Chapter 2
Foundations of Differentiability: Curves

The calculus of several variables deals with functions $f: \mathbb{R}^n \to \mathbb{R}^m$ or, more generally, with functions from an open subset of an n-dimensional normed vector space to an m-dimensional normed vector space. Several special cases leap to mind: $n = 1$, m arbitrary; n arbitrary, $m = 1$; $n = m$ arbitrary. In due course we shall address all of these, but the first special case permits us a gentle transition from Analysis I because the domain is one dimensional: One used to say we have only one independent variable.

1. Curves in Metric spaces and in Banach spaces

Indeed the simplest special case of functions $f: X \to Y$ with $X \subseteq \mathbb{K}^n$ and $Y \subseteq \mathbb{K}^m$ is that of $n = 1$, while the dimension m of the range space is arbitrary. Nevertheless, this special case is highly interesting. If the domain is one dimensional, we may allow more general circumstances with the range without creating additional difficulties. We know that $\mathbb{R}^m$ is a Banach space for any norm by 1.11 and 1.27. Thus we shall allow the range space to be an arbitrary Banach space E (≪espace≫) over $\mathbb{K} = \mathbb{R}$ or $= \mathbb{C}$. For many basic aspects we might just as well consider arbitrary metric spaces as range spaces. When we now discuss the differentiability of functions $f: I \to E$ where $I \subseteq \mathbb{R}$ is a real interval, then the theory is very close to the theory of differentiability of one variable calculus which we discussed in Analysis I. The first section of this chapter can therefore be considered as a bridge between Analysis I and general several variables differential calculus.

Definition 2.1. A *curve* f in a metric space X is a continuous function $f: I \to X$ from an interval $I \subseteq \mathbb{R}$ into the metric space X. This applies, in particular, to each Banach space E in place of X. □

Thus, in particular, a curve in $\mathbb{R}^n$ is nothing else than an n-tuple of continuous functions $f_j: I \to \mathbb{R}$ such that $f(t) = \big(f_1(t), \ldots, f_n(t)\big)$.

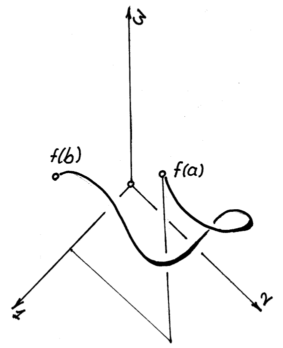

Figure 2.1

Often the number t in the argument of a curve is interpreted as a time variable and the vector $f(t) \in E$ as *position* of a point in the space E. Thus a curve represents a "motion" of a point in space. Since the metric space X and a Banach space E are very general concepts, the elements of X or E often represent the *state* of a complex system which may be characterised by several parameters describing physical, social, economic conditions that change in time. Then a curve describes the development of such a system in time. One often refers to it as a *dynamical system.*

In Theorem 1.35, for every Banach algebra A, for $a \in A$, and $T \geq 0$ we have seen a curve $t \mapsto \exp t{\cdot}a : [0,T] \to A$. In particular, if V is a finite dimensional Banach space such as $\mathbb{K}^n$ with some norm, for each $L\colon V \to V$ in $\mathrm{Hom}(V,V)$ we have curves $t \mapsto \exp t{\cdot}L : [0,T] \to \mathrm{Hom}(V,V)$, giving us curves $t \mapsto (\exp t{\cdot}L)(v_0) : [0,T] \to V$ for each $v_0 \in V$.

The simplicity of the intuition of curves is deceptive. There are some very strange curves. Peano and Hilbert discovered in the eighties of the 19th century that there are curves $f\colon [0,1] \to \mathbb{R}^2$ in the plane which fill out an entire triangle or square.

These curves are obtained as limits of a sequence of functions; the diagrams indicate how one might proceed to create a "space filling" curve from the unit interval onto the unit square as a uniform limit of a sequence of piecewise affine functions.

We shall see just a little later that a differentiable curve cannot be *space filling.*

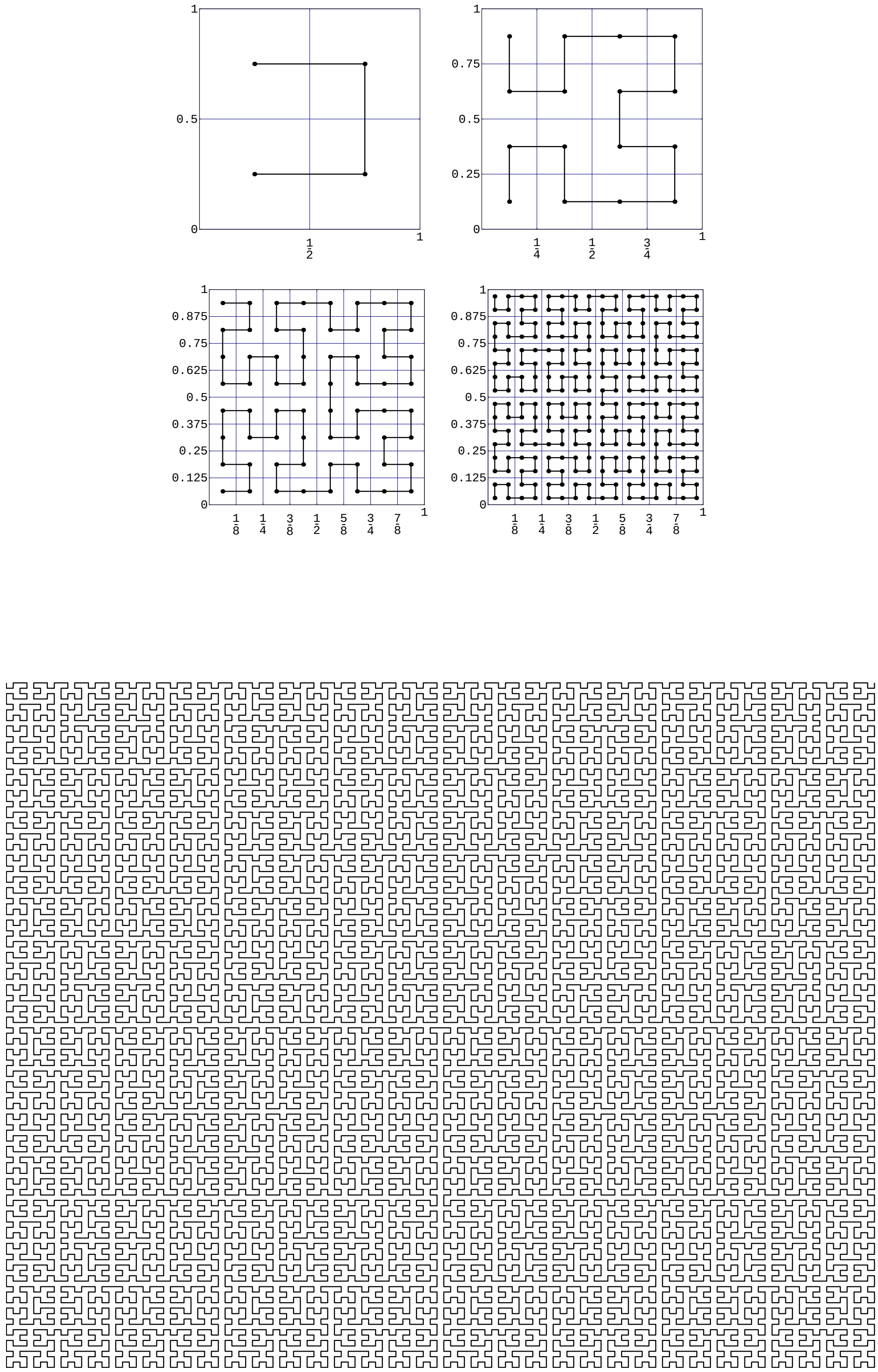
1
0.5
0
1/2
1
1
0.75
0.5
0.25
0
1/4
1/2
3/4
1
1
0.875
0.75
0.625
0.5
0.375
0.25
0.125
0
1/8
1/4
3/8
1/2
5/8
3/4
7/8
1
1
0.875
0.75
0.625
0.5
0.375
0.25
0.125
0
1/8
1/4
3/8
1/2
5/8
3/4
7/8
1

Figure 2.2

If a curve $f\colon I \to \mathbb{E}$ describes a "motion" of a "particle" in a normed space E such as $E = \mathbb{R}^3$, ordinary three space with the euclidean norm, then we would like to have a precise concept of a "velocity" with which the point progresses at a given time $t_0 \in E$, when the particle is at the point $f(t_0) \in E$.

Definition 2.2. Let $f\colon I \to E$ be a curve in a normed space. We say that it is *differentiable* at $t_0 \in I$ if there is a vector $v \in E$ and a curve $r\colon I \to E$ (both depending on t_0) such that

$$(1) \qquad f(t) = f(t_0) + (t - t_0)\cdot v + r(t) \text{ and } \lim_{\substack{t \to t_0 \\ t \neq 0}} (t - t_0)^{-1} r(t) = 0.$$

We shall say that f is *differentiable* if it is differentiable at all points of I. □

This says that near $f(t_0)$ the curve f is very close to the affine curve $t \mapsto t\cdot v + (f(t_0) - t_0\cdot v) : \mathbb{R} \to E$.

Proposition 2.3. *A curve $f\colon I \to E$ is differentiable at t_0 if and only if*

$$\lim_{\substack{t \to t_0 \\ t \neq t_0}} \frac{1}{t - t_0}\cdot\big(f(t) - f(t_0)\big)$$

exists. If this is the case then this limit and the vector $v \in E$ of (1) *agree. In particular, v is uniquely determined.*

Proof . Exercise. □

Exercise E2.1. Prove Proposition 2.3.
[Hint. The proof is easy. Its organisation may be modelled according to the proof of Theorem 4.7.]

We follow the elementary case situation of Definition 2.3 and call the vector v the *derivative* of the curve f at t_0 and denote it by $f'(t_0)$, $\dot{f}(t_0)$ or $\frac{df(t)}{dt}\Big|_{t=a}$. If the curve f is interpreted as a "motion" or a "dynamical system," we shall also call $\dot{f}(t_0)$ the *velocity*, respectively, the *rate of change* at time t_0.

If $E = \mathbb{K}^n$, then $f(t) = \big(f_1(t), \dots, f_n(t)\big)$ and $f'(t) = \big(f_1'(t), \dots, f_n'(t)\big)$.

As a first example we consider a point which moves uniformly on a spiral as indicated by the rule $f(t) = (\cos t, \sin t, t)$. At time t this point has the velocity $f'(t) = (-\sin t, \cos t, 1)$. The vector $f(2\pi) - f(0) = (0, 0, 2\pi)$ is not ever parallel to any of the velocity vectors $f'(t) = (-\sin t, \cos t, 1)$, $t \in [0, 2\pi]$. So this example shows, in particular that no 3-dimensional analog of the Mean Value Theorem can be expected. (See 4.29 and 4.53 and the comments following 4.53.)

Incidentally, the example we just considered may be seen in the context of the curves defined via the exponential function which we observed immediately after Definition 1.36 in the section on norms. Indeed let

$$L = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \text{ then } \exp t\cdot L = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and thus

$$\begin{pmatrix}\cos t\\ \sin t\\ t\end{pmatrix} = (\exp t{\cdot}L)\begin{pmatrix}1\\0\\0\end{pmatrix} + t{\cdot}\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

A consequence of the Mean Value Theorem for a continuous function $f\colon[a,b]\to\mathbb{R}$, differentiable on $]a,b[$, saying $(\exists u\in\,]a,b[)\; f(b)-f(a)=f'(u)(b-a)$, is that

$$|f(b)-f(a)|\le(b-a)\sup\{\|f'(t)\|:a<t<b\},$$

provided, of course, that the derivative is bounded on $]a,b[$. (Cf. Corollary 4.30(ii).) A generalisation of this theorem to curves, fortunately, remains valid and yields important information.

GENERALIZED MEAN VALUE THEOREM FOR CURVES

Theorem 2.4. *Assume that $f\colon[a,b]\to E$ is a curve which is differentiable on $]a,b[$, and that $\{f'(t):t\in\,]a,b[\}$ is bounded. Set*

(2) $$\|f'\|\stackrel{\text{def}}{=}\sup\{\|f'(t)\|:t\in\,]a,b[\}.$$

Then

(3) $$\|f(b)-f(a)\|\le(b-a){\cdot}\|f'\|.$$

For the proof of this theorem we prove the following lemma, which will quickly yield a proof of 2.4.

Lemma 2.5. *Let $f\colon[a,b]\to E$ and $g\colon[a,b]\to\mathbb{R}$ be continuous functions which are differentiable on $]a,b[$ and which satisfy*

(4) $$\|f'(t)\|\le g'(t)\quad\text{for}\quad a<t<b.$$

Then

(5) $$\|f(b)-f(a)\|\le g(b)-g(a).$$

Proof. Let $a<t_0<b$ and use the differentiability of f and g to find remainder functions $\rho\colon]a,b[\,\to E$ and $\sigma\colon]a,b[\,\to\mathbb{R}$ such that

(6) $$\lim_{\substack{t\to t_0\\ t\ne t_0}}|t-t_0|^{-1}\rho(t)=0,\qquad \lim_{\substack{t\to t_0\\ t\ne t_0}}|t-t_0|^{-1}\sigma(t)=0,\qquad\text{and}$$

(7) $$f(t)-f(t_0)=(t-t_0){\cdot}f'(t_0)+\rho(t),$$

(8) $$g(t)-g(t_0)=(t-t_0){\cdot}g'(t_0)+\sigma(t).$$

Hypothesis (4) implies $g'(t)\ge0$ for $a<t<b$. Thus g is isotone by 4.33. Given an $\varepsilon>0$ we find a t_1 with $t_0<t_1<b$ such that $t_0\le t\le t_1$ implies $(t-t_0)^{-1}\|\rho\|<\frac{\varepsilon}{2}$

and $(t-t_0)^{-1}\|\sigma\| < \frac{\varepsilon}{2}$. Hence for these t, in view of (4) we get

$$\begin{aligned}\|f(t)-f(t_0)\| \le& (t-t_0)\cdot\|f'(t_0)\| + \frac{\varepsilon}{2}(t-t_0)\\ \le& (t-t_0)g'(t_0) + \frac{\varepsilon}{2}(t-t_0), \qquad (9)\end{aligned}$$

$$g(t)-g(t_0) \ge (t-t_0)g'(t) - \frac{\varepsilon}{2}(t-t_0). \qquad (10)$$

Hence

(11) $(\forall\varepsilon>0)(\exists t_1\in\,]t_0,b[)(\forall t_0<t<t_1)\ \|f(t)-f(t_0)\| \le g(t)-g(t_0)+\varepsilon(t-t_0)$.

Now let $\varepsilon>0$ be arbitrary. We consider the functions φ, $\psi\colon[a,b]\to\mathbb{R}$ defined by

$$\varphi(t) = \|f(t)-f(a)\|-\varepsilon \quad \text{and} \quad \psi(t)=g(t)-g(a). \qquad (12)$$

We set

$$I = \{t\in[a,b] : (\forall s\in[a,t]) \quad \varphi(s)\le\psi(s)+\varepsilon(s-a)\}. \qquad (13)$$

Then, by definition (13), I is an interval containing a. Put $b'=\sup I$. By (12) and the continuity of the norm, the functions φ and ψ are continuous because of the continuity of f and g, and $\varphi(a)=-\varepsilon$ and $\psi(a)=0$. We deduce $a<b'$.

Now we claim that $b'=b$; once this is proved we have $\|f(b)-f(a)\|-\varepsilon = \varphi(b)\le\psi(b)+\varepsilon(b-a) = g(b)-g(a)+\varepsilon(b-a)$; since ε was arbitrary, (17) follows and this will complete the proof of the Lemma.

Now we suppose that the assertion is false and derive a contradiction. Thus, by assumption, $a<b'<b$. Let $s_n\in I$ be a sequence converging to b'; then $\varphi(s_n)\le\psi(s_n)+\varepsilon(s_n-a)$. Thus using the continuity of φ and ψ, passing to the limit we get

$$\varphi(b')\le\psi(b')+\varepsilon(b'-a).$$

Now by (11) we find a t_1 such that $b'<t_0<b$ and that

$$\begin{aligned}\varphi(t) =& \|f(t)-f(a)\|-\varepsilon \le \|f(t)-f(b')\| + \|f(b')-f(a)\|-\varepsilon\\ \le& g(t)-g(b')+\varepsilon(t-b')+\varphi(b')\\ \le& g(t)-g(b')+\varepsilon(t-b')+\psi(b')+\varepsilon(b'-a)\\ =& g(t)-g(b')+\varepsilon(t-b')+g(b')-g(a)+\varepsilon(b'-a)\\ =& g(t)-g(a)+\varepsilon(t-a) = \psi(t)+\varepsilon(t-a).\end{aligned}$$

It follows that $t_0\in I$ and thus $t_0\le\sup I=b'$, a contradiction which proves the claim $b'=b$. The proof is complete. □

For a proof of Theorem 2.4 we define $g(t)=\|f'\|\cdot t$ and thus get $\|f'(t)\|\le\|f'\|=g'(t)$. Then Lemma 2.5 yields $\|f(b)-f(a)\|\le g(b)-g(a)=\|f'\|(b-a)$, and this concludes the proof of 2.4. □

The proof is similar in spirit to the proof of the Théorème d' accroissements finis 4.34. There we proved an estimate from below, here we prove one from above.

Exercise E2.2. Prove the following assertion:
If, in Lemma 2.5 *and in* Theorem 2.4*, differentiability of f and g is assumed only with countably many exceptions, the conclusions remain nevertheless true.*
[Hint. Consider the hint for the proof of 4.34 in E4.14. Modify (5) above by replacing "$+\varepsilon$" by "$+\varepsilon\sigma(s)$" with the jump function σ of the proof of 4.34. Distinguish the cases (a) and (b) of the proof of 4.34; in case (a) proceed as in the proof above, in case (b) follow the idea suggested in the proof of 4.34.]

An easier generalisation is readily established:

Corollary 2.6. (i) *Assume that the curve $f\colon [a,b] \to E$ is continuous and piecewise differentiable, that is, there is a partition $a = t_0 < t_1 < \cdots < t_n = b$ of $[a,b]$ such that f is differentiable on each interval $]t_{j-1}, t_j[$, $j = 1, \ldots, n$. We assume, moreover, that the norms $\|f'(t)\|$ of the derivative are bounded so that $\|f'\| = \sup\left\{\|f'(t)\| : t \in [a,b] \setminus \{t_0, \ldots, t_n\}\right\}$ exists. Then*

$$\|f(b) - f(a)\| \le \|f'\|{\cdot}(b-a).$$

(ii) *If $f\colon [a,b] \to E$ is continuous and piecewise differentiable, and if $f'(t) = 0$ for all t in which $f'(t)$ exists, then f is constant.*

Proof. Exercise. □

Exercise E2.3. Prove Corollary 2.6

[Hint. For (i) use Theorem 2.4 and induction. Finally, (ii) is an easy consequence of (i).

The nomenclature "Generalized Mean Value Theorem" for 2.4 is not particularly well chosen. There is no "mean value" left in this theorem that would justify this name; it is chosen simply to remind us that it remains as the only available substitute for the Mean Value Theorem in one variable.
We now derive a lemma which will directly lead to a final generalization of the Mean Value Theorem in the next chapter.

We recall from Definition 4.36, that a metric space X is arc connected if for any pair of points $a, b \in X$ there is a curve $\gamma\colon [0,1] \to X$ connecting a and b, i.e. $\gamma(0) = a$ and $\gamma(1) = b$. In general, such a curve is not rectifiable. However, the proof that we have given for Proposition 4.39 applies to prove the following result

Proposition 2.7. *For an open subset X of a normed vector space E, the following statements are equivalent:*

(i) X is connected.
(ii) Each pair of points a, $b \in X$ is connected by a curve $\gamma\colon [\alpha, \beta] \to X$ such that for a suitable partition $t_0 = 0 < t_1 < \cdots < t_n = 1$ of the interval $[\alpha, \beta]$ the curve $\gamma|[t_{k-1}, t_k]$ is affine; i.e. $\gamma\big(t_{k-1} + \tau(t_k - t_{k-1})\big) = \gamma(t_{k-1}) + \tau\big(\gamma(t_k) - \gamma(t_{k-1})\big)$, $\tau \in [0,1]$, $k = 1, 2, \ldots, n$.

Proof. Exercise. □

Exercise E2.4. Prove Proposition 2.7.
[Hint. (ii)⇒(i) is straightforward (cf. 4.37). For a proof of (i) implies (ii) pick an $x_0 \in X$ and let $U = \{x \in X : x_0$ and x are connected by a curve γ such as in (ii). If $x \in U$ then since X is open, there is an open ball W of radius $r > 0$ with center $\gamma(\beta)$ which is contained in X. Let $w \in W$, say $\delta = d(\gamma(\beta), w) < r$. Define $\gamma_1 : [\alpha, \beta+\delta] \to X$ with $\gamma_1(t) = \gamma(t)$ for $t \in [\alpha, \beta]$ and $\gamma_1(\beta+\tau) = \gamma(\beta)+\tau\big(w-\gamma(\beta)\big)$ for $0 \leq \tau \leq \delta$. conclude that $W \subseteq X$ and thus that U is open. If $V \subseteq X$ is the set of all $x \in X$ which cannot be reached from x_0 by a curve γ such as in (ii), show that an entire neighborhood W of x_0 is in V, i.e. that V is open. Since $x_0 \in U$ and X is connected, $V = \emptyset$, i.e. $X = U$ follows.]

Definition 2.8. In accordance with 4.32, curves such γ in 2.7(ii) are called *piecewise affine.* For a piecewise affine curve $\gamma : [\alpha, \beta] \to X$ we say that $\|\gamma(t_1) - \gamma(t_0)\| + \|\gamma(t_2) - \gamma(t_1)\| + \cdots + \|\gamma(t_n) - \gamma(t_{n-1})\|$ is the *arc length* $L = L(\gamma)$ of γ.□

Let X be an open subset of a finite dimensional normed vector space E. For x, $y \in X$, set

$$d(x, y) = \inf\{L : \ L \text{ is the arc length of a piecewise affine curve from } x \text{ to } y\}.$$

Lemma 2.9. *We stay in the circumstances of* Definition 2.7*.*

(i) Given x, $y \in X$ and an $\varepsilon > 0$, we find a piecewise affine curve γ from x to y whose arc length L satisfies $L \leq d(x, y) + \varepsilon$.

(ii) The function $d : X \times X \to \mathbb{R}$ is a metric.

Proof . (i) is a consequence of 1.30 for infs, and Definition 2.7.

(ii) Let us prove the triangle inequality; the remainder is immediate. Let $x, y, z \in X$ and $\varepsilon > 0$. By (i) above, there is a piecewise affine curve γ_1 from x to y whose arc length L_1 satisfies $L_1 < d(x, y) + \frac{\varepsilon}{2}$. Likewise there is a piecewise affine curve γ_2 from y to z whose arc length L_2 satisfies $L_2 < d(y, z) + \frac{\varepsilon}{2}$. From γ_1 and γ_2 we construct a rectifiable arc γ from x to z with arc length $L_1 + L_2$. Then $d(x, z) \leq L_1 + L_2 + d(x, y) + \frac{\varepsilon}{2} + d(y, z) + \frac{\varepsilon}{2} = d(x, y) + d(y, z) + \varepsilon$. Since ε is arbitrary, $d(x, z) \leq d(x, y) + d(y + z)$ follows. □

Exercise E2.5. (i) Supply the missing details of the proof of 2.9(ii), notably the positive definiteness and symmetry of d and the explicit definition of γ from γ_1 and γ_2. Why is the ε-argument necessary?

(ii) Prove the following assertion:

If for all $t \in [0, 1]$ *we have* $(1 - t){\cdot}x + t{\cdot}y \in X$ *then* $d(x, y) = \|y - x\|$.

The metric d on X is called the *geodesic distance.*

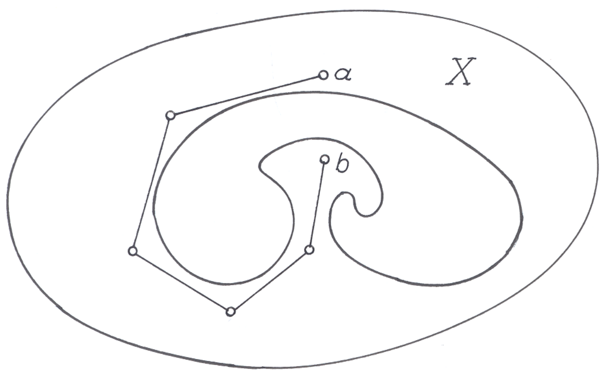

Figure 2.3

If any two points in X of a normed vector space E can be connected by a straight line segment, i.e. if X is *convex*, then $d(x, y) = \|y - x\|$, i.e. the geodesic distance agrees with the induced metric of E.

MEAN VALUE LEMMA

Lemma 2.10. *Let X be a connected open subset of a Banach space E. Let $f: X \to W$ be a differentiable function with values in a finite dimensional normed vector space W and assume that for each piecewise differentiable curve $\gamma: [a, b] \to U$ in U the curve $f \circ \gamma: [a, b] \to W$ is differentiable, and that theere is a number C such that $\|(f \circ \gamma)'(t)\| \le C{\cdot}\|\gamma'(t)\|$ for all such piecewise differentiable curves γ and all t in the domain of γ.*

$$(*) \qquad (\forall x,\, y \in X)\ \|f(x) - f(y)\| \le C.d(x, y).$$

If x and y are connected in X by a straight line segment, then

$$(**) \qquad \|f(x) - f(y)\| \le C{\cdot}\|x - y\|.$$

Proof . Let $x, y \in X$ and $\varepsilon > 0$. By 2.9(i) we find a piecewise affine curve $\gamma: [\alpha, \beta] \to X$ such that $\gamma(\alpha) = x$, $\gamma(\beta) = y$ and that its arc length $L(\gamma)$ satisfies $L < d(x, y) + \varepsilon$. Here $L(\gamma) = \sum_{k=1}^{n} \|\gamma(t_k) - \gamma(t_{k-1}\|$ where $t_0 = \alpha$, $t_n = \beta$, and where γ is affine on $[t_{k-1}, t_k]$, $k = 1, 2, \dots, t_n$. Then $\|\gamma(t_k) - \gamma(t_{k-1})\| = \|(t_k - t_{k-1}){\cdot}\gamma'(\tau)\| = (t_k - t_{k-1}){\cdot}\|\gamma'(\tau)\|$ for any $\tau \in \,]t_{k-1}, t_k[$. Since γ' is constant on $]t_{k-1}, t_k[$, we can write $\|\gamma'(\tau)\| = \sup\{\|\gamma'(t)\| : t_{k-1} < t < t_k\}$. By hypothesis, we have $\|(f \circ \gamma)'(t)\| \le C{\cdot}\|\gamma'(t)\|$ and thus for $\tau \in \,]t_{k-1}, t_k[$ we have $(t_k - t_{k-1}){\cdot}\|(f \circ \gamma)'\| \le C{\cdot}(t_k - t_{k-1}){\cdot}\sup\{\|gamma(t)\| : t_{k-1} < t < t_k\} = C{\cdot}\big(\gamma(t_k) - \gamma(t_{k-1})\big)$. Accordingly, by the Generalized Mean Value Theorem for Curves 2.4, for all $k = 1, \dots, n$, we have

$$(\dagger) \quad \|f\big(\gamma(t_k)\big) - f\big(\gamma(t_{k-1})\big)\| \le \|(f \circ \gamma)'\|{\cdot}(t_k - t_{k-1}) \le C{\cdot}\|\gamma(t_k) - \gamma(t_{k-1})\|.$$

Then

$$\begin{aligned}\|f(b)-f(a)\| &\le \sum_{k=1}^{n} \|f\big(\gamma(t_k)\big)-f\big(\gamma(t_{k-1})\big)\| \\ &\overset{(\dagger)}{\le} C\cdot \sum_{k=1}^{n} \|\gamma(t_k)-\gamma(t_{k-1})\| \\ &= C\cdot L(\gamma) \le C\cdot (d(a,b)+\varepsilon).\end{aligned}$$

Since this holds for all $\varepsilon > 0$, we conclude

$$\|f(b)-f(a)\| \le C\cdot d(a,b)$$

as asserted. □

Exponential curves

An example of curves that is of great theoretical interest is the example of exponential curves as we have already indicated. Let A be a Banach algebra with identity and $\exp: A \to A$ its exponential function (see 6.35). Let $a \in A$ and consider the curve $t \mapsto \exp t\cdot a : \mathbb{R} \to A$ in A.

Lemma 2.11. *For all elements a in a Banach algebra A, the curve $t \mapsto \exp t\cdot a : \mathbb{R} \to A$ is differentiable and its derivative is* $\left.\frac{d \exp t\cdot a}{dt}\right|_{t=t_0} = a \exp t_0\cdot a = (\exp t_0\cdot a)a.$

Proof. We compute $\exp(t_0+h)\cdot a - \exp t_0\cdot a = \exp a(\exp h\cdot a - 1) = (\exp h\cdot a - 1)\exp a$. Now we assume $0 < |h| \le 1$ and note $\|h^{-1}(\exp h\cdot a - 1) - a\| = \|\frac{h}{2!}a^{+}\cdots\| \le |h|(\frac{1}{2}\|a\|^2 + \frac{1}{3!}\|a\|^3 + \cdots) = |h|(e^{\|a\|} - 1 - \|a\|) \to 0$ for $0 \ne h \to 0$. It follows that $\lim_{\substack{h\to 0\\ h\ne 0}} \frac{1}{h}(\exp(t_0+h)\cdot a - \exp t_0\cdot a) = a \exp t_0\cdot a = (\exp t_0\cdot a)a$. □

Let V be a finite dimensional normed vector space such as $V = \mathbb{R}^n$ or $V = \mathbb{C}^n$ and let $L: V \to V$ be a linear map. In the Banach algebra $A \overset{\text{def}}{=} \operatorname{Hom}(V,V)$ equipped with the operator norm (see 6.34). Fix a vector t_0 and a real number t_0. Define a function $x: \mathbb{R} \to V$ by $x(t) = (\exp(t-t_0)\cdot L)x_0$. Using Lemma 2.20 we quickly see that $\dot{x}(t) = \lim_{\substack{h\to 0\\ h\ne 0}} h^{-1}\cdot\big(x(t+h) - x(t)\big) = (L \exp(t-t_0)\cdot L)x_0 = Lx(t)$. Also, we note that $x(t_0) = x_0$. Thus we have the following observation which we shall improve presently:

Remark. Assume that V is a finite dimensional normed vector space. If $x_0 \in V$, then the curve

$$x: I \to V, \quad x(t) = (\exp(t-t_0)L)x_0$$

is a solution of the initial value problem

$$\dot{x}(t) = Lx(t), \quad x(t_0) = x_0.$$

□

Rectifiable curves

Let $f: [a,b] \to X$ be a curve in a metric space X, and let $x_0 = a, x_1, \ldots, x_n = b$ be a finite subdivision of the interval. Then $F = \{x_0, \ldots, x_n\}$ is a finite subset of $[a,b]$, and conversely, every finite subset F of $[a,b]$ containing $\{a,b\}$ determines uniquely such a subdivition. If X is a Banach space V with $d(x,y)=\|y-x\|$, we can define uniquely a piecewise affine curve $f_F\colon[a,b]\to V$ such that

$$f_F(t)=\begin{cases} f(a), & \text{if } t=a,\\ f(x_{k-1})+\frac{t-x_{k-1}}{x_k-x_{k-1}}\cdot\big(f(x_k)-f(x_{k-1})\big), & \text{if } x_{k-1}<t\le x_k,\ k=1,\dots,n.\end{cases}$$

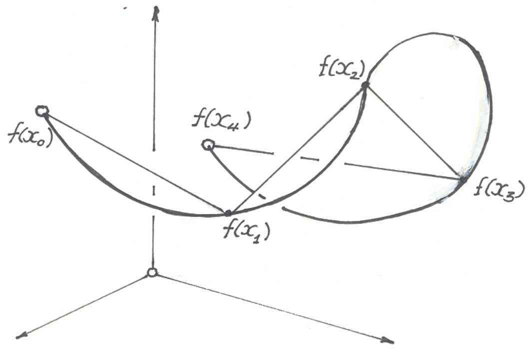

Figure 2.4

But even in a metric space we can define

$$L(f)_F=\sum_{k=1}^{n} d\big(f(x_{k-1}),f(x_k)\big).$$

In the case of a piecewise affine function this agrees with our previous notion of arc length: $L(f)_F=L(f_F)$.

Lemma 2.12. *If $\{a,b\}\subseteq F\subseteq G\subseteq[a,b]$ are two finite sets then $L(f)_F\le L(f)_G$.*

Proof. Exercise E2.6. □

Exercise E2.6. Prove Lemma 2.12.
[Hint. Let $x_0<\cdots x_n$ denote the elements of G and let $k_0=0<k_1<\cdots<k_p=n$ be such that $\{x_{k_0},x_{k_1},\dots,x_{k_p}\}=F$. By the triangle inequality and an iterated application of it where necessary, show that $d\big(f(x_{k_{j-1}}),f(x_{k_j})\big)\le\sum_{m=k_{j-1}+1}^{k_j} d\big(f(x_{m-1}),f(x_m)\big)$, and show that this leads to a proof of the claim.]

Thus if $\mathcal{F}=\mathcal{F}[a,b]$ denotes the set of all finite subsets of $[a,b]$ containing $\{a,b\}$, then the function $F\mapsto L(f)_F:\mathcal{F}\to\mathbb{R}^+$ is increasing in the sense that $F\subseteq G$ in $\mathcal{F}$ implies $L(f)_F\le L(f)_G$.

Definition 2.13. A curve $f\colon[a,b]\to V$ in a metric space X is called *rectifiable*, if the set $\{L(f)_F:F\in\mathcal{F}\}$ is bounded in $\mathbb{R}^+$. If f is rectifiable, then the nonnegative number

$$L(f)\overset{\text{def}}{=}\sup\{L(f)_F:F\in\mathcal{F}\}$$

is called the arc length of f. □

In view of the Characterisation Theorem for Sups 1.30, as an immediate consequence of the Definition we have the following

Lemma 2.14. *Let $f\colon [a,b] \to X$ be a rectifiable curve. Then for each $\varepsilon > 0$ there is a finite subset $F \in \mathcal{F}$ such that $L(f) - \varepsilon < L(f)_F \leq L(f)$.* □

Arc length behaves additively in the right way:

Lemma 2.15. *Let $f\colon [a,b] \to X$ and $g\colon [b,c] \to X$ be two curves in a metric space X such that $f(b) = g(c)$. Then we can form a new rectifiable curve $f\#g\colon [a,c] \to V$ by concatenation*

$$f\#g(x) = \begin{cases} f(x) & \text{if } x \in [a,b], \\ g(x) & \text{if } x \in [b,c], \end{cases}$$

and $L(f\#g) = L(f) + L(g)$.

Proof. Exercise E2.7. □

Exercise E2.7. Prove Lemma 2.15.
[Hint. Let $F \in \mathcal{F}[a,c]$. Then $F^* = F \cup \{b\} \in \mathcal{F}[a,c]$ and we know $L\big((f\#g)\big)_F \leq L(f\#g)_{F^*}\big) \leq L(f\#g)$. This shows that $f\#g$ is rectifiable. Also, $L(f\#g)_{F^*} = L(f)_{F^* \cap [a,b]} + L(g)_{F^* \cap [b,c]} \leq L(f) + L(g)$. If $\varepsilon > 0$ and $L(f\#g) - \varepsilon < L(f\#g)_F$ for suitable choice of F by Lemma 2.14, we get $L(f\#g) - \varepsilon < L(f) + L(g)$, whence $L(f\#g) \leq L(f) + L(g)$. Conversely let $F \in \mathcal{F}[a,b]$ and $G \in \mathcal{F}[b,c]$, then $F \cup G \in \mathcal{F}[a,c]$ and $L(f)_F + L(g)_G = L(f\#g)_{F \cup G} \leq L(f\#g)$. Conclude $L(f) + L(g) \leq L(f\#g)$.]

Exercise E2.8. Prove the following
Theorem. *The image of a rectifiable curve in $\mathbb{R}^2$ does not contain a square.*
[Hint. Let $f\colon [a,b] \to \mathbb{R}^2$ be a rectifiable curve. Suppose that $f([a,b])$ contains the square $Q = [A, A+h] \times [B, B+h]$. Find an affine map $\alpha\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $\alpha(Q) = [0,1] \times [0,1]$. Then $F = \alpha \circ f$ has the same properties as f. Replacing f by F, if necessary we may assuume $Q = [0,1]^2$. consider the set $M \subseteq Q$ of all points $(\frac{q}{n}, \frac{p}{n})$, $0 \leq q, p \leq n$. Choose a subset $D = \{x_1, \dots x_{(n+1)^2}\}$ of $[a,b]$ with $x_{j-1} < x_j$, $j = 2, \dots, (n+1)^2$ such that $f(D) = M$. The length of the arc $f([x_{j-1}, x_j])$ is $\geq \frac{1}{n}$ for $j = 2, \dots, (n+1)^2$ (why?). Thus the length L of the curve satisfies the estimate $L \geq \frac{(n+1)^2 - 1}{n} \geq n$. And this has to hold for all natural numbers n. Impossible!] The space filling curves of Peano and Hilbert therefore have no arc length.

Theorem 2.16. *A curve $f\colon [a,b] \to V$ in a Banach space is rectifiable, if f is differentiable on $]a,b[$ and $\{f'(t) : a < t < b\}$ is bounded in V.*

Proof . Let $\|f'\|$ denote $\sup\{\|f'(t)\| : a < t < b\}$. Let $F \in \mathcal{F}[a,b]$, say, $F = \{x_0, \ldots, x_n\}$ with $x_{k-1} < x_k$ for $k = 1, \ldots, n$. Then the Generalized Mean Value Theorem vor Curves 2.4 implies

$$L(f)_F = \sum_{k=1}^{n} \|f(x_k) - f(x_{k-1})\| \leq \sum_{k=1}^{n} (x_k - x_{k-1}) \cdot \|f'\| = \|f'\| \cdot (b-a)$$

This proves that f is rectifiable. □

By Corollary 2.6 it suffices that f be only piecewise differentiable on $]a,b[$. In fact, if we accept the result of Exercise E2.2 we get the following

Rectifiability of Differentiable Curves

Theorem 2.17. *Assume that the curve $f\colon [a,b] \to V$ in a Banach space is differentiable in all inner points of $[a,b]$ with the possible exception of a countable subset C of $]a,b[$, and assume further that $\{f'(x) : x \in]a,b[\setminus C\}$ is bounded. Then f is rectifiable and $L(f) \leq (b-a) \cdot \|f'\|$, $\|f'\| = \sup\{\|f'(x)\| : x \in]a,b[\setminus C\}$.* □

We can draw some additional conclusions on the rectifiability of curves in metric spaces which we do in the following.

Two curves $f_j\colon [a_j, b_j] \to X$, $j = 1, 2$ are called *equivalent* , if there is a strictly isotone surjective function $\sigma\colon [a_1, b_1] \to [a_2, b_2]$ such that $f_1 = f_2 \circ \sigma$.

Definition. An *equivalence relation* R on a set X is a binary relation $R \subseteq X \times X$ which is reflexive $(\forall x \in X)\, x\,R\,x)$, symmetric $(\forall x,\, y \in X)\, (x\,R\,y) \Leftrightarrow (y\,R\,y))$, and transitive $(\forall x \in X\big((x\,R\,y) \wedge (y\,R\,z)\big) \Rightarrow (x\,R\,z))$. □

Exercise E2.9. Show that the binary relation which we have defined on the set of all curves in a fixed metric space X is indeed an equivalence relation.

If $T_1 = \{t_{10}, \ldots, t_{1n}\}$ is a finite partition $a_1 = t_{10} < \cdots < t_{1n} = b_1$ of $[a_1, b_1]$, then $T_2 = \sigma(T_1)$ is a finite partition $\{a_2 = t_{20}, \ldots, t_{2n}\}$ of $[a_2, b_2]$ such that $t_{2,j} = \sigma(t_{1,j})$ for $j = 1, \ldots, n$. Thus $T_1 \mapsto \sigma(T_1)$ is a bijection from the set of finite partitions of $[a_1, b_1]$ onto the set of finite partitions of $[a_2, b_2]$. The numbers

$$L(T_1; f_1) \stackrel{\text{def}}{=} \sum_{j=1}^{n} d\big(f_1(t_{1(j-1)}), f_1(t_{1j})\big) \quad \text{and}$$

$$L(T_2; f_2) \stackrel{\text{def}}{=} \sum_{j=1}^{n} d\big(f_2(t_{2(j-1)}), f_1(t_{2j})\big)$$

are equal. Therefore f_1 is rectifiable if and only if f_2 is rectifiable, and the arc lengths of equivalent curves are equal. One way of expressing this fact is saying that *the arc length of a rectifiable curve is independent of its parametrisation.*

For a rectifiable curve $f\colon [a,b] \to X$ in a metric space, the restriction

$$f|[a,x]\colon [a,x] \to X,\ \ (f|[a,x])(t) = f(t) \text{ for } t \in [a,x]$$

is rectifiable as well, and we may define $S(a, x) = L(f|[a, x])$. If no confusion is possible, we shall abbreviate $S(a, x)$ by $S(x)$.

Theorem 2.18. (i) *For a rectifiable curve* $f\colon [a, b] \to X$ *in* X*, the function* $S = S(a, \cdot)\colon [a, b] \to [0, L(f)]$ *is an isotone function which satisfies the relation*

$$(*) \qquad (\forall a \le x < y < z \le b) \quad S(x, z) = S(x, y) + S(y, z).$$

(ii) $S\colon [a, b] \to [0, L(f)]$ *is continuous.*

(iii) *The curve* f *is constant on no proper subinterval iff the arc length function* $S\colon [a, b] \to [0, L(f)]$ *is injective and thus is bijective and has and inverse function.*

Proof . The proof of (i) is an exercise. Next we prove (ii). Suppose that (ii) fails, then by (i) and Proposition 3.54 of Analysis I, f has a jump. Suppose f has a jump at b; other cases are treated similarly. Then $s \stackrel{\text{def}}{=} \sup\{S(t) : t < b\} < S(b)$. Let $\varepsilon = \frac{S(b)-s}{2}$. Then

$$(a) \qquad (\forall a \le t < b)\, S(t) \le S(b) - 2\varepsilon.$$

By Lemma 2.14, there is a partition $T = \{x_0 = a < t_1 < \cdots < t_n = b\}$ such that

$$(b) \qquad S(b) - \varepsilon < L(f)_T.$$

Because of the continuity of f, we may assume that $d\big(f(t_{n-1}), f(b)\big) < \varepsilon$, for we may always add a partition point $t < b$ such that $d\big(f(t), f(b)\big) < \varepsilon$, thereby refining the partition without violating (b). Now

$$\begin{aligned} L(f)_T - \varepsilon &< \sum_{m=1}^{n} d\big(f(t_m), f(t_{m-1})\big) - d\big(f(t_{n-1}), f(b)\big) \\ &= \sum_{m=1}^{n-1} d\big(f(t_m), f(t_{m-1})\big) \le S(a, t_{n-1}) \quad \text{(by definition of } S(a, t_{n-1})) \\ &\le S(a, b) - 2\varepsilon \qquad \text{(by (a))}, \end{aligned}$$

whence

$$(c) \qquad L(f)_T < S(a, b) - \varepsilon.$$

Now (b) and (c) imply $S(a, b) - \varepsilon < L(f)_T < S(a, b) - \varepsilon$, a contradiction.

(iii): Let $t_1 < t_2$. In view of $S(t_2) = S(t_1) + S(t_1, t_2)$ we have $S(t_1) < S(t_2)$ iff $S(t_1, t_2) > 0$ iff there is a partition $t_1 = r_0 < r_1 < \cdots < r_n = t_2$ such that $0 < d(f(r_0), f(r_1)) + \cdots + d(f(r_{n-1}), f(r_n))$ by the Characterisation Theorem of Sups 1.30. This implies the existence of elements $t_1 < t_1' < t_2' < t_2$ such that $f(t_1') \ne f(t_2')$; conversely, if such t_1', t_2' exist, then $S(t_1, t_2) \ge S(t_1', t_2') \ge d(f(t_1'), f(t_2')) > 0$. This proves the main assertion of (iii); since S is surjective by (ii) and the Intermediate Value Theorem 3.17 of Analysis I, the existence of an inverse function of S is equivalent to the injectivity of S. □

It is often useful to reparametrize a given curve. This is particularly true if arc length is used as a new parameter. By this we mean the following. If $f\colon [a,b] \to X$ is a rectifiable curve in a metric space E, then the arc length $S(a,\cdot)\colon [a,b] \to [0,L]$ is a continuous nondecreasing surjective function by Theorem 2.18. In the following lemma we derive a few relevant facts of functions with these properties; this lemma only deals with functions between real compact intervals.

Lemma 2.19. *Let $\sigma\colon [a,b] \to [0,L]$ be a continuous isotone and surjective function between real compact intervals. Define*

$$\tau\colon [0,L] \to [a,b], \quad \tau(s) = \min\{t \in [a,b] : s \le \sigma(t)\} = \min \sigma^{-1}(s). \tag{19}$$

Then τ is strictly isotone and has the following properties
(i) $\sigma\big(\tau(s)\big) = s$ *for all* $s \in [0,L]$ *and* $\tau\big(\sigma(t)\big) \le t$ *for all* $t \in [a,b]$.
(ii) τ *is continuous from the left, that is* $\tau(s) = \sup\{\tau(s') : 0 \le s' < s\}$.
(iii) *For all* $s \in [0,L]$ *we have* $\tau_+(s) - \tau(s) = \max \sigma^{-1}(s) - \min \sigma^{-1}(s)$. *In particular, the function τ has a nonzero jump in a point s precisely when $\sigma^{-1}(s)$ is a nondegenerate interval. In particular, if σ is not constant on any nondegenerate interval, then τ is the inverse function of σ.*

Proof . Assertion (i) and the fact that τ is strictly isotone are direct consequences of the definition of τ.

(ii) Let $0 \le t_1 < \tau(s)$. Then $s_1 \stackrel{\text{def}}{=} \sigma(t_1) \le \sigma(\tau(s)) = s$. Then $\tau(s_1) \le t_1 < \tau(s)$ by (i), whence $s_1 < s$. Now let $s_1 < s' < s$. Then $t_1 < \tau(s')$, because $\tau(s') \le t_1$ would imply $s' = \sigma(\tau(s')) \le \sigma(t_1) = s_1$, a contradiction. This implies (ii).

(iii) The assertion is tantamount to $t \stackrel{\text{def}}{=} \max \sigma^{-1}(s) = \inf\{\tau(s') : s < s'\}$. Let $t < t' < t_1$ then $s = \sigma(t) \le \sigma(t') \stackrel{\text{def}}{=} s' \le \sigma(t_1)$; now $t \le \tau(s') \le \tau\big(\sigma(t_1)\big) \le t_1$. This proves the first assertion; the remaining ones are immediate. □

Now we prove the reparametrisation theorem via arc length.

Theorem 2.20. (Reparametrisation Theorem for Curves by Arc Length) *Let $f\colon [a,b] \to X$ be a rectifiable curve in a metric space. Let $S\colon [a,b] \to [0,L(f)]$ be the arc length $S(t) = L(f|[a,t])$ of f. Set $\tau(s) = \min S^{-1}(s)$ for $s \in [0,L]$. Define $F\colon [0,L] \to X$ by $F(s) = f(\tau(s))$. Then*
(i) *F is a rectifiable curve such that $f(t) = F(S(t))$ and $F([0,L(f)]) = f([a,b])$.*
(ii) *The arc length of F is $L(F|[0,s]) = s$ for all $s \in [0,L(f)]$. In particular, F is not constant on any nondegenerate interval of its domain.*
(iii) *If f is not constant on any subinterval of its domain, then f and F are equivalent curves.*

Proof . (i) By 2.18(ii) for each $0 < s \le L$ we have $\lim_{s'\to s,\, s'<s} F(s') = F(s)$. and by 2.18(iii) we know $\lim_{s'\to s,\, s<s'} F(s') = f(t_+)$ where $t_+ = \max S^{-1}(s)$ However, $S(t_+) = S(\tau(s))$ and this equation implies $f(t_+) = f(\tau(s)) = F(s)$. This proves

continuity of F. Now let $t \in [a, b]$. Then $f(t) = f(\tau(S(t)) = F(S(t))$ by the definition of τ. The relation $F([0, L]) = f([a, b])$ is trivial from the definition of F.

(ii) Let $0 = s_0 < s_1 \cdots < s_n = L$ be a partition of $[0, L]$. Set $t_k = \tau(s_k)$. Then $d(F(s_{k-1}), F(s_k)) = d(f(t_{k-1}, t_k)) = S(t_{k-1}, t_k) = S(t_k) - S(t_{k-1}) = s_k - s_{k-1}$. Thus $\sum_{=1}^{n} d(F(s_{k-1}), F(s_k)) = s_n - s_0 = L$. Hence F is rectifiable, and the overall arc length of F is $S_F(0, L) \leq L$. Now let $t_0 = a < t_1 \cdots < t_N = b$ be a partition of $[a, b]$. Since $f(t_{k-1}) = f(t_k)$ implies $d(f(t_{k-1}, f(t_k)) = 0$ we assume that $S(t_{k-1}) < S(t_k)$ for all $k = 1, \dots N$. Furthermore, $f(t_k) = f(\tau(S(t_k)))$ by the definition of τ. Now set $s_k = S(t_k)$. Then $a = s_0 < \cdots < s_N = L$ and $\sum_{k=1}^{N} d(f(t_{k-1}), f(t_k)) = \sum_{k=1}^{N} d(f(\tau(s_{k-1})), f(\tau(s_k))) = \sum_{k=1}^{N} d(F(s_{k-1}), F(s_k)) \leq S_F(0, L)$. By the definition of L as the sup over the sums on the left side of the inequality we obtain $L \leq S_F(0, L)$. Since this argument holds for each $s \in [0, L]$ in place of L we have $S_F(0, s) = s$.

(iii) If f is not constant on any proper subinterval of $[a, b]$, then S is injective by Theorem 2.18(iii), and τ is the inverse function of S. In particular, it is continuous. Hence the relations $F(s) = f(\tau(s))$ and $f(t) = F(S(t))$ show the equivalence of the two curves. □

This theorem shows that *every rectifiable curve in a metric space can be reparametrized in terms of its arc length.*

We emphasize that we have considered curves in *arbitrary metric spaces* with the sole exception of the existence theorems 2.16 and 2.17 of arc length for differentiable curves in Banach spaces.

In particular, we can always observe that the function $S\colon [a, b] \to [0, L] \subseteq \mathbb{R}$ is differentiable in $x \in [a, b]$ if and only if $\lim_{\substack{h \to 0 \\ h \neq 0}} \frac{1}{h}\big(S(a, x + h) - S(a, x)\big)$ exists, that is iff the following limits exist and agree:

$$\lim_{\substack{h \to 0 \\ h > 0}} S(x, x + h)/h, \qquad \lim_{\substack{h \to 0 \\ h > 0}} S(x - h, x)/h.$$

However, in order to link rectifiability and differentiability more profoundly, we assume that we have a curve $f\colon [a, b] \to E$ in a Banach space—or the very least in a metric space X contained in a Banach space E. Then we are ready for the second fundamental theorem on rectifiable curves.

Second Fundamental Theorem on Arc Length

Theorem 2.21. *Let $f\colon [a, b] \to E$ be a curve in a Banach space, which is differentiable on $]a, b[$. Assume that the set of nonnegative real numbers $\{\|f'(t)\| : a < t < b\}$ is bounded. Then the following conclusions hold.*

(i) *f is rectifiable.*

(ii) *The arc length $S(a, t)$ satisfies*

$$\underline{\int}_a^b \|f'(t)\| dt \leq S(a, t) \leq \overline{\int}_a^b \|f'(t)\| dt. \tag{20}$$

(iii) *If* $t \mapsto \|f'(t)\|$ *is integrable, then*

(21) $$S(a,t) = \int_a^t \|f'(\tau)\| d\tau = \int_a^t \|f'(\cdot)\|.$$

(iv) *If* $\|f'\|$ *is continuous, then* S *is differentiable and* $\left.\frac{dS(a,\tau)}{d\tau}\right|_{\tau=t} = \|f'(t)\|$.

Proof. (i) For a given partition $T = \{t_0, \ldots, t_p\}$ of $[a,b]$ we define a step function s_T whose value on $]t_{k-1}, t_k[$ for $k = 1, \ldots, p$ is exactly $\sup\{\|f'(t)\| : t_{k-1} < t < t_k\}$ and for all t_j we set $s(t_j) = \sup\{\|f'(t)\| : a < t < b\}$. Then

(22) $$\|f'(t)\| \leq s_T(t) \quad \text{for all} \quad t \in \,]a,b[\,.$$

By the Generalized Mean Value Theorem for Curves 2.4 we have

(23) $$\|f(t_k) - f(t_{k-1})\| \leq (t_k - t_{k-1}) \cdot s_T(\tau_k), \quad t_{k-1} < \tau_k < t_k.$$

Thus

(24) $$L(T) = \sum_{k=1}^{p} \|f(t_k) - f(t_{k-1})\| \leq \sum_{k=1}^{p} (t_k - t_{k-1}) \cdot s_T(\tau_k) = \int_a^b s_T.$$

If now $T_1 \subseteq T_2$ are two partitions of $[a,b]$, then one notices immediately that $s_{T_2} \leq s_{T_1}$ and that therefore

$$L(T_1) \leq L(T_2) \leq \int_a^b s_{T_2} \leq \int_a^b s_{T_1}.$$

If we fix a partition T, then $\int_a^b s_T$ is an upper bound for the set of all $L(T')$ for partitions T' refining T. The least upper bound of this set therefore exists by the Least Upper Bound Axiom 1.31, see 1.50. This least upper bound is $S(a,b)$. Thus

(25) $$S(a,b) \leq \int_a^b s_T \quad \text{for all partitions} \quad T.$$

(ii) Now let s be a step function on $[a,b]$ and T an associated partition (cf. definition following 5.1 in Analysis I). If $\|f'(t)\| \leq s(t)$ for all t, then $s_T \leq s$ by the definition of s_T (with the possible exception of the points of the partition). From (25) we conclude $S(a,b) \leq \int_a^b s_T \leq \int_a^b s$. Hence, if M denotes the set of all step functions $s\colon [a,b] \to \mathbb{R}$ satisfying $(\forall t \in \,]a,b[)\, \|f'(t)\| \leq s(t)$ we get

$$S(a,b) \leq \inf\{\int_a^b s : s \in M\} = \bar{\int}_a^b \|f'(\cdot)\| = \bar{\int}_a^b \|f'(\tau)\|\, d\tau.$$

Now we have to prove $\underline{\int}_a^b \|f'(\cdot)\| \leq S(a,b)$. For this purpose let $s\colon [a,b] \to \mathbb{R}$ be a step function with $s(t) \leq \|f'(t)\|$ for all $t \in [a,b]$ and let T be a partition associated with s. We denote by s_k the value of s on $]t_{k-1}, t_k[$ and claim that $(t_k - t_{k-1}) s_k \leq S(t_{k-1}, t_k)$; if this claim is proved, then $\int_a^b s = \sum_{k=1}^{p} (t_k - t_{k-1}) s_k \leq \sum_{k=1}^{p} S(t_{k-1}, t_k) = S(a,b)$. Therefore $S(a,b)$ is an upper bound for the set of all of these real numbers $\int_a^b s$; the least upper bound of this set, however, is $\underline{\int}_a^b \|f'(t)\|\, dt$; this will prove assertion (iii).

It all boils down to proving the following

Claim. *If $r \leq \|f'(t)\|$ for all $t \in [a,b]$, then $(b-a)r \leq S(a,b)$.* For $r = 0$ we have nothing to show. Thus let now be $r > 0$ and $\varepsilon > 0$. If $t \in]a,b[$ then, since f is differentiable at t, there is a $\delta_t > 0$ such that for all $\tau \in [a,b]$ mit $0 < |t-\tau| < \delta_t$ we have $\|\frac{1}{\tau - t}\big(f(\tau) - f(t)\big) - f'(t)\| < \varepsilon$, and so

$$\|f(\tau) - f(t)\| \geq |\tau - t|{\cdot}\|f'(t)\| - |\tau - t|\varepsilon \geq |\tau - t|(r - \varepsilon). \tag{26}$$

Now we define

$$I = \{t \in [a,b] : (\forall s \in [a,t])\,(s-a)(r-\varepsilon) \leq S(a,t)\}. \tag{27}$$

Then I is an interval containing a and being bounded above by b. Set $b' = \sup I$. Then $a \leq s_1 < s_2 < \ldots < b'$ and $b' = \lim_{n\to\infty} s_n$, as $S(a,\cdot)$ is isotone, implies $(s_n - a)(r-\varepsilon) \leq S(a, s_n) \leq S(a, b')$ and thus also

$$(b' - a)(r - \varepsilon) \leq S(a, b'). \tag{28}$$

Thus $b' \in I$ by (27). We claim $b' = b$. Suppose not. Then $b' < b$ and we find a $\tau \in]b', \min\{b' + \delta_{b'}, b\}[$; for this $\tau > b'$ we have $(\tau - b')(r - \varepsilon) \leq \|f(\tau) - f(b')\|$ by (26). From this relation and (27) we conclude $(\tau - a)(r-\varepsilon) = (b'-a)(r-\varepsilon) + (\tau - b')(r-\varepsilon) \leq S(a,b') + S(b',\tau) = S(a,\tau)$. By (27) this means $\tau \in I$, whence $\tau \leq \sup I = b'$, a contradiction to $b' < \tau$. This proves the claim and thus the proof of (ii) is complete.

(iii) By the definition, the function $\|f'(\cdot)\|\colon [a,b] \to \mathbb{R}$ integrable iff $\underline{\int}_a^b \|f'(\cdot)\| = \overline{\int}_a^b \|f'(\cdot)\|$ by 5.11(ii), and this implies $\underline{\int}_a^t \|f'(\cdot)\| = \overline{\int}_a^t \|f'(\cdot)\|$ for all $t \in [a,b]$ (cf. 5.16 and the discussions leading to 5.16). Hence (ii) implies (iii).

(iv) follows from (iii) by the Fundamental Theorem of Differential and Integral Calculus 5.18. □

The preceding theorem persists if the differentiability hypotheses on the curve f are satisfied only piecewise, that is, if there is a partition $T = \{t_0, \ldots, t_p\}$ of the domain $[a,b]$ of f that f' exists and is bounded on each of the intervals $]t_{k-1}, t_k[$.

Exercise E2.10 Prove the following result.
Theorem. *The circumference of a circle of radius r is $2\pi r$.*

In particular, *the circumference of the unit circle $\mathbb{S}^1$ is 2π.*
[Hint. In the euclidean plane, that is $\mathbb{R}^2$ equipped with the euclidean norm $\|\cdot\|_2$, we consider the differentiable curve $f(t) = (m_1 + r\cos t, m_2 + r\sin t)$. The circle of radius r around the point $M = (m_1, m_2)$ is $f([0, 2\pi])$, and the *circumference* of this circle is defined to be the arc length $S(0, 2\pi)$ for f. We note $f'(t) = (-r\sin t, r\cos t)$ and $\|f'(t)\| = r$. Thus $S(0,\theta) = \int_0^\theta r = \theta r$. With $\theta = 2\pi$, the assertion follows.]

This theorem provides us with the second geometric interpretation of the number π as the arc length of the unit semicircle; the first one was Corollary 5.44 showing that the area of the unit disk is π. The definition of the number π was given in 3.34.

Exercise E2.11. Use the ideas of the preceding exercise to prove:
Theorem. *Consider the complex absolute value as norm on* $\mathbb{C}$. *Then the arc length of the curve* $t \mapsto \exp it\colon [0,\theta] \to \mathbb{C}$ *for* $0 \le \theta \le 2\pi$ *is* θ.

This observation allows us to corroborate our theory of the concept of an angle in 3.39 ff. in Analysis I, because by the preceding theorem and 3.39 we have $\arg_0(\exp i\theta) = \theta$ is the arc length of the arc $t \mapsto \exp it : [0,\theta] \to \mathbb{C}$ on the unit circle $\mathbb{S}^1$.

Parametrizing curves in terms of arc length revisited

In the First Fundamental Theorem on Arc Length we have seen that it is (often) possible to reparametrize a given curve in terms of arc length: For any rectifiable curve f the curve $F = f \circ \tau$ is well defined, and the arc length of the curve F measured from $F(0)$ to $F(s)$ is s. Of course, the arc length function of F, being given by $S_F(s) = s$ is differentiable with derivative 1 in all points.

Theorem 2.22. *If* $f\colon [a,b] \to E$ *is a continuously differentiable curve in a Banach space, then the following conclusions hold in the terminology of* 2.17 *and* 2.19

(i) *If* f *is constant on no nondegenerate interval of* $[a,b]$ *then the curves* f *and* F *are equivalent.*
(ii) *Let* $I = \{s \in [0,L] : f'\big(\tau(s)\big) \neq 0\}$. *Then in all points* $s \in I$ *the curve* F *is differentiable and* $\|F'(s)\| = 1$.
(iii) $f'(t) = \|f'(t)\|{\cdot}F'\big(S(a,t)\big)$ *whenever* $f'(t) \neq 0$.

In particular, if $f'(t) \neq 0$ *for all* $t \in [a,b]$, *then* F *is differentiable for all* $s \in [0,L]$ *and* $\|F'(s)\| = 1$.

Proof. Since $f\colon [a,b] \to E$ is a continuously differentiable curve in a Banach space E, the continuous function $t \mapsto \|f'(t)\|$ is bounded on $[a,b]$ by the Theorem of the Maximum 3.52 in Analysis I. Then, by 2.21(i) the curve is rectifiable and $S(t) = \int_a^t \|f'(.)\|$, $S'(t) = \|f'(t)\|$.

(i) is 2.21(iii).

(ii) If $f'(t) \neq 0$, then the inverse function $\tau\colon [0,L] \to [a,b]$ is differentiable in $s = S(t)$ and $\tau'(s) = S'(\tau(s))^{-1} = \|f'(\tau(s))\|^{-1}$ by 4.19. of Analysis I. We claim that an appropriate Chain Rule applies readily to the composition of the functions $\tau\colon [0,L] \to [a,b]$ and $f\colon [a,b] \to E$, and yields that the function $F = f \circ \tau$ is differentiable in such a point s and that

$$\begin{aligned}
F'(s) &= \lim_{\substack{s' \to s \\ s' \neq s}} \frac{1}{s'-s}\big(F(s') - F(s)\big) \\
&= \lim_{\substack{s' \to s \\ s' \neq s}} \frac{\tau(s') - \tau(s)}{s'-s}{\cdot}\left(\frac{1}{\tau(s') - \tau(s)}{\cdot}\big(f\big(\tau(s')\big) - f\big(\tau(s)\big)\big)\right) \\
&= \tau'(s){\cdot}f'\big(\tau(s)\big).
\end{aligned}$$

Then $\|F'(s)\| = \tau'(s){\cdot}\|f'(\tau(s))\| = 1$ since $\tau'(s) \ge 0$. Setting $t = \tau(s)$ implies (iii). □

The condition that f is not constant on any nondegenerate interval does not suffice for $f'(t)$ to be nonzero for all $t \in [a,b]$. In fact, $f'(t)$ may still vanish on an uncountable set. (See E5.12 in Analysis I.)

We deduced an appropriate Chain Rule for the special situation; the conclusive form of the Chain Rule will be presented in Chapter 3.

If we assume that $f'(t) \neq 0$ for all $t \in [a,b]$, then in the sense of (i) and (ii), F describes the same curve as f. Statement (iv) says, that the norm of the velocity vector of F at $f(t) = F(s)$ where $f'(t) \neq 0$ is 1. Notice that *a differentiable curve* $F\colon [0,L] \to E$ *is parametrized by arc length if and only if* $\|F'(s)\| = 1$ *for all* $s \in [0,L]$. In this case, in a model representing the "motion" of a point in space, the point moves with uniform speed equal to 1, and $F'(s)$ tells us the direction of the velocity in the point $F(s)$. If $F'\colon [0,L] \to E$ itself is continuous, then the velocity vector $F'(s)$ runs through a curve on the surface $\mathbb{S}^{n-1} = \{x \in E : \|x\| = 1\}$ of the unit ball; this curve is sometimes called the *directional curve* of the given curve.

Higher derivatives of curves

We have immediate access to geometrically interesting aspects of the theory of differentiable curves. In a first overview of this theory the reader may jump to the next section.

If we have a curve $f\colon [a,b] \to \mathbb{R}^n$ it is not hard to interpret higher derivatives, because differentiating happens componentwise. So let us consider a curve $f\colon [a,b] \to E$ which is twice differentiable, where we interpret $f'(t)$ as an element of E as usual and thus obtain a function $f'\colon [a,b] \to E$ which we assume to be differentiable. In particular, $f'\colon [a,b] \to E$ a is a gain a curve in E the *velocity curve*, which, in case that f happened to be parametrized by arc length, is exactly the directional curve. The second derivative $f''(t)$ is a measure for the rate of change of the velocity f' at time t, and this rate of change is called *acceleration*. In the case of a curve $F\colon [0,L] \to \mathbb{R}^n$, which is parametrized by arc length and is twice differentiable, $F''(s)$ represents the rate of change of the directional curve at s. This rate of change is evidently 0 if $F(s) = v + s{\cdot}e$ with a unit vector e, that is if the motion is straight and uniform with speed 1. On the other hand, if we are in a Hilbert space and e_1 and e_2 are two perpendicular unit vectors, then $F(s) = (r\cos\frac{s}{r}){\cdot}e_1 + (r\sin\frac{s}{r}){\cdot}e_2$ is a circular motion with radius r; then $F'(s) = (-\sin\frac{s}{r}){\cdot}e_1 + (\cos\frac{s}{r}){\cdot}e_2$ and $F''(s) = -\frac{1}{r}{\cdot}\big((\cos\frac{s}{r}){\cdot}e_1 + (\sin\frac{s}{r}){\cdot}e_2 = -\frac{1}{r^2}{\cdot}F(s)$. The bigger the radius r the smaller $\|F''(s)\| = \frac{1}{r}$. Therefore $\|F''(s)\|$ is called the scalar *curvature* of the curve at the point $x = F(s)$, and we write $k(x) \stackrel{\text{def}}{=} \|F''(s)\|$. We might call $F''(s)$ itself the *curvature vector*, it contains directional information,too.

If $f\colon [a,b] \to E$ is a twice differentiable curve with $f'(t) \neq 0$ for all t, then we would like to know the curvature $k\big(f(t)\big)$. We set $F = f \circ \tau$ with the inverse function $\tau\colon [0,L] \to [a,b]$ of the arc length $t \mapsto S(a,t) : [a,b] \to [0,L]$ and assume through the remainder of this subsection that E is a Hilbert space whose norm is given via an inner product by $\|x\|^2 = (x|x)$. We shall need the *radial derivative*, that is, the derivative of $t \mapsto \|f(t)\|\colon [a,b] \to \mathbb{R}$. Since

$$\big((f(t+h) \mid f(t+h)\big) - \big(f(t)|f(t)\big) = \big(f(t+h) - f(t) \mid f(t)\big) + \big(f(t+h) \mid f(t+h) - f(t)\big),$$

the function $\varphi\colon [a,b] \to \mathbb{R}$, $\varphi(t) = \big(f(t) \mid f(t)\big)$ is differentiable and satisfies $\varphi'(t) = (f'(t) \mid f(t)\big) + \big(f(t) \mid f'(t)\big) = 2\big(f(t) \mid f'(t)\big)$. (A conclusive form of the Product Rule will be given below in Chapter 3. This together with the Chain Rule derived in the last portion of the proof of part (ii) of Theorem 2.21 yields

$$(29)\qquad \begin{aligned} \frac{d}{dt}\Big|_{t=t_0} \|f(t)\| &= \frac{d}{dt}\Big|_{t=t_0} \big(f(t) \mid f(t)\big)^{1/2} \\ &= \frac{d}{dt}\Big|_{t=t_0} \varphi(t)^{1/2} = \frac{\varphi'(t)}{2\varphi(t)} = \big(\|f(t_0)\|^{-1}{\cdot}f(t_0) \mid f'(t_0)\big). \end{aligned}$$

This gives us the projection of the velocity vector $f'(t_0)$ onto the straight line spanned by the unit vector $\|f(t_0)\|^{-1}{\cdot}f(t_0)$ of the position vector $f(t_0)$. We have assumed $f'(t) \neq 0$ and consider the *speed* $v\colon [a,b] \to \mathbb{R}$, $v(t) = \|f'(t)\|$ and its derivative $b(t) = v'(t)$, the scalar acceleration which by (29) is given by

$$(30)\qquad b(t) = v'(t) = \big(\|f'(t)\|^{-1}{\cdot}f'(t) \mid f''(t)\big) = \Big(F'\big(S(a,t)\big) \mid f''(t)\Big),$$

the projection of the acceleration vector $f''(t)$ onto the direction vector $F'\big(S(a,t)\big)$ of the velocity at time t. We now start from the relation $v\big(\tau(s)\big){\cdot}F'(s) = f'\big(\tau(s)\big)$ and differentiate with respect to arc length. We find

$$v'\big(\tau(s)\big)\tau'(s){\cdot}F'(s) + v\big(\tau(s)\big){\cdot}F''(s) = \tau'(s){\cdot}f''\big(\tau(s)\big).$$

We have $\tau'(s) = v\big(\tau(s)\big)^{-1}$ in view of the definition of v and the derivative of τ which we computed earlier. From (30) we get $v'\big(\tau(s)\big) = \Big(F'(s)|f''\big(\tau(s)\big)\Big)$ and obtain

$$(31)\qquad \Big(F'(s)|f''\big(\tau(s)\big)\Big){\cdot}F'(s) + v\big(\tau(s)\big)^2{\cdot}F''(s) = f''\big(\tau(s)\big).$$

We recall that f' and f'' are directly accessible to us while we are looking for $F''\big(S(a,t)\big)$. Hence we solve the equation (41) for $v(t)^2{\cdot}F''\big(S(a,t)\big) = f''(t) - \Big(F'(S(a,t))|f''(t)\Big){\cdot}F'\big(S(t)\big)$. We abbreviate the known direction vector $F'\big(S(t)\big) = v(t)^{-1}f'(t)$ of the velocity by $e(t)$. Thus for the desired curvature vector $F''\big(S(t)\big)$ we obtain the relation

$$(32)\qquad v(t)^2F''\big(S(a,t)\big) = f''(t) - \big(e(t)|f''(t)\big){\cdot}e(t),$$

the acceleration minus the projection of the acceleration onto the direction of the velocity. The square of the norm of the right side is $(f''|f'') - 2(f''|e)^2 + (e|f'')^2 = \|f''\|^2 - (f''|e)^2$, which yields the curvature in the point $x = f(t) = F\big(S(a,t)\big)$ as

$$(33)\qquad k\big(x\big) = v(t)^{-2}(\|f''\|^2 - (f''|e)^2)^{1/2},$$

where we abbreviated $f''(t)$ by f'' and $v(t)^{-1}f'(t)$ by e. The relation $\big(F'(s)|F'(s)\big) = 1$ for all $s \in [0, L]$, upon differentiating yields the equation $0 = 2\big(F'(s)|F''(s)\big)$, that is $\big(F''(s)|F'(s)\big) = 0$. The curvature vector is orthogonal to the direction vector. If we set $H(s) = \|F''(s)\|^{-1}{\cdot}F''(s)$, then the pair $\big(F'(s), H(s)\big)$ is an *orthonormal 2-frame*.

The Taylor expansion of a function has taught us that a function $g\colon I \to \mathbb{R}$ on an interval I of real numbers is twice differentiable in a point a if we have a representation $g(a+h) = g(a) + hg'(a) + \frac{h^2}{2}g''(a) + h^2R(h)$ such that $\lim_{h\to 0} R(h) = 0$. In a completely analogous way, for our curves in Hilbert space we have a representation

$$(34)\qquad F(s+h) = F(s) + hF'(s) + \frac{h^2}{2}F''(s) + h^2R(h) \qquad \lim_{h\to 0} R(h) = 0.$$

Indeed for $E = \mathbb{R}^n$ this is an immediate consequence from the scalar version of Taylor's Theorem; for the case of a Banach space E we leave the proof as an exercise. We shall address Taylor's Theorem in greater detail below.

If $F''(s) \neq 0$, then we set $\rho(s) = k(s)^{-1} = \|F''(s)\|^{-1}$. This allows us to write the expansion (34) in the following form:

$$F(s+h) - F(s) = h{\cdot}F'(s) + \frac{h^2}{2\rho(s)}{\cdot}H(s) + h^2R(h) \qquad \lim_{h\to 0} R(h) = 0.$$

Up to a remainder $h^2R(h)$ which is small "of the third order," in the vicinity of $F(s)$ the curve is "contained in" the plane which is spanned by $F'(s)$ and $H(s)$. It is called the *osculating plane* of the curve at $F(s)$. If we describe the points of the osculating plane via the bijection $(x, y) \mapsto x{\cdot}F'(s) + y{\cdot}H(s)$, then the image under this bijection of the graph of the function $x \mapsto x^2/2\rho(s)$ is very close to the trajectory of the curve near $F(s)$. We compare The parabola with the equation $y = x^2/2\rho(s)$ with the circle with center $\big(0, \rho(s)\big))$ and radius $\rho(s)$, that is, with the circle defined by the equation $x^2 + \big(y - \rho(s)\big)^2 = \rho(s)^2$. Write $r = \rho(s)$ for short. The lower semicircle is the graph of the function $x \mapsto y(x) = r - \sqrt{r^2 - x^2} = r(1-(1-(x/r)^2)^{1/2}) = r\Big(1 - (\sum_{n=0}^{\infty}(\binom{1/2}{n})(x/r)^{2n})\Big) = \frac{x^2}{2r} + \frac{x^4}{8r^3} - + \ldots$ (for $|x| < r$). The difference between the parabola and the circle is of order 4 and thus is extremely small near 0. One expresses this fact by saying that the parabola has at 0 the circle which we have constructed as *osculating circle*. Therefore we can now call the circle

$$\{F(s) + x{\cdot}F'(s) + y{\cdot}H(s) : x^2 + \big(y - \rho(s)\big)^2 = \rho(s)^2\}$$

as the *osculating circle* of the curve at $f(t) = F(s)$, $s = S(a, t)$, and call its radius $\rho(s) = \|F''(s)\|^{-1} = k(x)^{-1}$ as the *radius of curvature* of the curve at $x = F(s)$. The radius of curvature and the curvature are reciprocals of each other.

If the curve F is three times differentiable, then the curve $H\colon [0, L] \to E$ is defined, as soon as we assume that the second derivative F'' vanishes nowhere. Then the curve H is itself differentiable, and because of $\big(H(s)|H(s)\big) = 1$ the vectors $H(s)$ and $H'(s)$ are perpendicular. We also have $\big(H(s)|F'(s)\big) = 0$ and $\big(H'(s)|F'(s)\big) = -k(x)$, $x = F(s)$. Thus the vector $H'(s) + k(x)F'(s)$ is orthogonal to $H(s)$ as well as $F'(s)$ and therefore measures the rotation of the osculating plane about the axis spanned by $F'(s)$; this means the turning of the curve out of its osculating plane. The norm of this vector therefore is called the *torsion* of the curve.

In the euclidean space $\mathbb{R}^3$, some things become simpler due to the serendipity of special geometrical properties attached to dimension three. One such fortuitous concept is the *vector product* $(u, v) \mapsto [u, v]$ on $\mathbb{R}^3$, where formulae such as

$$\|[u, v]\|^2 = \|u\|^2\|v\|^2 - (u|v)^2$$

help to simplify formulae like (33).