UNIVERSITY OF
SASKATCHEWAN

The local differential geometry of curves in $\mathbb{R}^3$

A *curve* in $\mathbb{R}^{\mathrm{d}}$ is a map from a subset of $\mathbb{R}$ into $\mathbb{R}^{\mathrm{d}}$. A curve is *immersed* if it it has nowhere vanishing derivative (this condition excludes singularities such as corners). By the local differential geometry of a curve α we mean properties of the set

$$\operatorname{image}(\alpha) \coloneqq \bigl\{\alpha(t) \bigm| t \in \operatorname{domain}(\alpha)\bigr\}$$

that depend on arbitrarily small open intervals of the domain. For example,

$$\alpha(t) \coloneqq (\cos t, \sin t, 0) \qquad \text{and} \qquad \beta(t) \coloneqq (\cos 2t, \sin 2t, 0)$$

are immersed curves with image the circle of radius 1, the images are the same sets, and the distinction between the two is not the central focus here. The use of the term *geometry* is appropriate, because the study of circles as subsets of the plane is appropriately described as that.

You may drive a road at various speeds, according to conditions, but the road is the same from one day to the next. It curves, has inclinations — it has a geometry. As it turns out, there are only two local properties of curves in $\mathbb{R}^3$: curvature, quantifying how much the curve bends, and torsion, quantifying how far the road is from being in a single plane. The curvature and torsion inform your driving inputs — you might aim for less speed at higher curvature, and you will sense any change in the car's plane of motion. The fundamental theorem of curves states that curvature and torsion classify immersed curves, up to rotation and translation.

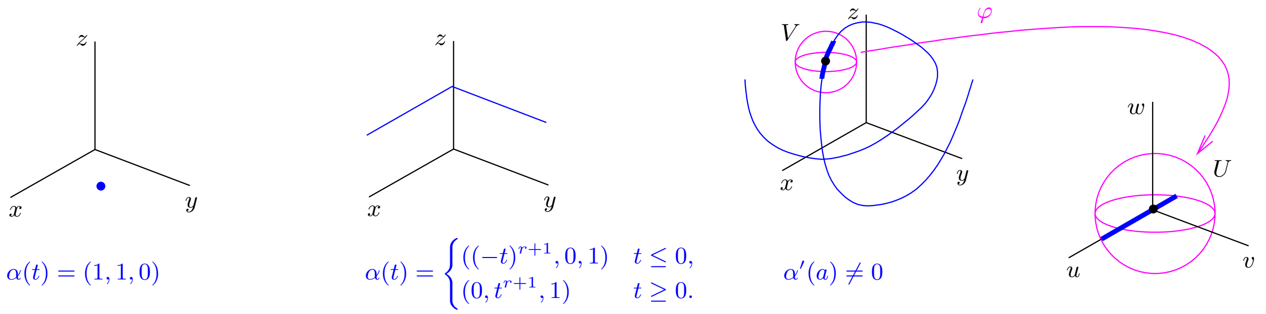
In my opinion, a large part of the value of many mathematical proofs can be conveyed in a few simply sketched central ideas. I have provided those in the main exposition, and accumulated the complete proofs in a technical appendix. The appendix is not intended to be read linearly — selecting the end-of-proof symbols toggles between the sketches and the complete proofs. For the purpose of reference and for convenience of a reader, some foundational results are collected in a second appendix.

§1. Immersed curves, arclength

The image of the curve $\alpha(t) \coloneqq (0,0,0)$ is the single point $(0,0,0)$, not a smooth curve. Another example: let $r \geq 1$ and define

$$\alpha(t) \coloneqq \alpha(t) = \begin{cases} ((-t)^{r+1}, 0, 1) & t \leq 0, \\ (0, t^{r+1}, 1) & t \geq 0. \end{cases}$$

This curve is C^r and its image is the union of the positive x-axis and the positive y-axis, shifted up by 1 to avoid obscuring the axes in the figure. The image cannot be regarded as smooth near $\alpha(0) = (0,0,0)$ because there is a



Left and center: the image of a smooth curve may not be a smooth and one dimensional. Right: the image is smooth near $t = a$ such that $\alpha'(a) \neq 0$, in the sense of becoming a line segment after a smooth change of coordinates.

corner there. Although the first function is C^∞ and the second is C^r, in neither case is the image a smooth curve. However, supposing $\alpha'(a) \neq 0$, the following theorem asserts coordinates which (locally) transform the image of into a line. *Thus, under local change of coordinates on the codomain, all immersed curves are line segments.*

§2. **Theorem** (Local normal form for immersed curves). *Let α be a C^r curve in $\mathbb{R}^3$, $r \geq 1$, and suppose $\alpha'(a) \neq 0$. Then there is a C^r diffeomorphism $\varphi\colon U \subseteq \mathbb{R}^3 \to V \subseteq \mathbb{R}^3$, and a $\delta > 0$, such that $\alpha(t) \in U$ and $\varphi \circ \alpha(t) = (t-a, 0, 0)$ for $a - \delta < t < a + \delta$.*

Proof. It is simple enough to use the functional form of the curve to map a line segment on an axis to the curve, eg $\psi(u,v,w) = \alpha_1(u) + \alpha_2(v) + \alpha_3(w)$, such that $\alpha_1 = \alpha(u+a)$, and $\alpha_2(0) = \alpha_3(0) = 0$. The inverse function theorem then provides a local inverse φ which necessarily maps the curve back to the segment. ■

Definition of arclength: Let $\alpha\colon I \to \mathbb{R}^3$ and let $a < b \in I$ be such that $[a,b] \subseteq I$. Choose a *partition* of $[a,b]$ ie choose $t_i \in (a,b)$, $i = 1, \ldots, n$, such that $a = t_0 < t_1 < t_2 < \ldots < t_n = b$. The length of the polygonal path of line segments obtained by joining in sequence the points $\alpha(t_i)$, is $\sum_{i=1}^n |\alpha(t_i) - \alpha(t_{i-1})|$. Intuitively, as the t_i become close that polygonal path becomes close to the image of α. By definition, the *length of α between $t = a$ to $t = b$* is $\text{length}(\alpha; a, b) = \lim \sum_{i=1}^n |\alpha(t_i) - \alpha(t_{i-1})|$, if that limit exists as the partition t_i becomes finer, in the same sense as definition of the Riemann integral. The curve α is called *rectifiable* if the limit exits for all choices of a, b.

The definition of arclength is not predicated on differentiability of the curve, but the most common computational formula is.

§3. **Proposition** (Arclength formula). *If $\alpha\colon I \to \mathbb{R}^3$ is C^1, then α is rectifiable and $[a,b] \subseteq I$ then*

$$\text{length}(\alpha; a, b) = \int_a^b |\alpha'(t)|\, dt\,.$$

Proof. The formula is the limit of the approximation

$$\sum_{i=1}^n |\alpha(t_i) - \alpha(t_{i-1})| = \sum_{i=1}^n \left| \frac{\alpha(t_i) - \alpha(t_{i-1})}{t_i - t_{i-1}} \right| (t_i - t_{i-1}) \approx \sum_{i=1}^n |\alpha'(t_i^*)| (t_i - t_{i-1}),$$

where $t_i^* \in [t_{i-1}, t_i]$, the last term being a Riemann sum for the stated integral. ■

Reparameterizations: Let $\alpha\colon I \to \mathbb{R}^3$. A *$C^r$ reparameterization* of $\alpha(t)$ is a C^r diffeomorphism $t = t(u)$, $t \in I$. The *reparameterized curve* is $\alpha(u) = \alpha(t(u))$. Because of the correspondence between t and u, the image set, and so the local differential geometry, are the same for a curve and a reparameterization. Arclength parametrizations are geometrically defined parametrizations, in which the many formulas become much simpler.

§4. **Definition**. A curve $\alpha\colon I \to \mathbb{R}^3$ is *parametrized by arclength* if $\text{length}(\alpha; a, b) = a - b$ for all $[a,b] \subseteq I$.

For an arclength parametrized curve, the parameter, usually denoted "s", coincides with length along the curve, as though points in the image of the curve are labeled by a ruler bent along the curve. There is a simple test for arclength parametrization, and, at least in theory, a curve can always be reparametrized by arclength.

§5. **Proposition**. *Let α be a C^r immersed curve defined on an open interval, $r \geq 1$.*

1. *α is parametrized by arclength if and only if it is unit speed ie $|\alpha'(t)| = 1$ for all t.*
2. *For any $a \in I$, $s(t) = \int_a^t |\alpha'(t)|\, dt$ is a C^r diffeomorphism to an open interval, the inverse of which is a C^r arclength reparametrization of α.*

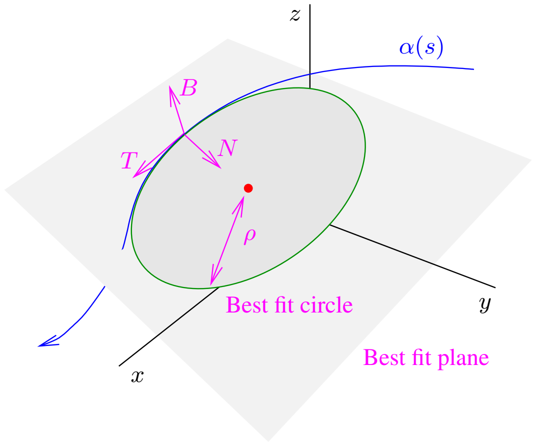
Proof. For (1), differentiation of the arclength formula obtains $|\alpha'(t)| = 1$, and the converse is a direct integration. For (2), $s(t)$ is strictly increasing because $s'(t) = |\alpha'(t)| > 0$, and so s is invertible. The resulting reparameterization is unit speed by the chain rule. ■

Arclength reparametrization is a procedure: find s as a function of t, solve that for t as a function of s, and then substitute $t = t(s)$ into the function that defined the curve. The square root involved in the arclength formula usually precludes a simple expression for $s(t)$ and its inverse $t(s)$, but the integral is easily calculated numerically, and the inverse $t(s)$ may be numerically calculated using the differential equation

$$\frac{dt}{ds} = \frac{1}{|\alpha'(t)|}\,.$$

§6. The fundamental theorem of curves

A best-fit circle or plane to a curve can be determined by three points on the curve which are made to coalesce, as in the calculus definition of a tangent line to the graph of a function as a limit of secant lines, or by matching derivatives, as in Taylor approximations. The best-fit circle to an arc-parametrized curve $\alpha(s)$ at $s = a$ has radius $1/|\alpha''(a)|$ and center $\alpha(a) + \alpha''(a)/|\alpha''(a)|^2$. This circle lies in the best fit plane to α at $s = a$, and that plane has normal $\alpha'(a) \times \alpha''(a)$ (remarks 18, 20, Thm. 22).



$T = \alpha'(a)$, $N = T'/|T'|$, $B = T \times N$,
$\kappa = |T'|$, $\rho = 1/\kappa$, $\tau = -N \cdot B'$

The best fit circle has radius ρ and center along N.
The best fit plane has normal B and contains the best-fit circle.
Plane curve $\Leftrightarrow \tau = 0$.

Serret-Frenet formulas: $T' = \kappa N$, $N' = -\kappa T + \tau B$, $B' = -\tau N$
Fundamental theorem of curves: $\kappa, \tau \Leftrightarrow \alpha$ up to $SE(3)$

$$\boldsymbol{v} = v\, T,\ \boldsymbol{a} = \frac{dv}{dt}\, T + v^2 \kappa\, N,$$
$$T = \frac{\boldsymbol{v}}{v},\ B = \frac{\boldsymbol{v} \times \boldsymbol{a}}{|\boldsymbol{v} \times \boldsymbol{a}|},\ N = B \times T,\ \kappa = \frac{|\boldsymbol{v} \times \boldsymbol{a}|}{v^3},\ \tau = \frac{(\boldsymbol{v} \times \boldsymbol{a}) \cdot \boldsymbol{a}'}{|\boldsymbol{v} \times \boldsymbol{a}|^2}$$

§7. **Definition** (Serret-Frenet data). Let $\alpha\colon (a, b) \to \mathbb{R}^3$ be a unit speed curve defined at $t = a$.

1. The *unit tangent vector* is $T(a) = \alpha'(a)$; the *unit normal vector* is $N(a) = \alpha''(a)/|\alpha''(a)|$; the *binormal vector* is $B(a) = T(a) \times N(a)$.
2. The *curvature* is $\kappa(a) = |\alpha''(a)|$; the *radius of curvature* is $\rho(a) = 1/\kappa(a)$.
3. The *torsion* is $\tau(a) = -N(a) \cdot B'(a)$.
4. The *Serret-Frenet frame* is the tuple $(T(a), N(a), B(a))$. The *Serret-Frenet data* is the tuple (κ, τ, T, N, B).

In terms of the Serret-Frenet data, the radius of the best-fit circle at $t = a$ is $\rho(a)$ and its center at $\alpha(a) + \rho(a)\, N(a)$. $B(a)$ is a unit normal to the plane of best fit, and the torsion is zero if that normal is constant. T and N are orthogonal, because $\alpha' \cdot \alpha'' = 0$ follows from differentiating $\alpha' \cdot \alpha' = 1$, and $T = \alpha'$ and $N = \alpha''/|\alpha''|$. Thus, for each s, $(T(s), N(s), B(s))$ is a right-handed orthonormal basis.

If a curve is rotated and translated, then its Serret-Frenet frame is also rotated, because that data is related to best fitting circles, and circles are sent to congruent circles under rotations and translations. The curve and its data are said to be *covariant*.

§8. **Theorem** (Covariance of the Serret-Frenet data). If $(A, a) \in \mathrm{SE}(3)$ and $\alpha(s)$ is a unit speed curve with Serret-Frenet data (κ, τ, T, N, B), then $A\alpha(s) + a$ is unit speed and has Serret-Frenet data $(AT, AN, AB, \kappa, \tau)$.

Proof. This is a direct calculation from the formula defining the Serret-Frenet data and from the rotational invariance of the dot and cross product. For example, $\tilde{\alpha} := A\alpha(s) + a$ is unit speed because $\tilde{\alpha}' = A\tilde{\alpha}'$ and

$$|\tilde{\alpha}'|^2 = \tilde{\alpha}' \cdot \tilde{\alpha}' = A\alpha' \cdot A\alpha' = |\alpha'|^2 = 1. \qquad ∎$$

Any vector can be written as a unique sum of vectors in an orthonormal frame. Thus, the derivatives of the Serret-Frenet frame can be decomposed relative to the frame itself, with coefficients that turn out to be the curvature and torsion.

§9. **Theorem** (Serret-Frenet equations). *If α is a C^3 unit speed curve and $\alpha''(t) \neq 0$ for all t, then*

$$\frac{dT}{ds} = \kappa N, \quad \frac{dN}{ds} = -\kappa T + \tau B, \quad \frac{dB}{ds} = -\tau N.$$

Proof. The expansion of any vector v in terms of T, N, and B is $v = (T \cdot v)T + (N \cdot v)N + (B \cdot v)B$. Putting $v = dT/ds$ obtains

$$\frac{dT}{ds} = \left(T \cdot \frac{dT}{ds}\right) T + \left(N \cdot \frac{dT}{ds}\right) N + \left(B \cdot \frac{dT}{ds}\right) B$$

and the inner products are easily calculated in terms of κ and τ. The formulas for dN/ds and dB/ds are similar. ∎

The Serret-Frenet equations are important because, given κ and τ, they are differential equations for T, N, and B, thus enabling an application of the powerful existence on uniqueness theory of ordinary differential equations.

§10. **Theorem** (Fundamental theorem of curves). *Let $r \geq 3$, and let $\kappa_0 > 0$ and τ_0 be C^{r-2} and C^{r-3} functions on and open interval I, respectively. Then there is a C^r immersed curve $\alpha \colon I \to \mathbb{R}^3$ with curvature $\kappa = \kappa_0$ and torsion $\tau = \tau_0$. If α_1 and α_2 are two such curves, then there is an element $(A, a) \in \mathrm{SE}(3)$ such that $\alpha_2(s) = A\alpha_1(s) + a$.*

Proof. The initial value problem of the Serret-Frenet formulas, with initial data the standard coordinate orthonormal frame, has a unique solution. The T, N, and B so obtained are an orthonormal frame because the relevant inner products satisfy a (different) initial value problem which has a (unique) constant solution. $\alpha' = T$, so the curve α is constructed from T by integration, and has the given curvature and torsion because those involve higher than first derivatives of α, so at least the first derivatives of the constructed T, N, and B, and those derivatives are determined by their defining differential equation. The uniqueness of the curve up to SE(3) follows from the uniqness of initial value problems and the invariance of the Serret-Frenet data. ∎

Calculation of Serret-Frenet data: For a curve $\alpha(t)$, not necessarily unit speed, the *velocity* is $\boldsymbol{v} := \alpha'$; the *speed* is $v := |\boldsymbol{v}|$; the *acceleration* is $\boldsymbol{a} := \alpha''$; the *scalar acceleration* is $a := |\boldsymbol{a}|$, and the Serret-Frenet data of the corresponding arc-length parameterized curve can be calculated from these, without explicitly calculating the arc-length. Conversely, the velocity $\boldsymbol{v}$ and acceleration $\boldsymbol{a}$ can be written in terms of the Serret-Frenet frame.

§11. **Theorem**. *For C^3 immersed curve $\alpha(t)$ (not necessarily unit speed) such that $\alpha''(t) \neq 0$ for all t,*

$$\boldsymbol{v} = v\,T, \quad \boldsymbol{a} = \frac{dv}{dt}\,T + v^2\kappa\,N, \quad T = \frac{\boldsymbol{v}}{v}, \quad B = \frac{\boldsymbol{v} \times \boldsymbol{a}}{|\boldsymbol{v} \times \boldsymbol{a}|}, \quad N = B \times T, \quad \kappa = \frac{|\boldsymbol{v} \times \boldsymbol{a}|}{v^3}, \quad \tau = \frac{(\boldsymbol{v} \times \boldsymbol{a}) \cdot \boldsymbol{a}'}{|\boldsymbol{v} \times \boldsymbol{a}|^2},$$

where (κ, τ, T, N, B) is the Serret-Frenet data of any arc-length parameterization.

Proof. The arc-length parametrized curve $\alpha(s)$ is related to the original curve $\alpha(t)$ by the equation $\alpha(t) = \alpha(s(t))$. The chain rule calculates the derivatives through this substitution eg $\boldsymbol{v} = d\alpha/dt = d\alpha/ds\,ds/dt = \boldsymbol{v}(t)\,T(s(t))$, and the formulas are obtained by further calculating the second and third derivatives of $\alpha(t)$ using the Serret-Frenet formulas. ∎

The formulas provide insight into three dimensional motion. The curvature is zero where $\boldsymbol{v}$ and $\boldsymbol{a}$ are parallel, and the torsion is arising from the component along B of the third derivative of the curve. The velocity $\boldsymbol{v}$ and acceleration $\boldsymbol{a}$ are in the best-fit plane, because they are both orthogonal to B. From the second formula, the tangent component of the acceleration is the rate of change of speed, and its normal component depends only on the speed and the curvature. This last aspect is a generalization of the common formula that the centripetal acceleration of an object in a circular motion is v^2/R. The scalar acceleration is *not* the <u>(unsigned)</u> rate of change of speed unless either the curvature is zero or the speed is zero, since $a = |\boldsymbol{a}| = \sqrt{(dv/dt)^2 + (v^2\kappa)^2}$.

Special curves: The fundamental theorem of curves determines the curve from the curvature and torsion, thus classifying curves up to rotation and translation as pairs of functions. Because of this classification, and simpler curves can be recognized from the pair (Thm. 28). In summary, a line is characterized by zero curvature, a plane curve by zero torsion, and a circle by constant curvature. A general helix, moving along an axis at a constant pitch, is characterized by a constant ratio of curvature and torsion. A curve on a sphere has a more-or-less complicated functional relative between curvature and torsion.

§12. Epilogue

The study of curves in $\mathbb{R}^3$ is the beginning of differential geometry, because, being dimension 1, these are the simplest nontrivial objects. The result is satisfyingly complete. Curves are classified, up to congruence, by two functions — curvature and torsion. The correspondence is *constructive*: a curve corresponding to given curvature κ and torsion τ can be found by solving the initial value problem

$$\begin{aligned} &\frac{d\boldsymbol{x}}{ds} = T, \qquad \frac{dT}{ds} = \kappa N, \qquad \frac{dN}{ds} = -\kappa T + \tau B, \qquad \frac{dB}{ds} = \tau N, \\ &\boldsymbol{x}(0) = 0, \qquad T(0) = \boldsymbol{i}, \qquad N(0) = \boldsymbol{j}, \qquad B(0) = \boldsymbol{k}, \end{aligned}$$

and such can be numerically solved to any precision. The development identifies quantities—the Serret-Frenet data—that are specific to the *geometry of the curve*, as opposed to being of the way that the curve is *moved along.* With those identified, motion along three dimensional curves can be separated into geometric and dynamic parts, and motions of objects in three space are better understood.

From the modest study of curves emerges far-reaching themes of differential geometry. Immersed curves are locally line segments — the stance is that *smooth* means locally diffeomorphic to linear, and otherwise there is a *singularity.* An arc-length parametrization can be thought of as an assignment of a number to each point of a curve, at least locally (because of the possibility that the curve might self intersect). This is the most elementary example of a *special, geometrically motivated coordinate system, in which formulas greatly simplify.* The most significant progress was made through the use of two powerful theorems of analysis: the *inverse function theorem* and *the existence and uniqueness of solutions of odes.*

Another emergent theme is the use of *geometry to guide and motivate definitions and analysis.* For example, to find the curvature and torsion, circles were fit to curves in geometrically meaningful ways, by best fit according to a geometrically meaningful distance, or by passing circles through coalescing points. Formally, none of this is required: simply define the Serret-Frenet data according to the formulas in Definition 7, and the development *will* be shorter. There is no real need to derive the arc-length formula from the more basic Riemann sums; the formula itself would adequately serve as the definition. The cost of abandoning geometric motivations is a loss of vision, and that can lead to loss of simplicity and elegance. Differential geometry brings vision to analysis. That is present even in the simplest study.

§13. Appendix 1: Technical

§14. **Theorem** (Local normal form for immersed curves). *Let α be a C^r curve in $\mathbb{R}^3$, $r \geq 1$, and suppose $\alpha'(a) \neq 0$. Then there is a C^r diffeomorphism $\varphi\colon U \subseteq \mathbb{R}^3 \to V \subseteq \mathbb{R}^3$, and a $\delta > 0$, such that $\alpha(t) \in U$ and $\varphi \circ \alpha(t) = (t - a, 0, 0)$ for $a - \delta < t < a + \delta$.*

Proof. Define the map ψ by

$$\psi(u, v, w) = \alpha(u + a) + \alpha_2(v) + \alpha_3(w)$$

where α_2 and α_3 are any C^k curves of (a,b) such that $\alpha_2(0)=\alpha_3(0)=0$ and $\alpha'(a),\alpha_2'(0),\alpha_3'(0)$ is a basis of $\mathbb{R}^3$. For example, one can complete the vector $\alpha'(a)$ to an orthonormal basis $\alpha'(a,e_2,e_3$ and choose $\alpha_2(t)=te_2$ and $\alpha_3(t)=te_3$. Then $\psi(0,0,0)=\alpha(a)$ and the derivative of ψ at $(0,0,0)$ is the 3×3 matrix $[\alpha'(a),\alpha_2'(0),\alpha_3'(0)]$, the rows of which are linearly independent. By the inverse function theorem, ψ has a C^k local inverse $\varphi\colon U\ni\alpha(a)\to V\ni(0,0,0)$. Since V is open, there is a $\delta>0$ such that $(u,0,0)\in V$ whenever $u\in(-\delta,\delta)$, and for such a u, $(u,0,0)=\varphi(\psi(u,0,0))=\varphi(\alpha(u+a))$. Substituting $t=u-a$ obtains $\varphi(\alpha(t))=(t+a,0,0)$ for $t\in(a-\delta,a+\delta)$. ■

§15. **Proposition** (Arclength formula). *If $\alpha\colon I\to\mathbb{R}^3$ is C^1, then α is rectifiable and $[a,b]\subseteq I$ then*

$$\operatorname{length}(\alpha;a,b)=\int_a^b|\alpha'(t)|\,dt\,.$$

Proof. By the definition of derivative, $\alpha(t+u)-\alpha(t)=\alpha'(t)u+R(t,u)$ where $\lim_{u\to 0}R(t,u)/u=0$.

$$|\alpha(t+u)-\alpha(t)|=\left|\alpha'(t)+\frac{R(t,u)}{u}\right|u=|\alpha'(u)|+e(t,u),\qquad e(t,u):=\left|\alpha'(t)+\frac{R(t,u)}{u}\right|-|\alpha'(t)|.$$

Since $|\alpha'|$ is continuous, it is Riemann integrable on $[a,b]$, and also $\lim_{u\to 0}e(t,u)=0$, so by Lecture Note 1

$$\begin{aligned}\lim\sum_i|\alpha(t_i)-\alpha(t_{i-1})|&=\lim\sum_i\big(|\alpha'(t_{i-1})|\Delta t_i+e(t_i,\Delta t_i)\big)\\&=\lim\sum_i|\alpha'(t_{i-1})|\Delta t_i+\lim\sum_i e(t_i,\Delta t_i)=\int_p^q|\alpha'(t)|\,dt.\end{aligned}$$ ■

§16. **Lemma**. *A strictly increasing C^r function, $r\geq 1$, that is defined on an open interval and has positive derivative, is a C^r diffeomorphism between open intervals.*

Proof. Let $f\colon(a,b)\to\mathbb{R}$ be continuous with positive derivative, and let $c:=\inf f(a,b)$ and $d:=\sup f(a,b)$. If $y\in(c,d)$ then $y>c$ $\big[y<d\big]$ and there is a $y^-\in f(a,b)$, $y^-=f(x^-)$ $\big[y^+\in f(a,b), y^+=f(x^+)\big]$ such that $c<y<y^-$ $\big[y<y^+<d\big]$. Then $y^-<y<y^+$ and by the intermediate value theorem there is an x such that $x^-<x<x^+$ and $f(x)=y$, so f is onto (c,d). If $f(x_1)\neq f(x_2)$ then either $x_1<x_2$ or $x_1>x_2$ and $f(x_1)<f(x_2)$ or $f(x_1)>f(x_2)$, respectively, and in either case $f(x_1)\neq f(x_2)$, so f is bijective. The inverse of f is C^r by the inverse function theorem. ■

§17. **Proposition**. *Let α be a C^r immersed curve defined on an open interval, $r\geq 1$.*

1. *α is parametrized by arclength if and only if it is unit speed ie $|\alpha'(t)|=1$ for all t.*
2. *For any $a\in I$, $s(t)=\int_a^t|\alpha'(t)|\,dt$ is a C^r diffeomorphism to an open interval, the inverse of which is a C^r arclength reparametrization of α.*

Proof. For (1), if α is parametrized by arclength, $a\in\operatorname{domain}(\alpha)$, and $h\geq 0$, then

$$\int_a^{a+h}|\alpha'(t)|\,dt=(a+h)-a=h,\qquad\frac{1}{h}\int_a^{a+h}|\alpha'(t)|\,dt=1,\qquad|\alpha'(a)|=\lim_{h\to 0^+}\frac{1}{h}\int_a^{a+h}|\alpha'(t)|\,dt=1,$$

while if $h\leq 0$ then

$$\int_{a+h}^{a}|\alpha'(t)|\,dt=a-(a+h)=-h,\qquad\frac{1}{h}\int_{a+h}^{a}|\alpha'(t)|\,dt=-1,\qquad-|\alpha'(a)|=\lim_{h\to 0^-}\frac{1}{h}\int_{a+h}^{a}|\alpha'(t)|\,dt=-1,$$

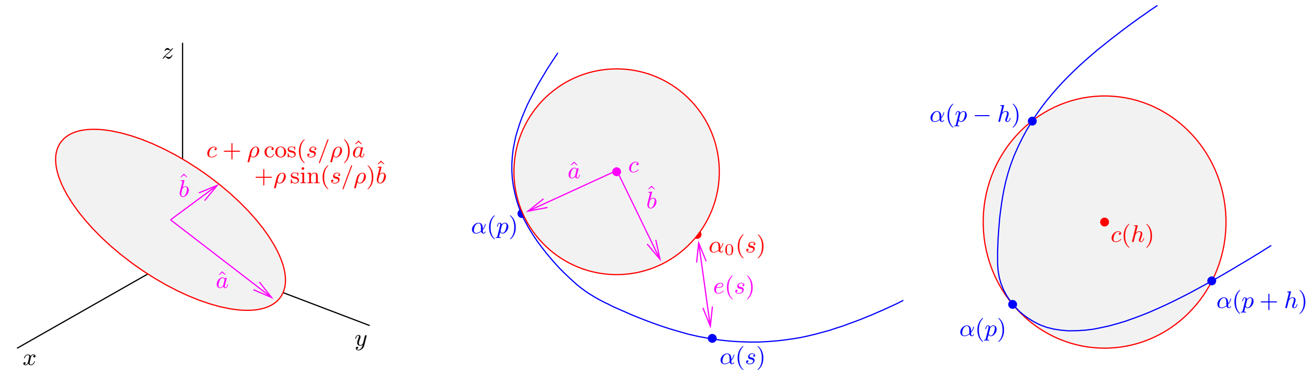
so in either case $|\alpha'(a)|=1$. Conversely, supposing $|\alpha'(t)|=1$ for all t,

$$\operatorname{length}(\alpha;a,b)=\int_a^b|\alpha'(t)|\,dt=\int_a^b 1\,dt=a-b.$$

For (2), $s(t)$ is C^1, and $s'(t)=|\alpha'(t)|$ by the fundamental theorem of calculus, so $s(t)$ is C^r since $|\alpha'(t)|$ is C^{r-1}. By Lem. 16 $s(t)$ is a C^r diffeomorphism to an open interval. The inverse $t(s)$ can be used to reparametrize $\alpha(t)$, and that is an arclength reparametrization because, by the chain rule

$$s'(t)=\left|\frac{d\alpha}{dt}\right|=\left|\frac{d\alpha}{ds}\frac{ds}{dt}\right|=\left|\frac{d\alpha}{ds}\right|s'(t)$$

and $s'(t)\neq 0$ implies $|d\alpha/ds|=1$. ■



Left: parametrization of a circle in general orientation in $\mathbb{R}^3$. Center: fitting that circle using the error defined as the distance between equal-arc points on the circle and the curve. Right: fitting by using three coalescing points.

§18. **Remark** (Fitting a circle to a given curve). Let $\hat{a}$ and $\hat{b}$ be orthogonal unit vectors, $\rho > 0$, and c be a point in $\mathbb{R}^3$. The (unit speed) curve

$$\beta(s) = c + \rho\cos(s/\rho)\,\hat{a} + \rho\sin(s/\rho)\,\hat{b}$$

is a circle of radius ρ, centered at c, in the plane which is parallel to $\hat{a}$ and $\hat{b}$. Because the interest is about the set $\{\alpha(s) \mid s \in I\}$, its is necessary to fit that set to the set $\{\beta(s) \mid s \in I\}$, near $\alpha(a)$. One way to do this is to measure the error between the two curves as the square of the distance $e(s)$ between points of the curve at equal-arc distances from $\alpha(a)$, ie

$$e(s) = |\boldsymbol{e}(s)|^2, \qquad \boldsymbol{e}(s) = \alpha(s) - \beta(s-a).$$

and set as many as possible derivatives of $e(s)$ to zero, because that corresponds to the smallest error for small (Figure 2, left and center). Since

$$e = \boldsymbol{e}\cdot\boldsymbol{e}, \qquad e' = 2\boldsymbol{e}\cdot\boldsymbol{e}', \qquad e'' = 2\boldsymbol{e}\cdot\boldsymbol{e}'' + 2\boldsymbol{e}'\cdot\boldsymbol{e}', \qquad e''' = 2\boldsymbol{e}\cdot\boldsymbol{e}''' + 6\boldsymbol{e}'\cdot\boldsymbol{e}'',$$
$$e^{(4)} = 2\boldsymbol{e}\cdot\boldsymbol{e}^{(4)} + 8\boldsymbol{e}'\cdot\boldsymbol{e}''' + 6\boldsymbol{e}''\cdot\boldsymbol{e}'',$$

this is the same as setting $\boldsymbol{e}(0) = \boldsymbol{e}'(a) = \boldsymbol{e}''(a) = 0$ ($e(a) = 0$ is the same as $\boldsymbol{e}(a) = 0$, and then $e'(a) = 0$ automatically, from which $e'' = 0$ is the same as $\boldsymbol{e}'(a) = 0$, and so on). So for the best fit of the two sets amounts to derivative matching of curves in the arc-length parametrization:

$$\boldsymbol{e}(a) = 0 \;\Leftrightarrow\; \alpha(a) = c + \rho\hat{a}, \qquad \boldsymbol{e}'(a) = 0 \;\Leftrightarrow\; \alpha'(a) = \hat{b}, \qquad \boldsymbol{e}''(a) = 0 \;\Leftrightarrow\; \alpha''(a) = (-1/\rho)\hat{a}. \tag{19}$$

Equations (19) are consistent with $a \cdot b = 0$, because $\alpha' \cdot \alpha'' = 0$, since $'$ is the derivative with respect to arc-length. From the third of (19), $\rho = 1/|\alpha''(a)|$, $\kappa = |\alpha''(a)|$, and $\hat{a} = -\alpha''(a)/|\alpha''(a)|$. The second of (19) implies $\hat{b} = \alpha'(0)$, and the first implies the center of the circle is located at $c = \alpha(a) - \rho\hat{a} = \alpha(a) + \alpha''(a)/|\alpha''(a)|^2$.

§20. **Remark** (Fitting a plane to a given curve). Let a be a nonzero vector and c be a point in $\mathbb{R}^3$. The plane perpendicular to a through the point c is $a \cdot (x - c) = 0$. To best-fit this plane to a given immersed curve $\alpha(t)$, measure the error between the curve and the plane as the signed distance $e(t)$ from points on the curve to the plane, i.e., $e(t) = \big(\alpha(t) - c\big) \cdot a$, and set as many as possible derivatives of $e(s)$ to zero, because that corresponds to the smallest error for small t:

$$e(a) = (\alpha(a) - c) \cdot a = 0, \qquad e'(a) = \alpha'(a) \cdot a = 0, \qquad e''(a) = \alpha''(a) \cdot a = 0,$$

The point c is ambiguous up to addition of vectors orthogonal to a, so one can take $c = \alpha(a)$. The next two equations imply a is orthogonal to $\alpha'(a)$ and $\alpha''(a)$, and one can take $a = \alpha'(a) \times \alpha''(a)$.

§21. **Lemma**. *There is a unique circle containing* $a_1, a_2 \in \mathbb{R}^3$ *and the origin if these three points are not collinear. The center of that circle is at* $\gamma a_1 + \delta a_2$, *where*

$$\gamma = \frac{1}{2D}\,|a_2|^2(|a_1|^2 - a_1\cdot a_2), \quad \delta = \frac{1}{2D}\,|a_1|^2(|a_2|^2 - a_1\cdot a_2), \quad D = |a_1 \times a_2|^2.$$

Proof. Let the center of the circle be at c. As the circle and its center lie in a the same plane, $c = \gamma a_1 + \delta a_2$ for some γ, δ. Since a_1, 0, and a_2 are all on the same circle, $|a_1 - c|^2 = |c|^2 = |a_2 - c|^2$, implying

$$0 = |a_1 - c|^2 - |c|^2 = |a_1|^2 - 2a_1\cdot c + |c|^2 - |c|^2 = |a_1|^2 - 2a_1\cdot c,$$

so $a_1 \cdot c = |a_1|^2/2$, and similarly $a_2 \cdot c = |a_2|^2/2$. By dotting the equation $c = \gamma a_1 + \delta a_2$ with a_1 and then a_2,

$$\frac{|a_1|^2}{2} = a_1\cdot c = \gamma|a_1|^2 + \delta a_1\cdot a_2, \qquad \frac{|a_2|^2}{2} = a_1\cdot c = \gamma a_1\cdot a_2 + \delta|a_2|^2.$$

The result is obtained by solving these linear equations for γ and δ.

§22. **Theorem**. *Let* α *be a* C^2 *unit speed curve and* $\alpha''(a) \neq 0$.

1. $\lim_{h\to 0^+} c(h) = \alpha(a) + \alpha''(a)/|\alpha''(a)|^2$, *where* $c(h)$ *is the center of the circle through* $\alpha(a-h)$, $\alpha(a)$, *and* $\alpha(a+h)$.
2. $\lim_{h\to 0} n(h) = \alpha'(a)\times\alpha''(a)/|\alpha'(a)\times\alpha''(a)|$, *where* $n(h)$ *is the unit normal of the plane passing though* $\alpha(p-h)$, $\alpha(a)$, *and* $\alpha(p+h)$, *defined as*

$$n(h) := \frac{a_1(h)\times a_2(h)}{|a_1(h)\times a_2(h)|}\,, \qquad a_1(h) := \alpha(p-h) - \alpha(a), \quad a_2(h) := \alpha(p+h) - \alpha(a).$$

Proof. By translating the curve, assume without loss of generality that $\alpha(a) = 0$. Let $a_1(h) = \alpha(s-h)$ and $a_2(h) = \alpha(s+h)$. Then Lem. 21 provides $\gamma(h)$, $\delta(h)$, $D(h)$, and $c(h)$, and it is required to show that $\lim\limits_{h\to 0^+} c(h) = \dfrac{\alpha''(a)}{|\alpha''(a)|^2}$. By Taylor's formula, $\alpha(p+h) = \alpha(a) + \alpha'(a)h + \frac{1}{2}\,\alpha''(a)h^2 + o(h^2)$, so, dropping the explicit dependence of functions on a (eg α' without evaluation means $\alpha'(a)$),

$$\begin{aligned}
&a_2(h) = \alpha' h + \frac{1}{2}\,\alpha'' h^2 + o(h^2), \quad a_1(h) = -\alpha' h + \frac{1}{2}\,\alpha'' h^2 + o(h^2),\\
&c(h) = \gamma(h)a_1(h) + \delta(h)a_2(h) = \frac{h^2}{2}\,\big(\delta(h)+\gamma(h)\big)\alpha'' + \big(\delta(h)-\gamma(h)\big)h\alpha' + \delta(h)o(h^2) + \gamma(h)o(h^2),
\end{aligned}$$

and it suffices to show

$$\lim_{h\to 0^+} \frac{h^2}{2}\,\big(\delta(h)+\gamma(h)\big) = \frac{1}{|\alpha''|^2}\,, \quad \lim_{h\to 0^+} \big(\delta(h)-\gamma(h)\big)h = 0, \quad \lim_{h\to 0^+} \delta(h)o(h^2) = 0, \quad \lim_{h\to 0^+} \delta(h)o(h^2) = 0.$$

Substitute $a_1(h)$ and $a_2(h)$, remembering $\alpha'(a)\cdot\alpha'(a) = 1$ and $\alpha'(a)\cdot\alpha''(a) = 0$:

$$\begin{aligned}
&|a_2|^2 = |\alpha' h|^2 + \frac{1}{4}\,|\alpha'' h^2|^2 + o(h^4) + (\alpha' h)\cdot(\alpha'' h^2) + (\alpha' h)\cdot o(h^2) + (\alpha'' h^2)\cdot o(h^2) = h^2 + o(h^3),\\
&|a_1|^2 = h^2 + o(h^3),\\
&a_1\cdot a_2 = -h^2 + o(h^3),\\
&a_1 \times a_2 = \alpha' \times \alpha'' h^3 + o(h^3),\\
&D = |a_1\times a_2|^2 = |\alpha'\times\alpha''|^2 h^6 + o(h^6) = |\alpha''|^2 h^6 + o(h^6),\\
&\gamma = \frac{1}{2D}\,|a_2|^2(|a_1|^2 - a_1\cdot a_2) = \frac{\big(h^2+o(h^3)\big)\big(2h^2+o(h^3)\big)}{2\big(|\alpha''|^2h^6 + o(h^6)\big)} = \frac{h^4 + o(h^5)}{|\alpha''|^2h^6 + o(h^6)} = \frac{1+o(h)}{|\alpha''|^2h^2}\,,\\
&\delta = \frac{1+o(h)}{|\alpha''|^2h^2}\,,
\end{aligned}$$

$$\begin{aligned}
&\lim_{h\to0^+}\frac{h^2}{2}\big(\delta(h)+\gamma(h)\big)=\lim_{h\to0^+}\frac{h^2}{2}\left(\frac{1+o(h)}{|\alpha''|^2h^2}+\frac{1+o(h)}{|\alpha''|^2h^2}\right)=\frac{1}{|\alpha''|^2}\\
&\lim_{h\to0^+}(\delta-\gamma)h=\lim_{h\to0^+}\frac{o(h^2)}{|\alpha''h^2|}=0\\
&\lim_{h\to0^+}\gamma(h)o(h^2)=\lim_{h\to0^+}\frac{o(h^2)+o(h^3)}{|\alpha''|^2h^2}=0,\\
&\lim_{h\to0^+}\delta(h)o(h^2)=\lim_{h\to0^+}\frac{o(h^2)+o(h^3)}{|\alpha''|^2h^2}=0.
\end{aligned}$$

For the second statement,

$$\begin{aligned}
\lim_{h\to0^+}n(h)&=\lim_{h\to0^+}\frac{a_1(h)\times a_2(h)}{|a_1(h)\times a_2(h)|}\\
&=\lim_{h\to0^+}\frac{\alpha'\times\alpha''h^3+o(h^3)}{\sqrt{|\alpha'\times\alpha''|^2h^6+o(h^6)}}\\
&=\lim_{h\to0^+}\frac{\alpha'\times\alpha''+o(h^0)}{\sqrt{|\alpha'\times\alpha''|^2+o(h^0)}}\\
&=\frac{\alpha'(a)\times\alpha''(a)}{|\alpha'(a)\times\alpha''(a)|}\,.
\end{aligned}$$

■

§23. **Theorem** (Covariance of the Serret-Frenet data). If $(A,a)\in\mathrm{SE}(3)$ and $\alpha(s)$ is a unit speed curve with Serret-Frenet data (κ,τ,T,N,B), then $A\alpha(s)+a$ is unit speed and has Serret-Frenet data (AT,AN,AB,κ,τ).

Proof. Let $\tilde\alpha(s)=A\alpha(s)+a$ and let the Serret-Frenet data of $\tilde\alpha$ be denoted similarly. Then

$$\tilde\alpha'(s)\cdot\tilde\alpha'(s)=\big(A\alpha'(s)\big)\cdot\big(A\alpha'(s)\big)=\alpha'(s)\cdot\alpha'(s)=1,$$

so $\tilde\alpha(s)$ is unit speed, and

$$\begin{aligned}
&\tilde T(s)=\tilde\alpha'(s)=A\alpha'(s)=AT(s),\\
&\tilde\kappa=|\tilde T'(s)|=|AT'(s)|=|T'(s)|=\kappa,\\
&\tilde N(s)=\tilde T'(s)/\tilde\kappa(s)=AT'(s)/\kappa(s)=AN(s),\\
&\tilde B(s)=\tilde T(s)\times\tilde N(s)=\big(AT(s)\big)\times\big(AN(s)\big)=A\big(T(s)\times N(s)\big)=AB(s),\\
&\tilde\tau(s)=-\tilde N(s)\cdot\tilde B'(s)=-\big(AN(s)\big)\cdot\big(AB(s)\big)'=-\big(AN(s)\big)\cdot\big(AB'(s)\big)=-N(s)\cdot B'(s)=\tau(s).
\end{aligned}$$

■

§24. **Theorem** (Serret-Frenet equations). *If α is a C^3 unit speed curve and $\alpha''(t)\neq0$ for all t, then*

$$\frac{dT}{ds}=\kappa N,\qquad\frac{dN}{ds}=-\kappa T+\tau B,\qquad\frac{dB}{ds}=-\tau N.$$

Proof. Let $u_1:=T$, $u_2:=N$, $u_3:=B$. By the expansion (31), $u_i'=\sum_{j=1}^3a_{ij}u_j$ where $a_{ij}=u_i'\cdot u_j$, and it is required to show that

$$[a_{ij}]=\begin{bmatrix}0&\kappa&0\\-\kappa&0&\tau\\0&-\tau&0\end{bmatrix}$$

The matrix $[a_{ij}]$ is antisymmetric because $0=(u_i\cdot u_j)=u_i'\cdot u_j+u_i\cdot u_j'=a_{ij}+a_{ji}$, so the diagonal elements are zero, and

$$a_{12}=N\cdot T'=N\cdot(\kappa N)=\kappa N\cdot N=\kappa,\quad a_{13}=B\cdot T'=B\cdot(\kappa N)=0,\quad a_{23}=-a_{32}=-N\cdot B'=\tau.$$

■

§25. **Theorem** (Fundamental theorem of curves). *Let $r \geq 3$, and let $\kappa_0 > 0$ and τ_0 be C^{r-2} and C^{r-3} functions on and open interval I, respectively. Then there is a C^r immersed curve $\alpha\colon I \to \mathbb{R}^3$ with curvature $\kappa = \kappa_0$ and torsion $\tau = \tau_0$. If α_1 and α_2 are two such curves, then there is an element $(A, a) \in \mathrm{SE}(3)$ such that $\alpha_2(s) = A\alpha_1(s) + a$.*

Proof. The *linear* differential equation

$$\frac{du_1}{ds} = \kappa_0(s)u_2, \quad \frac{du_2}{ds} = -\kappa_0(s)u_1 + \tau_0(s)u_3, \quad \frac{du_3}{ds} = -\tau_0(s)u_2 \tag{26}$$

is defined on triples of vectors (u_1, u_2, u_3). This is the differential equation

$$\frac{dp_i}{dt} = \sum_{j=1}^{3} a_{ij}u_j, \qquad [a_{ij}] \coloneqq \begin{bmatrix} 0 & \kappa_0 & \tau_0 \\ -\kappa_0 & 0 & \tau_0 \\ 0 & \tau_0 & \kappa_0 \end{bmatrix}.$$

Pick any $c \in I$. Since κ_0 and τ_0 are both C^{r-3}, and $r \geq 3$, there is a unique C^1 solution $\big(T_0(s), N_0(s), B_0(s)\big)$ defined on all I which satisfy the initial condition

$$T_0(c) = (1, 0, 0), \quad N_0(c) = (0, 1, 0), \quad B_0(c) = (0, 0, 1),$$

Let $\alpha(s) \coloneqq \displaystyle\int_c^s T_0(s)\, ds$.

By the existence and uniqueness theorem for differential equations, T_0, N_0, and B_0 are C^{r-3}. From the second equation of (26), dN_0/ds is the product of C^{r-3} functions, so N_0 is C^{r-2}. Since κ_0 is C^{r-2} and N_0 is C^{r-2}, the first equation of (26), implies dT_0/ds is the product of C^{r-2} functions, so T_0 is C^{r-1}. Since $\alpha' = T_0$, this shows that α' is C^r.

Define the 3×3 matrix $[p_{ij}] \coloneqq [u_i \cdot u_j]$ and note that

$$\frac{dp_{ij}}{dt} = \frac{du_i}{dt} \cdot u_j + u_i \cdot \frac{du_j}{dt} = \left(\sum_{k=1}^{3} a_{ik}\, u_k\right) \cdot u_j + u_i \cdot \left(\sum_{k=1}^{3} a_{jk}\, u_k\right)$$

so the p_{ij} satisfy the (linear) initial value problem

$$\frac{dp_{ij}}{dt} = \sum_{k=1}^{3} (a_{ik}\, p_{kj} + a_{jk}\, p_{ik}), \qquad [p_{ij}] = \mathbf{1}_{3\times 3}.$$

Constant p_{ij} with the same initial data is also a solution because substitution gives 0 on the left side and $a_{ij} + a_{ji} = 0$ on the right, so the vectors u_1 ,u_2, u_3 are orthonormal for all t, by uniqueness of initial value problems. The determinant of the 3×3 matrix $[u_1, u_2, u_3]$ is continuous in t and either $+1$ or -1, and is $+1$ at $t = c$, so it is $+1$ for all t, and u_1, u_2, u_3 is a right-handed orthonormal basis.

Let (κ, τ, T, N, B) be the Serret-Frenet data of $\alpha(s)$, which is unit speed because $\alpha'(s) = T_0(s)$ and $T_0(s)$ is a unit vector for all s. Then $T = \alpha' = T_0$ and $\kappa = |T'| = |T_0'| = |\kappa_0 T_0| = \kappa_0$. Also $N = T'/\kappa = T_0'/\kappa_0 = N_0$, and $B = T \times N = T_0 \times N_0 = B_0$ because (T_0, N_0, B_0) is right-handed and orthonormal, and $\tau = -B' \cdot N = -B_0' \cdot N_0 = -(-\tau_0 N_0) \cdot N_0 = \tau_0$. Thus the Serret-Frent data for α is $(\kappa_0, \tau_0, T_0, N_0, B_0)$ and α has the specified curvature and torsion.

Suppose α_1 and α_2 both have curvature κ_0 and torsion τ_0. The Serret-Frenet frames (T_i, N_i, B_i) of α_i , $i = 1, 2$, both satisfy the differential equation (26), possibly with different initial conditions at $s = c$. Let A be the 3×3 rotation matrix such that

$$A\,T_1(c) = T_2(c), \quad A\,N_1(c) = N_2(c), \quad A\,B_1(c) = B_2(c),$$

ie $A = A_2 A_1^{-1}$ where A_i are the matrices with column vectors $T_i(c), N_i(c), B_i(c)$. The curve $A\alpha_1(s)$ also has curvature κ_0 and torsion τ_0, so its Serret-Frenet frame $A\,\tilde{T}_1(s), A\,\tilde{N}_1(s), A\,\tilde{B}_1(s)$ satisfies the same initial value problem as the Serret-Frenet frame of α_2. By uniqueness of solutions to initial value problems, $A\,T_1(s) = T_2(s)$, so

$$A\alpha_1(s) - A\alpha_1(c) = \int_c^s \frac{d}{ds}\, A\,\alpha_1(s)\, ds = \int_c^s A\,T_1(s)\, ds = \int_c^s T_2(s)\, ds = \int_c^s \frac{d}{ds}\, \alpha_2(s)\, ds = \alpha_2(s) - \alpha_2(c),$$

ie $\alpha_2(s) = A\,\alpha_1(s) + a$, where $a \coloneqq \alpha_2(c) - A\,\alpha_1(c)$. ■

§27. **Theorem**. *For C^3 immersed curve $\alpha(t)$ (not necessarily unit speed) such that $\alpha''(t) \neq 0$ for all t,*

$$\boldsymbol{v} = v\,T, \quad \boldsymbol{a} = \frac{dv}{dt}\,T + v^2\kappa\,N, \quad T = \frac{\boldsymbol{v}}{v}, \quad B = \frac{\boldsymbol{v}\times\boldsymbol{a}}{|\boldsymbol{v}\times\boldsymbol{a}|}, \quad N = B\times T, \quad \kappa = \frac{|\boldsymbol{v}\times\boldsymbol{a}|}{v^3}, \quad \tau = \frac{(\boldsymbol{v}\times\boldsymbol{a})\cdot\boldsymbol{a}'}{|\boldsymbol{v}\times\boldsymbol{a}|^2},$$

where (κ, τ, T, N, B) is the Serret-Frenet data of any arc-length parameterization.

Proof.

$$\frac{ds}{dt} = \frac{d}{dt}\int_p^t \left|\frac{d\alpha}{dt}\right| dt = \left|\frac{d\alpha}{dt}\right| = v(t),$$

$$\boldsymbol{v} = \frac{d\alpha}{dt} = \frac{d\alpha}{ds}\frac{ds}{dt} = vT,$$

$$\boldsymbol{a} = \frac{d\boldsymbol{v}}{dt} = \frac{d}{dt}(vT) = \frac{dv}{dt}T + v\frac{dT}{ds}\frac{ds}{dt} = \frac{dv}{dt}T + \kappa v^2 N,$$

$$\boldsymbol{v}\times\boldsymbol{a} = (vT)\times\left(\frac{dv}{dt}T + \kappa v^2 N\right) = \kappa v^3 T\times N = \kappa v^3 B,$$

$$|\boldsymbol{v}\times\boldsymbol{a}| = |\kappa v^3 B| = \kappa v^3, \quad \text{so} \quad \kappa = \frac{|\boldsymbol{v}\times\boldsymbol{a}|}{v^3} \quad \text{and} \quad B = \frac{\boldsymbol{v}\times\boldsymbol{a}}{\kappa v^3} = \frac{\boldsymbol{v}\times\boldsymbol{a}}{|\boldsymbol{v}\times\boldsymbol{a}|},$$

$$B\times T = -T\times(T\times N) = -(T\cdot N)T + (T\cdot T)N = N,$$

$$(\boldsymbol{v}\times\boldsymbol{a})\cdot\frac{d\boldsymbol{a}}{dt} = |v\times a|B\cdot\frac{d}{dt}\left(\frac{dv}{dt}T + \kappa v^2 N\right) = |v\times a|B\cdot\left(\frac{d^2v}{dt^2}T + \frac{dv}{dt}\kappa N + \frac{d}{dt}(\kappa v^2)N + \kappa v^2\frac{dN}{dt}\right) = |v\times a|\kappa v^3\tau,$$

$$\tau = \frac{\boldsymbol{v}\times\boldsymbol{a}}{|v\times a|\kappa v^3} = \frac{\boldsymbol{v}\times\boldsymbol{a}}{|v\times a|^2}.$$

■

§28. **Theorem**.

1. *A line is an immersed curve in $\mathbb{R}^3$ which can be reparametrized to $\alpha(t) = a + bt$, where a and b are constant vectors, $b \neq 0$. For C^2 immersed curves in $\mathbb{R}^3$, the follow statements are equivalent: (A) α is a line; (B) T is constant; (C) $\kappa = 0$.*
2. *An immersed curve $\alpha(t)$ in $\mathbb{R}^3$ is planar if there is a nonzero constant vector a such that $\alpha(t)\cdot a$ is constant. For C^3 immersed curves α in $\mathbb{R}^3$ with κ never zero, the following statements are equivalent: (A) α is planar; (B) B is constant; (C) $\tau = 0$.*
3. *An immersed curve $\alpha(t)$ in $\mathbb{R}^3$ is a sphere curve (with center a) if there is a constant nonzero vector a such that $|\alpha(t) - a|$ is constant. For C^4 immersed curves α in $\mathbb{R}^3$ with κ never zero and τ never zero, the following statements are equivalent: (A) α is a sphere curve; (B) $\alpha + \rho N + (\rho'\sigma/v)B$ is constant; (C) $(\sigma\rho'/v)'/v + \rho\tau = 0$.*
4. *A circle is a planar sphere curve. For C^3 immersed curves α in $\mathbb{R}^3$ with κ never zero, the following statements are equivalent: (A) α is a circle; (B) $\alpha + \rho N$ is constant and B is constant; (C) κ is constant and $\tau = 0$.*
5. *A helix (with pitch $|a|$ and axial vector a) is an immersed curve in $\mathbb{R}^3$ which can be reparametrized to $\alpha(t) = \bar{\alpha}(t) + ta$, where $a \neq 0$ is constant and $\bar{\alpha}$ is unit speed and in a plane with normal a. For C^3 immersed curves α in $\mathbb{R}^3$ with κ never zero and τ never zero, the following statements are equivalent: (A) α is a helix; (B) there is a nonzero constant vector a such that $T\cdot a$ is a nonzero constant; (C) τ/κ is a nonzero constant.*
6. *A circular helix is an immersed curve in $\mathbb{R}^3$ which can be reparametrized to $\alpha(t) = \bar{\alpha}(t) + at$ where where $a \neq 0$ is constant and $\bar{\alpha}(t)$ is a circle in a plane with normal a. For C^3 immersed curves α in $\mathbb{R}^3$ with κ never zero and τ never zero, the following statements are equivalent: (A) α is a circular helix; (B) there is a nonzero vector a such that $T\cdot a$ is a nonzero constant, and $(\alpha + \rho v^2 N)\times a$ is constant; (C) κ and τ are nonzero constants.*

Proof. In the proofs of (1) through (4), assume without loss of generality that α is unit speed.

(1A)⇒(1B): Given $\alpha = a + bt$, $\alpha' = b$ and $T = \alpha'/|\alpha'| = b/|b|$, which is constant.

(1B)⇒(1C): Given that T is constant, $\boldsymbol{v} = vT$, so $\boldsymbol{a} = v'T + vT' = v'T$, and $\boldsymbol{v}\times\boldsymbol{a} = (vT)\times(v'T) = 0$ so $\kappa = |\boldsymbol{v}\times\boldsymbol{a}|/v^3 = 0$.

(1C)⇒(1A): Reparametrize α by arc-length. Then $0 = \kappa = |\alpha''| = |T'|$ implies $T' = 0$, so T is constant.

(2C)⇔(2B): $\tau = 0$ if and only if B is constant, because $B' = -\tau N$ and $N \neq 0$.

(2B)⇒(2A): Given B constant, $(B \cdot \alpha)' = B' \cdot \alpha + B \cdot \alpha' = B \cdot T = 0$ so $B \cdot \alpha$ is constant.

(2A)⇒(2B): Given $\alpha \cdot a = 0$, differentiate to obtain $T \cdot a = 0$ and $\kappa N \cdot a = 0$. So a is parallel to B for all s, so $B = \pm a/|a|$ and B is constant because it is continuous.

(3A)⇒(3C): Differentiate the defining property $(\alpha - a) \cdot (\alpha - a) =$ constant as many times as necessary in order to uncover the required functional relation between curvature and torsion:

$$\begin{aligned}
&(\alpha - a) \cdot (\alpha - a) = \text{constant} \\
&2\alpha' \cdot (\alpha - a) = 0, \quad \text{so} \quad T \cdot (\alpha - a) = 0 \\
&T' \cdot (\alpha - a) + T \cdot \alpha' = \kappa N \cdot (\alpha - a) + 1 = 0, \quad \text{so} \quad N \cdot (\alpha - a) = -\rho, \\
&N' \cdot (\alpha - a) + N \cdot \alpha' = (-\kappa T + \tau B) \cdot (\alpha - a) = -\rho' \quad \text{so} \quad B \cdot (\alpha - a) = -\sigma\rho', \\
&\big(-\sigma\rho'\big)' = \big(B \cdot (\alpha - a)\big)' = B' \cdot (\alpha - a) + B \cdot T = -\tau N \cdot (\alpha - a) = \rho\tau,
\end{aligned}$$

i.e., $\big(\sigma\rho'\big)' + \rho\tau = 0$.

(3C)⇒(3B): Assuming $\big(\sigma\rho'\big)' + \rho\tau = 0$,

$$\begin{aligned}
\big(\alpha + \rho N + (\rho'\sigma)B\big)' &= \alpha' + \rho' N + \rho N' + (\rho'\sigma)' B + (\rho'\sigma) B' \\
&= T + \rho' N + \rho(-\kappa N + \tau B) + (\rho'\sigma)' B + \rho'\sigma(-\tau N) \\
&= (1 - \kappa\rho)T + \rho'(1 - \tau\sigma)N + \big(\rho\tau + (\sigma\rho' + \rho\tau)'\big)B = 0.
\end{aligned}$$

(3B)⇒(3A): Assuming $\alpha + \rho N + (\rho'\sigma)B = a$, $\big((\alpha - a) \cdot (\alpha - a)\big)' = 2T \cdot (\alpha - a) = -T \cdot \big(\rho N + (\rho'\sigma)B\big) = 0$, so there is a constant distance from α to the constant a.

(4A)⇒(4C): Assume $|\alpha - a|$ and $\alpha \cdot b$ are both constant, $b \neq 0$, i.e., assume α is a sphere curve and α is planar, respectively. Without loss of generality, b is a unit vector. Note that any multiple of b can be added to a, since

$$|\alpha - (a + fb)|^2 = |\alpha|^2 + 2\alpha \cdot (a + fb) + |a + fb|^2 = |\alpha + a|^2 + 2f\alpha \cdot b + |a + fb|^2 - |a|^2,$$

and the left is constant if and only if the right is. Choose f so that a is in the same plane as α, i.e., $a \cdot b = \alpha \cdot b$, or equivalently, $(\alpha - a) \cdot b = 0$. As in (3A)⇒(3C), differentiating $(\alpha - a) \cdot (\alpha - a) = 0$ gives $T \cdot (\alpha - a) = 0$ and $N \cdot (\alpha - a) = \rho$. Also $(\alpha - a) \cdot b = 0$ gives $T \cdot b = N \cdot b = 0$ so $B = \pm b$, $B \cdot (\alpha - a) = 0$, and

$$|\alpha - a|^2 = \big(T \cdot (\alpha - a)\big)^2 + \big(N \cdot (\alpha - a)\big)^2 + \big(B \cdot (\alpha - a)\big)^2 = \rho^2,$$

implying $\kappa = 1/\rho$ is constant.

(4C)⇒(4A): Assume $\alpha(s)$ has constant nonzero curvature κ and torsion $\tau = 0$. The circle $c(s)$ defined by

$$x = \frac{1}{\kappa}\cos(\kappa s), \quad y = \frac{1}{\kappa}\sin(\kappa s), \quad z = 0, \qquad t \in (a, b)$$

has curvature κ and torsion $\tau = 0$. By the fundamental theorem for curves, $\alpha = Ac(s) + B$ for some $(A, b) \in \mathrm{SE}(3)$, and it follows easily that α is a circle.

(4B)⇔(4C): $B' = -\tau N$ so B is constant if and only if $\tau = 0$, so for either of (4B)⇒(4C) or (4B)⇐(4C) one can assume $\tau = 0$ and B is constant. Then $(\alpha + \rho N)' = T + \rho N' + \rho' N = T + \rho(-\kappa T + \tau B) + \rho' N = \rho' N$, which shows $\alpha + \rho N$ is constant if and only if ρ is constant, and $\kappa = 1/\rho$.

(5A)⇒(5B): Differentiating $\alpha(t) = \bar{\alpha}(t) + at$ gives $\alpha' = \bar{\alpha}' + a$ where $|\bar{\alpha}'| = 1$ and $\bar{\alpha}' \cdot a = 0$. Thus $|\alpha'|^2 = 1 + |a|^2$ and

$$T \cdot a = \frac{1}{|\alpha'|}\,\alpha' \cdot a = \frac{1}{\sqrt{1 + |a|^2}}\,\big(\bar{\alpha}' + a\big) \cdot a = \frac{|a|^2}{\sqrt{1 + |a|^2}}\,,$$

which is constant.

(5B)⇒(5A): Without loss of generality, α is unit speed and a is a unit vector. Let h be the constant $T\cdot a$ and define $\bar\alpha(s) = \alpha(s) - (ha)s$. Then $\alpha' = \bar\alpha' + ha$, so $\bar\alpha'\cdot a = \alpha'\cdot a - ha\cdot a = h - h = 0$, and $1 = |\alpha'|^2 = |\bar\alpha'|^2 + h^2$, i.e., $|\bar\alpha'| = \sqrt{1-h^2}$, a constant. If $\bar\alpha'(s)$ is ever zero, then it is zero for all s, and $\alpha(s) = (ha)s$, contradicting the assumption that the curvature of α is never zero. Thus $\bar\alpha$ is immersed and has constant, nonzero speed. Integrating $\bar\alpha'(s)\cdot a = 0$ implies $\bar\alpha\cdot a$ is constant, so $\bar\alpha(s)$ is planar. Reparametrize $\alpha(s)$ and $\bar\alpha(s)$ by $t = \sqrt{1-h^2}\,s$, so $\alpha(t) := \alpha(t/\sqrt{1-h^2})$ and $\bar\alpha(t) := \bar\alpha(t/\sqrt{1-h^2})$. Substituting into $\alpha(s) = \bar\alpha(s) + s(ha)$ gives $\alpha(t) = \bar\alpha(t) + (h/\sqrt{1-h^2}\,a)t$ and $\bar\alpha(t)$ is unit speed in a plane with normal a.

(5B)⇒(5C): Without loss of generality, assume $\alpha(s)$ is unit speed. Differentiating $T\cdot a$ constant gives $T'\cdot a = 0$, so $N\cdot a = \big(T'/\kappa\big)\cdot a = 0$, $N'\cdot a = 0$, and $0 = N'\cdot a = (-\kappa T + \tau B)\cdot a = -(T\cdot a)\kappa + (B\cdot a)\tau$. This implies κ and τ have constant ratio because $T\cdot a$ is constant, and $B\cdot a$ is constant follows from $(B\cdot a)' = B'\cdot a = -\tau N\cdot a = 0$.

(5C)⇒(5B): Let $h = \tau/\kappa$, a constant, and define $a = hT + B$. Then $a' = hT' + B' = h(\kappa N) + (-\tau N) = (h\kappa - \tau)N = 0$ so a is constant and $T\cdot a = T\cdot(hT+B) = h$.

(6A)⇒(6B): A circular helix is a helix so, by (5), $T\cdot a$ is constant. Assuming $\alpha(t) = \bar\alpha(t) + at$ where $a\cdot\alpha$ is constant and $\bar\alpha$ is a circle in a plane orthogonal to a, it follows that $\alpha' = \bar\alpha' + a$ and $\bar\alpha'\cdot a = 0$ so $|\alpha'|^2 = 1 + |a|^2$ and $v = |\alpha'|$ is constant. So $\boldsymbol{a} = \alpha'' = \kappa v^2 N = \bar\alpha'' = \bar\kappa\bar N$, so $\kappa v^2 = \bar\kappa$ and $N = \bar N$, where the bars indicate the Serret-Frenet data of $\bar\alpha$. Thus $a\times(\alpha + \rho v^2 N) = a\times(\bar\alpha + at + \rho/v^2 N) = a\times(\bar\alpha + \bar\rho\bar N + at) = a\times(\bar\alpha + \bar\rho\bar N)$ which is constant since $\bar\alpha$ is a circle.

(6B)⇒(6A): If $T\cdot a$ is constant then by (5) α is a helix, and $\alpha(t) = \bar\alpha(t) + at$ where $\bar\alpha$ is in a plane with normal a. As in (6A)⇒(6B), $v = |\alpha'|$ is constant, $\kappa v^2 = \bar\kappa$, and $N = \bar N$. Since $a\times(\alpha + \rho/v^2 N)$ is constant, it follows that $\alpha + \rho/v^2 N = c + f(t)a$ where c is constant and $f(t)$ is a function of t. Then $\bar\alpha = \alpha - at = c + fa - \rho/v^2 N - at = c + fa - \bar\rho\bar N - at$, and $a\cdot\bar\alpha = a\cdot c + (f - t)|a|^2$ is constant since α in in a plane with normal a. Thus $f = t$ and $\bar\alpha = c - \bar\rho\bar N$, i.e., $\bar\alpha + \bar\rho\bar N$ is constant.

(6A)⇒(6C): As in (6A)⇒(6B), from $\alpha = \bar\alpha + at$ follows $v = |\alpha'|$ is constant and $\kappa v^2 = \bar\kappa$. Thus κ is constant, and τ/κ is constant by (5), so τ is also constant.

(6C)⇒(6A): If κ and τ are constant then so is τ/κ, so α is a helix which can be parametrized as $\alpha(t) = \bar\alpha(t) + at$, as in (5). As in (6A)⇒(6B), v is constant and $\kappa v^2 = \bar\kappa$, so $\bar\kappa$ is constant. Thus $\bar\alpha$ is a circle. $\blacksquare$

§29. Appendix 2: Foundations

Differential geometry is a generalization of multivariate calculus. Some essential elements of multivariate calculus are reviewed here.

Many central results of differential geometry are applications three general theorems:

1. *the inverse function theorem*, which asserts the existence of smooth solutions to equations;
2. *the existence and uniqueness theorem for ordinary differential equations*, which asserts unique solutions for the initial value problems of systems of ordinary differential equations; and
3. *the Frobenius theorem*, which, generalizing the computation of a potential from a conservative vector field, asserts unique solutions to a particular kind of system of partial differential equations.

§30. Euclidean space, and the vector space $\mathbb{R}^n$

Euclidean space $\mathbb{R}^n$: By $\mathbb{R}^n$ is meant the set of n-tuples $x = (x_1, x_2, \ldots, x_n)$, where $x_i \in \mathbb{R}$. The tuples are thought of as *places*. No particular meaning can generally be assigned to the sum of two places eg to the sum of map coordinates of, say, Los Angeles and New York, so addition of n-tuples is undefined when they are thought this way. In this context, $\mathbb{R}^n$ is a *metric space*: the *distance* between two n-tuples x and y is $\sqrt{(x_1-y_1)^2 + \cdots + (x_n-y_n)^2}$.

The vector space $\mathbb{R}^n$: $\mathbb{R}^n$ also denotes the vector space of column vectors with n rows with componentwise vector addition and scalar plication. In this context $\mathbb{R}^n$ is an *inner product space*, with inner product $v\cdot w := v^T w = v_1w_1 + \cdots v_nw_n$. The *length* of a vector v is $|v| := \sqrt{v\cdot v}$. The *Cauchy-Schwartz inequality* states $|v\cdot w| \le |v||w|$; the *triangle inequality* states that $|v + w| \le |v| + |w|$. The following fact is often used: if $v_1, \ldots, v_n \in \mathbb{R}^n$ are pairwise orthonormal unit vectors ie $v_i\cdot v_j$ is 1 if $i = j$ and 0 otherwise, then any vector $x\in\mathbb{R}^n$ is uniquely a sum of the v_i by the formula

$$x = (v_1\cdot x)v_1 + (v_2\cdot x)v_2 + \cdots + (v_n\cdot x)v_n. \tag{31}$$

Such a tuple $(v_1, \ldots, v_n)$ of vectors is called an *orthonormal basis*.

The Cauchy-Schwartz inequality follows by expanding $(tv+w)\cdot(tv+w) \geq 0$ and putting $t = -v\cdot w/|v|^2$ (which is the value of t for which that expression has a global minimum). The triangle inequality follows from the Cauchy-Schwartz via $|v+w|^2 = (v+w)\cdot(v+w) = |v|^2+2v\cdot w+|w|^2$. Pairwise orthonormal vectors v_i are *linearly independent*: the inner product of $c_1v_1+c_2v_2+\cdots+c_nv_n = 0$ and v_i gives $c_1v_i\cdot v_1+c_2v_i\cdot v_2+\cdots+c_nv_i\cdot v_n = c_i = 0$. n linearly independent vectors in $\mathbb{R}^n$ form a *basis*: any vector is uniquely a linear combination of them. Dotting $x = a_1v_1+a_2v_2+\cdots+a_nv_n$ with v_i gives $v_i\cdot x = a_1v_i\cdot v_1 + a_2v_i\cdot v_2 + \cdots + a_nv_i\cdot v_n = a_i$.

The basic constructs in differential geometry — curves, surfaces, manifolds, and submanifolds — are locally Euclidean. Linearization is an essential theme, the natural carrier of which is a vector space. Every vector space is the same as $\mathbb{R}^n$ after choosing a basis. The triangle inequality is an often used tool to bound a sum of vectors in terms of the summands.

Summing places is nonsensical, but the difference of two places may not be: it does make sense to regard the difference of places x and y in $\mathbb{R}^n$ as the vector v, where $v_i = y_i - x_i$. That is the *vector from x to y* and it is typically visualized as an arrow with tail at x and head at y. The distance between x and y is the length of the vector v. It also makes sense to sum a vector v and a place x to obtain the new place $y = x + v$. The vectors like v are steps between the places. While the conceptual distinction between steps and places is significant, both have the same name $\mathbb{R}^n$, and elements of $\mathbb{R}^n$ in either of the two meanings are sometimes written as a tuple, to save vertical space on the page, and sometimes written as columns, so that the usual matrix multiplication can be applied to them with the matrix on the left. A linear mapping α on vectors with values in $\mathbb{R}$ is regarded as a row vector, so that its value on a vector v is the matrix product αv. The set of such *dual vectors* is a vector space which is conceptually different from the metric space of places, but those are both n-tuples written as rows. *Differential geometry is conceptual. You have to get used to carrying the burden of intepretation. Elements of $\mathbb{R}^n$ may represent conceptually different objects, and you have to fully understand what element is representing which kind of objects.*

Open and closed sets; continuous functions: The *open balls* and *closed balls* around $x \in \mathbb{R}^n$ of radius r are respectively

$$B_r(x) \coloneqq \{y \in \mathbb{R}^n \bigm| |y-x| < r\}, \qquad \bar{B}_r(x) \coloneqq \{y \in \mathbb{R}^n \bigm| |y-x| \leq r\}.$$

$A \subseteq \mathbb{R}^n$ is an *open subset* of $\mathbb{R}^n$ if it contains an open ball around each of its points ie if for all $x \in A$ there is an r such that $B_r(x) \subseteq A$. $K \subseteq \mathbb{R}^n$ is a *closed subset* if $\mathbb{R}^n \setminus K$ is open. A function $f\colon A \to \mathbb{R}^m$ is *continuous* at a if $|f(x) - f(a)|$ is arbitrarily small whenever $|x - a|$ is sufficiently small; f is *continuous* if it continuous at all a. Some useful facts:

1. The empty set and the whole of $\mathbb{R}^n$ are open, the union af open sets is open, and the intersection of finitely many open sets is open.
2. The empty set and the whole of $\mathbb{R}^n$ are closed, the intersection af closed sets is closed, and the union of finitely many closed sets is closed.
3. K is a closed subset of $\mathbb{R}^n$ if and only if all convergent sequences in K converge to an element of K itself ie if $x_i \in K$ and $\lim_{i\to\infty} x_i = x$ then $x \in K$.
4. If $f\colon A \to \mathbb{R}$ is continuous, and A is open, then $\{x \in A \mid f(x) > 0\}$ is open.
5. If $f\colon A \to \mathbb{R}$ is continuous, and A is closed, then $\{x \in A \mid f(x) \geq 0\}$ is closed.

Open sets express *membership stability*: if A is open, $a \in A$, and a is moved a little, then still $a \in A$. Also, it is often important that a point in a set is surrounded by other points in the same set. For example, the derivative of a function $f(x)$ at $x = a$ is defined as

$$\frac{df}{dx} = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h},$$

where the limit means that for all $\epsilon > 0$ there is a $\delta > 0$ such that $\bigl|f(a+h) - f(a)\bigr| < \epsilon$ whenever $x \in \operatorname{domain} f$ and $\bigl|h\bigr| < \delta$. If there is a disk about $x = a$ which meets the domain of f only at $x = a$ then the limit definition is vacuously satisfied for any value of the limit and the derivative is not unique. So, the usual expositions restrict the definition of the partial derivative to functions defined only on open intervals where the limit is unique if it exists at all, rather than have to postulate such uniqueness in the definition of the derivative itself. An important use of closed sets: if elements of a convergent sequence have a property, and the set of all elements with that property is known to be closed, then the limit of the sequence also has the property.

These basic topological notions and results may be found in standard advanced calculus texts, such as [5, 6, 7]. The proofs are easy: For (3), if K is closed, and $x_i \in K$ and $\lim_{i\to\infty} x_i = x$, then $x \neq K$ is a contradiction because $\mathbb{R}^n \setminus K$ is open and x_i could not be within some radius of x. For (4), suppose $f(a) > 0$, choose $0 < \epsilon < f(a)$, find $\delta_1 > 0$ such that $|f(x) - f(a)| < \epsilon$ whenever $|x - a| < \delta_1$, and find $\delta_2 > 0$ such that $B_{\delta_2}(a) \subseteq A$. Setting $\delta \coloneqq \min(\delta_1, \delta_2)$, if $x \in B_\delta(a)$ then $x \in A$ and $f(x) > f(a) - \epsilon > 0$.

The special Euclidean group: An *affine map* of $\mathbb{R}^3$ is a linear map followed by a translation ie a map of the form $x \to Ax + b$ where A is a 3×3 matrix and b is a vector. An affine map $Ax + b$ is a *Euclidean isometry* if it preserves distance ie if $|(Ax + b) - (Ay + b)| = |A(x - y)| = |x - y|$. A tuple (v_1, v_2, v_3) of vectors in $\mathbb{R}^3$ is *right-handed* if the 3×3 matrix $[v_1\, v_2\, v_3]$ with columns v_1, v_2, v_3 has positive determinant, and *left-handed* if it is negative. There are exactly three kinds of such tuples: left-handed, right-handed, and, where the determinant is zero, *degenerate*. An affine map $Ax + b$ is *proper* if it preserves handedness ie (Av_1, Av_2, Av_3) is right handed whenever (v_1, v_2, v_3) is. The set of proper Euclidean isometries of $\mathbb{R}^3$ is called the *special euclidean group* and it is denoted $SE(3)$. Note that $A^T A = \mathbf{1}$ implies $A^T = A^{-1}$ and hence $AA^T = AA^{-1} = \mathbf{1}$, and also note that $A^T A = \mathbf{1}$ is equivalent to $(Au) \cdot (Av) = u \cdot v$ of all u and v.

§32. **Proposition**.

1. *An affine map $Ax + b$ is proper if and only if $\det A > 0$. It is a Euclidean isometry if and only if $A^T A = \mathbf{1}$, where $\mathbf{1}$ is the identity matrix. It is a proper Euclidean isometry if and only if $A^T A = \mathbf{1}$ and $\det A = 1$.*
2. *For a 3×3 matrix A, $\det A = 1$ and $A^T A = \mathbf{1}$ if and only if there is an orthonormal basis with respect to which A has the matrix*

$$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Proof.

For (1), if (v_1, v_2, v_3) is a right-handed tuple then

$$\det([Av_1, Av_2, Av_3]) = \det(A\,[v_1, v_2, v_3]) = \det(A)\det([v_1, v_2, v_3]),$$

so $Ax + b$ is a proper affine map if and only if $\det A > 0$. If $Ax + b$ is an isometry then the distance between 0 and any $u \in \mathbb{R}^3$ is preserved, so

$$|u|^2 = |u - 0|^2 = |(Au + b) - (A\,0 + b)|^2 = |Au|^2,$$

and $u^T u = |Au|^2 = (Au)^T(Au) = u^T(A^T A)u$ for all u. It follows that, if $u, v \in \mathbb{R}^3$ then

$$\begin{aligned} 0 &= (u + v)^T(A^T A)(u + v) - (u + v)^T(u + v) \\ &= (u^T(A^T A)u + u^T(A^T A)v + v^T(A^T A)u + v^T(A^T A)v) - (u^T u + u^T v + v^T u + v^T v) \\ &= 2u^T(A^T A - \mathbf{1})v. \end{aligned}$$

Setting u and v to the coordinate vectors $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$, respectively, obtains that the general (i, j) element of $A^T A - \mathbf{1}$ is zero, so $A^T A = \mathbf{1}$. The converse, that if $A^T A = \mathbf{1}$ then $Ax + b$ is a Euclidean isometry, is a direct calculation. Also, if $A^T A = \mathbf{1}$, then $\det(A)^2 = \det(A^T A) = \det \mathbf{1} = 1$, so $\det A > 0$ if and only if $\det A = 1$.

For (2), using $A^T A = \mathbf{1}$ and $\det A = 1$,

$$p(\lambda) = \det(\lambda\mathbf{1} - A) = \det(\lambda\mathbf{1} - A)^T = \det(\lambda\mathbf{1} - A^T) = \det(-\lambda A^{-1}((1/\lambda)\mathbf{1} - A)) = -\frac{1}{\lambda^3}\,p(1/\lambda),$$

and in particular $p(1) = 0$, so there is a unit vector, say u_3, such that $Au_3 = u_3$. Let u_1 and u_2 be such that (u_1, u_2, u_3) is a right handed orthonormal basis. The matrix $U \coloneqq [u_1, u_2, u_3]$ is also in $SO(3)$, and $u_3 \cdot Au_1 = Au_3 \cdot Au_1 = u_3 \cdot u_1 = 0$, $u_1 \cdot Au_3 = u_1 \cdot u_3 = 0$, and similarly $u_3 \cdot (Au_2) = 0$ and $u_2 \cdot (Au_3) = 0$, so the matrix of A with respect to the basis (u_1, u_2, u_3) is

$$U^T AU \coloneqq \begin{bmatrix} u_1 \cdot Au_1 & u_1 \cdot Au_2 & u_1 \cdot Au_3 \\ u_2 \cdot Au_1 & u_2 \cdot Au_2 & u_2 \cdot Au_3 \\ u_3 \cdot Au_1 & u_3 \cdot Au_2 & u_3 \cdot Au_3 \end{bmatrix}^T = \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Premultiplication by the transpose gives $a^2+c^2=b^2+d^2=1$ and $ab+cd=0$. Choose θ such that $a=\cos\theta$ and $c=\sin\theta$. The (unit) 2-vector (b,d) is orthogonal to (a,c), so there are the two possibilities

$$(b,d)=(-\sin\theta,\cos\theta) \quad \text{or} \quad (b,d)=(\sin\theta,-\cos\theta),$$

but the second of these gives $\det(U^TAU)=-1$. ∎

Euclidean isometries are everywhere: every time you go somewhere, your new position is approximately a proper Euclidean isometry applied to your prior. They are also the foundation of Euclidean geometry eg two triangles are *congruent* if and only if they can be mapped to one another using a proper Euclidean isometry.

§33. Multivariable calculus

C^r functions: The function $y\colon A\subseteq\mathbb{R}^m\to\mathbb{R}^n$ given by n real valued *component functions*

$$y_1=y_1(x_1,\dots,x_m),\ \dots\ ,y_n=y_n(x_1,\dots,x_m).$$

is called C^r, $r\geq 1$, if A is open and the partial derivatives of the component functions exist and are continuous up to and including order r. C^0 is synonymous with continuous, and C^∞ means C^r for all $r\geq 0$. A *C^r diffeomorphism* is a C^r map with a C^r inverse. A *homeomorphism* is any continuous map with a continuous inverse (A does not have to be open).

Derivatives: The derivative of $y\colon A\subseteq\mathbb{R}^m\to\mathbb{R}^n$ is the $n\times m$ matrix $Dy:=\left[\dfrac{\partial y_i}{\partial x_j}\right]$. If $z(y)$ is is another function

$$z_1=z_1(y_1,\dots,y_n),\dots,z_p=z_p(y_1,\dots,y_n),$$

and both z and y are C^r, then, by the chain rule, the result of substituting the functions y_j into z is also C^r, and

$$\frac{\partial z_k}{\partial x_i}=\frac{\partial z_k}{\partial y_1}\frac{\partial y_1}{\partial x_i}+\dots+\frac{\partial z_k}{\partial y_n}\frac{\partial y_n}{\partial x_i},$$

where the meaning is that equality is obtained after substituting the y in the functions $\partial z_k/\partial y_j$.

Any function from $A\subseteq\mathbb{R}^m$ to $\mathbb{R}^m$ defined by a rule consisting of the operations $+$, $-$, $\times$, $\div$, and substitution, and the including functions a^x, $\log_a x$, trigonometric functions, power series, and any other C^r function, is continuous on its domain and C^r on any open subset of $\mathbb{R}^m$ contained in its domain. Care has to be exercised with functions such as $\sqrt{x^2}$, the rule for which consists of such ordinary operations, but which is not differentiable at $x=0$ (it is equal to $|x|$, the classic example of a non-differentiable continuous function). To understand this, remember that x^a is defined as $e^{a\ln x}$, so $\sqrt{x^2}=e^{\frac{1}{2}\ln(x\times x)}$ which is C^∞ on any open subset on which it is defined, and so is C^∞ on $\{x\mid x\neq 0\}$. The differentiability of $\sqrt{x^2}$ at $x=0$ is not assured by the formula $\sqrt{x^2}=e^{\frac{1}{2}\ln(x\times x)}$ since that is not defined at $x=0$. Of course, with any of its common meanings for non-integral a, x^a is continuous on its domain of definition, C^∞ except at $x=0$, and, for $a>0$, C^k at $x=0$, where $k=\lfloor a\rfloor$, if it is defined for both positive and negative x.

If $f(x,y):=xy/\sqrt{x^2+y^2}$ for $(x,y)\neq 0$, and $f(0,0):=0$, then $\partial f/\partial x$ and $\partial f/\partial y$ exist at all (x,y) but are not continuous at $(0,0)$; and also, if the graph is rotated by 45° then the partial derivatives of the rotated graph do not exist at $(0,0)$ (along the line $y=x$ the function is $f(x,x)=x^2/\sqrt{2x^2}=|x|/\sqrt{2}$). So mere existence of partial derivatives is *coordinate dependent*. On the other hand, composition of a C^1 function with a C^1 diffeomorphism is also C^1 (so in particular composition with a linear change of coordinates, such as a rotation, does not destroy the C^1 property). *The C^r property is independent of coordinates: if a function has that property in one coordinate system, then it has that same property in all coordinate systems.*

The class of C^r functions provides a convenient framework for many results; for example, a standard result in multivariate calculus is that C^1 functions are differentiable, in the sense that they are well approximated by their tangent planes. Also, if a function is C^r, $r\geq 2$, and $0\leq k\leq r$, then any partial derivative of order k is independent of the order of differentiation. The chain rule is needed to compute derivatives through substitutions, something which happens alot.

Taylor expansions, big and little oh notation: If $f(x)$ is C^r, $r \geq 1$, then the Taylor formula is

$$f(x+h) = f(x) + \frac{f^{(1)}(x)}{1!}\, h + \cdots + \frac{f^{(r-1)}(x)}{(r-1)!}\, h^{r-1} + \left(\int_0^1 \frac{(1-t)^{r-1}}{(r-1)!}\, f^{(r)}(x+th)\, dt \right) h^r.$$

There are many variations — it is sharper to rely on this most basic one, adapting it to various purposes. The Taylor formula is obtained by repeated integration by parts ($u = f^{(r)}(x+th)$ and $dv = (1-t)^{r-1}/(r-1)!\, dt$), stopping at the order $r-1$ polynomial when the differentiability of the function f under the integral is exhausted. An order r polynomial can be recovered from the identity

$$\frac{f^r(x)}{r!} = \int_0^1 \frac{(1-t)^{r-1}}{(r-1)!}\, f^{(r)}(x)\, dt\,,$$

and then adding and subtracting

$$f(x+h) = f(x) + \frac{f^{(1)}(x)}{1!}\, h + \cdots + \frac{f^{(r)}(x)}{r!}\, h^r + R_r(x,h),$$

where the *remainder* is

$$R_r(x,h) \coloneqq h^r \int_0^1 \frac{(1-t)^{r-1}}{(r-1)!} \big(f^{(r)}(x+th) - f^{(r)}(x) \big)\, dt.$$

The integrand is zero when $h = 0$ and so the remainder is expected to be small when h is small.

Often, the detailed functional form of the remainder is not of much interest. Since $R_r(x,h)h^r/h^r \to 0$ as $h \to 0$, Taylor's formula may be written

$$f(x+h) = f(x) + \frac{f^{(1)}(x)}{1!}\, h + \cdots + \frac{f^{(r)}(x)}{r!}\, h^r + o(h^r),$$

where $o(h^r)$ means some (unspecified) function of x and h which satisfies $o(h^r)/h^r \to 0$ as $h \to 0$ (little oh notation). If $f(x)$ is C^{r+1} then applying Taylor's formula at order $r+1$

$$\begin{aligned} f(x+h) &= f(x) + \frac{f^{(1)}(x)}{1!}\, h^r + \cdots + \frac{f^{(r)}(x)}{r!}\, h^r + \left(\frac{f^{(r+1)}(x)}{r!} + R_{r+1}(x,h) \right) h^{r+1} \\ &= f(x) + \frac{f^{(1)}(x)}{1!}\, h + \cdots + \frac{f^{(r)}(x)}{r!}\, h^r + O(h^{r+1}), \end{aligned}$$

where $O(h^{r+1})$ denotes a continuous function times h^{r+1} (big oh notation). The big oh notation has the advantage that it explicitly incorporates the power h^{r+1}, but it requires one more degree of smoothness than the little oh. There are more-or-less obvious manipulations: for example,

$$\begin{aligned} &h^2 o(h^3) = o(h^5) && \text{because} && h^2 o(h^3)/h^5 = o(h^3)/h^3 \to 0, \\ &o(h^3)^2 = o(h^6) && \text{because} && o(h^3)^2/h^6 = \big(o(h^3)/h^3\big)^2 \to 0, \\ &o(h^3) + o(h^5) = o(h^3) && \text{because} && \big(o(h^3) + o(h^5)\big)/h^3 = o(h^3)/h^3 + h^2\, o(h^5)/h^5 \to 0. \end{aligned}$$

However $o(h^r) - o(h^r) = 0$ is false because the two little oh's generally represent different functions.

The remainder in Taylor's theorem is essential in understanding the local error in polynomial approximations, or where information about a function is to be extracted from its derivatives eg a positive Hessian at a critical point implies that the actual function has a local minimum. The remainder is also be important fir intuition when locally best-fitting two curves. For example, in elementary calculus, the tangent line is the best linear approximation to a curve at a point because it is the approximation where the error falls faster than linear. The error of the tangent approximation is $o(h)$, whereas any other linear approximation has error $o(h^0)$. The ratio of the first kind of error to the second kind of error falls to zero for small h, so the tangent line approximation is *infinitely better for small* h than any other linear approximation.

Riemann integral: Let f be a function defined on $[a, b]$. A *Riemann sum* is a sum of the form $\sum_{i=1}^{n} f(x_i^*)\,\Delta x_i$ where $x_i \in [a, b]$, $i = 0, \ldots, n$ are such that $a = x_0 < x_1 < x_2 < \ldots < x_n = b$, $x_i^* \in [x_{i-1}, x_i]$, and $\Delta x_i \coloneqq x_i - x_{i-1}$. f is *Riemann integrable* if the limit of its Riemann sums exists, in the sense that they are arbitrarily near some finite number if all the Δx_i are sufficiently small, and then the limit is the *integral* of f. The central result is that any continuous function is Riemann integrable. The following *approximation result* is useful:

§34. **Lemma**. *If $e(x, u)$ is continuous, defined on an open set containing $[a, b] \times \{0\}$, and $e(x, 0) = 0$ for $x \in [a, b]$, then $\lim \sum_{i=1}^{n} e(x_i, \Delta x_i)\,\Delta x_i = 0$, where the limit is in the sense of the Riemann integral.*

The Riemann integral is a basic tool. The approximation lemma is convenient for deriving the arc-length formula for curves.

Lemma 34 is useful in the many situations where one wants to recognize a sum of a local physical quantities as an integral, and one can approximate,

$$\text{local physical quantity} \approx f(x)\,\Delta x,$$

or more precisely

$$\text{local physical quantity} = f(x)\,\Delta x + (\text{error depending on } x,\ \Delta x).$$

For example, by definition, mass is linear density times length, *in the case that the linear density is constant.* In the presence of a *variable* density $\varrho(x)$, the mass of a length from x to $x + \Delta x$ is $\varrho(x)$ times Δx for small Δx, plus some small error accounting for the variation of the density over Δx, and the total mass is the sum of such. If the error is $e(x, \Delta x)\,\Delta x$, then by Lemma 34

$$\begin{aligned}
\text{total physical quantity} &= \lim \sum_i \Big(f(x_i)\,\Delta x_i + e(x_i, \Delta x_i)\,\Delta x \Big) \\
&= \lim \sum_i f(x_i)\,\Delta x_i + \lim \sum_i e(x_i, \Delta x_i)\,\Delta x \\
&= \int_a^b f(x)\,dx,
\end{aligned}$$

where the limits are the sense of Riemann sums, and so the total physical quantity is recognized as an integral. This same argument proves the fundamental theorem of calculus: if $f(x)$ is C^1 then, by the definition of the derivative, $f(x + u) - f(x) = f'(x)u + e(x, u)$ where $\lim_{u \to 0} e(x, u)/u = 0$, and so

$$\begin{aligned}
f(b) - f(a) &= \lim_{x \to b^-} f(x) - f(a) \\
&= \lim \sum_i f(x_i) - f(x_{i-1}) \\
&= \lim \sum_i \big(f'(x_i)\,\Delta x_i + e(x_{i-1}, \Delta x_i)\,\Delta x_i \big) \\
&= \lim \sum_i f'(x_i)\,\Delta x_i + \lim \sum_i e(x_{i-1}, \Delta x_i)\,\Delta x_i \\
&= \int_a^b f'(x)\,dx.
\end{aligned}$$

Expositions of the Riemann integral can be found in any first year calculus text. The approximation result relies on uniform continuity of a continuous function on a closed and bounded set: Given any $\epsilon > 0$, uniform continuity of $e(x, u)$ on a set of the form $[a, b] \times [0, c]$ implies a δ such that $|e(x, u)| < \epsilon$ for all $x \in [a, b]$ and all $0 \leq u < \delta$. For $\max_i \Delta x_i < \delta$,

$$\left| \sum_{i=1}^{n} e(t_i, \Delta x_i)\,\Delta x_i \right| \leq \sum_{i=1}^{n} \epsilon\,\Delta x_i = (b - a)\epsilon,$$

which is arbitrarily small for sufficiently small ϵ.

Multivariable Riemann integral: If $A \subseteq \mathbb{R}^n$ is closed and $A \subseteq [a,b] \times [c,d]$, and

$$\begin{aligned} &a = x_0 < x_1 < x_2 < \ldots < x_m = b, \qquad && x_i^* \in [x_{i-1}, x_i], \qquad && \Delta x_i := x_i - x_{i-1}, \\ &c = y_0 < y_1 < y_2 < \ldots < y_n = d, && y_j^* \in [y_{j-1}, y_j], && \Delta y_j := y_j - y_{j-1}, \end{aligned}$$

and $p_{ij} := (x_{ij}^*, y_{ij}^*) \in [x_{i-1}, x_i] \times [y_{j-1}, y_j]$, then the corresponding *Riemann sum* is

$$\sum_{p_{ij} \in A} f(p_{ij})\, \Delta x_i\, \Delta y_{ij},$$

and f is *Riemann integrable* if the limit of its Riemann sums exists, in the sense that they are arbitrarily near some finite number if all the Δx_i and Δy_j are small. The principal result is that f is integrable if f is continuous and the boundary of A can be contained by finitely many rectangles of arbitrarily small total area. The definitions and results are analogous for any number of variables.

§35. The three pillars of differential geometry

§36. **Theorem** (Inverse function theorem). *Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be C^r, $r \geq 1$, let $a \in A$ and $b = f(a)$, and suppose that the matrix $Df(a)$ is nonsingular. Then there are open neighbourhoods $U \ni a$ and $V \ni b$ such that $f|U$ is a C^r diffeomorphism between U and V. Moreover, for all $x \in U$, $D(f|U)^{-1}\big(f(x)\big) = \big(Df(x)\big)^{-1}$.*

For example, the map defined by the equations

$$u = x^2 + \sin(x+y), \qquad v = y^3 + \tan(xy)$$

is C^∞ from $A := \big\{(x,y) \,\big|\, xy \neq (2n+1)\pi/2\big\}$ to $\mathbb{R}^2 = \big\{(u,v)\big\}$, since (1) A is open because it is in the inverse image by the continuous $(x,y) \mapsto xy$ of the open set $\cdots(-\frac{5\pi}{2}, -\frac{3\pi}{2}) \cup (-\frac{3\pi}{2}, -\frac{\pi}{2}) \cup (\frac{\pi}{2}, \frac{3\pi}{2})\cdots$, and (2) it is defined by a formula involving algebraic operation and trigonometric functions. *But, could this map be a diffeomorphism?* To analyze this directly one would have the difficult task of solving the equations for u and v in terms of x and y, arriving at a formula also involving algebraic operations and trigonometric functions. The inverse function theorem provides a partial, local, answer: a map is a diffeomorphism near to some point in its domain if its derivative is invertible at that point.

Given $f\colon A \subset \mathbb{R}^m \to \mathbb{R}^m$ and $a \in A$, choose y near $a := f(b)$ and define $g(x) := f(x) - y$, so that inverting f at y is the same as solving the equation $g(x) = 0$. The existence part of the proof of the inverse function theorem is obtained by showing the convergence of the (multidimensional) Newton method $x_{i+1} = x_i - Dg(a)^{-1}\, g(x_i)$, with start $x_1 = a$. The proof that the inverse has the same differentiability as the original function is more difficult, and relies on estimates of changes in the inverse directly from its definition via the Newton iteration, inductively on r [5, 6].

Vector fields and differential equations: A *(time dependent) (parameterized) vector field* is a map

$$X\colon A \subseteq \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m = \{(x,t,\lambda)\} \to \mathbb{R}^n.$$

If $(x,t,\lambda) \in A$ then x is though of as a *place*, t as a *time*, λ as a parameter, and the the values of $X(x,t,\lambda)$ as a *velocity*; the vector field is a time and parameter dependent assignment of a *velocity* to *place*. An *integral curve at parameter λ of X starting at x at time s* is a differentiable curve function $c\colon I \to U$, where I is an *open* interval, such that

$$\frac{dc}{dt}(t) = X\big(t, c(t), \lambda\big), \qquad c(s) = x. \tag{37}$$

In words, the velocity of an integral curve must agree with that specified by the vector field. Equation (37) is the generic system of n differential equations, and the term integral curve is synonymous with the solution of the initial value problem. A *maximal integral curve* is an integral curve $c\colon I \to \mathbb{R}^n$ which does not admit an extension ie c is maximal if $\tilde{I} = I$ whenever $\tilde{c}\colon \tilde{I} \to \mathbb{R}^n$ is an integral curve with $I \subset \tilde{I}$ and $\tilde{c}|I = c$.

§38. **Theorem** (Ordinary differential equations existence and uniqueness). Suppose $U \subseteq \mathbb{R}^n$ is open, $I \subseteq \mathbb{R}$ is an open interval, $X(t, x, \lambda)$, $(t, x, \lambda) \in I \times U \times V$, is continuous in (t, x) (for each fixed λ), and the partial derivatives $\partial X/\partial x_i$ exist and are continuous. Then for every $(x, s, \lambda) \in U \times I \times V$ there is a unique maximal integral curve $c_{s,x,\lambda} \colon I_{s,x,\lambda} \to \mathbb{R}^n$ of X, at parameter λ starting at x at time s. Moreover

1. if X is C^r then the map defined by $(t, s, x, \lambda) \to c_{s,x,\lambda}(t)$ is also C^r; and
2. if $U = \mathbb{R}^n$ and X is linear in the variable x for each fixed λ and t, then $I_{(s,x,\lambda)} = (a, b)$ ie the maximal integral curve is defined for all the times that X is defined.

Vector fields provide a geometrically vivid way to work with systems of differential equations and their solutions, and so have a central role in scientific modeling. Also, there are a number of objects in differential geometry that are defined as the solutions to special differential equations. Simple differential equations can have remarkably complex solutions; a number of results in differential geometry rely on the construction of complicated maps defined as the solutions of relatively simple differential equations.

Local existence and uniqueness of integral curves is equivalent to local existence and uniqueness of a system of ordinary differential equations. Existence is proved by establishing the convergence of the Picard iteration

$$c_{i+1}(t) = x + \int_s^t X\big(u, c_i(u), \lambda\big)\, du, \qquad c_1(t) = x.$$

Uniqueness for times near s is proved using a technical result called Gronwald's Lemma, which is an (exponentially in time) estimate of the growth of deviations of initial conditions. Existence and uniqueness of maximal integral curves follows from local existence and uniqueness, and the connectedness of intervals. The standard proof of smoothness of the solution in the initial conditions and the parameter relies on an induction which calculates the derivatives as solutions of related differential equations. Expositions of the theory can be found in many texts [1, 2, 4, 8].

Invariant sets for differential equations: Let $X \colon U \times (a, b) \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a time dependent vector field and let $B \subseteq \mathbb{R}^n$. B is *X-invariant* if all integral curves of x starting at any point of B at any time remain in B for all times ie if for all integral curves $x(t)$, $x(t) \in B$ for one t implies $x(t) \in B$ for all t. A vector v is *tangent to B at $b \in B$* if $v = c'(0)$ for some curve $c(t) \in B$. X is *tangent to B* if, for all x and t, $X(x, t)$ is tangent to B at x. The following is quite useful in proving that a set is invariant:

§39. **Theorem**. *Let $X(x, t)$ be C^r vector field, $r \geq 1$. Then $B \subseteq \mathbb{R}^n$ is X-invariant if B is closed and X is tangent to B.*

It is common to have to show that solutions of a differential equation have some property if the initial condition has that property. For any particular differential equation and property there are usually a variety of ways to do this, but Theorem 39 provides a systematic and geometrically vivid method when it is possible to define B as the set of (x, t) with the property.

The proof uses the same Gronwald's Lemma that is used to show uniqueness of integral curves [3].

Given functions $f_{\alpha i}(x_1, \ldots, x_m, y_1, \ldots, y_n)$, the system of partial differential equations

$$\frac{\partial y_\alpha}{\partial x_i} = f_{\alpha i}(x_1, \ldots, x_m, y_1, \ldots, y_n) \qquad 1 \leq i \leq m,\ 1 \leq \alpha \leq n \tag{40}$$

for functions $y_\alpha(x_1, \ldots x_m)$ occurs importantly in a number of contexts. In these equations, *all first derivatives of some number of functions are specified as prescribed functions of their independent variables and the functions themselves.* In contrast, the more complicated partial differential equation (the Laplace equation)

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

specifies the value of a mixture of partial derivatives. Equality of mixed partial derivatives implies a constraint on those $f_{\alpha i}$ for which solutions to (40) exist: substituting a solution and differentiating gives

$$\frac{\partial^2 y_\alpha}{\partial x_j\, \partial x_i} = \frac{\partial f_{\alpha i}}{\partial x_j} + \sum_\beta \frac{\partial f_{\alpha i}}{\partial y_\beta} \frac{\partial y_\beta}{\partial x_j} = \frac{\partial f_{\alpha i}}{\partial x_j} + \sum_\beta f_{\beta j} \frac{\partial f_{\alpha i}}{\partial y_\beta}\,,$$

so that existence of solutions implies, for all α, i, and j, that

$$\frac{\partial f_{\alpha i}}{\partial x_j} + \sum_\beta f_{\beta j}\,\frac{\partial f_{\alpha i}}{\partial y_\beta} = \frac{\partial f_{\alpha j}}{\partial x_i} + \sum_\beta f_{\beta i}\,\frac{\partial f_{\alpha j}}{\partial y_\beta}\,. \tag{41}$$

The *Frobenius theorem* asserts that this necessary condition is also sufficient to solve the system (40).

§42. **Theorem** (Frobenius theorem). *Let $f_{\alpha i}\colon A \times B = \{(x,y)\} \subseteq \mathbb{R}^m \times \mathbb{R}^n$ be C^r functions, $r \geq 1$, and suppose conditions (41). Let $a \in A$ and $b \in B$. Then there are open neighbourhoods $U \ni a$ and $V \ni b$ such that, for all $u \in U$ and $v \in V$, there are unique C^r functions $y_\alpha\colon U \to \mathbb{R}^n$ satisfying equations (40) and $y_\alpha(u) = v_\alpha$.*

The Frobenius Theorem has great impact in applications: for example, this is the theorem used to show that a spacetime is locally flat under the condition of vanishing Riemann tensor. The system of partial differential equations which arises as a specification of all derivatives is natural and it is not so surprising that it occurs.

The Frobenius theorem is a generalization of the standard problem of finding a local potential φ for a vector field X eg finding $\varphi(x,y,z)$ such that

$$\frac{\partial\varphi}{\partial x} = P(x,y,z), \qquad \frac{\partial\varphi}{\partial y} = Q(x,y,z), \qquad \frac{\partial\varphi}{\partial z} = R(x,y,z). \tag{43}$$

Both (40) and (43) are systems of partial differential equations; for the potential problem (43), the functions of the right side do not involve the potential itself, but rather only involve independent variables. *Most potential problems do not have solutions, because equality of mixed partial derivative implies that, at any point (x,y,z) at which a C^2 solution exists,*

$$\frac{\partial P}{\partial y} = \frac{\partial^2\varphi}{\partial y\,\partial x} = \frac{\partial^2\varphi}{\partial x\,\partial y} = \frac{\partial Q}{\partial x}\,, \qquad \frac{\partial P}{\partial z} = \frac{\partial^2\varphi}{\partial z\,\partial x} = \frac{\partial^2\varphi}{\partial x\,\partial z} = \frac{\partial R}{\partial x}\,, \qquad \frac{\partial Q}{\partial z} = \frac{\partial^2\varphi}{\partial z\,\partial x} = \frac{\partial^2\varphi}{\partial x\,\partial z} = \frac{\partial R}{\partial z}\,.$$

For a vector field X, $\nabla \times X = 0$ is necessary for X to be conservative, and, if $\nabla \times X = 0$ then a potential can be obtained from X by a succession of simple integrations and substitutions. The proof of the Frobenius theorem is analogous, just successive solutions of ordinary differential equations replaces successive simple integrations.

There is some terminology surrounding Frobenius systems: System (40) is *integrable* or *exact* if, for all $a \in A$ and $b \in B$, there are functions y_α defined on an open neighbourhood of a such that $y_{\alpha i}(a) = b_\alpha$. It is *closed* or *involutive* if, for all $a \in A$ and all $b \in B$, there is a solution y_α defined on an open subset of a and satisfying $y_{\alpha i}(a) = b$. Then Theorem 42 asserts that a C^1 Frobenius system (40) is involutive if and only if it is integrable. As in eg [9], the modern form of the theorem is written in terms of vector fields and Lie-brackets.

References

[1] R. Abraham, J. E. Marsden, and T. S. Ratiu. *Manifolds, tensor analysis, and applications*, volume 75 of *Applied Mathematical Sciences*. Springer-Verlag, second edition, 1988.

[2] H. Amann. *Ordinary differential equations. An introduction to nonlinear analysis*. Walter de Gruyter, 1990. Translated from the German by Gerhard Metzen.

[3] H. Brezis. On a characterization of flow-invariant sets. *Comm. Math. Phys.*, 23:261–263, 1970.

[4] C. Chicone. *Ordinary differential equations with applications*. Springer, 2003.

[5] P. M. Fitzpatrick. *Advanced Calculus*. Brooks/Cole, second edition, 2006.

[6] J. E. Marsden and M. J. Hoffman. *Elementary classical analysis*. W. H. Freeman, second edition, 1993.

[7] J. E. Marsden and A. Tromba. *Vector Calculus*. W. H. Freeman, fifth edition, 2003.

[8] L. Perko. *Differential equations and dynamical systems*. Springer-Verlag, third edition, 2001.

[9] F. W. Warner. *Foundations of differentiable manifolds and Lie groups*. Spinger-Verlag, 1983.