Chapter 1

Curves and Surfaces in $\mathbb{R}^3$

The study of curves and surfaces have a long history in mathematics and have been some of the principal objects of study. In the classical era, we are all familiar with the Euclid's elements with its emphasis on parallel lines, straightedge constructions, triangles, and proportions. These are not very curvy, but perhaps this is because his work on conics (circles, ellipses, parabolas, and hyperbolas) is lost. But surviving works of Archimedes (Measurement of a Circle, On the Sphere and Cylinder, Quadrature of the Parabola) and Apollonius show that curves were a topic of interest and understood in this time.

Though were was some development in the middle ages, particularly in connection with the cubic equation (Khayyam, Viète), and also indirectly in map making (Mercator), general curves were not really considered until the arrival of Cartesian coordinates in the 17th century. A notable early application of this newfound analytical power is seen in the solution to the brachistochrone curve. This problem, posed by Johann Bernoulli in 1696, challenged mathematicians to find the curve along which an object influenced only by gravity would travel between two points in the least amount of time. According to his niece Newton solved the problem literally overnight using tangents and analysis-type reasoning. In the generations that follow there was an explosion in the search for finding curves with various special properties.

In the 18th and 19th centuries the tools of calculus were turned to the study of surfaces, notably with Euler and Gauss exploring different notions of curvatures. A natural question of Lagrange, asking for the surfaces with the least area, had to wait until non-trivial examples could be found which gave hints towards general methods.

But perhaps the most important contribution, and the most important for us in this course, was that of Riemann. All previous mathematicians, as we will do in this first chapter, considered curves and surfaces inside regular old euclidean three dimensional space. Riemann (1868) gave the definition for abstract spaces, which he called manifolds. This definition splits the properties of a space into two type: intrinsic (depending only on the abstract space) and extrinsic (depending how that object is positioned in space). For example, consider two points on a piece of paper. The distance between those points *along the paper* does not depend on whether the piece of paper is laid flat or bent in an arch, whereas the distance *through space* clearly does. The former is an intrinsic property and the latter extrinsic. This program was carried to completion by the 'Italians', who we will meet in later chapters, by around 1900. This was just in time (or possibly a precondition) for Einstein to use differential geometry in its mature form as a basis the general theory of relativity: our universe is a manifold.

That's enough history; let's see it in action.

1.1 Space Curves and Length

Remark 1.1. In this chapter, and throughout this script unless otherwise stated, we will assume that parameterisations are smooth and injective.

Definition 1.2. *A curve is a smooth function $\alpha : (a, b) \to M$, for some target space M. A path is the restriction of a curve to a closed interval $[\tilde{a}, \tilde{b}] \subset (a, b)$.*

We begin with the example of a helix $\alpha : \mathbb{R} \to \mathbb{R}^3$

$$\alpha(t) = (a \cos t, a \sin t, bt),$$

for constants a, b that describe the size and steepness. In this chapter we will generally consider a parameterised curve $\alpha : [a, b] \to \mathbb{R}^3$, called a *space curve*. Sometimes we will distinguish a parameterised curve α from an un-parameterised curve $\operatorname{img} \alpha$, but other times not. A suitable first question for this curve is to determine its length (or more specifically, the length of any segment of it). In turn, we then must ask how to define the length of a curve. We know how to calculate the distance between points in $\mathbb{R}^3$, so an approximation would be to choose points on the curve, compute the distance between those points, and then add up to total. This approximation will be less than the length, because we are 'cutting corners'. But as we take more and more points into our approximation, it should approach the true value. This leads to

Definition 1.3. *The* length *of a path $\alpha : [a, b] \to \mathbb{R}^3$ is*

$$L(\alpha) = \sup \left\{ \sum_{i=0}^{m-1} \operatorname{dist}(\alpha(t_i), \alpha(t_{i+1})) \mid m \in \mathbb{N}, a = t_0 \leq t_1 \leq \cdots \leq t_m = b \right\}.$$

Let's see if we can use this definition to compute one turn of the helix, $t \in [0, 2\pi]$. The key part of the calculation is the distance between points

$$\begin{aligned}
\operatorname{dist}(\alpha(s), \alpha(t))^2 &= (a \cos s - a \cos t)^2 + (a \sin s - a \sin t)^2 + (bs - bt)^2 \\
&= a^2(2 - 2 \cos s \cos t - 2 \sin s \sin t) + b^2(s - t)^2 \\
&= a^2(2 - 2 \cos(s - t)) + b^2(s - t)^2 \\
&= 4a^2 \sin^2 \tfrac{1}{2}(s - t) + b^2(s - t)^2.
\end{aligned}$$

This only depends on the difference in parameter values $s - t$. Thus if we break choose the points t_i to be equally space between 0 and 2π in terms of the parameter, each term of the sum will be the same. For this choice we have

$$\sum_{i=0}^{m-1} \operatorname{dist}(\alpha(t_i), \alpha(t_{i+1})) = m \times \sqrt{4a^2 \sin^2 \tfrac{\pi}{m} + b^2 \left(\tfrac{2\pi}{m}\right)^2} = \sqrt{4a^2 m^2 \sin^2 \tfrac{\pi}{m} + 4\pi^2 b^2}.$$

Taking the limit as $m \to \infty$ gives $\sqrt{4\pi^2 a^2 + 4\pi^2 b^2} = 2\pi\sqrt{a^2 + b^2}$.

Exercise 1.4. Complete the proof that this is the length, by showing it is an upper bound.

Though the calculation was straightforward from this example, we should a develop a better method to calculate the length. What we see is that the length will be given by the sum of many small pieces, which should remind you of an integral. Indeed

Theorem 1.5 (Speed)**.** *Let $\alpha : [a, b] \to \mathbb{R}^3$ be a continuously differentiable function. Then*

$$L(\alpha) = \int_a^b \|\alpha'(t)\| \, dt \,.$$

Although it appears we have a simple way to calculate the length of any path, be aware that this integral is often not elementary, with the square root in the norm of the vector being the culprit. This means we must resort to methods to approximate the integral (numerical integration). An example of these difficulties is the length of an ellipse.

Proof. We first show $L(\alpha) \leq \int_a^b \|\alpha'(t)\| \, dt$. Consider therefore a partition $a = t_0 \leq t_1 \leq \ldots \leq t_m = b$ of the interval. We have then the inequality

$$\begin{aligned}\sum_{k=0}^{m-1} \operatorname{dist}(\alpha(t_k), \alpha(t_{k+1})) = \sum_{k=0}^{m-1} \|\alpha(t_{k+1}) - \alpha(t_k)\| &= \sum_{k=0}^{m-1} \left\| \int_{t_k}^{t_{k+1}} \alpha'(t) \, dt \right\| \\ &\leq \sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} \|\alpha'(t)\| \, dt = \int_a^b \|\alpha'(t)\| \, dt.\end{aligned}$$

Because $L(\alpha)$ is the supremum over all partitions, this gives an upper bound for $L(\alpha)$. This same argument shows that for the restriction of the curve $\alpha|_{[t_1,t_2]}$ (with $a \leq t_1 < t_2 \leq b$) we have

$$(*) \qquad L(\alpha|_{[t_1,t_2]}) \leq \int_{t_1}^{t_2} \|\alpha'(t)\| \, dt.$$

Consider now the following two functions

$$\begin{aligned} s &: [a, b] \to \mathbb{R}, \ t \mapsto L(\alpha|_{[a,t]}) \\ \widetilde{s} &: [a, b] \to \mathbb{R}, \ t \mapsto \int_a^t \|\alpha'(u)\| \, du \,. \end{aligned}$$

These are meant to capture the length at the parameter t from the start of the path, measured in two ways. Our strategy to finish the proof is not to show the reverse inequality directly. Rather we will show that these two functions are equal. Clearly they are equal at $t = a$.

Observe that s has the property that $L(\alpha|_{[t_1,t_2]}) = s(t_2) - s(t_1)$, and likewise $\int_a^t \|\alpha'(u)\| \, du = \widetilde{s}(t_2) - \widetilde{s}(t_1)$. Inequality (*) above then says $s(t_2) - s(t_1) \leq \widetilde{s}(t_2) - \widetilde{s}(t_1)$. It then follows that

$$\|\alpha(t_2) - \alpha(t_1)\| = \operatorname{dist}(\alpha(t_1), \alpha(t_2)) \leq L(\alpha|_{[t_1,t_2]}) = s(t_2) - s(t_1) \leq \widetilde{s}(t_2) - \widetilde{s}(t_1).$$

After dividing by $t_2 - t_1$, we arrive at

$$\left\| \frac{\alpha(t_2) - \alpha(t_1)}{t_2 - t_1} \right\| \leq \frac{s(t_2) - s(t_1)}{t_2 - t_1} \leq \frac{\widetilde{s}(t_2) - \widetilde{s}(t_1)}{t_2 - t_1}.$$

When we take the limit as $t_2 \to t_1$, the left and right terms both tend to $\|\alpha'(t_1)\|$, by the definition of derivative and the fundamental theorem of calculus respectively. By the squeeze law $s'(t) = \widetilde{s}'(t)$ and therefore $s(t) = \widetilde{s}(t)$. □

The function s in the above proof (using either definition) is called the *arc-length* function with respect to the parameterisation α. We see that its derivative is $\|\alpha'\|$, which we call the speed of the parameterisation. Clearly s is weakly monotonically increasing, since it it the integral of a non-negative function. If it is strongly monotonically increasing, then it is a bijective function from $[a, b]$ to $[0, L(\alpha)]$. In this case we can use it to give a new parameterisation of the same path. The advantage of this new parameterisation is that to find the length between two points, we can just subtract their parameter values. For obvious reasons this is called the arc-length parameterisation of a curve.

Example 1.6. In the case of the helix, the speed is

$$\alpha'(t) = (-a\sin t, a\cos t, b)$$
$$\|\alpha'(t)\| = \sqrt{a^2 + b^2},$$

which is constant. The arc-length function is simply $s(t) = t\sqrt{a^2 + b^2}$, and the inverse is $t(s) = s/\sqrt{a^2 + b^2}$. Hence the helix with arc-length parameterisation is

$$\alpha(s) = \left(a\cos\frac{s}{\sqrt{a^2+b^2}}, a\sin\frac{s}{\sqrt{a^2+b^2}}, \frac{bs}{\sqrt{a^2+b^2}}\right).$$

Note that it is common practice to reuse the name of the curve, even though strictly speaking it is a new function, $\tilde{\alpha} = \alpha \circ t(s)$.

Exercise 1.7. Reparameterise the following planar curves by arc-length:

a. The catenary $\alpha(t) = (t, \cosh t, 0)$.

b. The astroid $\alpha(t) = (\cos^3 t, \sin^3 t, 0)$ for $t \in (0, \pi/2)$.

c. The cardioid $\alpha(t) = (2(1 - \cos t)\cos t, 2(1 - \cos t)\sin t, 0)$.

The arc-length is an intrinsic property of the curve. One can imagine an ant crawling along a piece of string, counting the distance as it goes. In fact this is the only intrinsic invariant for a curve, though we do not yet have a clear definition of intrinsic, so we cannot yet prove this.

A sufficient condition that there exists an arc-length parameterisation is that $\|\alpha'(t)\| \neq 0$ for any t. We say that a curve with such a parameterisation is *regularly parameterised.* This condition also serves to rule out some other curves that have undesirable properties.

Example 1.8. Consider $\beta(t) = (t^2, t^3)$. This is called the cusp curve. We have that $\|\beta'\|^2 = 4t^2 + 9t^3$, and in particular vanishes for $t = 0$. If you look at a plot of this curve, you see that it has a cusp singularity at the origin. This is an example of a curve we would like to avoid.

1.2 Osculating Circles

We have already identified $\|\alpha'\|$ with the speed of the parameterisation, but what is α' itself? Naturally it is the tangent vector and it spans the tangent line to the curve. Recall the tangent line is the limit of the line that passes through the points $\alpha(t)$ and $\alpha(t+h)$ as $h \to 0$. Let us consider the generalisation of this to three points $\alpha(t), \alpha(t+h), \alpha(t-h)$ as $h \to 0$. In $\mathbb{R}^3$, three points span a plane, so long as they do not happen to all lie on the same line. The plane which is the limit of these planes is called the *osculating plane* of the curve.

The key to describing the osculating plane is to find two linearly independent vectors that lie in it. Clearly the tangent vector lies in it. The vectors $\alpha(t+h) - \alpha(t)$ and $\alpha(t-h) - \alpha(t)$ lie in the plane, hence their sum does too. We compute

$$\lim_{h\to 0} \frac{\alpha(t+h) - 2\alpha(t) + \alpha(t-h)}{h^2} = \lim_{h\to 0} \frac{\frac{\alpha(t+h)-\alpha(t)}{h} - \frac{\alpha(t)-\alpha(t-h)}{h}}{h} = \alpha''(t).$$

Thus the osculating plane is spanned by the first and second derivative. Moreover, if we use the arc-length parameterisation the we know that the tangent vector always has length 1. If we differentiate the equation $\alpha' \cdot \alpha' = 1$ then we get $\alpha'' \cdot \alpha' = 0$. In words, in this parameterisation the first and second derivatives are an orthogonal basis of the osculating plane.

In fact we can extract even more information from this three point construction. Three points determine not just a plane, but a circle within that plane. The limit of this circle as these three points come together is called the *osculating circle*. We think of it as the ‘tangent circle’, just like we have a tangent line.

How should we calculate the osculating circle? As for any circle, we should find its center and radius. Conceptually we can find the center by considering the chords $\alpha(t+h) - \alpha(t)$ and $\alpha(t-h) - \alpha(t)$, taking their perpendicular bisectors, and finding the intersection point. Practically, the difficulty is writing down the perpendicular bisector. Let's put this difficulty aside for a moment, and suppose that we have an operator R_h that rotates by a right angle the plane in spanned by the three points with origin $\alpha(t)$. Then the center is the intersection point:

$$\begin{aligned} c &= \alpha(t) + \frac{1}{2}(\alpha(t+h) - \alpha(t)) + uh^{-1}R_h(\alpha(t+h) - \alpha(t)) \\ &= \alpha(t) + \frac{1}{2}(\alpha(t-h) - \alpha(t)) - vh^{-1}R_h(\alpha(t-h) - \alpha(t)). \end{aligned}$$

This is a vector equation in a plane, so u and v are determined by this equation. Really c, u and v are functions of h, since for every h we have a different plane. Taking the limit $h \to 0$ we obtain $c(0) = \alpha(t) + u(0)R_0\alpha'(t)$, so the radius of the osculating circle is $u(0)\|\alpha'(t)\|$. It remains to find $u(0)$.

Rearranging the intersection equation gives

$$\frac{\alpha(t+h) - \alpha(t-h)}{2} = -vh^{-1}R_h(\alpha(t-h) - \alpha(t)) - uh^{-1}R_h(\alpha(t+h) - \alpha(t)).$$

If we just take the limit $h \to 0$ we see that

$$0 = -v(0)R_0(-\alpha'(t)) - u(0)R_0(\alpha'(t)).$$

Since $\alpha'(t) \neq 0$ it must be that $u(0) = v(0)$ (this justifies the choice of sign and the h^{-1} in the equations of the bisectors). If we instead first divide by h and then take the limit, we obtain

$$\begin{aligned}
\alpha'(t) &= -\lim_{h\to 0} vR_h \frac{\alpha(t-h)-\alpha(t)}{h^2} + uR_h \frac{\alpha(t+h)-\alpha(t)}{h^2} \\
&= -\lim_{h\to 0} (v-u)R_h \frac{\alpha(t-h)-\alpha(t)}{h^2} + uR_h \frac{\alpha(t-h)-\alpha(t)+\alpha(t+h)-\alpha(t)}{h^2} \\
&= -\lim_{h\to 0} \frac{v-u}{h} R_h \frac{\alpha(t-h)-\alpha(t)}{h} + uR_h \frac{\alpha(t-h)-2\alpha(t)+\alpha(t+h)}{h^2} \\
&= -(v'(0)-u'(0))R_0(-\alpha'(t)) + u(0)R_0\alpha''(t).
\end{aligned}$$

Dot product both sides with $\alpha'(t)$:

$$\begin{aligned}
\alpha'(t)\cdot\alpha'(t) &= u(0)\alpha'(t)\cdot R_0\alpha''(t) \\
u(0) &= \frac{\alpha'(t)\cdot\alpha'(t)}{\alpha'(t)\cdot R_0\alpha''(t)}
\end{aligned}$$

Thus we have determined $u(0)$, up to the rotation operator R_0. If the curve is parameterised by arc-length, then $\|\alpha'\| = 1$, so the radius of the osculating circle is $u(0)$. Moreover α'' is perpendicular to α', so $\alpha'(t)\cdot R_0\alpha''(t)$ is just the length of α''. In summary we have proved

Theorem 1.9 (Curvature)**.** *For a regular arc-length parameterised curve α, the radius of the osculating is κ^{-1}, where*

$$\kappa(s) = \left\| \frac{d^2\alpha}{ds^2} \right\|,$$

a quantity called the curvature*. The radius of the osculating circle is also called the* radius of curvature*.*

Example 1.10. Let's apply this to the helix. We use the arc-length parameterisation

$$\begin{aligned}
\alpha(s) &= \left(a\cos\frac{s}{\sqrt{a^2+b^2}}, a\sin\frac{s}{\sqrt{a^2+b^2}}, \frac{bs}{\sqrt{a^2+b^2}}\right) \\
\alpha'(s) &= \frac{1}{\sqrt{a^2+b^2}}\left(-a\sin\frac{s}{\sqrt{a^2+b^2}}, a\cos\frac{s}{\sqrt{a^2+b^2}}, b\right) \\
\alpha''(s) &= -\frac{a}{a^2+b^2}\left(\cos\frac{s}{\sqrt{a^2+b^2}}, \sin\frac{s}{\sqrt{a^2+b^2}}, 0\right) \\
\kappa(s) &= \frac{a}{a^2+b^2}.
\end{aligned}$$

Consider the special case $b = 0$, then our helix is a circle and the curvature is a^{-1}. Increasing a decreases the curvature and vice versa. This matches our intuition, a car driving on a large circle only needs to turn slowly. For the general case, we see that increasing either a or b decreases curvature. We also see that α'' points towards the central axis of the helix.

Exercise 1.11. Notice in the above argument that the rotation operator R_h is defined as a rotation of the plane spanned by the three points and only applied to vectors that lie that plane. We could extend R_h to a linear operator on $\mathbb{R}^3$ by declaring that it preserves vectors perpendicular to the plane. Then the operator can be applied to any vector. Simplify the above calculation by using this observation and the second order Taylor polynomial $\alpha(t+h) = \alpha(t) + h\alpha'(t) + \frac{1}{2}h^2\alpha''(t) + O(h^3)$.

Exercise 1.12. Suppose that α is a regular curve but do not assume that it is parameterised by arc-length. Show that the curvature can be calculated with the formula

$$\kappa(t) = \frac{\|\alpha'(t) \times \alpha''(t)\|}{\|\alpha'(t)\|^3}.$$

1.3 Frenet-Serret Equations

Let us return to our thought experiment of an ant on a piece of string. The ant cannot see the curvature of a piece of string. If the string was a helix, but then you straighten it out into a line, none of the distances along the string have changes. Therefore it must be that curvature is an extrinsic property of the curve. But the curvature is invariant under proper euclidean motions (translations and rotations). This is easy to prove, if $\beta(s) = O\alpha(s) + b$ where O is a rotation and b is a vector, then $\|\beta'(s)\| = \|O\alpha'(s)\| = 1$ shows that β is also parameterised by arc-length and $\beta''(s) = O\alpha''(s)$ shows that the curvatures are equal, because rotation does not change the length of a vector.

It turns out for curves in $\mathbb{R}^3$ up to proper euclidean motions there are only two extrinsic properties: curvature and torsion. The goal of this section is to find the torsion and prove that there are no other extrinsic invariants.

We have seen that for an arc-length parameterised curve, the first and second derivatives form an orthogonal basis of the osculating plane. It is customary to normalise them to an orthonormal basis. $T(s) := \alpha'(s)$ is already unit length and we set $N(s) := \kappa(s)^{-1}\alpha''(s)$. They are called the unit tangent and unit normal vectors respectively. Additionally we define $B(s) = T(s) \times N(s)$, called the unit binormal vector, so that we have an orthonormal basis of $\mathbb{R}^3$. This basis is also known as the Frenet frame. We can also recover a curve given knowledge of this basis by integrating $T(s)$, up to a translation

$$\alpha(s) = \int_a^s T(u)\,du + \alpha(a).$$

The advantage of using a basis that comes from the curve, is that if the curve is rotated, this basis is rotated too. If we use this basis to measure the curve, then we are using the curve to measure itself. We have seen this already in the fact that the curvature can be computed as the length of the second derivative. Let us then investigate the derivatives of the other basis vectors. All these vectors are unit length, so they are perpendicular to their derivatives, eg $N \cdot N = 1$ implies $2N \cdot N' = 0$. Further

$$\begin{aligned}
0 = T \cdot N \qquad &\Rightarrow \qquad 0 = T' \cdot N + T \cdot N' = \kappa + T \cdot N' \\
0 = T \cdot B \qquad &\Rightarrow \qquad 0 = T' \cdot B + T \cdot B' = 0 + T \cdot B' \\
0 = N \cdot B \qquad &\Rightarrow \qquad 0 = N' \cdot B + N \cdot B'.
\end{aligned}$$

From the second equation, we see that B' is perpendicular to T as well as B. Therefore it is a scalar of N.

Definition 1.13. *The torsion τ of a curve is defined by the formula $B'(s) = -\tau(s)N(s)$. The minus sign is the most common convention, but be aware some authors choose the opposite sign.*

The binormal B is perpendicular to the osculating plane and the tangent vector is always in it. Therefore the torsion tells us how quickly the osculating plane is rotating around the tangent vector. For a curve that lies entirely in a plane, the torsion is zero.

Example 1.14. For the helix, we already know

$$T(s) = \frac{1}{\sqrt{a^2+b^2}}\left(-a\sin\frac{s}{\sqrt{a^2+b^2}}, a\cos\frac{s}{\sqrt{a^2+b^2}}, b\right)$$
$$N(s) = -\left(\cos\frac{s}{\sqrt{a^2+b^2}}, \sin\frac{s}{\sqrt{a^2+b^2}}, 0\right).$$

Therefore

$$B(s) = \frac{1}{\sqrt{a^2+b^2}}\left(b\sin\frac{s}{\sqrt{a^2+b^2}}, -b\cos\frac{s}{\sqrt{a^2+b^2}}, a\right).$$

We should now differentiate this

$$B'(s) = \frac{b}{a^2+b^2}\left(\cos\frac{s}{\sqrt{a^2+b^2}}, \sin\frac{s}{\sqrt{a^2+b^2}}, 0\right).$$

As expected, this is a scalar multiple of the normal vector. We can read off the torsion

$$\tau(s) = \frac{b}{a^2+b^2}.$$

Helices with $b > 0$ and $b < 0$ are mirror images of one another. Ones that have $b > 0$ are called right-handed, in accordance with the 'right hand rule'. Here we see that our sign convention of torsion gives right-handed helices positive torsion and left-handed helices negative torsion.

We were not yet done with the derivatives of the basis vectors. From the third equation, we have $N' \cdot B = -N \cdot (-\tau N) = \tau$ and from the first $N' \cdot T = -\kappa$. Thus $N'(s) = -\kappa T + \tau B$. Together these derivatives are called the Frenet-Serret formulas

$$\frac{d}{ds}\begin{pmatrix} T \\ N \\ B \end{pmatrix} = \begin{pmatrix} \kappa N \\ -\kappa T + \tau B \\ -\tau N \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix}\begin{pmatrix} T \\ N \\ B \end{pmatrix}$$

Theorem 1.15 (Local uniqueness of curves). *Let α, β be two smooth regular arc-length parameterised curves in $\mathbb{R}^3$ with the same curvature and torsion. For simplicity assume that the curvature is always positive. Then there is a proper euclidean motion (translation and rotation) that takes one to the other.*

Proof. Both bases $\{T_\alpha(0), N_\alpha(0), B_\alpha(0)\}$ and $\{T_\beta(0), N_\beta(0), B_\beta(0)\}$ are right-handed orthonormal bases of $\mathbb{R}^3$. Therefore there is a rotation O that transforms one into the other. Let $b = \beta(0) - O\alpha(0)$. Define $\gamma(s) = O\alpha(s) + b$. It is also arc-length parameterised and $\gamma(0) = O\alpha(0) + \beta(0) - O\alpha(0) = \beta(0)$. Moreover, it has the same curvature and torsion as α and β. By differentiating, we have $T_\gamma(0) = OT_\alpha(0) = T_\beta(0)$ and $\kappa_\gamma N_\gamma = \kappa_\alpha O N_\alpha$, which implies $N_\gamma(0) = ON_\alpha(0) = N_\beta$. By the definition of binormals, $B_\gamma(0) = B_\beta(0)$.

It remains to show that $\beta(s) = \gamma(s)$ for all s. We give two proofs of this fact. The first lies in the observation that the Frenet-Serret formulas are in fact a nine-dimensional system of ODEs (three coordinates for each of the three vectors). By Picard-Lindelöff we know that the initial value problem, which both β and γ satisfy, has a unique solution. In particular $T_\beta(s) = T_\gamma(s)$ for all s. But we can integrate this to see that $\beta(s) = \gamma(s) + c$ for some constant c. Evaluation at $s = 0$ shows that $c = 0$.

We can also use apply the Frenet-Serret formulas directly to show uniqueness

$$\begin{aligned}
&\frac{1}{2}\frac{d}{ds}\left(\|T_\beta - T_\gamma\|^2 + \|N_\beta - N_\gamma\|^2 + \|B_\beta - B_\gamma\|^2\right)\\
&\quad= (T_\beta - T_\gamma)\cdot(T'_\beta - T'_\gamma) + (N_\beta - N_\gamma)\cdot(N'_\beta - N'_\gamma) + (B_\beta - B_\gamma)\cdot(B'_\beta - B'_\gamma)\\
&\quad= (T_\beta - T_\gamma)\cdot\kappa(N_\beta - N_\gamma) + (N_\beta - N_\gamma)\cdot\Big(-\kappa(T_\beta - T_\gamma) + \tau(B_\beta - B_\gamma)\Big)\\
&\qquad\quad + (B_\beta - B_\gamma)\cdot(-\tau)(N_\beta - N_\gamma)\\
&\quad= 0.
\end{aligned}$$

Therefore the sum of squares of the differences is constant. Because it is zero at $s = 0$, it must stay zero for all s. In particular, the tangent vectors are equal. The argument can be finished similarly to the other proof. □

This shows that a curve is uniquely determined in $\mathbb{R}^3$ up to proper euclidean motion by κ and τ. This proves our assertion that these are the only two invariants. Thus the only curves whose curvature and torsion are constant functions are helices. As a special case, the circle is the only curve in the plane with constant curvature (torsion is zero).

In the first half of this chapter dealing with curves, we have already seen several important themes. First there is the difference between intrinsic and extrinsic quantities. Second is the use of special coordinates, in this case arc-length parameterisations. And finally is the idea of measuring the change of vector fields to learn about a space. All three of these ideas will occur repeatedly throughout this course.

Exercise 1.16. Show for a helix

$$a = \frac{\kappa}{\kappa^2 + \tau^2}, \qquad b = \frac{\tau}{\kappa^2 + \tau^2}.$$

1.4 Surfaces

In the second half of this chapter, we consider the example of the helicoid $\Phi : \mathbb{R}^2 \to \mathbb{R}^3$

$$\Phi(u, v) = (u \cos v, u \sin v, bv).$$

We see that for any fixed value of u that we have a helix. Conversely, for any fixed value of v we have a straight line in a plane with $z = bv$. Surfaces which are 'made up of' straight lines are called *ruled surfaces* and were a major topic of study in the theory of surfaces, though we will not go too far down that path. We will also use Φ to represent a general parameterised surface from an open subset $U \subset \mathbb{R}^2$ in $\mathbb{R}^3$. Again, we blur the distinction between a parameterised surface Φ and an un-parameterised surface $\Sigma = \operatorname{img} \Phi$.

We will try to reproduce many of the useful tools for curves in the situation of surfaces. Given a curve $\tilde{\alpha}$ in U, that is a function $\tilde{\alpha} : (a, b) \to U$, we can compose it with Φ to get a curve on $\mathbb{R}^3$ on the surface. We will call this curve $\alpha = \Phi \circ \tilde{\alpha}$ a space curve and $\tilde{\alpha}$ a curve in coordinates if it necessary to distinguish them. The helices and straight lines of the previous sections are then $\tilde{\alpha}_v(t) = (t, v)$ and $\tilde{\beta}_u(t) = (u, t)$. The tangent vector of a general curve $\Phi \circ \tilde{\alpha}$ on the surface can be computed using the chain rule

$$\frac{d}{dt}(\Phi \circ \tilde{\alpha}) = \frac{\partial \Phi}{\partial u}\frac{d\tilde{\alpha}^1}{dt} + \frac{\partial \Phi}{\partial v}\frac{d\tilde{\alpha}^2}{dt} = \left(\tfrac{\partial \Phi}{\partial u} \quad | \quad \tfrac{\partial \Phi}{\partial v} \right) \tilde{\alpha}'.$$

In this way we see that $\frac{\partial \Phi}{\partial u}$ and $\frac{\partial \Phi}{\partial v}$ are a basis for the tangent vectors to the surface, called the coordinate basis vectors.

There are two considerations to make now. Just as we restrict ourselves to regular curves, so too should we restrict ourselves to regular surfaces.

Definition 1.17. *A parameterised surface $\Phi : U \to \mathbb{R}^3$ is called* regular *if $J\Phi$ (the Jacobian of Φ, the matrix of partial derivatives) has rank two at every point. Equivalently, if the vectors $\frac{\partial \Phi}{\partial u}$ and $\frac{\partial \Phi}{\partial v}$ are linearly independent.*

The relation to regular curves should be clear. If the two vectors are linearly dependent at some point $\Phi(c^1, c^2)$, i.e. $w^1\frac{\partial \Phi}{\partial u} + w^2\frac{\partial \Phi}{\partial v} = 0$, then the curve $\tilde{\alpha}(t) = (w^1 t + c^1, w^2 t + c^2)$ would produce a curve α on the surface that was not regular.

The second consideration would be to try to make an 'arc-length' parameterisation. However this is not possible for a surface, for deep reasons that we will explore in the chapter on curvature. For now we can gain a simple understanding through a thought experiment on the sphere. Suppose that there existed a parameterisation like longitude-latitude coordinates on the unit sphere, but instead of angle it used distance. Choose a point on the equator. We can walk the distance π east along a latitude. This is half-way around the sphere. Then we can go a short distance ε north, walk distance π west, and then ε south. In the coordinate chart, this is a rectangle and we are back where we started. But on the sphere, the line of latitude north of the equator is shorter than the equator, so when we walked a distance π on it, we walked too far. At the end we ended up west of our starting point. You have probably already experienced this problem, because no map of the earth can represent the distances correctly to scale. Some maps represent the distances accurately in one direction (cylindrical equidistant projection preserves distance on lines of longitudes, azimuthal equidistant projection preserves distance on radial lines), but most maps do not try to represent distance at all and instead try to preserve angle or area.

So if we can't find a parameterisation of a surface that is arc-length in every direction, what data do we need to be able to calculate angle and distance in a given parameterisation? We know from Theorem 1.5 that to calculate length of a curve we only need the length of the tangent vectors. In $\mathbb{R}^3$ both the length of vectors and the angle between two vectors is given by the dot product

$$\|v\| = \sqrt{v \cdot v}, \qquad \operatorname{ang}(v, w) = \frac{v \cdot w}{\|v\| \, \|w\|}.$$

Abstractly the dot product is an example of an inner product: a function on pairs of vectors that is bilinear, symmetric, and positive definite. The restriction of an inner product to a subspace is again an inner product. Thus we can restrict the inner product of $\mathbb{R}^3$ (the dot product) to an inner product on the tangent space of the surface at any point. This is called the *first fundamental form*. It has the symbol I or g. The word 'form' is an old fashioned term for a function from vectors to scalars, which still appears in certain names.

Practically, how can we describe g at some point? We know that every tangent vector to the surface is in the span of $\frac{\partial \Phi}{\partial u}$ and $\frac{\partial \Phi}{\partial v}$. So we compute

$$\begin{aligned}
& g\left(v^1 \tfrac{\partial \Phi}{\partial u} + v^2 \tfrac{\partial \Phi}{\partial v}, w^1 \tfrac{\partial \Phi}{\partial u} + w^2 \tfrac{\partial \Phi}{\partial v}\right) \\
&= v^1 w^1 g\left(\tfrac{\partial \Phi}{\partial u}, \tfrac{\partial \Phi}{\partial u}\right) + v^1 w^2 g\left(\tfrac{\partial \Phi}{\partial u}, \tfrac{\partial \Phi}{\partial v}\right) + v^2 w^1 g\left(\tfrac{\partial \Phi}{\partial v}, \tfrac{\partial \Phi}{\partial u}\right) + v^2 w^2 g\left(\tfrac{\partial \Phi}{\partial v}, \tfrac{\partial \Phi}{\partial v}\right) \\
&= \begin{pmatrix} v^1 & v^2 \end{pmatrix} \begin{pmatrix} \frac{\partial \Phi}{\partial u} \cdot \frac{\partial \Phi}{\partial u} & \frac{\partial \Phi}{\partial u} \cdot \frac{\partial \Phi}{\partial v} \\ \frac{\partial \Phi}{\partial v} \cdot \frac{\partial \Phi}{\partial u} & \frac{\partial \Phi}{\partial v} \cdot \frac{\partial \Phi}{\partial v} \end{pmatrix} \begin{pmatrix} w^1 \\ w^2 \end{pmatrix} \\
&=: v^T \begin{pmatrix} g_{1\,1} & g_{1\,2} \\ g_{2\,1} & g_{2\,2} \end{pmatrix} w.
\end{aligned}$$

This is called writing g with respect to the coordinate basis. The symmetry of the dot product means that $g_{i\,j}$ is a symmetric matrix, $g_{1\,2} = g_{2\,1}$. The point here is not that we have some short-cut to avoid taking dot products; when you are doing an example it will often be fast to use the dot product in $\mathbb{R}^3$. The point is that we only need part of the information of the dot product of $\mathbb{R}^3$, namely how it acts on the tangent space of the surface, to compute lengths and angles on the surface. We are trying to separate intrinsic information from extrinsic information. If we allow the point on the surface to vary, then we obtain functions $g_{i\,j}(u, v) : U \to \mathbb{R}$.

Remark 1.18. Observe here that v^i and w^i have superscripts instead of subscripts: these are not powers! This is part of a larger notational convention in the field. It's a little annoying at first, especially when you have to write $(v^2)^2$, but it's worth it in the long run. Roughly speaking, coordinates and components of vectors should use superscripts, and forms should use subscripts.

Example 1.19. For the helicoid, we have remarked that for constant v we have straight lines $u \mapsto (u \cos v, u \sin v, bv)$, so

$$\tfrac{\partial \Phi}{\partial u} = (\cos v, \sin v, 0).$$

Similarly we remarked that for constant u we have helices. We have already computed the tangent vector of a helix, though it is simple enough to repeat it

$$\tfrac{\partial \Phi}{\partial v} = (-u \sin v, u \cos v, b).$$

With respect to this basis, we have

$$\begin{aligned} g_{1\,1}(u,v) &= \tfrac{\partial\Phi}{\partial u}\cdot\tfrac{\partial\Phi}{\partial u} = 1,\\ g_{1\,2}(u,v) = g_{2\,1}(u,v) &= \tfrac{\partial\Phi}{\partial u}\cdot\tfrac{\partial\Phi}{\partial v} = 0,\\ g_{2\,2}(u,v) &= \tfrac{\partial\Phi}{\partial v}\cdot\tfrac{\partial\Phi}{\partial v} = u^2+b^2. \end{aligned}$$

Notice that $g_{1\,2}$ is always zero for this example, so the coordinate basis vectors are perpendicular at every point of the surface.

Let us introduce one more tool inspired by the previous sections before we dive into the geometry of surfaces. We saw how useful the Frenet frame was, and it would be nice to have something similar for surfaces. We already have two vectors that span the tangent space, although they are not necessarily unit length or orthogonal. The cross product $\frac{\partial\Phi}{\partial u}\times\frac{\partial\Phi}{\partial v}$ is perpendicular to both coordinate basis vectors and is non-zero because these vectors are linearly independent. We define the *surface normal* ν to the unit length rescaling of this cross product.

Notice that there are two possible unit length vectors perpendicular to the tangent plane. Per our definition, which one we choose comes down to which coordinate we call u and which we call v. This is effectively the choice of an orientation for the surface, and we will refer to this choice of sign of ν as the orientation. Intuitively this can be thought of as choosing one side of the surface to be the 'plus' side. The choice of sign is determined if we think of ν as a function in the coordinates, but it is potentially ambiguous if we think of it as a function from the unparameterised surface Σ. Fortunately, many definitions using ν do not depend on the sign or do so only in a trivial way. Mirroring the notation for curves on the surface, we will use $\nu:\Sigma\to\mathbb{R}^3$ and $\tilde\nu=\nu\circ\Phi:U\to\mathbb{R}^3$ for the two closely related functions.

Finally, the output of ν is a unit length vector in $\mathbb{R}^3$. These are the points of the unit sphere $\mathbb{S}^2$. It is customary therefore to write the target of ν as a sphere and not $\mathbb{R}^3$. Particularly when we think of $\nu:\Sigma\to\mathbb{S}^2$ it is common to call it the *Gauss map*.

Example 1.20. We can compute the surface normal of the helicoid

$$\begin{aligned} \tfrac{\partial\Phi}{\partial u}\times\tfrac{\partial\Phi}{\partial v} &= (b\sin v,-b\cos v,u)\\ \left\|\tfrac{\partial\Phi}{\partial u}\times\tfrac{\partial\Phi}{\partial v}\right\|^2 &= b^2+u^2\\ \tilde\nu = \frac{\frac{\partial\Phi}{\partial u}\times\frac{\partial\Phi}{\partial v}}{\left\|\frac{\partial\Phi}{\partial u}\times\frac{\partial\Phi}{\partial v}\right\|} &= \frac{1}{\sqrt{b^2+u^2}}(b\sin v,-b\cos v,u). \end{aligned}$$

Notice that this is basically the same formula as the binormal B of the helix. This is because $\frac{\partial\Phi}{\partial u}$ is the negative of the normal of the helix, but we are also taking the cross product in the other order.

The normal, or rather $\frac{\partial\Phi}{\partial u}\times\frac{\partial\Phi}{\partial v}$, also gives a definition of surface area. The example of the Schwarz lantern shows that for a surface unlike a curve, we cannot make a definition just by taking straight line approximations. Instead we approximate a surface by the parallelogram spanned by the coordinate vector basis. The area of this parallelogram is the length of the cross product. Thus we make the definition that the surface area is

$$\text{Area} = \int_U \left\|\tfrac{\partial\Phi}{\partial u}\times\tfrac{\partial\Phi}{\partial v}\right\|\;du\,dv.$$

1.5 Curvatures

But let us turn to our main focus: curvature. We want to use curves on the surface to say something about the curvature of the surface itself. But this is not as simple as it might first appear. Consider the example of a plane. This is a surface that has no curvature (by any reasonable definition). But we can take circles of any size in the plane, and therefore there are curves with any amount of curvature. The question is how to distinguish curvature that arises because of the choice of curve from curvature that is forced by the surface itself. The answer is to define the normal curvature of a curve on a surface and prove Meusnier's theorem.

Definition 1.21. *Let $\alpha = \Phi \circ \tilde{\alpha}$ be a curve on a regular surface Σ with normal ν. Suppose that α is parameterised by arc-length as a curve in $\mathbb{R}^3$. The* normal curvature *is $\kappa_n := \alpha'' \cdot \nu$.*

Recall that for an arc-length parameterised curve $\alpha'' = \kappa N$, where N is the normal of the curve. So the normal curvature is $\kappa \cos\theta$ for θ the angle between the normal of the curve and the normal of the surface. This shows that the normal curvature is at most the curvature. There is also the possibility of a sign, but this depends on the choice of orientation.

Example 1.22. We mentioned already the case of a plane Σ, of which $\Phi(u, v) = (u, v, 0)$ is an example. The normal to the surface is $\nu = (0, 0, 1)$, which is constant with respect to the point on the surface. Since Σ is a plane, it must be the osculating plane of every curve on it. Hence N is always perpendicular to ν, and the normal curvature of every curve is zero.

Example 1.23. Next we consider a sphere of radius R. Let's argue geometrically so we don't have to do any calculations. Consider the curve which is the equator of the sphere. It is a circle in the plane $z = 0$, so this plane is its osculating plane and the circle is its osculating circle. Its curvature is $\kappa = R^{-1}$. The normal of the curve is unit vector that points towards the center of the sphere. The normal of the surface is the unit vector points either towards or away from the center of the sphere. Therefore the normal curvature of the equator is $\pm R^{-1}$. This argument applies not just to the equator, but to any great circle of the sphere.

Example 1.24. Our main example for this chapter is the helicoid, so of course we must examine its normal curvatures. We have considered two special sets of curves on the helicoid: the helices and the radial lines. The radial lines are lines, and so have zero curvature. Thus their normal curvature is also zero. The helix $\alpha_u(t) = (u\cos t, u\sin t, bt)$ has curvature $u(u^2 + b^2)^{-1}$. But the normal of this curve is $-\frac{\partial\Phi}{\partial u}$, as we remarked upon in Example 1.20 and therefore the dot product with the surface normal is zero. This shows that the helices also have zero normal curvature in the helicoid.

Let us examine several curves at once, all of which pass through the point $\Phi(1, 0) = (1, 0, 0)$. Consider

$$\alpha(t) = \Phi(w^1 t + 1, w^2 t) = \Big((w^1 t + 1)\cos(w^2 t), (w^1 t + 1)\sin(w^2 t), bw^2 t\Big).$$

There is not a nice arc-length parameterisation for this curve. Indeed

$$\begin{aligned}\alpha'(t) &= w^1 \tfrac{\partial\Phi}{\partial u} + w^2 \tfrac{\partial\Phi}{\partial v} \\ &= w^1\Big(\cos(w^2 t), \sin(w^2 t), 0\Big) + w^2\Big(-(w^1 t + 1)\sin(w^2 t), (w^1 t + 1)\cos(w^2 t), b\Big)\end{aligned}$$

shows that the length of the tangent vector of this curve is quite a complicated function. However, we are mainly interested in the behaviour at the point with $t = 0$, where

$$\alpha'(0) = \Big(w^1, w^2, w^2 b\Big).$$

We can simplify the calculations a little if we choose the constants w^1, w^2 such that this is a unit length. That means $\|\alpha'(0)\|^2 = (w^1)^2 + (w^2)^2(1+b^2) = 1$. The osculating plane of the curve is spanned by the first and second derivatives, regardless of the parameterisation, since

$$\frac{d^2\alpha}{dt^2} = \frac{d}{dt}\big(\|\alpha'\|\, T\big) = \frac{dt}{ds}\frac{d}{ds}\big(\|\alpha'\|\, T\big) = \|\alpha'\| \left(\frac{d\|\alpha'\|}{ds} T + \|\alpha'\|\kappa N\right). \tag{1.25}$$

In fact, with this formula we almost have the answer, because T is orthogonal to the surface normal ν. Taking the dot product on both sides

$$\alpha'' \cdot \nu = \|\alpha'\|^2 \kappa N \cdot \nu = \|\alpha'\|^2 \kappa_n.$$

Thus it only remains to carry out this calculation, using the surface normal N from Example 1.20:

$$\begin{aligned}
\alpha''(t) &= 2w^1w^2\Big(-\sin(w^2t), \cos(w^2t), 0\Big) - (w^2)^2\Big((w^1t+1)\cos(w^2t), (w^1t+1)\sin(w^2t), 0\Big) \\
\alpha''(0) &= \Big(-(w^2)^2, 2w^1w^2, 0\Big) \\
\tilde{\nu}(0,1) &= \frac{1}{\sqrt{1+b^2}}(0, -b, 1) \\
\kappa_n &= \|\alpha'(0)\|^{-2}\alpha''(0)\cdot\tilde{\nu}(0,1) = \frac{1}{\sqrt{1+b^2}}\big(0 - 2bw^1w^2 + 0\big) = \frac{-2b}{\sqrt{1+b^2}}\, w^1w^2.
\end{aligned}$$

This last example shows that the calculation for the normal curvature is not too bad. But we can simplify the calculation in such a way that we don't even have to calculate α''! Consider any curve α on the surface Σ. Because the surface normal is unit length, if we differentiate $\nu(\alpha(t)) \cdot \nu(\alpha(t)) = 1$ we obtain

$$2\nu \cdot \frac{d}{dt}\Big(\nu(\alpha(t))\Big) = 0.$$

Therefore the derivative of the surface normal lies in the tangent plane. First we do a little parameter function shuffle: $\nu \circ \alpha = \nu \circ \Phi \circ \tilde{\alpha} = \tilde{\nu} \circ \tilde{\alpha}$. Now when we apply the chain rule the 'middle' stage is the coordinate chart $U \subset \mathbb{R}^2$:

$$\frac{d}{dt}\Big(\nu(\alpha(t))\Big) = \frac{d}{dt}\Big(\tilde{\nu}(\tilde{\alpha}(t))\Big) = \frac{\partial\tilde{\nu}}{\partial u}\frac{d\tilde{\alpha}^1}{dt} + \frac{\partial\tilde{\nu}}{\partial v}\frac{d\tilde{\alpha}^2}{dt}$$

Likewise we know that along the curve $\alpha'(t) \cdot \nu(\alpha(t)) = 0$, so we can differentiate this relation and obtain

$$\begin{aligned}
\kappa_n &= \|\alpha'\|^{-2}\,\alpha''\cdot\nu = -\|\alpha'\|^{-2}\,\alpha'\cdot\frac{d}{dt}\Big(\nu(\alpha(t))\Big)\\
&= -\|\alpha'\|^{-2}\left(\frac{\partial\Phi}{\partial u}\frac{d\tilde\alpha^1}{dt}+\frac{\partial\Phi}{\partial v}\frac{d\tilde\alpha^2}{dt}\right)\cdot\left(\frac{\partial\tilde\nu}{\partial u}\frac{d\tilde\alpha^1}{dt}+\frac{\partial\tilde\nu}{\partial v}\frac{d\tilde\alpha^2}{dt}\right)\\
&= \|\alpha'\|^{-2}\begin{pmatrix}\frac{d\tilde\alpha^1}{dt} & \frac{d\tilde\alpha^2}{dt}\end{pmatrix}\begin{pmatrix}-\frac{\partial\Phi}{\partial u}\cdot\frac{\partial\tilde\nu}{\partial u} & -\frac{\partial\Phi}{\partial u}\cdot\frac{\partial\tilde\nu}{\partial v}\\ -\frac{\partial\Phi}{\partial v}\cdot\frac{\partial\tilde\nu}{\partial u} & -\frac{\partial\Phi}{\partial v}\cdot\frac{\partial\tilde\nu}{\partial v}\end{pmatrix}\begin{pmatrix}\frac{d\tilde\alpha^1}{dt}\\ \frac{d\tilde\alpha^2}{dt}\end{pmatrix}
\end{aligned}$$

This matrix defines the *second fundamental form* on the tangent plane of the surface, notated with $\mathrm{I\!I}$ or h. We have proved

Theorem 1.26 (Meusnier)**.** *All curves on a regular surface Σ having at some point the same tangent vector w have at that point the same normal curvature. Their normal curvature is given by*

$$\kappa_n = g(w,w)^{-1}h(w,w).$$

Because of this theorem it makes sense speak of the normal curvature $\kappa_n(w)$ of a surface in a direction w. For this reason we say that the normal curvature is telling us something about the curvature of the surface itself, rather than the curves on the surface.

Exercise 1.27. Prove the follow corollaries of Meusnier's theorem. Try as much as possible to argue geometrically rather than relying on calculation. Fix a tangent vector w to the regular surface Σ and let κ_n be the normal curvature in this direction.

a. Consider the plane P spanned by w and ν. Argue that the intersection $\Sigma\cap P$ is a curve with tangent vector w whose curvature is $|\kappa_n|$.

b. Extend this argument to show that for every $\kappa \geq |\kappa_n|$ there is a curve α_κ with tangent vector w and curvature κ.

c. Argue that the union of the osculating circles of every curve with tangent vector w form a sphere with radius κ_n^{-1}.

Let us investigate the second fundamental form a little more. We know that $\frac{\partial\Phi}{\partial u}\cdot\tilde\nu = \frac{\partial\Phi}{\partial v}\cdot\tilde\nu = 0$. Differentiating these with respect to the coordinates u and v give the relations

$$\begin{aligned}
&\frac{\partial^2\Phi}{\partial u\partial u}\cdot\tilde\nu+\frac{\partial\Phi}{\partial u}\cdot\frac{\partial\tilde\nu}{\partial u}=0 & h_{1\,1}&=\frac{\partial^2\Phi}{\partial u\partial u}\cdot\tilde\nu,\\
&\frac{\partial^2\Phi}{\partial u\partial v}\cdot\tilde\nu+\frac{\partial\Phi}{\partial u}\cdot\frac{\partial\tilde\nu}{\partial v}=0 & h_{1\,2}&=\frac{\partial^2\Phi}{\partial u\partial v}\cdot\tilde\nu,\\
&\frac{\partial^2\Phi}{\partial v\partial u}\cdot\tilde\nu+\frac{\partial\Phi}{\partial v}\cdot\frac{\partial\tilde\nu}{\partial u}=0 & h_{2\,1}&=\frac{\partial^2\Phi}{\partial u\partial v}\cdot\tilde\nu=h_{1\,2},\\
&\frac{\partial^2\Phi}{\partial v\partial v}\cdot\tilde\nu+\frac{\partial\Phi}{\partial v}\cdot\frac{\partial\tilde\nu}{\partial v}=0 & h_{2\,2}&=\frac{\partial^2\Phi}{\partial v\partial v}\cdot\tilde\nu.
\end{aligned}$$

Not only do these relations give an easier method to calculate h, but they show that the second fundamental form is a symmetric bilinear form. We have already seen in Example 1.24 that the normal curvature can be both positive and negative for different directions at the same point, therefore the second fundamental form is not positive definite in general.

Example 1.28. We use these formulas to calculate the second fundamental form of the helicoid. Recall

$$\begin{aligned}\frac{\partial\Phi}{\partial u} &= \Big(\cos v, \sin v, 0\Big)\\ \frac{\partial\Phi}{\partial v} &= \Big(-u\sin v, u\cos v, b\Big)\\ \tilde{\nu} &= \frac{1}{\sqrt{b^2+u^2}}(b\sin v, -b\cos v, u)\end{aligned}$$

so

$$\begin{aligned}\frac{\partial^2\Phi}{\partial u\partial u} &= \Big(0,0,0\Big)\\ \frac{\partial^2\Phi}{\partial u\partial v} &= \Big(-\sin v, \cos v, 0\Big)\\ \frac{\partial^2\Phi}{\partial v\partial v} &= \Big(-u\cos v, -u\sin v, 0\Big).\end{aligned}$$

Hence

$$\begin{aligned}h_{1\,1} &= \frac{\partial^2\Phi}{\partial u\partial u}\cdot\tilde{\nu} = 0\\ h_{1\,2} = h_{2\,1} &= \frac{\partial^2\Phi}{\partial u\partial v}\cdot\tilde{\nu} = \frac{-b}{\sqrt{b^2+u^2}}\\ h_{2\,2} &= \frac{\partial^2\Phi}{\partial v\partial v}\cdot\tilde{\nu} = 0.\end{aligned}$$

The fact that $h_{1\,1} = h_{2\,2} = 0$ explains the behaviour in Example 1.24 that the normal curvature in the direction of the coordinate basis vectors was zero. Indeed, we easily reproduce the result from that example for any point on the helicoid not just $(1,0,0)$:

$$\kappa_n = g(w,w)^{-1}h(w,w) = 1\begin{pmatrix}w^1 & w^2\end{pmatrix}\begin{pmatrix}0 & \frac{-b}{\sqrt{b^2+u^2}}\\ \frac{-b}{\sqrt{b^2+u^2}} & 0\end{pmatrix}\begin{pmatrix}w^1\\ w^2\end{pmatrix} = \frac{-2b}{\sqrt{b^2+u^2}}w^1w^2$$

for all w with $g(w,w) = (w^1)^2 + (u^2+b^2)(w^2)^2 = 1$.

The formula for the normal curvature shows that it is invariant under rescaling of the tangent vector w. The geometric explanation is of course that the normal curvature was defined using an arc-length parameterised curve. This motivates us to consider at each point of the surface the normal curvature as a function of unit-length tangent vectors. The set of unit-length tangent vectors is a circle in the tangent plane. So we have a continuous function from a circle to $\mathbb{R}$, hence it must have a maximum and minimum. Actually this function is extremely well-behaved

Theorem 1.29 (Euler)**.** *Fix a point on a regular surface Σ and let κ_1 be the maximum of the normal curvatures and κ_2 the minimum. Let e_1 be a unit-length tangent vector such that $\kappa_n(e_1) = \kappa_1$. Let e_2 be a unit-length tangent vector perpendicular to e_1. Then for any unit-length tangent vector $w = \cos\varphi\, e_1 + \sin\varphi\, e_2$ the normal curvature is*

$$\kappa_n(w) = \kappa_1\cos^2\varphi + \kappa_2\sin^2\varphi.$$

There are some special cases to observe. If the normal curvature is constant at a point, such as for the sphere, then this formula still works, as $\kappa_1 \cos^2\varphi + \kappa_1 \sin^2\varphi = \kappa_1$. Points where this is the case are called *umbilic points.* For umbilic points, every direction can be chosen for e_1. If the normal curvature is not constant at a point, then we see that the maximum and minimum occur exactly twice, $\varphi = 0, \pi$ for the maximum and $\varphi = \pi/2, 3\pi/2$ for the minimum. Essentially there is one maximum and one minimum direction (up to sign) and they are perpendicular to one another. They are called the *principal directions* and κ_1, κ_2 are call the *principal curvatures.*

Proof. Let e_1, e_2, w be as in the statement of the theorem. We expand using bilinearity and symmetry

$$h(w,w) = \kappa_1 \cos^2\varphi + 2h(e_1,e_2)\cos\varphi\sin\varphi + h(e_2,e_2)\sin^2\varphi.$$

Taking the derivative with respect to φ gives

$$\left.\frac{dh(w,w)}{d\varphi}\right|_{\varphi=0} = 0 + 2h(e_1,e_2)(0+1) + 0.$$

Thus e_1 is a maximum point only if $h(e_1,e_2) = 0$. Further

$$h(w,w) = \kappa_1\cos^2\varphi + h(e_2,e_2)\sin^2\varphi \geq h(e_2,e_2)\cos^2\varphi + h(e_2,e_2)\sin^2\varphi = h(e_2,e_2).$$

again using that κ_1 is the maximum value. This shows that $h(e_2,e_2) = \kappa_2$ is the minimum normal curvature. □

Remark 1.30. The proof shows, particularly the part where $h(e_1,e_2) = 0$, that in this orthonormal basis e_1, e_2 that the second fundamental form is diagonalised with the principal curvatures on the diagonal.

Remark 1.31. Suppose that you have a curve α in a principal direction and you investigate the derivative of the surface normal $\nu(\alpha(t))$. Then it transpires that this derivative, which must lie in the tangent plane, is in the principal direction. This is an alternative method of characterising principal directions, and perhaps the more common one in textbooks.

Example 1.32. For the plane and the sphere of radius R, the normal curvature is constant and equal at every point, respectively 0 and R^{-1}.

Example 1.33. For the helicoid we have already computed the normal curvature function, but we have not determined the principal curvatures and directions. Write a unit-length vector as $w^1 = \cos\phi, w^2 = \frac{1}{\sqrt{u^2+b^2}}\sin\phi$ with respect to the coordinate vector basis. Then

$$\kappa_n = \frac{-2b}{u^2+b^2}\cos\phi\sin\phi = \frac{-b}{u^2+b^2}\sin(2\phi) = \frac{b}{u^2+b^2}\cos(2\varphi) = \frac{b}{u^2+b^2}\Big(\cos^2\varphi - \sin^2\varphi\Big),$$

for $\phi = \varphi - \pi/4$. Therefore we see that the principal curvatures of the helicoid are $\pm\frac{b}{u^2+b^2}$, and the principal directions are half-way between the helix and radial directions.

If this were a course purely about curves and surfaces, we would spend a lot more time here investigating special surfaces. For example, the only surfaces where every point is umbilic are (all or parts of) the plane or the sphere. One can also try to find curves on the surface whose tangent vector is always a principal direction, a so-called *line of curvature.* And so on. This would lead us naturally to the definitions of elliptic and hyperbolic points of a surface, and to Gauss curvature. Instead we will give definitions and claim some properties before moving on.

From Euler's theorem, we see that the normal curvatures at a point are completely characterised by the principal curvatures. From these we define two types of curvature. Conversely, if we know the Gauss and mean curvatures, it is possible to solve for the principal curvatures. Thus the normal curvatures at a point are equivalently described by these two quantities.

Definition 1.34. *The* Gauss curvature *of a surface at a point is $K = \kappa_1\kappa_2$ and the* mean curvature *is $H = \frac{1}{2}(\kappa_1 + \kappa_2)$.*

We include a useful formula.

Lemma 1.35. *Let v, w are an orthonormal basis for T_pM and write h with respect to this basis. Then the Gauss curvature at p is $h_{1\,1}h_{2\,2} - h_{1\,2}^2$ and the mean curvature is $\frac{1}{2}(h_{1\,1} + h_{2\,2})$.*

Proof. Be careful: this looks like a determinant but the formula is only true for an orthonormal basis. Although the determinant of the matrix of a linear map is basis-independent, this is not true of the matrix of a bilinear form.

Any unit-length vector of T_pM can be written as $v\cos\theta + w\sin\theta$. Meusnier's theorem and bilinearity tells us

$$\begin{aligned}
\kappa_n &= h_{11}\cos^2\theta + 2h_{12}\cos\theta\sin\theta + h_{22}\sin^2\theta \\
&= h_{11}\frac{1+\cos 2\theta}{2} + h_{12}\sin 2\theta + h_{22}\frac{1-\cos 2\theta}{2} \\
&= \frac{1}{2}(h_{11} + h_{22}) + \frac{1}{2}(h_{11} - h_{22})\cos 2\theta + h_{12}\sin 2\theta.
\end{aligned}$$

To put this into the simplest form, or the form of Euler's theorem, we should try to combine both terms with θ into a single cos. This can be done through polar coordinates: find R and ϕ such that

$$(R\cos\phi, R\sin\phi) = \left(\frac{1}{2}(h_{11} - h_{22}), h_{12}\right).$$

In particular $R^2 = \frac{1}{4}(h_{11} - h_{11})^2 + h_{12}^2$. Then

$$\begin{aligned}
\kappa_n &= \frac{1}{2}(h_{11} + h_{22}) + R\cos\phi\cos 2\theta + R\sin\phi\sin 2\theta \\
&= \frac{1}{2}(h_{11} + h_{22}) + R\cos(2\theta - \phi).
\end{aligned}$$

We see that this obtains its maximum and minimum for $\cos(2\theta - \phi) = \pm 1$. Hence

$$\begin{aligned}
K &= \kappa_1\kappa_2 = \frac{1}{4}(h_{11} + h_{22})^2 - R^2 = h_{11}h_{22} - h_{12}^2, \\
H &= \frac{1}{2}(\kappa_1 + \kappa_2) = \frac{1}{2}(h_{11} + h_{22}).
\end{aligned}$$

What we have effectively done is to deduce the angle between our orthonormal basis and the principal directions, and make a change of basis for the second fundamental form. The principal direction with highest normal curvature is $v\cos\phi/2 + w\sin\phi/2$, for example, because $\theta = \phi/2$ makes κ_n maximum. $\square$

Despite the fact that these our formulas for these quantities depend on the first and second fundamental form, it turns out that it is possible to calculate the Gauss curvature using only the first fundamental form. We say that the Gauss curvature is intrinsic to the surface, whereas the mean curvature is extrinsic. This is known as Gauss' *Theorem Egregium* (the extraordinary theorem) and we will prove a generalisation of it in Chapter 5. In fact, we are finally in a position where we can give good definitions of intrinsic and extrinsic.

Definition 1.36. *Suppose we have two parameterised regular surfaces Φ and $\tilde{\Phi}$ from the same open subset $U \subset \mathbb{R}^2$ to $\mathbb{R}^3$. If both first fundamental forms $g_{i\,j}, \tilde{g}_{i\,j}$, considered as functions on U, are equal, then we say that the surfaces are* isometric. *Geometrically, this means that the distances between corresponding points on both surfaces are equal. Quantities that depend only on the first fundamental form are said to be intrinsic.*

Example 1.37. The classic example of an isometric transformation is rolling a sheet of paper into a cylinder. Consider $U = (-\pi, \pi) \times \mathbb{R}$ and the two parameterisations $\Phi(u, v) = (1, u, v)$ and $\tilde{\Phi}(u, v) = (\cos u, \sin u, v)$. The domain has been chosen so that both parameterisations are injective, as we require. We compute the first fundamental forms

$$\begin{aligned} \tfrac{\partial \Phi}{\partial u} &= (0,1,0),\ \tfrac{\partial \Phi}{\partial v} = (0,0,1) & \qquad g_{1\,1} = 1,\ g_{1\,2} = g_{2\,1} = 0,\ g_{2\,2} = 1, \\ \tfrac{\partial \tilde{\Phi}}{\partial u} &= (-\sin u, \cos u, 0),\ \tfrac{\partial \tilde{\Phi}}{\partial v} = (0,0,1) & \qquad \tilde{g}_{1\,1} = 1,\ \tilde{g}_{1\,2} = \tilde{g}_{2\,1} = 0,\ \tilde{g}_{2\,2} = 1. \end{aligned}$$

So these are indeed isometric surfaces.

From Example 1.22 we know that the normal curvature of the plane is identically zero. Thus so too are the Gauss and mean curvatures.

For the cylinder, we compute the normal and second fundamental form

$$\nu = (\cos u, \sin u, 0),\ \tfrac{\partial^2 \tilde{\Phi}}{\partial u^2} = (-\cos u, -\sin u, 0),\ \tfrac{\partial^2 \tilde{\Phi}}{\partial v \partial u} = \tfrac{\partial^2 \tilde{\Phi}}{\partial v^2} = 0$$
$$\tilde{h}_{1\,1} = -1,\ \tilde{h}_{1\,2} = \tilde{h}_{2\,2} = 0.$$

This is already diagonalised, so we see that the principal curvatures are -1 and 0. (The minus sign is because our cylinder has the outward pointing normal ν but the curve in the surface have an inward pointing normal N.) The Gauss curvature is everywhere zero, same as the plane, but the mean curvature is everywhere $-\frac{1}{2}$.

1.6 Minimal Surfaces

In the final section of this chapter we will indulge my tastes and look at a class of special surfaces related to my own research. A minimal surface is one that has the smallest surface area for a given boundary. To simplify matters, we will consider only graphs. Let U be a bounded region of the plane with smooth boundary, and let $g : \partial U \to \mathbb{R}$ be a continuous function. We consider the set of functions

$$\mathcal{F}_g = \{f \in \mathcal{C}^2(\overline{U}) \mid f|_{\partial U} = g\},$$

and their graphs $\Phi_f(u, v) = (u, v, f(u, v))$. Graphs are always regular surfaces. The area is

$$\text{Area}(f) = \int_U \|\tfrac{\partial\Phi}{\partial u} \times \tfrac{\partial\Phi}{\partial v}\|\, du\, dv = \int_U \|\left(\tfrac{\partial f}{\partial u}, \tfrac{\partial f}{\partial v}, 1\right)\|\, du\, dv = \int_U \sqrt{\left(\tfrac{\partial f}{\partial u}\right)^2 + \left(\tfrac{\partial f}{\partial v}\right)^2 + 1}\, du\, dv.$$

If a surface is a minimal surface, then it must be a critical point for the area. That means, for all variations that don't change the boundary $h \in \mathcal{F}_0$ we have

$$\frac{d}{ds}\text{Area}(f + sh)\Big|_{s=0} = 0.$$

We compute

$$\begin{aligned}
&\frac{d}{ds}\text{Area}(f + sh)\Big|_{s=0} = \int_U \frac{d}{ds}\sqrt{(\partial_u f + s\partial_u h)^2 + (\partial_v f + s\partial_v h)^2 + 1}\Big|_{s=0} du\, dv \\
&= \int_U \frac{\partial_u f \partial_u h + \partial_v f \partial_v h}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}}\, du\, dv \\
&= \int_U \frac{\partial_u f}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}}\, \partial_u h\, du\, dv + \int_U \frac{\partial_v f}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}}\, \partial_v h\, du\, dv \\
&= -\int_U \frac{\partial}{\partial u}\left(\frac{\partial_u f}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}}\right) h\, du\, dv - \int_U \frac{\partial}{\partial v}\left(\frac{\partial_v f}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}}\right) h\, du\, dv,
\end{aligned}$$

using partial integration and the fact that h is zero on the boundary. The only way that this can be zero for all $h \in \mathcal{F}_0$ is if

$$\frac{\partial}{\partial u}\left(\frac{\partial_u f}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}}\right) + \frac{\partial}{\partial v}\left(\frac{\partial_v f}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}}\right) = 0.$$

This can also be written in vector notation as

$$\nabla \cdot \left(\frac{\nabla f}{\|\nabla f\|^2 + 1}\right) = 0. \tag{1.38}$$

This is called the minimal graph equation, which is treated in the course *Partial Differential Equations.* If you expand out the derivatives you obtain

$$\begin{aligned}
&\frac{\partial_u^2 f + \partial_v^2 f}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}} - \frac{1}{2}\frac{\partial_u f\,(2\partial_u f \partial_u^2 f + 2\partial_v f \partial_u \partial_v f) + \partial_v f\,(2\partial_u f \partial_u \partial_v f + 2\partial_v f \partial_v^2 f)}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}^3} \\
&= \frac{(\partial_u^2 f + \partial_v^2 f)((\partial_u f)^2 + (\partial_v f)^2 + 1) - \partial_u f\,(\partial_u f \partial_u^2 f + \partial_v f \partial_u \partial_v f) - \partial_v f\,(\partial_u f \partial_u \partial_v f + \partial_v f \partial_v^2 f)}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}^3} \\
&= \frac{\partial_u^2 f((\partial_v f)^2 + 1) + \partial_v^2 f((\partial_u f)^2 + 1) - 2\partial_u f \partial_v f \partial_u \partial_v f}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}^3}.
\end{aligned}$$

We pause our calculation here to compute the mean curvature of a graph. As in the previous section, we need the second derivatives of the parameterisation

$$\frac{\partial^2 f}{\partial u^2} = (0, 0, \partial_u^2 f), \ \frac{\partial^2 f}{\partial u \partial v} = (0, 0, \partial_u \partial_v f), \ \frac{\partial^2 f}{\partial v^2} = (0, 0, \partial_v^2 f),$$
$$h = \frac{1}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}} \begin{pmatrix} \partial_u^2 f & \partial_u \partial_v f \\ \partial_u \partial_v f & \partial_v^2 f \end{pmatrix}$$

Unfortunately, there is no reason that this will be a diagonal matrix in general, so how should we find the principal curvatures? We will use a little bit of linear algebra. We define a linear transformation A on the tangent space using the formula $h(v, w) = g(v, Aw)$. This is well-defined because g is positive definite. If we use the basis e_1, e_2 of principal directions then we have

$$0 = h(e_1, e_2) = g(e_1, Ae_2), \qquad 0 = h(e_2, e_1) = g(e_2, Ae_1).$$

We conclude that Ae_2 is orthogonal to e_1. Therefore it must be a multiple λ_2 of e_2. Likewise $Ae_1 = \lambda_1 e_1$. Further

$$\kappa_i = h(e_i, e_i) = g(e_i, Ae_i) = \lambda_i g(e_i, e_i) = \lambda_i.$$

In other words, the principal curvatures are the eigenvalues of this matrix A. Hence the mean curvature is $H = \frac{1}{2}(\kappa_1 + \kappa_2) = \frac{1}{2} \operatorname{tr} A$. In terms of the coordinate basis, the matrix A is the product of the inverse of g with h.

$$\begin{aligned} A &= \begin{pmatrix} 1 + (\partial_u f)^2 & \partial_u f \partial_v f \\ \partial_u f \partial_v f & 1 + (\partial_v f)^2 \end{pmatrix}^{-1} \frac{1}{\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}} \begin{pmatrix} \partial_u^2 f & \partial_u \partial_v f \\ \partial_u \partial_v f & \partial_v^2 f \end{pmatrix} \\ &= \frac{1}{\det(g)\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}} \begin{pmatrix} 1 + (\partial_v f)^2 & -\partial_u f \partial_v f \\ -\partial_u f \partial_v f & 1 + (\partial_u f)^2 \end{pmatrix} \begin{pmatrix} \partial_u^2 f & \partial_u \partial_v f \\ \partial_u \partial_v f & \partial_v^2 f \end{pmatrix} \\ &= \frac{1}{*} \begin{pmatrix} (1 + (\partial_v f)^2)\partial_u^2 f - \partial_u f \partial_v f \partial_u \partial_v f & * \\ * & -\partial_u f \partial_v f \partial_u \partial_v f + (1 + (\partial_u f)^2)\partial_v^2 f \end{pmatrix}, \end{aligned}$$

where we have used $*$ to abbreviate expressions that are too long and not important. Hence

$$H = \frac{1}{2} \operatorname{tr} A = \frac{(1 + (\partial_v f)^2)\partial_u^2 f - 2\partial_u f \partial_v f \partial_u \partial_v f + (1 + (\partial_u f)^2)\partial_v^2 f}{\det(g)\sqrt{(\partial_u f)^2 + (\partial_v f)^2 + 1}}.$$

But notice, the numerator of H is exactly the same as the numerator in the minimal graph equation. Thus one is zero if and only if the other is too. In summary, a surface solves the minimal graph equation if and only if it has zero mean curvature.

We can turn this around to give a geometric characterisation of mean curvature: it is a measure of how far a surface deviates from being locally area minimising. For example, the plane has zero mean curvature because if you draw a loop on the plane, the least area surface with that boundary is a plane. The same is true if you draw a small circle on a sphere: a flat circle would have less area then the spherical cap. This is a geometric argument that a sphere has non-zero mean curvature.

Remarkably, the helicoid is a minimal surface! We have seen in Example 1.33 that the principal curvatures are $\pm\frac{b}{u^2+b^2}$. Therefore the mean curvature is zero. Minimal surfaces have a rich and fascinating theory. Just one example would be that the helicoid along with the catenoid belongs to a family of minimal surfaces, all of which are (locally) isometric to one another.