1} - x_2\frac{\partial}{\partial x_2}$, $Y = \frac{\partial}{\partial x_2}$. Then,

$$\alpha\beta(X,Y) = \frac{1}{2}\big((x_1+x_2)x_1x_2 + x_1x_2^2\big) = \frac{1}{2}(x_1^2x_2 + 2x_1x_2^2)$$

2. The dot product of tangent vectors in $T_p\mathbb{R}^n$ is given by[9]

$$(\mathrm{d}s)^2 = (\mathrm{d}x_1)^2 + \cdots + (\mathrm{d}x_n)^2$$

I.e. if $X = 3\frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2}$ and $Y = 2\frac{\partial}{\partial x_1} + 5\frac{\partial}{\partial x_2}$, then

$$(\mathrm{d}s)^2(X,Y) = \mathrm{d}x_1^2(X,Y) + \mathrm{d}x_2^2(X,Y) = \mathrm{d}x_1(X)\mathrm{d}x_1(Y) + \mathrm{d}x_2(X)\mathrm{d}x_2(Y) = 6 - 5 = 1$$

That is, $\begin{pmatrix}3\\-1\end{pmatrix} \cdot \begin{pmatrix}2\\5\end{pmatrix} = 1.$

If 1-forms α, β are vector-valued (e.g. $\alpha(X) \in \mathbb{E}^3$) then we can still define symmetric products, but we must specify how the vectors are to be multiplied. For example

$$(\alpha \cdot \beta)(X,Y) = \frac{1}{2}(\alpha(X)\cdot\beta(Y) + \alpha(Y)\cdot\beta(X))$$

requires multiplication of vectors by the dot product, as in the following definitions.

[9] $s = \sqrt{x_1^2 + \cdots + x_n^2}$ is the distance of a point from the origin.

Definition 2.24. Given a local parameterized surface $\mathbf{x} : U \to \mathbb{E}^3$, we define the *first* I *and second* $\mathbb{I}$ *fundamental forms*

$$\mathrm{I} = \mathrm{d}\mathbf{x} \cdot \mathrm{d}\mathbf{x}, \qquad \mathbb{I} = -\mathrm{d}\mathbf{x} \cdot \mathrm{d}\mathbf{U}$$

Here $\mathrm{d}\mathbf{U}$ is the differential of the unit normal field. The fundamental forms are symmetric bilinear forms on each tangent space T_pU.

By the discussion in the previous aside, if $\mathbf{y}(s,t) = \mathbf{x}\big((u(s,t), v(s,t))\big)$, then $\mathrm{d}\mathbf{y} \cdot \mathrm{d}\mathbf{y} = \mathrm{d}\mathbf{x} \cdot \mathrm{d}\mathbf{x}$ so that the first fundamental form is independent of co-ordinates. Similarly the second fundamental form is independent of co-ordinates up to a change of sign if the orientation is reversed.

Before worrying about the meanings of the fundamental forms, we compute a few.

Examples

1. The sphere of radius a.

$$\mathrm{d}\mathbf{x} = a \begin{pmatrix} \cos\phi\cos\theta \\ \cos\phi\sin\theta \\ -\sin\phi \end{pmatrix} \mathrm{d}\phi + a \begin{pmatrix} -\sin\phi\sin\theta \\ \sin\phi\cos\theta \\ 0 \end{pmatrix} \mathrm{d}\theta$$

In this case $\mathrm{I} = a^2\,\mathrm{d}\phi^2 + a^2\sin^2\phi\,\mathrm{d}\theta^2$. Computing $\mathbb{I}$ is very easy here, because the normal vector is simply the scaled position vector:

$$\mathbf{U} = \frac{1}{a}\mathbf{x} \implies \mathrm{d}\mathbf{U} = \frac{1}{a}\,\mathrm{d}\mathbf{x} \implies \mathbb{I} = -\frac{1}{a}\mathrm{I} = -a\,\mathrm{d}\phi^2 - a\sin^2\phi\,\mathrm{d}\theta^2$$

2. If $\mathbf{x}$ is a surface of revolution around the z-axis, recall that

$$\mathrm{d}\mathbf{x} = \begin{pmatrix} f'(u)\cos\phi \\ f'(u)\sin\phi \\ 1 \end{pmatrix} \mathrm{d}u + f(u) \begin{pmatrix} -\sin\phi \\ \cos\phi \\ 0 \end{pmatrix} \mathrm{d}\phi$$

Thus $\mathrm{I} = (1 + f'(u))\mathrm{d}u^2 + f(u)^2\mathrm{d}\phi^2$. The unit normal vector is

$$\mathbf{U} = \frac{\mathbf{x}_u \times \mathbf{x}_\phi}{|\mathbf{x}_u \times \mathbf{x}_\phi|} = \frac{1}{\sqrt{f'(u)^2 + 1}} \begin{pmatrix} -\cos\phi \\ -\sin\phi \\ f'(u) \end{pmatrix}$$

and so

$$\begin{aligned} \mathbb{I} &= -\mathrm{d}\mathbf{x} \cdot \mathrm{d}\mathbf{U} \\ &= -\left(\begin{pmatrix} f'(u)\cos\phi \\ f'(u)\sin\phi \\ 1 \end{pmatrix} \mathrm{d}u + f(u) \begin{pmatrix} -\sin\phi \\ \cos\phi \\ 0 \end{pmatrix} \mathrm{d}\phi \right) \\ &\qquad \cdot \frac{1}{\sqrt{f'(u)^2+1}} \left(\begin{pmatrix} 0 \\ 0 \\ f''(u) \end{pmatrix} \mathrm{d}u + \begin{pmatrix} \sin\phi \\ -\cos\phi \\ 0 \end{pmatrix} \mathrm{d}\phi \right) \\ &= \frac{1}{\sqrt{f'(u)^2+1}} (-f''(u)\mathrm{d}u^2 + f(u)\mathrm{d}\phi^2) \end{aligned}$$

The first fundamental form I transfers the dot product from $\mathbb{E}^3$ to U. Indeed if $\mathrm{d}\mathbf{x}(z_i) = \mathbf{x}_i$, $i = 1, 2$, then $\mathrm{I}(z_1, z_2) = \mathbf{x}_1 \cdot \mathbf{x}_2$. This is in the spirit of transferring calculations from $\mathbb{E}^3$ back to U. The first fundamental form I encodes 'infinitessimal distances' on a surface. For example, the distance s measured between two very close points on the surface of the sphere of radius a is given by,

$$s^2 \approx a^2(\Delta\phi)^2 + a^2 \sin^2 \phi (\Delta\theta)^2$$

where $\Delta\phi, \Delta\theta$ are the small changes in the co-ordinates. If all you want to do on a surface is to measure distances, then all you need to know is the first fundamental form—you don't even need to know the surface!

The second fundamental form measures how the unit normal vector of a surface is changing in comparison to the surface itself. Since $\mathbf{U}$ has unit length, $\mathrm{d}\mathbf{U}|_p (X)$ is orthogonal to $\mathbf{U}$ for any $X \in T_pU$ and is thus a vector in the tangent space to the surface at $\mathbf{x}(p)$. The second fundamental form $\mathbb{I} = -\mathrm{d}\mathbf{x} \cdot \mathrm{d}\mathbf{U}$ thus measures how the unit normal changes relative to the surface.

Curves in surfaces

One of the best ways of understanding what the fundamental forms $\mathrm{I}, \mathbb{I}$ are measuring is to consider curves lying in a surface.

Let $\mathbf{x} : U \to \mathbb{E}^3$ be a local surface. If $z(t) = (u(t), v(t))$ is a curve in U then the composition $t \mapsto \mathbf{x}(z(t))$ describes a curve in $\mathbb{E}^3$ lying in the surface. It is often useful to think of $z(t)$ as being a curve in the surface, even though it is only a curve in U. We will abuse notation by thinking of $\mathbf{x}' = \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(z(t))$ as the velocity vector of the curve in $\mathbb{E}^3$. In a similar way, define $\mathbf{U}' = \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{U}(z(t))$.

Example Consider the curve $z(t) = (\phi(t), \theta(t)) = (2t, 5t)$ on the surface of the unit sphere. In this case,

$$\mathbf{x}(t) = \begin{pmatrix} \sin 2t \cos 5t \\ \sin 2t \sin 5t \\ \cos 2t \end{pmatrix}$$

Lemma 2.25. *Considering the fact that $z'(t) = u'(t)\frac{\partial}{\partial u} + v'(t)\frac{\partial}{\partial v}$, we have*

$$\mathbf{x}' = \mathrm{d}\mathbf{x}(z'), \qquad \mathbf{U}' = \mathrm{d}\mathbf{U}(z')$$

Proof. $\dfrac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(z(t)) = \mathbf{x}_u u' + \mathbf{x}_v v' = \mathrm{d}\mathbf{x}\left(u'(t)\dfrac{\partial}{\partial u} + v'(t)\dfrac{\partial}{\partial v}\right) = \mathrm{d}\mathbf{x}(z')$ ■

If $\mathbf{x}(z(t))$ is a curve in the surface $\mathbf{x}$, then the speed of the curve is

$$v(t) = \sqrt{\mathbf{x}' \cdot \mathbf{x}'} = \sqrt{\mathrm{d}\mathbf{x}(z') \cdot \mathrm{d}\mathbf{x}(z')} = \sqrt{\mathrm{I}(z', z')}$$

We've therefore proved the following:

Theorem 2.26. *Let $\mathbf{x}(z(t))$ be a curve in a surface $\mathbf{x} : U \to \mathbb{E}^3$ with first fundamental form* I. *Then the arc-length of $\mathbf{x}(z(t))$ between $\mathbf{x}(z(a))$ and $\mathbf{x}(z(b))$ is given by*

$$\int_a^b \sqrt{\mathrm{I}(z', z')}\mathrm{d}t$$

Examples

1. Consider the unit sphere with the curve given above $z(t) = (2t, 5t)$. Here $z'(t) = 2\frac{\partial}{\partial\phi} + 5\frac{\partial}{\partial\theta}$. Recalling that the first fundamental form of the unit sphere is $\mathrm{I} = \mathrm{d}\phi^2 + \sin^2\phi\,\mathrm{d}\theta^2$, we have

   $$\mathrm{I}(z', z') = 4 + 25\sin^2 5t$$

   The arc-length of the curve over $a \le t \le b$ is therefore given by[10]

   $$\int_a^b \sqrt{4 + 25\sin^2 5t}\,\mathrm{d}t$$

   Generally for a curve $z(t) = (\phi(t), \theta(t))$ on a sphere, the arc-length between $t = a, b$ is given by,

   $$\int_a^b \sqrt{(\phi')^2 + \sin^2\phi(\theta')^2}\,\mathrm{d}t$$

2. Suppose that a surface satisfying $y > 0$ has first fundamental form $\mathrm{I} = \frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2}$. Find the arc-length over the circular path $z(t) = (\cos t, \sin t)$ between t_0 and t_1.
   Here $z'(t) = -\sin t\frac{\partial}{\partial x} + \cos t\frac{\partial}{\partial y}$ so that $\mathrm{I}(z', z') = \frac{1}{\sin^2 t}$. Thus the arc-length is

   $$\int_{t_0}^{t_1} \frac{1}{\sin t}\mathrm{d}t = \left[-\ln\frac{1 + \cos t}{\sin t}\right]_{t_0}^{t_1} = \ln\frac{\sin t_1(1 + \cos t_0)}{\sin t_0(1 + \cos t_1)}$$

   Notice that as $t_0 \to 0^+$ or as $t_1 \to \pi^-$ the arc-length becomes infinite.[11]

The examples show the advantages of our approach:

- If you know I, then you can *change the curve* lying in a given surface and still compute the arc-length with minimal effort. Before this, you would have had to go back to the start: compute $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(z(t))$ again, then it's length... Since applications often involve studying all curves simultaneously, this approach is essential for being able to quickly compare the outcomes.
- You don't need to know the surface explicitly, all you need is I. Example 2 is a description of the Poincare half-plane model of hyperbolic space: it is simply a domain U (here the upper half-plane) with an abstract first fundamental form. There is no surface $\mathbf{x} : U \to \mathbb{E}^3$ and so no $\mathbb{I}$ to worry about.

Interpreting $\mathbb{I}$ Suppose you were skateboarding in a bowl: how much effort are your legs going to have to make in order to keep you standing up? The second fundamental form along a curve measures exactly this.

Theorem 2.27. *For a curve $\mathbf{x}(z(t))$ lying in a surface $\mathbf{x}$ with unit normal $\mathbf{U}$ and second fundamental form $\mathbb{I}$,*

$$\mathbf{U} \cdot \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathbf{x}(z(t)) = \mathbb{I}(z', z').$$

[10] Don't try to compute this elliptic integral explicitly!

[11] You may see this example again as *hyperbolic space*.

Proof. Write $\mathbf{x}' = \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(z(t))$ as before. Since the tangent vector to the curve is always perpendicular to the unit normal, we have

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{x}' \cdot \mathbf{U}) = \mathbf{x}'' \cdot \mathbf{U} + \mathbf{x}' \cdot \mathbf{U}' = \mathbf{x}'' \cdot \mathbf{U} + \mathrm{d}\mathbf{x}(z') \cdot \mathrm{d}\mathbf{U}(z') \\ &= \mathbf{x}'' \cdot \mathbf{U} - \mathbb{I}(z', z') \end{aligned}$$

∎

Theorem 2.27 says that the second fundamental form along a curve measures the *acceleration* in the normal direction to the surface that is required in order to keep the curve lying in the surface.

$\mathrm{I}, \mathbb{I}$ in co-ordinates

Suppose that u, v are co-ordinates on $U \subseteq \mathbb{R}^2$. We can write the fundamental forms of a surface $\mathbf{x}$ in terms of easily calculated derivatives.

Proposition 2.28. *Define functions E, F, G and l, m, n on $\mathbf{x} : U \to \mathbb{E}^3$ by*

$$\begin{aligned} E &= \mathbf{x}_u \cdot \mathbf{x}_u & F &= \mathbf{x}_u \cdot \mathbf{x}_v & G &= \mathbf{x}_v \cdot \mathbf{x}_v \\ l &= \mathbf{x}_{uu} \cdot \mathbf{U} = -\mathbf{x}_u \cdot \mathbf{U}_u & m &= \mathbf{x}_{uv} \cdot \mathbf{U} = -\mathbf{x}_u \cdot \mathbf{U}_v & n &= \mathbf{x}_{vv} \cdot \mathbf{U} = -\mathbf{x}_v \cdot \mathbf{U}_v \end{aligned}$$

Then

$$\mathrm{I} = E\mathrm{d}u^2 + 2F\mathrm{d}u\mathrm{d}v + G\mathrm{d}v^2 \quad \textit{and} \quad \mathbb{I} = l\mathrm{d}u^2 + 2m\mathrm{d}u\mathrm{d}v + n\mathrm{d}v^2$$

These should be clear: I is just the formula $\mathrm{d}\mathbf{x} \cdot \mathrm{d}\mathbf{x}$, while the expressions for $\mathbb{I}$ come from differentiating $\mathbf{x}_u \cdot \mathbf{U} = 0 = \mathbf{x}_v \cdot \mathbf{U}$.

Examples

1. Consider the graph of a function $z = f(x, y)$. Parameterizing in the usual way gives us,

$$\mathbf{x}_u = \begin{pmatrix} 1 \\ 0 \\ f_u \end{pmatrix}, \quad \mathbf{x}_v = \begin{pmatrix} 0 \\ 1 \\ f_v \end{pmatrix}, \quad \mathbf{U} = \frac{1}{\sqrt{1 + f_u^2 + f_v^2}} \begin{pmatrix} -f_u \\ -f_v \\ 1 \end{pmatrix},$$

$$E = 1 + f_u^2, \quad F = f_u f_v, \quad G = 1 + f_v^2,$$

$$l = \frac{f_{uu}}{\sqrt{1 + f_u^2 + f_v^2}}, \quad m = \frac{f_{uv}}{\sqrt{1 + f_u^2 + f_v^2}}, \quad n = \frac{f_{vv}}{\sqrt{1 + f_u^2 + f_v^2}}$$

Combining these gives us,

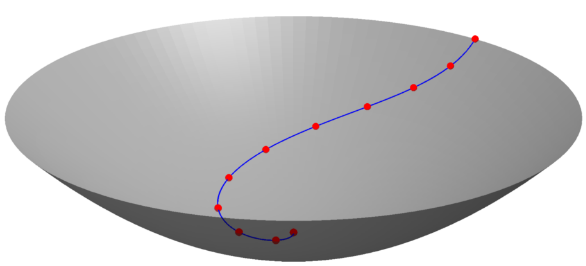
$$\begin{aligned} \mathrm{I} &= (1 + f_u^2)\,\mathrm{d}u^2 + 2f_u f_v\,\mathrm{d}u\mathrm{d}v + (1 + f_v^2)\,\mathrm{d}v^2, \\ \mathbb{I} &= \frac{1}{\sqrt{1 + f_u^2 + f_v^2}} \left\{ f_{uu}\,\mathrm{d}u^2 + 2f_{uv}\,\mathrm{d}u\mathrm{d}v + f_{vv}\,\mathrm{d}v^2 \right\} \end{aligned}$$

2. A skater follows a path parameterized by $z(t) = (r(t), \theta(t)) = (1 - t, 4t^2)$ in the paraboloidal bowl $z = \frac{1}{2}r^2$. Compute:

   (a) The fundamental forms of the surface:

   (b) The speed of the skater at time t.

   (c) The normal acceleration experienced by the skater at time t.



The surface can be parameterized $\mathbf{x}(r,\theta) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \\ \frac{1}{2}r^2 \end{pmatrix}$ which yields

$$\begin{aligned}
&\mathbf{x}_r = \begin{pmatrix} \cos\theta \\ \sin\theta \\ r \end{pmatrix}, \quad \mathbf{x}_\theta = \begin{pmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{pmatrix}, \quad \mathbf{U} = \frac{1}{\sqrt{1+r^2}}\begin{pmatrix} -r\cos\theta \\ -r\sin\theta \\ 1 \end{pmatrix}, \\
&E = 1+r^2, \quad F = 0, \quad G = r^2, \\
&l = \frac{1}{\sqrt{1+r^2}}, \quad m = 0, \quad n = \frac{r^2}{\sqrt{1+r^2}}, \\
&\mathrm{I} = (1+r^2)\,\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2, \qquad \mathbb{I} = \frac{1}{\sqrt{1+r^2}}(\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2)
\end{aligned}$$

For the skater's path, we have $z'(t) = -\frac{\partial}{\partial r} + 8t\frac{\partial}{\partial\theta}$. Therefore

$$\mathrm{I}(z',z') = (1+(1-t)^2) + 64t^2(1-t)^2$$

The skater's speed is therefore

$$v(t) = \sqrt{\mathrm{I}(z',z')} = \sqrt{1+(64t^2+1)(1-t)^2}$$

The second fundamental form along the path is

$$\mathbb{I}(z',z') = \frac{1}{\sqrt{1+(1-t)^2}}(1+64t^2(1-t)^2)$$

Theorem 2.27 says that this is the normal acceleration of the skater and thus proportional (via Newton's second law $F = ma$) to the force experienced by the skater.

2.4 Principal Curvatures

The idea of the next section is to think about finding directions on a surface with special properties: this amounts to finding co-ordinates with respect to which the fundamental forms look particularly simple. The most obvious way to may the forms simple is to insist that there be no dudv term.

Definition 2.29. Co-ordinates u, v on U are *orthogonal* for a surface $\mathbf{x} : U \to \mathbb{E}^3$ if the first fundamental form is *diagonal*: that is

$$\mathrm{I} = E\,\mathrm{d}u^2 + G\,\mathrm{d}v^2$$

where there is no dudv term. Otherwise said, $\mathbf{x}_u \cdot \mathbf{x}_v = 0$ so the tangent vectors generated by the co-ordinate vector fields $\frac{\partial}{\partial u}, \frac{\partial}{\partial v}$ are always orthogonal.
In addition, u, v are *curvature-line* co-ordinates if the second fundamental form is diagonal:

$$\mathbb{I} = l\,\mathrm{d}u^2 + n\,\mathrm{d}v^2$$

While the meaning of *orthogonal* is clear, to understand why we use the term *curvature-line* will need a little work.

Examples

1. Recall the sphere using spherical polar co-ordinates: since both $\mathrm{I}, \mathbb{I}$ are diagonal, these are curvature line co-ordinates.[12]

2. The graph of a function $z = f(x, y)$ has its usual parameterization $(u, v) = (x, y)$ in orthogonal co-ordinates if and only if $f(x, y)$ is a function only of one of the variables ($f_x = 0$ or $f_y = 0$). In this case the co-ordinates are also curvature-line. The surface is necessarily ruled, with a straight line in the surface parallel to either the x or the y axis.

3. A surface of revolution parameterized using (u, ϕ) where u is the distance along the axis of rotation and ϕ the angle round the axis, has

$$\mathrm{I} = \mathrm{I} = (1 + f'(u))\mathrm{d}u^2 + f(u)^2\mathrm{d}\phi^2, \qquad \mathbb{I} = \frac{1}{\sqrt{1 + f'(u)^2}}(-f''(u)\mathrm{d}u^2 + f(u)\mathrm{d}\phi^2)$$

These are curvature-line co-ordinates.

Finding Curvature Directions: A Little Linear Algebra

Finding curvature-line co-ordinates is difficult in general. It starts with finding curvature directions at a point, which amounts to finding directions with respect to which both I and $\mathbb{I}$ are *diagonal*. First we need some background linear algebra to extend the idea of eigenvalues and eigenvectors.

Definition 2.30. Let A, B be two $n \times n$ matrices. The *eigenvalues* of B with respect to A are the roots of the polynomial

$$\det(B - \lambda A) = 0$$

A vector $\mathbf{v}$ is an *eigenvector* of B with respect to A and with eigenvalue λ if $(B - \lambda A)\mathbf{v} = \mathbf{0}$

Theorem 2.31. *Suppose that A is an $n \times n$ symmetric, positive definite[13] matrix and that B is $n \times n$ symmetric. Then the eigenvalues of B with respect to A are real, and there exists a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbb{R}^n$ such that each $\mathbf{v}_k$ is an eigenvector of B with respect to A.*

You should skip the proof if you haven't studied the spectral theorem in linear algebra.

Proof. The spectral theorem states that there exists an orthogonal basis $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of $\mathbb{R}^n$ of eigenvectors of A. In particular, since A is positive definite, all of the eigenvalues are positive and we may choose $\mathbf{x}_1, \ldots, \mathbf{x}_n$ to be orthonormal with respect to A: i.e.

$$\mathbf{x}_i^T A \mathbf{x}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

With respect to the basis $\mathbf{x}_1, \ldots, \mathbf{x}_n$ the matrix of A is the identity matrix.
Now let $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be the matrix with column vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then $X^T A X = I$, whence

$$A = (X^T)^{-1} X^{-1}$$

[12] This is something of a cheat: *every* pair of orthogonal co-ordinates on the sphere is curvature-line (exercise).
[13] $\mathbf{v}^T A \mathbf{v} \geq 0$ for all $\mathbf{v}$, with equality if and only if $\mathbf{v} = \mathbf{0}$. This says that $(\mathbf{v}, \mathbf{w}) := \mathbf{v}^T A \mathbf{w}$ is an *inner product*.

It follows that

$$\det(B - \lambda A) = \det\big((X^T)^{-1}(X^T BX - \lambda I)X^{-1}\big) = 0 \iff \det(X^T BX - \lambda I) = 0$$

$X^T BX$ is a symmetric matrix and so has real eigenvalues $\lambda_1, \ldots, \lambda_n$, and n distinct orthogonal eigenvectors $\mathbf{w}_1, \ldots, \mathbf{w}_n$. Define $\mathbf{v}_k = X\mathbf{w}_k$: these form a basis of $\mathbb{R}^n$ such that $\mathbf{v}_k$ is an eigenvector of B with respect to A with eigenvalue λ_i.

■

Example Let $A = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ 1 & 3 \end{pmatrix}$. Then the eigenvalues of A are $7 \pm \sqrt{48}$ so that A is positive definite. Now

$$\det(B - \lambda A) = \begin{vmatrix} -2\lambda & 1 - 3\lambda \\ 1 - 3\lambda & 3 - 5\lambda \end{vmatrix} = \lambda^2 - 1 = 0 \iff \lambda = \pm 1$$

The eigenvalues of B with respect to A are thus ± 1. When $\lambda = 1$ we have

$$(B - A)\mathbf{v}_1 = \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix} \mathbf{v}_1 = \mathbf{0} \iff \mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

When $\lambda = -1$ we obtain $\mathbf{v}_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$.

The eigenvectors of B with respect to A are $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ with eigenvalues $1, -1$ respectively.

Diagonalizing the Fundamental Forms Suppose we have co-ordinates u, v on U, then for each $p \in U$ we have a basis $\frac{\partial}{\partial u}\Big|_p, \frac{\partial}{\partial v}\Big|_p$, with respect to which $\mathrm{I}, \mathrm{I\!I}$ are symmetric bilinear forms with matrices

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}, \qquad \begin{pmatrix} l & m \\ m & n \end{pmatrix}$$

respectively. Think about what this means: if $Y = y_1 \frac{\partial}{\partial u}\Big|_p + y_2 \frac{\partial}{\partial v}\Big|_p$ and $Z = z_1 \frac{\partial}{\partial u}\Big|_p + z_2 \frac{\partial}{\partial v}\Big|_p$ are tangent vectors at p, then

$$\mathrm{I}(Y, Z) = (y_1\ y_2) \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \qquad \mathrm{I\!I}(Y, Z) = (y_1\ y_2) \begin{pmatrix} l & m \\ m & n \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

Indeed the matrix of I is positive definite, since

$$\mathrm{I}(Y, Y) = \mathrm{d}x(Y) \cdot \mathrm{d}x(Y) = |\mathrm{d}x(Y)|^2 \geq 0$$

According to Theorem 2.31, there is a basis

$$X_1|_p = X_1^{(1)} \frac{\partial}{\partial u}\Big|_p + X_1^{(2)} \frac{\partial}{\partial v}\Big|_p, \qquad X_2|_p = X_2^{(1)} \frac{\partial}{\partial u}\Big|_p + X_2^{(2)} \frac{\partial}{\partial v}\Big|_p$$

of T_pU and scalars k_1, k_2 such that $X_i|_p$ is an eigenvector of $\mathbb{I}$ with respect to I with eigenvalue k_i. Otherwise said,

$$\left(\begin{pmatrix} l & m \\ m & n \end{pmatrix} - k_i \begin{pmatrix} E & F \\ F & G \end{pmatrix}\right) \begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \end{pmatrix} = 0, \quad \text{for } i = 1, 2$$

Since the fundamental forms depend smoothly on the position $p \in U$, there is nothing stopping us from considering k_1, k_2 as smooth functions $U \to \mathbb{R}$ and X_1, X_2 as vector fields on U.

Definition 2.32. If $\mathbf{x} : U \to \mathbb{E}^3$ is a surface, the functions $k_1, k_2 : U \to \mathbb{R}$ are the *principal curvatures* of $\mathbf{x}$. The vector fields X_1, X_2 are the *principal curvature directions*.
A point $\mathbf{x}(p)$ is *umbilic* if $k_1(p) = k_2(p)$. At an umbilic point, all tangent directions are principal curvature directions.

Corollary 2.33. *Suppose that X_1, X_2 are the principal curvature directions on $\mathbf{x}$. Then, at any non-umbilic point,* $\mathrm{I}(X_1, X_2) = 0 = \mathbb{I}(X_1, X_2)$.

Proof. Suppose that $\mathbf{x}(p)$ is non-umbilic and fix a basis of T_pU so that the matrices of $\mathrm{I}, \mathbb{I}$ are the symmetric matrices A, B in Theorem 2.31. Viewing $X_1 = \begin{pmatrix} X_1^{(1)} \\ X_1^{(2)} \end{pmatrix}, X_2 = \begin{pmatrix} X_2^{(1)} \\ X_2^{(2)} \end{pmatrix}$ as column vectors with respect to the basis of T_pU, we see that $BX_i = k_iAX_i$ for each i. It is then straightforward to compute:

$$\begin{aligned} \mathbb{I}(X_1, X_2) &= X_1^T B X_2 = X_1^T (k_2 A X_2) = k_2 \mathrm{I}(X_1, X_2) \\ \| \quad & \\ \mathbb{I}(X_2, X_1) &= X_2^T B X_1 = X_2^T (k_1 A X_1) = k_1 \mathrm{I}(X_1, X_2) \end{aligned}$$

Since $k_1 \neq k_2$ it follows that $\mathrm{I}(X_1, X_2) = 0 = \mathbb{I}(X_1, X_2)$. ∎

Examples

1. The sphere of radius a has fundamental forms related by $\mathbb{I} = -\frac{1}{a}\mathrm{I}$. It follows that $k_1(p) = k_2(p) = -\frac{1}{a}$ at all points: *every point on the sphere is umbilic!*

2. Suppose that a surface[14] has fundamental forms

$$\mathrm{I} = u^2\,\mathrm{d}u^2 + v^2\,\mathrm{d}v^2, \qquad \mathbb{I} = u^2\,\mathrm{d}u^2 + 2uv\,\mathrm{d}u\mathrm{d}v + v^2\,\mathrm{d}v^2$$

Then the principal curvatures are the solutions λ of the equation

$$\det\left(\begin{pmatrix} u^2 & uv \\ uv & v^2 \end{pmatrix} - \lambda \begin{pmatrix} u^2 & 0 \\ 0 & v^2 \end{pmatrix}\right) = 0$$

Multiplying this out, we obtain

$$u^2v^2(1-\lambda)^2 - u^2v^2 = 0 \iff (1-\lambda)^2 = 1 \iff \lambda = 0,\ 2$$

whence the principal curvatures are $k_1 = 0, k_2 = 2$. We can also compute the eigenvectors:

[14]For example $\mathbf{x}(u,v) = \frac{1}{2}\begin{pmatrix} \cos(\frac{u^2+v^2}{\sqrt{2}}) \\ \sin(\frac{u^2+v^2}{\sqrt{2}}) \\ \frac{u^2-v^2}{\sqrt{2}} \end{pmatrix}$ though it is unimportant.

$$k_1 = 0 \quad \begin{pmatrix} u^2 & uv \\ uv & v^2 \end{pmatrix} \begin{pmatrix} X_1^{(1)} \\ X_1^{(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} X_1^{(1)} \\ X_1^{(2)} \end{pmatrix} = \begin{pmatrix} v \\ -u \end{pmatrix}$$

$$k_2 = 2 \quad \begin{pmatrix} -u^2 & uv \\ uv & -v^2 \end{pmatrix} \begin{pmatrix} X_2^{(1)} \\ X_2^{(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} X_2^{(1)} \\ X_2^{(2)} \end{pmatrix} = \begin{pmatrix} v \\ u \end{pmatrix}$$

We therefore have curvature directions

$$X_1 = v\frac{\partial}{\partial u} - u\frac{\partial}{\partial v}, \qquad X_2 = v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}$$

You can easily checked that $\mathrm{I}(X_1, X_1) = 2u^2v^2$, etc., and that the matrices of I and $\mathbb{I}$ with respect to this basis are

$$[\mathrm{I}] = 2u^2v^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad [\mathbb{I}] = 2u^2v^2 \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

in which the principal curvatures $k_1 = 0$ and $k_2 = 2$ are clearly visible.

In fact we can get more than just curvature directions, we can always locally assume the existence of curvature *co-ordinates.*

Theorem 2.34 (Frobenius: easy version)**.** *There exist curvature-line co-ordinates on a neighborhood of any non-umbilic point.*

A more thorough discussion is given below in the aside. Two things are important for us:

- We can only guarantee the existence of curvature-line co-ordinates *locally.*
- If s_1, s_2 are curvature-line co-ordinates for $\mathbf{x}$, and X_1, X_2 are curvature directions, then $X_1 = f_1\frac{\partial}{\partial s_1}$ and $X_2 = f_2\frac{\partial}{\partial s_2}$ for some functions f_1, f_2. That is, the vector fields $\frac{\partial}{\partial s_i}$ and X_i are *parallel* but not necessarily equal!

Example, mark II In the above example, $s = u^2 - v^2$ and $t = u^2 + v^2$ are curvature line co-ordinates. You should check the following.

- The co-ordinate vector fields are

$$\frac{\partial}{\partial s} = \frac{1}{u}\frac{\partial}{\partial u} - \frac{1}{v}\frac{\partial}{\partial v} = \frac{1}{uv}X_1 \quad \text{and} \quad \frac{\partial}{\partial t} = \frac{1}{u}\frac{\partial}{\partial u} + \frac{1}{v}\frac{\partial}{\partial v} = \frac{1}{uv}X_2$$

  which are parallel but not equal to X_1, X_2.

- The fundamental forms really are diagonalized by s, t:

$$\mathrm{I} = 2\mathrm{d}s^2 + 2\mathrm{d}t^2 \qquad \mathbb{I} = \mathrm{d}t^2$$

We finish by considering totally umbilic surfaces: the proof is left to the homework.

Theorem 2.35. *Suppose that all points of* $\mathbf{x} : U \to \mathbb{E}^3$ *are umbilic. Then* $\mathbf{x}$ *is part of a plane or a sphere.*

Aside: The Frobenius Theorem and Curvature-line co-ordinates

As seen above, we can find curvature directions X_1, X_2, using linear algebra. What we'd really like are curvature *co-ordinates*: functions $s, t : U \to \mathbb{R}$ such that $\frac{\partial}{\partial s} = X_1$ and $\frac{\partial}{\partial t} = X_2$. Unfortunately this is usually too much too ask! A necessary condition is immediate: since mixed partial derivatives should be equal, we'd need

$$X_1X_2 = \frac{\partial^2}{\partial s \partial t} = \frac{\partial^2}{\partial t \partial s} = X_2X_1$$

That this condition is locally sufficient, and the consequent existence of curvature co-ordinates is given by the following result, which provides a deep link between differential geometry and the theory of partial differential equations.

Theorem 2.36 (Frobenius: hard version)**.** 1. *Suppose that X, Y are linearly independent vector fields on an open set U containing p.*

   (a) If the Lie bracket $[X, Y] = XY - YX$ of the vector fields is zero, then there exists a neighborhood $V \subseteq U$ of p and functions $s, t : V \to \mathbb{R}$ such that

   $$X = \frac{\partial}{\partial s}, \qquad Y = \frac{\partial}{\partial t}$$

   (b) There exists a neighborhood $W \subseteq U$ of p and functions $f, g : W \to \mathbb{R}$ such that $[fX, gY] = 0$ on W. Part (a) then says there exists $V \subseteq W$ and co-ordinates $s, t : W \to \mathbb{R}$ such that $\frac{\partial}{\partial s} = fX$ and $\frac{\partial}{\partial t} = gY$ are parallel to X, Y.

2. *If $\mathbf{x}(p)$ is a non-umbilic point of a surface $\mathbf{x} : U \to \mathbb{E}^3$, then there exist curvature-line co-ordinates for $\mathbf{x}$ on some neighborhood of p. Indeed if X, Y are curvature directions such that $[X, Y] = 0$, let α, β be the dual 1-forms to X, Y, then (locally) $\alpha = \mathrm{d}s$ and $\beta = \mathrm{d}t$.*

The proof is a little too technical for this course: at its heart is the usual existence and uniqueness theorem for differential equations and a bit more technology on forms than we've yet developed. It is moreover completely generalizable to arbitrary dimensions.

Example, mark III Let's revisit our example again! Frobenius' Theorem actually tells us how to find curvature co-ordinates. We give a sketch, though in practice this involves invoking the existence of solutions to certain differential equations!

1. Hunt for functions f, g such that the vector fields $\widetilde{X}_1 = fX_1, \widetilde{X}_2 = gX_2$ satisfy the Lie bracket condition:

   $$[\widetilde{X}_1, \widetilde{X}_2] = 0$$

   Multiply this out (it's messy) to see that it is equivalent to being able to solve

   $$\begin{cases} v\frac{\partial f}{\partial u} + u\frac{\partial f}{\partial v} = -f\frac{u^2+v^2}{uv} \\ v\frac{\partial g}{\partial u} - u\frac{\partial g}{\partial v} = -g\frac{u^2-v^2}{uv} \end{cases} \tag{$*$}$$

   That this system can be solved is essentially the theorem. Indeed you can check that $f = g = \frac{1}{uv}$ does the job!

2. The co-ordinate direction vector fields

$$\widetilde{X}_1 = \frac{1}{u}\frac{\partial}{\partial u} - \frac{1}{v}\frac{\partial}{\partial v} \quad \text{and} \quad \widetilde{X}_2 = \frac{1}{u}\frac{\partial}{\partial u} + \frac{1}{v}\frac{\partial}{\partial v}$$

satisfy the Lie bracket condition. The dual basis 1-forms to these fields are easy to find using linear algebra:

$$\alpha = \frac{1}{2}(u\,\mathrm{d}u - v\,\mathrm{d}v) \quad \text{and} \quad \beta = \frac{1}{2}(u\,\mathrm{d}u + v\,\mathrm{d}v)$$

3. Observe that $\alpha = \mathrm{d}s$ and $\beta = \mathrm{d}t$ where $s = u^2 - v^2$ and $t = u^2 + v^2$. We therefore have curvature line co-ordinates s and t.

There are sometimes short-cuts to this procedure if you are feeling lucky, but in the abstract this is what is needed. For instance, we could have gone straight to α, β by noting that we have to have $\alpha(X_2) = 0 = \beta(X_1)$: by luck we would observe that α, β are the exterior derivatives of some functions and we're done! In practice, all this trick does is transform the problem to an equivalent system of PDE. In general, given independent vector fields X_1, X_2 on $U \subseteq \mathbb{R}^2$, we know that there exist functions a, b such that

$$[X_1, X_2] = aX_1 + bX_2$$

Hunting for functions f, g such that $[fX_1, gX_2] = 0$ is equivalent to f, g satisfying the PDEs

$$X_2[f] = af \qquad X_1[g] = -bg$$

This is the system $(*)$ in our example, and the problem of finding curvature-line co-ordinates essentially boils down to solving this. The theorem assures us that solutions *exist*, so we can always *assume* the existence of local curvature co-ordinates. Finding them explicitly may be essentially impossible!

2.5 Gauss and mean curvature

The principal curvatures of a surface are usually rearranged in two combinations which have a deeper geometric meaning.

Definition 2.37. Suppose that $\mathbf{x} : U \to \mathbb{E}^3$ has principal curvatures k_1, k_2. The *mean curvature* H and *Gauss curvature* K of $\mathbf{x}$ are defined respectively by

$$H = \tfrac{1}{2}(k_1 + k_2) \qquad K = k_1 k_2$$

Example The sphere of radius a (with outward pointing normal) has $k_1 = k_2 = -\frac{1}{a}$. Thus $H = -\frac{1}{a}$ and $K = \frac{1}{a^2}$.

Proposition 2.38. *In co-ordinates, the mean and Gauss curvatures are given by,*

$$H = \frac{lG + nE - 2mF}{2(EG - F^2)}, \qquad K = \frac{ln - m^2}{EG - F^2}$$

Proof. In the basis $\frac{\partial}{\partial u}, \frac{\partial}{\partial v}$, I and II have matrices $\left(\begin{smallmatrix} E & F \\ F & G \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} l & m \\ m & n \end{smallmatrix}\right)$ respectively. Now expand the determinant,

$$\det\left(\begin{pmatrix} l & m \\ m & n \end{pmatrix} - \lambda \begin{pmatrix} E & F \\ F & G \end{pmatrix}\right) = 0$$

to obtain

$$(EG - F^2)\lambda^2 - (lG + nE - 2mF)\lambda + (ln - m^2) = 0$$

K is the product and H half the sum of the roots of this equation. ∎

Examples

1. A surface of revolution found by rotating the curve $x = f(z)$ around the z-axis, and parameterized in the usual way, has principal curvatures

   $$k_1 = -\frac{f''(u)}{(1 + f'(u)^2)^{3/2}}, \qquad k_2 = \frac{1}{f(u)(1 + f'(u)^2)^{1/2}}$$

   whence

   $$H = \frac{1 + f'(u)^2 - f(u)f''(u)}{2f(u)(1 + f'(u)^2)^{3/2}}, \qquad K = -\frac{f''(u)}{f(u)(1 + f'(u)^2)^2}$$

   The simplest example would be a cylinder, where the radius $f(u)$ is constant. We immediately see that a cylinder has zero Gauss curvature and mean curvature half the radius.

2. If $z = f(x, y)$ is the graph of a function then, parameterizing in the usual way, we have

   $$H = \frac{f_{vv}(1 + f_u^2) + f_{uu}(1 + f_v^2) - 2f_u f_v f_{uv}}{2(1 + f_u^2 + f_v^2)^{3/2}}, \qquad K = \frac{f_{uu}f_{vv} - f_{uv}^2}{(1 + f_u^2 + f_v^2)^2}$$

The Gauss and mean curvatures are extremely important objects, the Gauss curvature especially so. In particular, note that these are invariants of a surface:

- The Gauss curvature is independent of parameterization, while the mean curvature is invariant under orientation-preserving reparameterizations.

Many of the applications of these curvatures will have to wait until another course: here are a few to whet your appetite.

Definition 2.39. *Constant mean curvature* (CMC) surfaces are surfaces for which H is constant. A *minimal surface* is a surface for which H is identically zero.

CMC surfaces have many useful applications and appear throughout mathematics. Minimal surfaces are so-called because they minimize areas. For instance, suppose that $\mathbf{y}$ is a closed curve in $\mathbb{E}^3$. Imagine filling in the curve so that you have some sort of warped disc surface whose edge is $\mathbf{y}$. There are infinitely many choices, but the one with minimal area will have $H = 0$. The same holds for surfaces joining a pair of closed curves. As such, minimal surfaces are often the solutions to many practical questions: for example, the shape assumed by a soap film is typically a minimal surface as this will be the surface which minimizes the total tension in the film.

It can be shown that constant Gauss curvature surfaces are spheres if $K > 0$, and planes, cones or cylinders when $K = 0$.

Definition 2.40. A *pseudosphere* is a surface with constant negative Gauss curvature.

It is an exercise to show that the surface of revolution defined by a tractrix[15] defines a pseudosphere with $K = -1$, and that the surface obtained by rotating a catenary $f(u) = a\cosh(a^{-1}u + c)$ is a minimal surface.

Perhaps the most important fact regarding Gauss curvature is that it *depends only on the first fundamental form.* This is Gauss' famous Theorem Egregium, whose proof will also have to wait for a little more technology. It is therefore entirely detectable to any inhabitant of a surface who can only measure distance and angle, but is incapable of understanding the normal direction. This concept is important when extended to several dimensions and the theory of relativity and spacetime: we cannot travel 'normal' to the Universe in order to measure curvature! Indeed the Gauss curvature is the 2-dimensional version of the more general *Riemann curvature tensor* which rules the discussion of relativity.

2.6 Power series expansion and Euler's theorem

We can analyze the properties of a surface at a fixed point in terms of the curvatures described in the previous sections. Essentially we view a surface as a graph of a function nearby a given point. The analysis will tell us what happens at that point and approximately nearby, but not globally.

Theorem 2.41. *Let s be a point on a regular surface S. Choose axes such that s is the origin and the (x_1, x_2)-plane is tangent to the surface. The surface is then locally the graph of a function $x_3 = f(x_1, x_2)$. Parameterizing in the usual way we have that* at the origin,

$$\mathrm{I}_{(0,0)} = \mathrm{d}u^2 + \mathrm{d}v^2$$
$$\mathrm{I\!I}_{(0,0)} = f_{uu}(0,0)\mathrm{d}u^2 + 2f_{uv}(0,0)\mathrm{d}u\mathrm{d}v + f_{uv}(0,0)\mathrm{d}v^2$$

Proof. That the surface is locally a graph is essentially Theorem 2.21 after rotation and translation of the surface to the origin in such a way that $\mathbf{U}|_s$ becomes the vertical basis vector $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Parameterizing in the usual way, we take (u, v) to be the standard orthonormal co-ordinates on the tangent plane T_sS such that $(0,0)$ corresponds to s and obtain a parameterization $\mathbf{x}(u,v) = \begin{pmatrix} u \\ v \\ f(u,v) \end{pmatrix}$. Section 2.3 shows the fundamental forms for such a surface:

$$\mathrm{I} = (1 + f_u)\,\mathrm{d}u^2 + 2f_u f_v\,\mathrm{d}u\mathrm{d}v + (1 + f_v^2)\,\mathrm{d}v^2$$
$$\mathrm{I\!I} = \frac{1}{\sqrt{1 + f_u^2 + f_v^2}}\left\{f_{uu}\,\mathrm{d}u^2 + 2f_{uv}\,\mathrm{d}u\mathrm{d}v + f_{vv}\,\mathrm{d}v^2\right\}$$

from which evaluation at $(0,0)$ is immediate. ■

Note that we're only computing the fundamental forms *at the origin*: in particular, although the co-ordinates u, v will extend at least nearby on the surface, the first fundamental form need not be diagonal anywhere except at the origin.

[15] Recall the definition from our discussion of curves.

Definition 2.42. The matrix of $\mathbb{I}(0,0)$ with respect to the co-ordinates (u,v) is the *Hessian* of f at $(0,0)$:given by

$$\operatorname{Hess} f(0,0) = \begin{pmatrix} f_{uu} & f_{uv} \\ f_{uv} & f_{vv} \end{pmatrix}\Big|_{(0,0)}$$

It is immediate that the Gauss and mean curvatures satisfy

$$K(s) = \det \operatorname{Hess} f(0,0) \quad \text{and} \quad H(s) = \frac{1}{2} \operatorname{tr} \operatorname{Hess} f(0,0)$$

Suppose now that we rotate the (x_1,x_2)-plane such that the axes point in the principal directions at $(0,0)$. Then $\mathbb{I}(0,0)$ is also diagonal ($f_{uv}(0,0) = 0$). We have the following result.

Theorem 2.43. *Suppose that* $\mathbf{x}(u,v) = \left(\begin{smallmatrix} u \\ v \\ f(u,v) \end{smallmatrix}\right)$ *is a local surface defined as a graph* $z = f(x,y)$*, such that the* (x,y)*-plane is tangent at* $\mathbf{x}(0,0)$ *and both* $\mathrm{I}(0,0)$ *and* $\mathbb{I}(0,0)$ *are diagonal. Then the principal curvatures at* $(0,0)$ *are given by*

$$k_1 = f_{uu}(0,0) \quad \textit{and} \quad k_2 = f_{vv}(0,0)$$

Furthermore, the Taylor expansion of $x_3 = f(x_1,x_2)$ *about the origin is*

$$Tx_3 = \frac{k_1}{2}x_1^2 + \frac{k_2}{2}x_2^2 + \textit{ higher order terms}$$

Proof. The principal curvatures are clear since both fundamental forms are diagonal. For the latter, recall Taylor series for multi-valued functions:

$$\begin{aligned} Tx_3 &= f(0,0) + (x_1\ x_2)\ \nabla f|_{(0,0)} + \frac{1}{2}(x_1\ x_2) \operatorname{Hess} f|_{(0,0)} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \text{ higher order terms} \\ &= f(0,0) + (x_1\ x_2) \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{2}(x_1\ x_2) \begin{pmatrix} f_{uu} & f_{uv} \\ f_{uv} & f_{vv} \end{pmatrix}\Big|_{(0,0)} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \text{ higher order terms} \\ &= \frac{1}{2}(x_1\ x_2) \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \text{ higher order terms} \\ &= \frac{k_1}{2}x_1^2 + \frac{k_2}{2}x_2^2 + \text{ higher order terms} \end{aligned}$$

■

Again note that we are not claiming that u,v are curvature co-ordinates on the surface: we only have to have diagonal fundamental forms *at the origin*. Nearby there is no requirement that u,v be curvature-line.

The Theorem is the surface analogy of a result we saw earlier regarding curves: a regular curve in $\mathbb{E}^2$ passing through the origin horizontally at $t = 0$ has its graph given locally by

$$y = \frac{\kappa(0)}{2}x^2 + \text{higher order terms} \qquad *$$

We can put this correspondence to work by considering the curvature of certain curves passing through a point on a surface.

Definition 2.44. Let $\mathbf{v}$ be a tangent vector at a point s to a surface S in $\mathbb{E}^3$. Let $\Pi_s(\mathbf{v})$ be the plane through s spanned by $\mathbf{v}$ and $\mathbf{U}|_s$. Its intersection $\Pi_s(\mathbf{v}) \cap S$ with the surface contains a connected component which describes a curve lying in the surface passing through s. The *normal curvature* of S in the direction $\mathbf{v}$ is the curvature at s of this curve.

The normal curvature is surprisingly easy to compute.

Theorem 2.45 (Euler)**.** *Suppose that* $\mathbf{v}$ *is a tangent vector to* S *making angle* θ *with the first principal curvature direction. Then the normal curvature of* S *in the direction* $\mathbf{v}$ *is*

$$\kappa = k_1 \cos^2\theta + k_2 \sin^2\theta$$

Proof. Choose axes such that the principal curvature directions at s are $\left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right)$, and the unit normal is $\mathbf{U}|_s = \left(\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right)$. Consider the position vector

$$\mathbf{x} = \begin{pmatrix} t\cos\theta \\ t\sin\theta \\ z \end{pmatrix}$$

We can think about this in two ways:

1. Let θ be the angle made by $\mathbf{v}$ and the principal direction $\left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right)$. Then $\mathbf{x}(t,z)$ parameterizes the plane $\Pi_s(\mathbf{v})$. By observation $(*)$, the plane curve in the intersection $\Pi_s(\mathbf{v}) \cap S$ has the form

$$\mathbf{y}(t) = \begin{pmatrix} t\cos\theta \\ t\sin\theta \\ \frac{1}{2}\kappa t^2 + \text{ higher order terms} \end{pmatrix}$$

   where κ is the normal curvature.

2. By Theorem 2.41, the surface is locally a graph $z = f(t,\theta)$, where t,θ are polar co-ordinates in the horizontal plane. Theorem 2.43 says that

$$\begin{aligned} z &= \tfrac{1}{2}k_1(t\cos\theta)^2 + \tfrac{1}{2}k_2(t\sin\theta)^2 + \text{ higher order terms} \\ &= \frac{1}{2}(k_1\cos^2\theta + k_2\sin^2\theta)t^2 + \text{ higher order terms} \end{aligned}$$

Comparing the two expressions for z yields the result. ∎

A straightforward consequence is that the normal curvature always lies *between* the two principal values k_1 and k_2. The principal curvature are therefore the extremes of curvature at any given point.

2.7 Elliptic and hyperbolic points: the Dupin Indicatrix

Theorem 2.43 provides a description of a surface near a point s.

Definition 2.46. Suppose that k_1, k_2, K, H are the principal, Gauss and mean curvatures of a surface S at a given point s. We say that s is:

1. *Elliptic* $\iff K > 0 \iff k_1, k_2$ are non-zero and have the same sign
   In this case, H has the same sign as the principal curvatures: it is positive if and only if the surface bends towards the unit normal.
2. *Hyperbolic* $\iff K < 0 \iff k_1, k_2$ are non-zero and have opposite signs
3. *Parabolic* $\iff K = 0$ and $H \neq 0 \iff$ exactly one of k_1, k_2 are zero
4. *Planar* $\iff K = H = 0 \iff k_1 = k_2 = 0$

Recalling Theorem 2.43, we see that a surface may be written locally as

$$z \approx \frac{1}{2}k_1 x^2 + \frac{1}{2}k_2 y^2$$

Near an elliptic point, level curves (z =constant) of the surface are approximately *ellipses*. Near a hyperbolic point they are approximately *hyperbolæ.* The correspondence doesn't extend to the other types of point. Near a parabolic point a surface looks approximately like a parabolic cylinder, which near a planar point a surface looks approximately like part of a plane.

Examples

1. The (elliptic) paraboloid $\mathbf{x}(u,v) = \begin{pmatrix} u \\ v \\ u^2+v^2 \end{pmatrix}$ has

   $$\mathrm{I} = (1+4u^2)\mathrm{d}u^2 + 8uv\,\mathrm{d}u\mathrm{d}v + (1+4v^2)\mathrm{d}v^2 \qquad \mathbb{I} = \frac{2}{\sqrt{1+4u^2+4v^2}}(\mathrm{d}u^2 + \mathrm{d}v^2)$$

   from which we can calculate[16]

   $$K = \frac{4}{(1+4u^2+4v^2)^2} \quad \text{and} \quad H = \frac{1+2u^2+2v^2}{(1+4u^2+4v^2)^{3/2}}$$

   Both are positive everywhere, and so all points are elliptic with the surface always bending towards the unit normal.

2. The (hyperboloic) paraboloid $\mathbf{x}(u,v) = \begin{pmatrix} u \\ v \\ u^2-v^2 \end{pmatrix}$ has

   $$K = -\frac{4}{(1+4u^2+4v^2)^2} \quad \text{and} \quad H = \frac{4(v^2-u^2)}{(1+4u^2+4v^2)^{3/2}}$$

   so that all points are hyperbolic.

[16] Revisit Proposition 2.38 and the examples following to do this quickly.

3. The surface $\mathbf{x}(u,v) = \begin{pmatrix} u \\ v \\ u^2+v^4 \end{pmatrix}$ has

$$K = \frac{24v^2}{(1+4u^2+16v^6)^2} \quad \text{and} \quad H = \frac{6v^2(1+4u^2)+1+16v^6}{(1+4u^2+16v^6)^{3/2}}$$

   Away from $v = 0$, $\mathbf{x}$ is entirely elliptic and curves towards the unit normal vector. The points $\mathbf{x}(u,0)$ are parabolic. This doesn't mean that the surface contains a straight line, indeed all points $\mathbf{x}(u,v)$ where $(u,v) \neq (u_0,0)$ lie on the same side of the tangent plane at $(u_0,0)$. To see this, note that

$$\mathbf{U}(u_0,0) = \frac{1}{(1+4u_0^2)^{1/2}} \begin{pmatrix} -2u_0 \\ 0 \\ 1 \end{pmatrix}$$

   and so

$$\begin{aligned}(\mathbf{x}(u,v) - \mathbf{x}(u_0,0)) \cdot \mathbf{U}(u_0,0) &= \frac{1}{(1+4u_0^2)^{1/2}} \begin{pmatrix} u-u_0 \\ v \\ u^2+v^2-u_0^2 \end{pmatrix} \cdot \begin{pmatrix} -2u_0 \\ 0 \\ 1 \end{pmatrix} \\ &= \frac{1}{(1+4u_0^2)^{1/2}}((u-u_0)^2+v^2) > 0\end{aligned}$$

   for all $(u,v) \neq (u_0,0)$. As this example shows, a parabolic point can still be convex (or indeed a saddle).

Definition 2.47. Recall from Euler's theorem that the normal curvature takes values at each point between the two principal values. A tangent vector $\mathbf{v}$ at a point on a surface is *asymptotic* if the normal curvature defined by $\mathbf{v}$ is zero.

Examples

1. The elliptic paraboloid has no asymptotic directions at any point: the principal curvatures are both positive at all points, whence the normal curvature at an point is $k_1 \cos^2\theta + k_2 \sin^2\theta > 0$.

2. The hyperbolic paraboloid has asymptotic directions at every point. The principal curvatures have opposite signs, whence there are two angles $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ such that $k_1\cos^2\theta + k_2\sin^2\theta = 0$: i.e.

$$\tan^2\theta = -\frac{k_1}{k_2}$$

   Computing this more specifically is awkward, as the principal curvatures themselves are ugly. At the origin, however, things are easier: $k_1 = 2 = -k_2$, so the asymptotic directions satisfy $\tan^2\theta = 1 \iff \theta = \pm\frac{\pi}{4}$. In this case the asymptotic directions point half way between the principal curvature directions.

3. in the third example, there are only asymptotic directions at the parabolic points when $v = 0$. These point parallel to the y-axis.

In keeping with our theme of transferring calculations back to the parameterization space U, we prove theorem.

Theorem 2.48. *Suppose that $\mathbf{x} : U \to \mathbb{E}^3$ is a local surface and let $v \in T_pU$ be a non-zero tangent vector. Then $\mathrm{d}\mathbf{x}(v)$ is asymptotic iff $\mathbb{I}(v, v) = 0$.*

Because of the theorem we tend to call the tangent vector v asymptotic in its own right.

Proof. Let $v \in T_pU$ be a non-zero tangent vector and let Π be the plane at $\mathbf{x}(p)$ spanned by $\mathbf{U}(p)$ and $\mathrm{d}\mathbf{x}(v)$. Let $\mathbf{y}$ be the curve defined by the intersection of the surface and Π, parameterized such that $\mathbf{y}(0) = \mathbf{x}(p)$. Then the curvature of $\mathbf{y}$ at $\mathbf{x}(p)$ is precisely the normal curvature of the surface in the direction $\mathrm{d}\mathbf{x}(v)$. Now recall Theorem 2.27, from which we see that

$$\mathbf{y}''(0) \cdot \mathbf{U} = \mathbb{I}(v, v)$$

But $\mathbf{y}''(0) \cdot \mathbf{U}$ is zero iff the curvature of $\mathbf{y}$ is zero at $t = 0$ which is iff the normal curvature in the direction $\mathrm{d}\mathbf{x}(v)$ vanishes. ∎

Definition 2.49. Let $\mathbf{x} : U \to \mathbb{E}^3$ be a parameterized surface. The *Dupin indicatrix* at $p \in U$ is the set of tangent vectors $v \in T_pU$ such that $\mathbb{I}(v, v) = \pm 1$.

We can easily derive an equation for the Dupin indicatrix. Suppose that

$$v = x_1 X_1 + x_2 X_2$$

where X_1, X_2 are unit-length principal curvature vectors.[17] Then

$$\begin{aligned}\mathbb{I}(v, v) &= x_1^2 \mathbb{I}(X_1, X_1) + 2x_1x_2 \mathbb{I}(X_1, X_2) + x_2^2 \mathbb{I}(X_2, X_2) \\ &= k_1 x_1^2 + k_2 x_2^2 = \pm 1\end{aligned}$$

defines a curve in the tangent space T_pU. This curve depends on the signs of k_1, k_2.

1. If k_1, k_2 have the same sign, then the Dupin indicatrix is an ellipse (with respect to the basis X_1, X_2). This motivates the name *elliptic point*. In the special case of an umbilic point $k_1 = k_2$ the indicatrix is a circle (with respect to X_1, X_2).

2. If k_1, k_2 are non-zero with opposite signs, then the indicatrix is a pair of hyperbolae, hence *hyperbolic point*.

3. If one of k_1, k_2 is zero (a parabolic point) the indicatrix is a pair of lines.

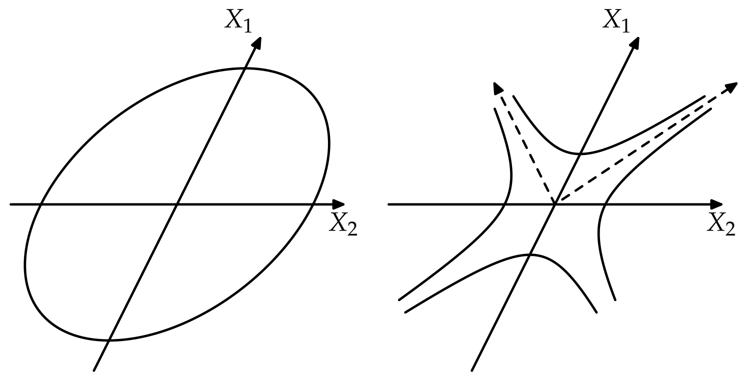
4. If both curvatures are zero, the indicatrix is empty.

When we say that the indicatrix is an ellipse or a circle, etc., with respect to the basis X_1, X_2, what we really mean is with respect to the first fundamental form. Remember that T_pU, although a vector space, has no inherent notion of length or angle (it is not a Euclidean space). It is the first fundamental form that *defines* what is meant by length and angle in T_pU.

The Dupin indicatrix (at least for elliptic and hyperbolic points) gives a visualization of what the surface looks like near a point. Near an elliptic point, planar slices through the surface orthogonal to the unit normal are approximately ellipses. Near a hyperbolic point planar slices through the surface

[17] I.e. $\mathrm{I}(X_1, X_1) = 1 = \mathrm{I}(X_2, X_2)$ and $\mathrm{I}(X_1, X_2) = 0$.

orthogonal to the unit normal are approximately hyperbolae. Typical indicatrices for an elliptic and a hyperbolic point are shown below, with the asymptotic directions indicated by dashed arrows for the hyperbolic point. The first is based on an umbilic point $k_1 = k_2 = 0.9$ (this is a circle with respect to the chosen principal curvature directions X_1, X_2), while the second has $k_1 = -k_2 = 0.3$.



Corollary 2.50. *A point s on a surface has* $\begin{cases} 0 \\ 1 \\ 2 \\ \infty \end{cases}$ *asymptotic directions if s is* $\begin{cases} \textit{elliptic} \\ \textit{parabolic} \\ \textit{hyperbolic} \\ \textit{planar} \end{cases}$.

Proof. The equation for asymptotic directions v is $\mathrm{I\!I}(v, v) = 0$. This requires solving the singular conic

$$\mathrm{I\!I}(v, v) = k_1 x_1^2 + k_2 x_2^2 = 0$$

The options are:

- Elliptic point: $(x_1, x_2) = (0, 0)$ if k_1, k_2 are the same sign. We obtain no asymptotic directions $v = x_1 X_1 + x_2 X_2$.

- Hyperbolic point: if k_1, k_2 have opposite signs, the conic factors as two distinct lines $x_2 = \pm\sqrt{-k_1/k_2}\, x_1$ and we obtain two distinct asymptotic directions.

- Parabolic point: if $k_1 = 0$, then the conic is the x_1-axis ($x_2 = 0$). We have a single asymptotic direction along this axis. ($k_2 = 0$ is similar.)

- Planar point: if $k_1 = k_2 = 0$ then *all* vectors v satisfy $\mathrm{I\!I}(v, v) = 0$ whence any direction is asymptotic.

∎

There are thus at most two asymptotic directions at each non-planar point on a surface.