Chapter 1
Parametrized Plane Curves

1.1 Definitions

We start with some definitions.

Definition 1.1 A (parametrized plane) curve is a continuous mapping $m : I \to \mathbb{R}^2$, where $I = [a, b]$ is an interval.

A curve m is closed if $m(a) = m(b)$.

A curve m is a Jordan curve if it is closed and has no self-intersection: $m(x) = m(y)$ only for $x = y$ or $\{x, y\} = \{a, b\}$.

A curve is piecewise C^1 if it has everywhere left and right derivatives, which coincide except at a finite number of points.

The range of a curve m is the set $m([a, b])$. It will be denoted by $\mathcal{R}_m$.

Notice that we have defined curves as functions over bounded intervals. Their range must therefore be a compact subset of $\mathbb{R}^2$ (this forbids, in particular, curves with unbounded branches).

A Jordan curve is what we can generally accept as a definition of the outline of a shape. An important theorem [292] states that the range of a Jordan curve partitions the plane $\mathbb{R}^2$ into two connected regions: a bounded one, which is the interior of the curve, and an unbounded one (the exterior). The proof of this rather intuitive theorem is quite complex (see, for example [184] for an argument using Brouwer's fixed point theorem).

However, requiring only continuity for curves allows for more irregularities than what we would like to handle. This is why we will always restrict ourselves to piecewise C^1, generally Jordan, curves. We will in fact often ask for more, and consider curves which are regular (or piecewise regular).

Definition 1.2 A C^1 curve $m : I \mapsto \mathbb{R}^2$ is a regular curve if $\partial m \neq 0$ for all $u \in I$. If m is only piecewise C^1, we extend the definition by requiring that all left and right derivatives are non-vanishing.

L. Younes, *Shapes and Diffeomorphisms*, Applied Mathematical Sciences 171,
https://doi.org/10.1007/978-3-662-58496-5_1

Here, and in the rest of the book, we will use either ∂m or $\dot{m}$ to denote the derivative of a function $u \mapsto m(u)$.

The previous definition is fundamental. It avoids, in particular, curves which are smooth functions (C^∞, for example) but with a range having geometric singularities. Consider the following example: let

$$m(u) = \begin{cases} (\varphi(u), 0), & u \in [0, 1/2] \\ (1, \varphi(u - 1/2)), & u \in [1/2, 1] \end{cases}$$

with $\varphi(u) = 16u^2(1-u)^2$, $u \in [0, 1]$. It is easy to check that m is continuously differentiable, whereas the range of m is the corner $[0, 1] \times \{0\} \cup \{1\} \times [0, 1]$.

We will say that a curve $m : [a, b] \to \mathbb{R}^2$ is C^p if it is p times continuously differentiable, including all right derivatives at a and left derivatives at b up to order p. If the curve is closed, we will implicitly require that the derivatives at a and b coincide. More precisely, a closed curve is C^p if and only if m is C^p when restricted to the open interval (a, b), and so is the curve $\tilde{m}$ defined on (a, b) by $\tilde{m}(u) = m(u + \varepsilon)$ if $u \in (a, b - \varepsilon]$ and $\tilde{m}(u) = m(u + \varepsilon - b + a)$ if $u \in [b - \varepsilon, b)$ (for some $0 < \varepsilon < b - a$).

Alternatively (and more conveniently), closed curves can be handled by considering the interval $[a, b]$ closed onto itself after identifying a and b (which provides a one-dimensional torus). We will denote this torus by $[a, b]_*$ and let $a \sim b \in [a, b]_*$ denote a or b after the identification. For $u, v \in [a, b]_*$, we let

$$d_*(u, v) = \min(|u - v|, (b - a) - |u - v|). \tag{1.1}$$

This is a distance on $[a, b]_*$ (if u (or v) are equal to $\tilde{ab}$, the result does not depend on which value is chosen to compute the expression).

If $\delta \in \mathbb{R}$ and $u \in [a, b]_*$, we define $u +_* \delta \in [a, b]_*$ by $u + \delta - (u +_* \delta) = k(b - a)$ for some integer k (so that we consider addition modulo $b - a$). A function $f : [a, b]_* \to \mathbb{R}^d$ is continuous on $[a, b]_*$ if and only if, for all $u \in [a, b]_*$, $|f(u +_* \delta) - f(u)| \to 0$ when $\delta \to 0$, which is equivalent to $f(u) = \hat{f}(u)$ for some function $\hat{f}$ continuous on $[a, b]$ satisfying $\hat{f}(a) = \hat{f}(b)$. One defines derivatives of functions by

$$\partial f(u) = \lim_{\delta \to 0} \frac{f(u +_* \delta) - f(u)}{\delta}$$

when the right-hand side exists and higher derivatives are defined accordingly. With this notation, it is easy to see that C^p closed curves are functions $m : [a, b]_* \to \mathbb{R}^2$ with at least p continuous derivatives.

We also use integrals along $[a, b]_*$ as follows: if $u_0, u_1 \in [a, b]$ and $f : [a, b] \to \mathbb{R}^d$ is continuous, then

$$\int_{u_0}^{u_1} f(v)dv_* := \begin{cases} \displaystyle\int_{u_0}^{u_1} f(v)dv \text{ if } u_0 \le u_1 \\ \displaystyle\int_a^b f(v)dv - \int_{u_1}^{u_0} f(v)dv \text{ if } u_1 \le u_0. \end{cases} \tag{1.2}$$

This is the integral of f along the "positively oriented arc" going from u_0 to u_1 in $[a, b]_*$. Note that

$$\int_{u_0}^{u_1} f(v)dv_* \neq -\int_{u_1}^{u_0} f(v)dv_*$$

in general. For example $\int_b^a f(v)dv_* = 0$ for all f.

The length of the positively oriented arc going from u_0 to u_1 is

$$\ell_*(u_0, u_1) = \int_{u_0}^{u_1} dv_* := \begin{cases} |u_1 - u_0| \text{ if } u_0 \le u_1 \\ (b-a) - |u_1 - u_0| \text{ if } u_1 \le u_0. \end{cases}$$

With this notation, $d_*(u_0, u_1) = \min(\ell_*(u_0, u_1), \ell_*(u_1, u_0))$.

1.2 Reparametrization Equivalence

1.2.1 Open Curves

Definition 1.3 Let $m : I \to \mathbb{R}^2$ be a plane curve. A change of parameter for m is a function $\psi : I' \to I$ such that:
(i) I' is a bounded interval;
(ii) ψ is continuous, increasing (strictly) and onto.

From (ii), ψ is one-to-one and onto, hence invertible. Its inverse, ψ^{-1} is also a change of parameter (the proof being left to the reader). In particular, ψ is a *homeomorphism* (a continuous invertible function with a continuous inverse).

The new curve $\tilde{m} = m \circ \psi$ is called a reparametrization of m. The ranges $\mathcal{R}_m$ and $\mathcal{R}_{\tilde{m}}$ coincide.

When m belongs to a specific smoothness class, the same properties will be implicitly required for the change of parameter. For example, if m is (piecewise) C^1, ψ will also be assumed to be C^1 (in addition to the previous properties). When working with regular curves, the following assumption will be made.

Definition 1.4 If I, I' are bounded intervals, a regular change of parameter is a C^1 function $\psi : I' \to I$ which is onto and satisfies $\dot{\psi} > 0$ everywhere (including left and right limits at the bounds).

A piecewise regular change of parameter is continuous, piecewise C^1 and such that its left and right derivatives (which coincide everywhere except at a finite number of points) are all strictly positive.

It is easy to see that the property of two curves being related by a change of parameter is an equivalence relation. This is called "parametric equivalence." We will denote the parametric equivalence class of m by $[m]$. A property, or a quantity, which only depends on $[m]$ will be called parametrization-invariant. For example, the range of a curve is parametrization-invariant.

Note that the converse is not true. If two curves have the same range, they are not necessarily parametrically equivalent: the range of the piecewise C^1 curve defined on $I = [0, 1]$ by $m(t) = (2t, 0), t \in [0, 1/2]$ and $m(t) = (2 - 2t, 0),\ t \in [1/2, 1]$ is the segment $[0, 1] \times 0$, but this curve is not equivalent to $\tilde{m}(t) = (t, 0), t \in [0, 1]$, even though they have the same range (the first one travels back to its initial point after reaching the end of the segment). Also, if m is a curve defined on $I = [0, 1]$, then $\tilde{m}(t) = m(1 - t)$ has the same range, but is not equivalent to m, since we have required the change of parameter to be increasing (changes of orientation are not allowed).

Changes of parameter will always be assumed to match the class of curves that is being considered: (piecewise) regular reparametrizations for (piecewise) regular curves, or, when more regularity is needed, C^p regular reparametrizations for C^p regular curves.

1.2.2 Closed Curves

Changes of parameters for closed curves must be slightly more general than for open curves, because the starting point of the parametrization is not uniquely defined. Using representations over tori, we will say that a continuous mapping $\psi : [a', b']_* \to [a, b]_*$ is increasing if, for all u, one can write, for small enough δ,

$$\psi(u +_* \delta) = \psi(u) +_* \varepsilon(\delta),$$

where $\varepsilon : (-\delta_0, \delta_0) \to \mathbb{R}$ can be defined for some $\delta_0 > 0$ as a (strictly) increasing function such that $\varepsilon(\delta) \to 0$ if $\delta \to 0$. This says that ψ moves in the same direction as u.

A change of parameter is then a continuous, increasing, one-to-one transformation ψ from $[a', b']_*$ onto $[a, b]_*$ (and its inverse is then continuous too). The main difference with the open case is that such a transformation does not necessarily satisfy $\psi(a') = a$: it can start anywhere and wrap around to return to its initial point. We will then say that the change of parameter is regular if it is C^1 with $\dot{\psi} > 0$ everywhere, as in the open case.

Letting $c' = \psi^{-1}(b)$ (recall that $a = b \in [a, b]_*$) and taking $\hat{\psi} : [a', b'] \to [a, b]$ to be such that $\hat{\psi}(u') = \psi(u')$ if $u' \neq c'$ and $\hat{\psi}(u') = b$ otherwise, the definition is equivalent to requiring that $\hat{\psi}$ is increasing over $[a', c']$ and over $(c', b']$, continuously differentiable over these intervals, with left and right derivatives coinciding at c', and the right derivative at a' coinciding with the left derivative at b'.

1.3 Unit Tangent and Normal

If $M \subset \mathbb{R}^d$ is an arbitrary set, we will say that a vector $v \in \mathbb{R}^d$ is tangent to M at a point p in M if one can find points x in M that are arbitrarily close to p and such that v is arbitrarily close to the half line $\mathbb{R}^+(x - p)$. This is formalized in the following definition (see [107]):

Definition 1.5 If $M \subset \mathbb{R}^d$, and $p \in M$, a vector $v \in \mathbb{R}^d$ is an oriented tangent to M at p if, for any $\varepsilon > 0$, there exist $x \in M$ and $r > 0$ such that $|x - p| < \varepsilon$ and $|v - r(x - p)| < \varepsilon$.

The set of oriented tangents to M at p will be denoted by $T_p^+ M$, and the set of (unoriented) tangents by $T_p M$, so that $v \in T_p M$ if either v or $-v$ belongs to $T_p^+ M$. Taking $x = p$, one sees that $v = 0$ always belong to $T_p^+ M$, which is therefore never empty.

Let $m : I \to \mathbb{R}^2$ be a regular curve (here, I can be a closed interval or a torus). The unit tangent at $u \in I$ is the vector

$$T_m(u) = \frac{\dot{m}(u)}{|\dot{m}(u)|}.$$

We then have

Proposition 1.6 *If $m : I \to \mathbb{R}^2$ is regular, and $p \in \mathcal{R}_m$, then*

$$T_p \mathcal{R}_m = \{\lambda T_m(u) : \lambda \in \mathbb{R}, m(u) = p\}.$$

Note that, when m is regular the set of parameters u such that $m(u) = p$ is necessarily finite. (Each such u is necessarily isolated because $\dot{m} \neq 0$ and any family of isolated points in a compact set must be finite.)

Proof Let $I = [a, b]$ (the case of a closed curve being addressed similarly). Take $p \in \mathcal{R}_m$ and u such that $p = m(u)$. Fix $\lambda \in \mathbb{R}$ and $\varepsilon > 0$. One has $m(u + \delta) - m(u) - \delta|\dot{m}(u)|T_m(u) = o(\delta)$. So, taking δ small enough so that $|m(u + \delta) - m(u)| < \varepsilon$ and

$$\left| \lambda T_m(u) - \frac{\lambda}{\delta|\dot{m}(u)|}(m(u + \delta) - m(u)) \right| < \varepsilon$$

one gets $|x - p| < \varepsilon$ and $|\lambda T_m(u) - r(x - p)| < \varepsilon$ with $x = m(u + \delta)$ and $r = \lambda/(\delta|\dot{m}|)$. If $u \in (a, b)$, one can ensure that $r > 0$ by choosing the sign of δ appropriately. If $u = a$ one must take $\delta > 0$ and $r > 0$ only if $\lambda > 0$. Similarly, if $u = b$, one needs $\lambda < 0$. In any case, either $\lambda T_m(u)$ or $-\lambda T_m(u)$ belongs to $T_p^+\mathcal{R}_m$, so that $\lambda T_m(u) \in T_p\mathcal{R}_m$.

Conversely, if $v \in T_p\mathcal{R}_m$, there exists sequences (u_n) and (r_n) with $u_n \in I$ and $r_n > 0$ such that $|m(u_n) - p| < 1/n$ and $|v - r_n(m(u_n) - p)| < 1/n$. Taking a subsequence if needed, one can assume that $u_n \to u \in I$, necessarily with $p = m(u)$. If $u_n = u$ for an infinite number of v, then $|v| \leq 1/n$ for these n, which implies $v = 0$. Otherwise, remove these values from the sequence to ensure $u_n \neq u$ for all n and use the fact that

$$v - r_n(u_n - u)\frac{m(u_n) - p}{u_n - u} \to 0$$

with $(m(u_n) - p)/(u_n - u) \to \dot{m}(u) \neq 0$ to prove that $r_n(u_n - u)$ converges to some $\lambda \in \mathbb{R}$. We then have $v = \lambda \dot{m}_u$, which completes the proof. □

The unit normal is the unique vector $N_m(u)$ which extends $T_m(u)$ to a positively oriented orthonormal basis of $\mathbb{R}^2$: $(T_m(u), N_m(u))$ is orthonormal and $\det[T_m(u), N_m(u)] = 1$. The subscript m is generally dropped in the absence of ambiguity.

The frame (T, N) is parametrization-invariant in the following sense: if $\varphi : I \to \tilde{I}$ is a regular change of parameter, and $m = \tilde{m} \circ \varphi$, then $T_{\tilde{m}}(\varphi(u)) = T_m(u)$ and similarly for the normal.

1.4 Embedded Curves

Letting I be either an interval or a torus, a C^1 function $m : I \to \mathbb{R}^2$ such that $\dot{m}(u) \neq 0$ everywhere is a special case of an immersion (see Definition B.13), and regular curves are also sometimes called *immersed curves*. Among immersed curves, one also distinguishes *embedded curves* which are furthermore assumed to be non-intersecting (so that closed embedded curves are regular Jordan curves). For embedded curves, $T_m(u)$ is (up to a sign change) the only unit element of $T_{m(u)}M$. Moreover, if $m : I \mapsto \mathbb{R}^2$ is an embedding, the inverse map $m^{-1} : \mathcal{R}_m \to I$, which is well defined by assumption, is continuous: if $p_n \in \mathcal{R}_m$ is a sequence that converges to $p \in \mathcal{R}_m$, then, for some u_n and u, $p_n = m(u_n)$ and $p = m(u)$. Any limit v of a subsequence of u_n (recall that I is compact, so that any sequence has at least a convergent subsequence, and any limit of a subsequence belongs to I) must satisfy, by continuity, $m(v) = p$, which implies $v = u$. This implies that $m^{-1}(p) = u$.

If two embedded curves have the same range, they can be deduced from one another through a change of parameters, possibly after reorientation (this is not true for regular curves). Letting $m : I \mapsto \mathbb{R}^2$ and $m' : I' \mapsto \mathbb{R}^2$ be two such curves, $\psi = m^{-1} \circ m'$ is a homeomorphism (continuous, with a continuous inverse) between I' and I. If v is any point in the case of closed curves, or $v \in (a', b')$ for open curves, one can apply the implicit function theorem to the identity $m' = m \circ \psi$ to prove that ψ is differentiable with

$$\dot{\psi}\, \dot{m} \circ \psi = \dot{m}',$$

implying

$$\dot{\psi} = \frac{(\dot{m}')^T \, \dot{m} \circ \psi}{|\dot{m} \circ \psi|^2}.$$

For open curves, one shows that ψ has non-zero right and left derivatives at a' and b' by passing to the limit, the detailed argument being left to the reader. Because of this, any parametrization-invariant quantity only depends on the (oriented) range of the curve when restricted to embeddings.

With some abuse of terminology, we will say that a subset $\mathcal{R} \subset \mathbb{R}^2$ is an embedded curve if there exists an embedded curve m such that $\mathcal{R} = \mathcal{R}_m$. Such a curve m is then defined up to a change of parameter.

1.5 The Integral Along a Curve and Arc Length

Let $m : [a, b] \to \mathbb{R}^2$ be a parametrized curve. If $\sigma = (a = u_0 < u_1 < \cdots < u_n < u_{n+1} = b)$ is a subdivision of $[a, b]$, one can approximate m by the polygonal line m_σ with vertices $(m(u_0), \ldots, m(u_{n+1}))$. The length of m_σ is the sum of lengths of the segments that form it, namely

$$L_{m_\sigma} = \text{length}(m_\sigma) = \sum_{i=1}^{n+1} |m(u_i) - m(u_{i-1})|.$$

One then defines the length of m as

$$L_m = \sup_\sigma L_{m_\sigma},$$

where the supremum (which can be infinite) is over all possible subdivisions σ of $[a, b]$.

One then has the following proposition.

Proposition 1.7 *If $m : [a, b] \to \mathbb{R}^2$ is C^1, then*

$$L_m = \int_a^b |\dot{m}(t)| dt < \infty.$$

Proof The fact that the integral is finite results from the derivative being bounded on the compact interval $[a, b]$ (because the curve is C^1). If $\sigma = (a = u_0 < u_1 < \cdots < u_n < u_{n+1} = b)$ is a subdivision of $[a, b]$, then one has

$$|m(u_{i+1}) - m(u_i)| = \left| \int_{u_i}^{u_{i+1}} \dot{m}(t) dt \right| \leq \int_{u_i}^{u_{i+1}} |\dot{m}(t)| dt.$$

Summing over i yields the fact that $L_{m_\sigma} \leq \int_a^b |\dot{m}(t)|dt$ and taking the supremum on the right-hand side implies the same inequality for L_m.

On the other hand, given any σ, the finite increment theorem implies that, for all i, there exists $v_i \in (u_i, u_{i+1})$ such that $m(u_{i+1}) - m(u_i) = \dot{m}(v_i)(u_{i+1} - u_i)$. Using this, we see that

$$L_{m_\sigma} = \sum_{i=0}^{n} |\dot{m}(v_i)|(u_{i+1} - u_i),$$

which is a Riemann sum for $\int_a^b |\dot{m}(t)|dt$, and can therefore be made arbitrarily close to the integral by taking fine enough subdivisions. So for any ε, one can find σ such that

$$\int_a^b |\dot{m}(t)|dt \leq L_{m_\sigma} + \varepsilon$$

and since the upper-bound is less than $L_m + \varepsilon$, we find $\int_a^b |\dot{m}(t)|dt \leq L_m$ by letting ε tend to 0. This completes the proof of the proposition. □

If $f : I \to \mathbb{R}$ is a continuous function, one defines the integral of f along m by

$$\int_m f\, d\sigma_m = \int_a^b f(u)\, |\dot{m}(u)|\, du. \tag{1.3}$$

The definition is parametrization-independent: if $\psi : [a', b'] \to [a, b]$ is a change of parameters, then, using a change of variable,

$$\begin{aligned}\int_{a'}^{b'} f(\psi(u'))|\partial(m \circ \psi)(u')|du' &= \int_{a'}^{b'} f(\psi(u'))|\dot{m} \circ \psi(u')|\, \dot{\psi}(u')\, du' \\ &= \int_a^b f(u)|\dot{m}(u)|du\end{aligned}$$

so that

$$\int_{m\circ\psi} f \circ \psi\, d\sigma_{m\circ\psi} = \int_m f\, d\sigma_m.$$

The same result holds if $\psi : [a', b']_* \to [a, b]_*$ is a change of parameter between closed curves. In that case, taking c' such that $\psi(c') = a \sim b$ and letting $\psi(a' \sim b') = c$, we have

$$\begin{aligned}\int_{a'}^{b'} f \circ \psi(u')du' &= \int_{a'}^{c'} f \circ \psi(u')|\partial(m \circ \psi)(u')|du' \\ &\quad + \int_{c'}^{b'} f \circ \psi(u')|\partial(m \circ \psi)(u')|du' \\ &= \int_c^b f(u)|\dot{m}(u)|du + \int_a^c f(u)|\dot{m}(u)|du\end{aligned}$$

$$= \int_a^b f(u)|\dot{m}(u)|du.$$

These results imply that, when m is an embedding, the integral along m only depends on the range $\mathcal{R}_m$. This allows us to define the integral of a function over $\mathcal{R}_m$ by

$$\int_{\mathcal{R}_m} f \, d\sigma_{\mathcal{R}_m} = \int_m f \, d\sigma_m,$$

which does not depend on how $\mathcal{R}_m$ is parametrized.

We now give the following important definition.

Definition 1.8 Let $m : I \to \mathbb{R}^2$ be a (piecewise) C^1 curve, where I is either $[a, b]$ or $[a, b]_*$. A change of parameter $\sigma : I \to [0, L_m]$ (or $[0, L_m]_*$) is an arc-length reparametrization of m if

$$\dot{\sigma} = |\dot{m}|.$$

One says that m is parametrized by arc length if $m : [0, L_m] \to \mathbb{R}^2$ satisfies $|\dot{m}| = 1$.

If m is regular, then σ is a regular change of parameter and $m \circ \sigma^{-1}$ is an arc-length reparametrization of m. When $I = [a, b]$, the arc-length reparametrization is unique and given by

$$\sigma_m(u) = \int_a^u |\dot{m}(v)|dv. \tag{1.4}$$

When $I = [a, b]_*$, the parametrization is unique once the starting point $c = \sigma^{-1}(0)$ is chosen, and is given by (following (1.2))

$$\sigma_{m,c}(u) = \int_c^u |\dot{m}(v)|dv_*. \tag{1.5}$$

The arc length is parametrization-invariant: if m is a curve, with arc-length reparametrization σ, and $\tilde{m} = m \circ \psi$ is another parametrization of m, then $\sigma \circ \psi$ is an arc-length parametrization of $\tilde{m}$ (this is obvious, since $\tilde{m} \circ (\sigma \circ \psi)^{-1} = m \circ \sigma^{-1}$).

When a curve is parametrized by arc length, it is customary to denote its parameter by s instead of u, and we will follow this convention. From our definition of integrals, we clearly have in that case

$$\int_m f d\sigma_m = \int_0^{L_m} f(s)ds$$

(or ds_* in the case of closed curves).

We will also use the notion of derivative with respect to the arc length. For open curves, this corresponds to the limit of the ratio

$$\frac{g(u+\varepsilon)-g(u)}{\sigma_m(u+\varepsilon)-\sigma_m(u)}$$

as $\varepsilon \to 0$ (for closed curves, replace $+$ by $+_*$) therefore leading to the following definition.

Definition 1.9 Let $m : I \to \mathbb{R}^2$ be a C^1 regular curve. The operator ∂_{s_m} transforms a C^1 function g over I into the function $\partial_{s_m} g$, which is defined over I by

$$\partial_{s_m} g(u) = \frac{\dot{g}(u)}{|\dot{m}(u)|}. \tag{1.6}$$

We will write ∂_s if there is no ambiguity concerning the curve m. Note that, if m is parametrized by arc length (so that $u = s$), this notation coincides with the usual derivative with respect to s and therefore introduces no conflict.

The next proposition expresses that the derivative with respect to the arc length is parametrization invariant.

Proposition 1.10 *Let $m : I \to \mathbb{R}^2$ be a regular curve and $\psi : I' \to I$ be a change of parameter, with $\tilde{m} = m \circ \psi$. Then, for any C^1 function g defined on I,*

$$(\partial_{s_m} g) \circ \psi = \partial_{s_{\tilde{m}}}(g \circ \psi).$$

Proof This derives from the definition and from the chain rule, namely

$$\partial_{s_{\tilde{m}}}(g \circ \psi) = \frac{\partial_{\tilde{u}}(g \circ \psi)}{|\partial_{\tilde{u}}(m \circ \psi)|} = \frac{(\partial_u g) \circ \psi}{|(\partial_u m) \circ \psi|} = (\partial_{s_m} g) \circ \psi$$

(the positive term $\partial_{\tilde{u}}\psi$ cancels in the ratio). □

Note that, with this definition, one can rewrite the definition of the unit tangent as $T_m = \partial_{s_m} m$.

The following proposition shows how the arc length parametrization can be used to stitch several local parametrizations of a set to a global one forming an embedding.

Proposition 1.11 *Let $\mathcal{R} \subset \mathbb{R}^2$ be compact and connected. Assume that there exists a family $V_1, \dots, V_n$ of open sets in $\mathbb{R}^2$, and a family $m_i : [a_i, b_i] \to \mathbb{R}^2$ of embeddings, such that $\mathcal{R} \subset \bigcup_{i=1}^n V_i$ and, for every $i = 1, \dots, n$, $\mathcal{R} \cap V_i = m_i((a_i, b_i))$. Then, there exists a closed embedding $m : [a, b]_* \to \mathbb{R}^2$ such that $\mathcal{R} = \mathcal{R}_m$.*

Proof Note that, since $\mathcal{R}$ is compact (hence closed), it contains each extremity $m_i(a_i)$ or $m_i(b_i)$. Also, assume, without loss of generality, that each curve is parametrized by arc length so that $a_i = 0$ and $b_i = L_i$ (the length of m_i). Let $I^{(1)} = [0, L_1]$ and $m^{(1)} = m_1$, and define the following iterative construction.

Given the current interval $I_n = [0, \ell_n]$ and embedding $m : I_n \to \mathbb{R}^2$ such that $\mathcal{R}_m \subset \mathcal{R}$, choose an index j such that $m(\ell_n) \in \mathcal{R} \cap V_j$ and $\mathcal{R}_{m_j} \not\subset \mathcal{R}_m$. Let $\mathcal{R}_m^0 = m((0, \ell_n))$, the set $\mathcal{R}_m$ without its extremities.

Define $u_j \in (0, L_j)$ by $m_j(u_j) = m(\ell_n)$. Let A_j be the connected component of $m^{-1}(\mathcal{R}_m^0 \cap V_j)$ that contains u_j: A_j is a sub-interval of $(0, L_j)$ taking either the form (x_j, u_j), $0 < x_j < u_j$ or (u_j, y_j), $u_j < y_j < L_j$. Reorienting m_j if needed, assume that $\mathcal{R} \cap V_j = (x_j, u_j)$.

We now consider two cases.

(i) $m_j([u_j, L_j]) \cap \mathcal{R}_m = \emptyset$. Define $\ell_{n+1} = \ell_n + L_j - u_j$ and extend m to $[0, \ell_{n+1}]$ by $m(u) = m_j(u - \ell_n + u_j)$ for $u > \ell_n$. Then m is an embedding and the construction can continue.

(ii) $m_j([u_j, L_j]) \cap \mathcal{R}_m \neq \emptyset$. Let $v_j > u_j$ be the first parameter such that $m_j(v_j) \in \mathcal{R}_m$. If $m(v_j) \neq m_1(0)$, then, by construction, there exists a V_i, $i \neq j$ such that $m(v_j) \in \mathcal{R}_m \cap V_i$. This implies that m_j coincides with m_i in $m_j((u_j, v_j) \cap V_i$, but this contradicts the fact that v_i was the first point of self-intersection.
So we have $m(v_i) = m_1(0)$ and we conclude the construction with $\ell_{n+1} = \ell_n + v_j - u_j$, extending m to $[0, \ell_{n+1}]$ by $m(u) = m_j(u - \ell_n + u_j)$ for $u > \ell_n$.

Note that we always reach case (ii) (there are at most n steps). After case (ii) is completed, $\mathcal{R}_m$ is an embedded closed curve which is necessarily equal to $\mathcal{R}$, which is connected. □

1.6 Curvature

The curvature of a C^2 regular curve $m : I \to \mathbb{R}^2$ is a function $\kappa_m : I \to \mathbb{R}$, related to the arc-length derivative of the tangent through the formula:

$$\partial_{s_m} T_m = \kappa_m N_m. \tag{1.7}$$

Note that $T_m^T T_m = 1$ implies that $T_m^T \partial_{s_m} T_m = 0$ so that $\partial_{s_m} T_m$ is collinear to N_m and κ_m is the coefficient of collinearity. From the remark made at the end of the previous section, one also has

$$\kappa_m N_m = \partial_{s_m}^2 m, \tag{1.8}$$

the second derivative of the curve with respect to its arc length. This implies that

$$\kappa_m = N_m^T \partial_{s_m}^2 m = \det(T_m, \partial_{s_m}^2 m). \tag{1.9}$$

Assume that T_m can be expressed as $T_m(u) = (\cos\theta_m(u), \sin\theta_m(u))$ (so that $N_m = (-\sin\theta_m, \cos\theta_m)$) where θ is differentiable in u (we will show below that this is always true). Then, from a direct computation, $\partial_{s_m} T_m = \partial_{s_m}\theta N_m$, from which we deduce an alternative interpretation of κ_m:

$$\kappa_m(u) = \partial_{s_m}\theta_m(u), \tag{1.10}$$

where θ_m is a C^1 version of the angle between T_m and the "horizontal axis."

The same kind of easy computation yields

$$\partial_{s_m} N_m = -\kappa_m T_m \tag{1.11}$$

and Eqs. (1.7) and (1.11) together form what are called the *Frénet formulas* for the curve m.

Since it is defined as a double arc-length derivative, the curvature is parametrization invariant. Indeed, if $\tilde{m} = m \circ \psi$, then, applying Proposition 1.10 twice,

$$\partial^2_{s_{\tilde{m}}} \tilde{m} = \partial_{s_{\tilde{m}}}((\partial_{s_m} m) \circ \psi) = (\partial^2_{s_m} m) \circ \psi$$

so that $\kappa_{\tilde{m}} = \kappa_m \circ \psi$.

When $\kappa_m(u) \neq 0$, one defines the radius of curvature $\rho_m(u) = 1/|\kappa_m(u)|$ and the center of curvature $c_m(u) = m(u) + N_m(u)/\kappa_m(u)$. The circle with center $c_m(u)$ and radius $\rho_m(u)$ is called the *osculating circle* of the curve at $m(u)$.

We now prove the fact that a smooth version of the tangent angle θ exists as a consequence of the following lemma.

Lemma 1.12 *Let $I = [a, b]$ or $[a, b]_*$ and $f : I \to \mathbb{R}^2$ be a C^p function satisfying $|f(u)| = 1$ for all $u \in I$, with $p \geq 0$. Assume that for all $u \in I$, there is a small neighborhood $J_u \subset I$ and a C^p function $\tau_u : J_u \to \mathbb{R}$ such that $f(u') = (\cos\tau(u'), \sin\tau(u'))$ for $u' \in J_u$. Then there exists a C^p function $\tau : I \to \mathbb{R}$ such that $f = (\cos\tau, \sin\tau)$.*

Proof Since I is compact, we can find a finite number of $u_1, \ldots u_n$ such that $I = \bigcup_{i=1}^n J_{u_i}$. The result can then be proved by induction on n. There is nothing to prove if $n = 1$. Assume that $n > 1$ and that the result is true for $n - 1$. Then there must exist a subset J_{u_j} with $j \neq n$ such that $J_{u_j} \cap J_{u_n} \neq \emptyset$. Assume without loss of generality that $j = n - 1$. There must exist an integer k such that, for any u in this intersection, $\tau_{u_n}(u) = \tau_{u_{n-1}}(u) + 2k\pi$. Define $\tilde{J}_{u_{n-1}} = J_{u_{n-1}} \cup J_{u_n}$ and $\tilde{\tau}_{u_{n-1}}(u) = \tau_{u_{n-1}}(u)$ on $J_{u_{n-1}}$ and $\tilde{\tau}_{u_{n-1}}(u) = \tau_{u_n}(u) - 2k\pi$ on J_{u_n}, so that $\tilde{\tau}$ is C^p on $\tilde{J}_{u_{n-1}}$. Then we can apply the induction hypothesis to $J_{u_1}, \ldots, J_{u_{n-2}}, \tilde{J}_{u_{n-1}}$ with associated functions $\tau_{u_1}, \ldots, \tau_{u_{n-2}}, \tilde{\tau}_{u_{n-1}}$. □

To prove the existence of a differentiable $\theta(u)$, the lemma needs to be applied with $p = 1$, $f = T_m$, $\tau = \theta$ and $\tau_u(u') = \theta_0(u) + \arcsin(\det(T_m(u), T_m(u')))$.

1.7 Expression in Coordinates

1.7.1 Cartesian Coordinates

To provide explicit formulas for the quantities that have been defined so far, we introduce the space coordinates (x, y) and write, for a curve m: $m(u) = (x(u), y(u))$.

The first, second and higher derivatives of x will be denoted by $\dot{x}, \ddot{x}, x^{(3)}, \ldots$ and similarly for y. The tangent and the normal vector expressions in coordinates are

$$T = \frac{1}{\sqrt{\dot{x}^2 + \dot{y}^2}} \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix}, \quad N = \frac{1}{\sqrt{\dot{x}^2 + \dot{y}^2}} \begin{pmatrix} -\dot{y} \\ \dot{x} \end{pmatrix}.$$

The arc length is $ds = \sqrt{\dot{x} + \dot{y}}\, du$ and the curvature is

$$\kappa = \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}. \tag{1.12}$$

The last formula is proved as follows. Since $\kappa N = \partial_s T$, we have

$$\kappa = N^T \partial_s T.$$

Using $T = \partial_s m = \dot{m}/\dot{s}$, we have

$$N^T \partial_s T = \dot{s}^{-2} \ddot{m}^T N + \dot{s}^{-1} \partial \left(\dot{s}^{-1}\right) \dot{m}^T N.$$

The last term vanishes, and the first one gives (1.12) after introducing the coordinates.

1.7.2 Polar Coordinates

Let (Oxy) be a fixed frame. A point m in the plane can be characterized by its distance, r, to the origin, O, and by θ, the angle between the horizontal axis (Ox) and the half-line Om. (Notice that this is different from the angle of the tangent with the horizontal, for which we also used θ. Unfortunately, this is the standard notation in both cases.) The relation between the Cartesian coordinates (x, y) of m and its polar coordinates (r, θ) is $(x = r \cos\theta, y = r \sin\theta)$. This representation is unique, except for $m = O$, for which θ is undetermined.

A polar parametrization of a curve $u \mapsto m(u)$ is a function $u \mapsto (r(u), \theta(u))$. Often, the parameter u coincides with the angle θ and the parametrization reduces to a function $r = f(\theta)$. Some shapes have very simple polar coordinates, the simplest being a circle centered at O for which the equation is $r = \text{const}$.

Let us compute the Euclidean curvature from such a parametrization. Let $\tau = (\cos\theta, \sin\theta)$ and $\nu = (-\sin\theta, \cos\theta)$. We have $m = r\tau$, and

$$\dot{m} = \dot{r}\tau + r\dot{\theta}\nu,$$

$$\ddot{m} = (\ddot{r} - r\dot{\theta}^2)\tau + (2\dot{r}\dot{\theta} + r\ddot{\theta})\nu.$$

Therefore,

$$\kappa = \frac{\det[\dot{m}, \ddot{m}]}{|\dot{m}|^3} = \frac{r^2(\dot{\theta})^3 - r\ddot{r}\dot{\theta} + 2\dot{r}^2\dot{\theta} + r\dot{r}\ddot{\theta}}{(\dot{r}^2 + r^2\dot{\theta}^2)^{3/2}}.$$

When the curve is defined by $r = f(\theta)$, we have $\theta = u$, $\dot{\theta} = 1$ and $\ddot{\theta} = 0$, so that

$$\kappa = \frac{r^2 - r\ddot{r} + 2\dot{r}^2}{(\dot{r}^2 + r^2)^{3/2}}.$$

The polar representation does not have the same invariance properties as the arc length (see the next section), but still has some interesting features. Scaling by a factor λ simply corresponds to multiplying r by λ. Making a rotation with center O and angle α simply means replacing θ by $\theta + \alpha$. However, there is no simple relation for a translation. This is why a curve is generally expressed in polar coordinates with respect to a curve-dependent origin, such as its center of gravity.

1.8 Euclidean Invariance

The arc length and the curvature have a fundamental invariance property. If a curve is transformed by a rotation and translation, both quantities are invariant. The rigorous statement of this is as follows. Let R be a planar rotation and b a vector in $\mathbb{R}^2$. Define the transformation $g : \mathbb{R}^2 \to \mathbb{R}^2$ by $g(p) = Rp + b$. Then, if $m : I = [a, b] \to \mathbb{R}^2$ is a plane curve, one can define $g \cdot m : I \to \mathbb{R}^2$ by $(g \cdot m)(u) = g(m(u)) = Rm(u) + b$. Then, the statements are:

(i) $\sigma_{g \cdot m}(u) = \sigma_m(u)$, and in particular $L_{g \cdot m} = L_m = L$.
(ii) The curvatures κ_m and $\kappa_{g \cdot m}$, reparametrized over $[0, L]$ (as functions of the arc length), coincide.

The proof of (i) is straightforward from the definition of σ_m (see Eq. (1.4)). For (ii), use $\partial^2_{s_m}(g \cdot m) = R\partial^2_{s_m} m$, $N_{g \cdot m} = RN_m$ and (1.9).

Note that in this discussion we have taken $I = [a, b]$, an interval, for which the arc length reparametrization is uniquely defined by (1.4). If one wants to consider "wrapped intervals" $[a, b]_*$, arc lengths should be compared with the same inverse image of 0 (c in (1.5)).

We now state and prove the converse statement of (ii).

Theorem 1.13 (Characterization Theorem) *If two C^2 regular plane curves m and $\tilde{m}$ have the same curvature as a function of the arc length, denoted $\kappa : [0, L] \to \mathbb{R}$, then there exist R and b, and a change of parameter, ψ, such that $\tilde{m} = Rm \circ \psi + b$.*

With our notation, the assumption means that

$$\kappa = \kappa_m \circ \sigma_m^{-1} = \kappa_{\tilde{m}} \circ \sigma_{\tilde{m}}^{-1}$$

and implicitly implies that the lengths of the two curves coincide (with L).

Proof Let m^* and $\tilde{m}^*$ be m and $\tilde{m}$ reparametrized with arc length. We prove that

$$\tilde{m}^* = Rm^* + b$$

for some R and b, which implies the statement of the theorem after reparametrization. Equivalently, we assume without loss of generality that both m and $\tilde{m}$ are parametrized by arc length.

Now, let $\kappa : [0, L] \to \mathbb{R}$ be an integrable function. We build all possible curves m that are parametrized by arc length over $[0, L]$ and have κ as curvature and prove that they all differ by a rotation and translation. By definition, the angle θ_m, defined over $[0, L]$, must satisfy:

$$\dot{\theta}_m = \kappa \text{ and } \dot{m} = (\cos\theta_m, \sin\theta_m).$$

Let $\theta(s) = \int_0^s \kappa(u)du$. The first equality implies that, for some $\theta_0 \in [0, 2\pi)$, we have $\theta_m(s) = \theta(s) + \theta_0$ for all $s \in [0, L]$. The second implies that, for some $b \in \mathbb{R}^2$,

$$m(s) = \int_0^s (\cos(\theta(u) + \theta_0), \sin(\theta(u) + \theta_0))du + b.$$

Introduce the rotation $R = \begin{pmatrix} \cos\theta_0 & -\sin\theta_0 \\ \sin\theta_0 & \cos\theta_0 \end{pmatrix}$. From standard trigonometric formulas, we have

$$R\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} \cos(\theta + \theta_0) \\ \sin(\theta + \theta_0) \end{pmatrix}$$

so that, letting $\hat{m}(s) = \int_0^s (\cos\theta(u), \sin\theta(u))du$, we have $m = R\hat{m} + b$. Since $\hat{m}$ is uniquely defined by κ, we obtain the fact that m is uniquely defined up to a rotation and translation. □

1.9 The Frénet Frame

If m is a C^2 regular plane curve, its Frénet frame is defined by

$$F_m(u) = \begin{pmatrix} T_m(u) & N_m(u) \end{pmatrix}.$$

Considering T_m and N_m as column vectors, F_m is a rotation matrix satisfying $F_m^T F_m = \mathrm{Id}$ and $\det(F_m) = 1$. It is a *moving frame* along the curve.

Equations (1.7) and (1.11), which, put together, form the *Frénet formulas* for plane curves, can be summarized in matrix form as

$$\partial_{s_m} F_m = F_m S_m \tag{1.13}$$

with

$$S_m = \begin{pmatrix} 0 & -\kappa_m \\ \kappa_m & 0 \end{pmatrix}. \tag{1.14}$$

Note that, applying ∂_{s_m} to $F_m^T F_m = \mathrm{Id}$, we get

$$(\partial_{s_m} F_m)^T F_m + F_m^T \partial_{s_m} F_m = 0,$$

which states that the matrix $F_m^T \partial_{s_m} F_m$ (which is equal to S_m) must be skew-symmetric ($S_m^T = -S_m$). This implies that Eqs. (1.13) and (1.14) can be used as alternative definitions of the curvature, via the Frénet formulas.

The advantage of this construction is that it generalizes to arbitrary dimensions (cf. Sect. 3.1), and to more general forms of moving frames (like affine, or projective frames). It also leads to an alternative proof of the Characterization Theorem, as detailed below.

Proof (Alternative proof of Theorem 1.13) If one applies a rotation, R, and a translation to a curve m, the Frénet frame of the new curve, $\tilde{m}$, is $F_{\tilde{m}} = RF_m$, and using $R^T R = \mathrm{Id}$ and $\sigma_m = \sigma_{\tilde{m}}$, we have

$$S_{\tilde{m}} = F_{\tilde{m}}^T \partial_{s_{\tilde{m}}} F_{\tilde{m}} = F_m^T \partial_{s_m} F_m = S_m.$$

We therefore retrieve the fact that κ_m is invariant under rotation. The invariance by change of parameter is again a consequence of the invariance of the arc-length derivative.

We now prove the converse, assume that m and $\tilde{m}$ are such that $S_m = S_{\tilde{m}} =: S$ with both curves parametrized by arc length (as in the first proof of Theorem 1.13, it suffices to restrict to this case).

Let $G_m(s) = F_m(0)^T F_m(s)$ and $G_{\tilde{m}}(s) = F_{\tilde{m}}(0)^T F_{\tilde{m}}(s)$, so that

$$\begin{cases} \dot{G}_m = G_m S \\ \dot{G}_{\tilde{m}} = G_{\tilde{m}} S. \end{cases}$$

Both G_m and $G_{\tilde{m}}$ are therefore solutions of the differential equation $\dot{G} = GS$. We have, in addition $G_{\tilde{m}}(0) = G_m(0) = \mathrm{Id}$, and the theory of differential equations states that two functions that satisfy the same linear differential equation with the same initial condition must coincide. Thus $G_{\tilde{m}} = G_m$, which yields $F_{\tilde{m}} = RF_m$ with $R = F_{\tilde{m}}(0) F_m(0)^T$. This implies, in particular, that $T_{\tilde{m}} = RT_m$, and, since $T_m = \dot{m}_s$ for curves parametrized with arc length,

$$\tilde{m}(s) - \tilde{m}(0) = \int_0^s T_{\tilde{m}}(u)du = \int_0^s RT_m(u)du = Rm(s) - Rm(0)$$

so that $\tilde{m} = Rm + b$ with $b = \tilde{m}(0) - Rm(0)$. □

1.10 Enclosed Area and the Green (Stokes) Formula

When a closed curve m is embedded, its enclosed area can be computed with a single integral instead of a double integral. Let Ω_m be the bounded connected component of $\mathbb{R}^2 \setminus \mathcal{R}_m$. We assume that m is defined on $I = [a, b]_*$, and that *the curve is oriented so that the normal N points inward*, which means that for any $u \in [a, b]_*$, there exists an $\varepsilon > 0$ such that $m(u) + tN(u) \in \Omega_m$ for $0 < t < \varepsilon$. Since this is a convention that will be used repeatedly, we state it as a definition.

Definition 1.14 A closed regular curve oriented so that the normal points inward is said to be positively oriented.

For a circle, positive orientation corresponds to moving counter-clockwise.

We have the following proposition:

Proposition 1.15 *Using the notation above, and assuming that m is positively oriented, we have*

$$\text{Area}(\Omega_m) = \int_{\Omega_m} dx\, dy = -\frac{1}{2}\int_a^b N(u)^T m(u)\, |\dot{m}(u)|\, du. \tag{1.15}$$

Note that the last integral can also be written as $-(1/2)\int_m (N^T m)$, as defined in Sect. 1.5. We also have $N^T m = -\det(m(s), T(s))$ which provides an alternative expression. Indeed, we have

$$\begin{aligned} -(1/2)\int_m (N^T m)\, d\sigma_m &= (1/2)\int_m \det(m, T)\, d\sigma_m \\ &= (1/2)\int_a^b \det(m(u), T(u))|\dot{m}(u)|du \end{aligned}$$

so that, using $T(u) = \dot{m}(u)/|\dot{m}(u)|$,

$$\text{Area}(\Omega_m) = (1/2)\int_a^b \det(m(u), \dot{m}(u))du. \tag{1.16}$$

We will not prove Proposition 1.15, but simply remark that (1.15) is a particular case of the following important theorem.

Theorem 1.16 (Divergence theorem) *If $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a smooth function (a vector field), then*

$$\int_a^b N(u)^T f(m(u))\, |\dot{m}(u)|\, du = -\int_{\Omega_m} \text{div} f\, dx\, dy \tag{1.17}$$

where, letting $f(x, y) = (\alpha(x, y), \beta(x, y))$, one has $\text{div} f = \partial_1\alpha + \partial_2\beta$.

(Here ∂_i denotes the derivative with respect to the ith coordinate.)

Equation (1.17) is called Green's formula. To retrieve Eq. (1.15) from it, take $f(x, y) = (x, y)$, for which $\operatorname{div} f = 2$. Note that Green's formula is sometimes given with a plus sign, N being chosen as the *outward* normal.

Formula (1.15) can also be nicely interpreted as the limit of an algebraic sum of triangle areas. For this, consider a polygonal discretization, say $\tilde{m}$, of m with vertices $p_1, \ldots, p_N$. Let O be an arbitrary point in $\mathbb{R}^2$.

First consider the simple case in which the segment Op_k is included in the region $\Omega_{\tilde{m}}$ for all k (the polygonal curve is said to be star shaped with respect to O). In this case, the area enclosed by the polygon is the sum of the areas of the triangles. The area of (O, p_k, p_{k+1}) is $|\det(p_k p_{k+1}, Op_k)|/2$.[1] Assuming that the discretization is counterclockwise, which is consistent with the fact that the normal points inward, the vectors Op_k and $p_k p_{k+1}$ make an angle between 0 and π, which implies that their determinant is positive. We therefore get

$$\operatorname{Area}(\Omega_{\tilde{m}}) = \frac{1}{2} \sum_{k=1}^{N} \det[Op_k, p_k p_{k+1}]. \tag{1.18}$$

Since this can be written as $\frac{1}{2} \sum_{k=1}^{N} \det(Op_k, p_k p_{k+1}/|p_k p_{k+1}|)|p_k p_{k+1}|$, this is consistent with the continuous formula

$$\frac{1}{2} \int_a^b \det(Om(u), T(u))|\dot{m}(u)|du.$$

The interesting fact is that (1.18) is still valid for polygons which are not star shaped around the origin. In this case, the determinant may take negative values, which provides a necessary correction because, for general polygons, some triangles can intersect $\mathbb{R}^2 \setminus \Omega_m$.

Finally, we mention a classical inequality comparing the area and the perimeter of a simple closed curve.

Theorem 1.17 (Isoperimetric Inequality) *It m is a simple closed curve with perimeter L and area A, then*

$$4\pi A \leq L^2 \tag{1.19}$$

with equality if and only if m is a circle.

1.11 The Rotation Index and Winding Number

Let m be a closed, C^1, plane curve, defined on $I = [a, b]$. Express $T : [a, b] \to S^1$ (the unit circle) as a function $t \mapsto (\cos\theta(t), \sin\theta(t))$ where θ is a *continuous* function (cf. Lemma 1.12).

---

[1] The general expression of the area of a triangle (A, B, C) is $|\det(AB, AC)|/2$, half the area of the parallelogram formed by the two vectors.

Since m is closed, we must have $T(b) = T(a)$, which implies that $\theta(b) = \theta(a) + 2r_m\pi$, where r_m is an integer called the *rotation index* of the curve.

The rotation index is parametrization-invariant, since it is defined in terms of T, which is itself parametrization-invariant. If the curve is regular and C^2, then, taking the arc length parametrization, we find, using $\kappa = \dot{\theta}$,

$$\theta(L) - \theta(0) = \int_0^L \kappa(s)ds$$

or

$$r_m = \frac{1}{2\pi}\int_0^L \kappa(s)ds.$$

The rotation index provides an algebraic count of the number of loops in the curve: a loop is counted positively if it is parametrized counter-clockwise (normal inward), and negatively otherwise. The figure "8", for example, has a rotation index equal to 0. This also provides an alternative definition of a positively oriented curve: *a simple closed curve is positively oriented if and only if its rotation index is* $+1$.

A similar notion is the winding number of a curve. It depends on a reference point $p_0 \in \mathbb{R}^2$, and is based on the angle between $p_0m(t)/|p_0m(t)|$ and the horizontal axis, which is again assumed to be continuous in t. Denoting this angle by $\alpha_{p_0}(t)$, the winding number of m around p_0 is

$$w_{p_0}(m) = (\alpha_{p_0}(b) - \alpha_{p_0}(a))/2\pi.$$

It provides the number of times the curve loops around p_0. Again, it depends on the curve orientation.

If a curve is *simple* (i.e., it has no self-intersection), then it is intuitively obvious that it can loop only once. This is the statement of the theorem of turning tangents, which says that *the rotation index of a simple closed curve is either 1 or* -1. However, proving this statement is not so easy (even in the differentiable case we consider) – the reader may refer to [86] for a proof.

1.12 More on Curvature

There is an important relationship between positive curvature (for positively oriented curves) and convexity. One says that a simple closed curve is convex if the bounded region it outlines is convex (it contains all line segments between any two of its points). Another characterization of convexity is that the curve lies on a single side of any of its tangent lines. The relation between convexity and curvature is stated in the next theorem.

Theorem 1.18 *A positively oriented C^2 curve is convex if and only if its curvature is everywhere nonnegative.*

We only provide a partial justification of the only if part. Assume that m is positively oriented and that its interior, Ω_m, is convex. For a fixed arc length, s and ε small enough, we have (since m is positively oriented): $m(s) + \varepsilon N(s) \in \Omega_m$ if $\varepsilon > 0$ and $\in \overline{\Omega}_m^c$ if $\varepsilon < 0$. Now, using a second-order expansion around s, we get

$$\frac{1}{2}(m(s+h) + m(s-h)) = m(s) + \frac{h^2}{s}\kappa(s)N(s) + o(h^2)$$

and this point cannot be in Ω_m if h is small and $\kappa(s) < 0$.

The local extrema of the curvature are also of interest. They are called the vertices of the curve. The four-vertex theorem, which we also state without proof, is another classical result for plane curves [63, 212, 228].

Theorem 1.19 *Every simple closed C^2 curve has at least four vertices.*

1.13 Discrete Curves and Curvature

1.13.1 Least-Squares Approximation

Because it involves a ratio of derivatives, the numerical computation of the curvature is unstable (very sensitive to noise). We give here a brief account of how one can deal with this issue.

Assume that the curve is discretized as a finite sequence of points, say $m(1), \dots, m(N)$. The usual finite-difference representation of derivatives are:

$$m'(k) = (m(k+1) - m(k-1))/2;$$
$$m''(k) = m(k+1) - 2m(k) + m(k-1).$$

The simplest formula for the approximate curvature is then

$$\kappa(k) = \frac{\det(m'(k), m''(k))}{|m'(k)|^3}.$$

This is however very sensitive to noise. A small variation in the position of $m(k)$ can have large consequences on the value of the estimated curvature. To be robust, curvature estimation has to include some kind of smoothing. As an example of such an approach, we describe a procedure in which one fits a curve of order 2 at each point.

Fix an approximation scale $\Delta \geq 1$, where Δ is an integer. For each k, compute three two-dimensional vectors $a(k)$, $b(k)$, $c(k)$ in order to have, for $-\Delta \leq l \leq \Delta$:

$$m(k+l) \simeq a(k)\frac{l^2}{2} + b(k)l + c(k).$$

Once this is done, $b(k)$ will be our approximation of the first derivative of m and $a(k)$ our approximation of the second derivative. The curvature will then be approximated by

$$\kappa(k) = \frac{\det[b(k), a(k)]}{|b(k)|^3}.$$

We will use least-squares estimation to compute a, b, c. First, build the matrix

$$A = \begin{pmatrix} \sum_{l=-\Delta}^{\Delta} \frac{l^4}{4} & 0 & \sum_{l=-\Delta}^{\Delta} \frac{l^2}{2} \\ 0 & \sum_{l=-\Delta}^{\Delta} l^2 & 0 \\ \sum_{l=-\Delta}^{\Delta} \frac{l^2}{2} & 0 & 2\Delta + 1 \end{pmatrix}$$

which is the matrix of second moments for the "variables" $l^2/2$, l and 1. They can be computed in closed form as a function of Δ, since

$$\sum_{l=-\Delta}^{\Delta} l^2 = \frac{\Delta}{3}(2\Delta^2 + 3\Delta + 1) \text{ and } \sum_{l=-\Delta}^{\Delta} l^4 = \frac{\Delta}{15}(6\Delta^4 + 15\Delta^3 + 10\Delta^2 - 1).$$

The second computation is, for all k:

$$z_0(k) = \sum_{l=-\Delta}^{\Delta} m(k+l),$$

$$z_1(k) = \sum_{l=-\Delta}^{\Delta} l m(k+l),$$

$$z_2(k) = \sum_{l=-\Delta}^{\Delta} \frac{l^2}{2} m(k+l).$$

Given this, the vectors $a(k)$, $b(k)$, $c(k)$ are provided by the row vectors of the matrix

$$A^{-1} \begin{pmatrix} z_2(k) \\ z_1(k) \\ z_0(k) \end{pmatrix}$$

where z_0, z_1, z_2 are also written as row vectors. As shown in Fig. 1.1, this method gives reasonable results for smooth curves. However, if the curve has sharp angles, the method will oversmooth and underestimate the curvature.

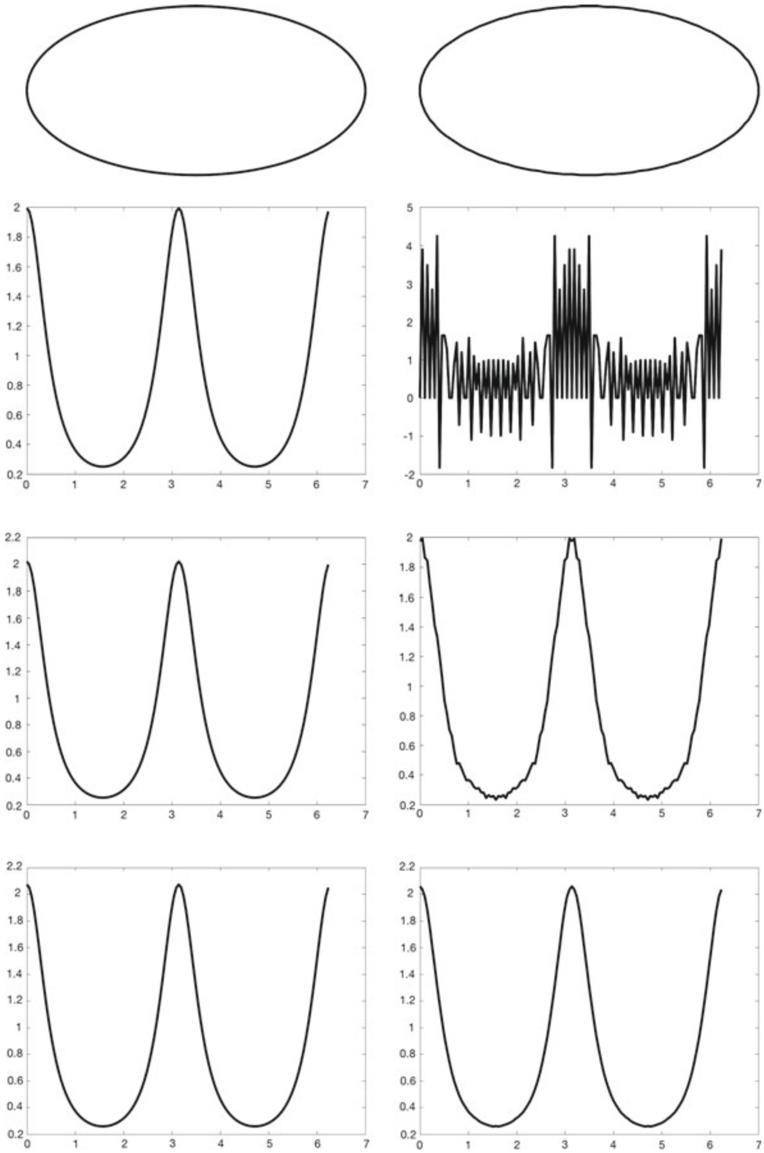

Fig. 1.1 Noise and curvature. The first curve on the left is an ellipse discretized over 125 points. The second on the right is the same ellipse, with coordinates rounded to two decimal points. The difference is almost imperceptible. However, the second row shows the result of estimating the curvature without smoothing, on the first and the second ellipse, with a very strong noise effect. The third (resp. fourth) row shows the result of the second-order approximation with $\Delta = 5$ (resp. $\Delta = 10$). The computed curvature for the truncated curve progressively improves while that of the original curve is minimally affected

1.13.2 Curvature and Distance Maps

If $A \subset \mathbb{R}^2$, one defines the distance map to A as

$$d_A(p) = \mathrm{dist}(p, \mathcal{A}) = \inf\{|p-q|, q \in \mathcal{A}\}.$$

If A is a closed set (which we will assume in the following), then for any $p \in \mathbb{R}^2$ there exists a $q \in A$ such that $d_A(p) = |p-q|$ (i.e., the infimum is a minimum). This is because any minimizing sequence q_n such that $|p-q_n| \to d_A(p)$ is necessarily bounded, and therefore has, according to the Heine–Borel theorem, a convergent subsequence, with limit $q \in A$ (because A is closed) and such that $|p-q| = d_A(p)$.

The optimal q is not always unique. For example, all points in a circle are closest to its center. The set of points $p \in \mathbb{R}^2$ for which there exists a unique $q \in A$ such that $|p-q| = d_A(p)$ will be denoted by $\mathcal{U}_A$, and we let $\pi_A : \mathcal{U}_A \to A$ be the projection, uniquely defined by $|p - \pi_A(p)| = d_A(p)$.

For $p \in \mathbb{R}^2$, we let $B(p,r) = \left\{q \in \mathbb{R}^2 : |p-q| < r\right\}$ denote the (open) disc with center p and radius r. For $q \in A$, define

$$r(A,q) = \sup\{r : B(q,r) \subset \mathcal{U}_A\}$$

and $r(A) = \inf\{r_A(q) : q \in A\}$, which is called the *reach* of A, and also has the following alternative definition.

Proposition 1.20

$$r(A) = \sup\{r : d_A(p) < r \Rightarrow p \in \mathcal{U}_A\}. \tag{1.20}$$

Proof Denote temporarily by $r'(A)$ the right-hand side of (1.20). Assume that $r \le r'(A)$. If $q \in A$ and $p \in B(q,r)$, then $d_A(p) \le |p-q| < r$ so that $p \in \mathcal{U}_A$ by definition of r'_A. Therefore, $B(q,r) \subset \mathcal{U}_A$ and $r \le r(A,q)$ for all $q \in A$, which implies that $r \le r(A)$. Taking the maximum in r, we get $r'(A) \le r(A)$.

Assume now that $r \le r(A)$. If $d_A(p) < r$, then $p \in B(\pi_A(p), r)$, and since $r(A) \le r(A, \pi_A(p))$, we have $p \in \mathcal{U}_A$. This proves that $r \le r'(A)$, and taking the maximum in r, we get $r(A) \le r'(A)$, which concludes the proof. □

We have the following proposition.

Proposition 1.21 *The distance map is 1-Lipschitz, i.e., for all $p, p' \in \mathbb{R}^2$, one has*

$$|d_A(p) - d_A(p')| \le |p - p'| \tag{1.21}$$

and the projection π_A is continuous on its domain.

Proof One has, for all $p, p' \in \mathbb{R}^2$ and $q \in A$,

$$d_A(p) \le |p-q| \le |p'-q| + |p-p'|.$$

Taking the inf of the right-hand side, we get $d_A(p) \le d_A(p') + |p - p'|$. By symmetry, we also have $d_A(p') \le d_A(p) + |p - p'|$ and (1.21) holds.

Now, take $p \in \mathcal{U}_A$ and a sequence $p_n \in \mathcal{U}_A$ such that $|p_n - p| \to 0$. Let $q_n = \pi_A(p_n)$, $q = \pi_A(p)$ and assume that there exists a subsequence of q_n (that we will still denote by q_n) and $\varepsilon > 0$ such that $|q_n - q| \ge \varepsilon$. Because $|p - q_n| \le |p - p_n| + d_A(p_n) \le 2|p - p_n| + d_A(p)$, which is bounded, q_n has a convergent subsequence (still called q_n), with limit $q' \in A$. But $|p - q'| = \lim_n |p_n - q_n| = \lim_n d_A(p_n) = d_A(p)$. Since $p \in \mathcal{U}_A$, this implies $q = q'$, a contradiction to the fact that $|q_n - q| \ge \varepsilon$ for all n. The latter condition being impossible implies that π_A is continuous. □

Proposition 1.22 *Assume that d_A is differentiable at $p \in \mathring{\mathcal{U}}_A$ (the interior of $\mathcal{U}_A$). Then, if $p \notin A$,*

$$\nabla d_A(p) = \frac{p - \pi_A(p)}{|p - \pi_A(p)|}. \tag{1.22}$$

Proof To see this, first note that, letting $q = \pi_A(p)$, one has $p_t := q + t(p - q) \in \mathcal{U}_A$ for all $t \in [0, 1]$, with $\pi_A(p_t) = q$. Indeed, if $q' \in A$, $q' \ne q$, one has $|p - q| < |p - q'| \le |p - p_t| + |p_t - q'| = (1 - t)|p - q| + |p_t - q'|$. This yields

$$|p_t - q| = t|p - q| < |p_t - q'|$$

so that $p_t \in \mathcal{U}_A$ with $q = \pi_A(p_t)$. This also implies that $d_A(p_t) = t|p - q|$ and taking the derivative with respect to t at $t = 1$, we get

$$\nabla d_A(p)^T (p - q) = |p - q|.$$

However, (1.21) implies that $|\nabla d_A(p)| \le 1$. This is only possible for $\nabla d_A(p)$ given by (1.22).

One can use the fact that the gradient of d_A is prescribed in $\mathring{\mathcal{U}}_A \setminus A$ whenever d_A is differentiable, in combination with Rademacher's theorem [107], which states that Lipschitz functions are differentiable *almost everywhere*, to prove that d_A is actually differentiable on the whole set $\mathring{\mathcal{U}}_A \setminus A$. Similarly, d_A^2 is differentiable on $\mathring{\mathcal{U}}_A$, with $\nabla(d_A^2)(p) = 2(p - \pi_A(p))$. This general fact is proved below in the special case $A = \mathcal{R}_m$, where m is a C^2, closed, regular curve with no self-intersection. Note that our definitions, so far, and Propositions 1.20–1.22 are valid for arbitrary closed sets, and in any dimension (and so is the differentiability of d_A on $\mathring{\mathcal{U}}_A \setminus A$).

We now specialize to the case $A = \mathcal{R}_m$, and we will write $d_m = d_{\mathcal{R}_m}$, $\mathcal{U}_m = \mathcal{U}_{\mathcal{R}_m}$, etc.

Proposition 1.23 *Let m be a simple closed C^2 regular curve. Then, we have the following statements.*

(i) If $|p - m(s)| = d_m(p)$, then $p = m(s) + tN_m(s)$ with $|t| = d_m(p)$ and $t\kappa_m(s) \le 1$.
(ii) Let

$$\rho_m = \max\left\{\frac{2\,|(m(\tilde{s}) - m(s))^T N_m(s)|}{|m(\tilde{s}) - m(s)|^2} : s, \tilde{s} \in [0, L]_*, s \neq \tilde{s}\right\}. \tag{1.23}$$

Then $\rho_m < \infty$ *and* $r(\mathcal{R}_m) \geq 1/\rho_m > 0$*. In particular,* $\mathring{\mathcal{U}}_m$ *is not empty.*

(iii) The distance map is differentiable on $\mathring{\mathcal{U}}_m$*.*

Proof Assume that m is parametrized by arc length over the wrapped interval $[0, L]_*$. The function $f : u \mapsto |p - m(s +_* u)|^2$ has by assumption a global minimum at $u = 0$. We therefore have $\dot{f}(0) = 0$ and $\ddot{f}(0) \geq 0$. Since $\dot{f}(0) = -2(p - m(s))^T T_m(s)$, we get the fact that $p - m(s)$ is normal to m, so that $p = m(s) + tN_m(s)$ with $|t| = d_m(p)$. We also have $\ddot{f}(0) = 2 - 2(p - m(s))^T N_m(s)\kappa_m(s) = 2(1 - t\kappa_m(s))$ yielding $t\kappa_m(s) \leq 1$. This proves (i).

We now prove that ρ_m is finite. If $m(s_n) \neq m(\tilde{s}_n)$ are such that

$$c_n := \frac{2\,|(m(\tilde{s}_n) - m(s_n))^T N_m(s_n)|}{|m(\tilde{s}_n) - m(s_n)|^2}$$

tends to infinity, then, necessarily, $m(\tilde{s}_n) - m(s_n) \to 0$. We can assume (taking subsequences if needed) that both s_n and $\tilde{s}_n$ converge, necessarily to the same limit (say s) because m is non-intersecting. Assume that $s \neq 0$ so that $s_n \in (0, L)$ for large enough n (otherwise, just reparametrize m with another starting point). We have, making a Taylor expansion,

$$m(\tilde{s}_n) = m(s_n) + (\tilde{s}_n - s_n)T_m(s_n) + \kappa(s_n)\frac{(\tilde{s}_n - s_n)^2}{2}N_m(s_n) + o((\tilde{s}_n - s_n)^2),$$

$$\left|(m(\tilde{s}_n) - m(s_n))^T N_m(s_n)\right| = |\kappa(s_n)|\frac{(\tilde{s}_n - s_n)^2}{2} + o((\tilde{s}_n - s_n)^2)$$

and

$$|m(\tilde{s}_n) - m(s_n)|^2 = (\tilde{s}_n - s_n)^2 + o((\tilde{s}_n - s_n)^2).$$

Thus $c_n \to |\kappa_m(s)|$, which is a contradiction, proving that ρ_m is finite. Note that the same limit argument also proves that $\rho_m \geq \|\kappa_m\|_\infty := \max_s |\kappa_m(s)|$.

Now, take $q \in \mathbb{R}^2$ with $d_m(q) = t < 1/\rho_m$ and assume that it has two closest points, so that there exists $s_0 \neq s_1$ such that $t = |q - m(s_0)| = |q - m(s_1)|$. Then $q = m(s_0) + t_0 N_m(s_0) = m(s_1) + t_1 N_m(s_1)$ with $|t_0| = |t_1| = t$. Moreover,

$$\begin{aligned}|m(s_1) - m(s_0)|^2 = |t_0 N_m(s_0) - t_1 N_m(s_1)|^2 &= 2t_0^2 - 2t_1 t_0 N_m(s_0)^T N_m(s_1)\\ &= 2t\,|t_0 - t_1 N_m(s_0)^T N_m(s_1)|\end{aligned}$$

and

$$|(m(s_1) - m(s_0))^T N_m(s_0)| = |t_0 - t_1 N_m(s_0)^T N_m(s_1)|.$$

We therefore get

$$\frac{2|(m(s_1) - m(s_0))^T N_m(s_0)|}{|m(s_1) - m(s_0)|^2} = \frac{1}{t}.$$

By definition, the right-hand side is less than or equal to ρ_m, which contradicts our assumption that $t < 1/\rho_m$. Therefore, $q \in \mathcal{U}_m$. This proves that $r(\mathcal{R}_m) \geq 1/\rho_m$.

Conversely, take $t < r(\mathcal{R}_m)$. By definition, we have $B(m(s), t + \varepsilon) \subset \mathcal{U}_A$ for all $s \in [0, L]_*$ and some ε such that $t + \varepsilon < r(\mathcal{R}_m)$. Therefore, letting $q_+ = m(s) + tN_m(s)$ and $q_- = m(s) - tN_m(s)$, we have $|m(\tilde{s}) - q_+| > t$ and $|m(\tilde{s}) - q_-| > t$ for all $\tilde{s} \neq s$. Developing the expression $|m(\tilde{s}) - m(s) \mp tN_m(s)| - t^2$ yields

$$|m(\tilde{s}) - m(s)|^2 \geq \pm 2t(m(\tilde{s}) - m(s))^T N_m(s)$$

so that

$$\frac{1}{t} \geq \frac{2\left|(m(\tilde{s}) - m(s))^T N_m(s)\right|}{|m(\tilde{s}) - m(s)|^2}$$

for all $s \neq \tilde{s}$, i.e., $t \leq 1\rho_m$. Taking the maximum in t implies $r(\mathcal{R}_m) \leq 1/\rho_m$.

We now prove (iii). Take $q \in \mathring{\mathcal{U}}_m$ and $m(s) = \pi_m(q)$. Write $q = m(s) + rN_m(s)$ with $|r| = d_m(q)$. Since $B(q, \varepsilon) \subset \mathring{\mathcal{U}}_m$ for ε small enough, we have $m(s) + tN(s) \in \mathring{\mathcal{U}}_m$ for $t \in (r - \varepsilon, r + \varepsilon)$, and $\pi_m(m(s) + tN(s)) = m(s)$. From (i), we get $\kappa_m(s)t \leq 1$ for $t \in (r - \varepsilon, r + \varepsilon)$, which implies $\kappa_m(s)r < 1$.

Take $\delta > 0$ such that $\kappa_m(u)t < 1$ if $|s - u| < \delta$ and $t \in (r - \varepsilon/2, r + \varepsilon/2)$. Consider the mapping $\varphi : (s - \delta, s + \delta) \times (r - \varepsilon/2, r + \varepsilon/2) \to \mathbb{R}^2$ defined by $\varphi(u, t) = m(u) + tN(u)$. Then $\partial_1\varphi(u, t) = (1 - t\kappa_m(u))T_m(u)$ and $\partial_2\varphi(u, t) = N_m(u)$ so that $\det(d\varphi) = 1 - t\kappa_m(u) \neq 0$. The inverse function theorem implies that φ (possibly restricted to a smaller open neighborhood of (s, r)) is invertible with a differentiable inverse. So, there exists a neighborhood of q in $\mathring{\mathcal{U}}_m$ such that $\varphi^{-1}(p) = (\pi_m(p), t(p))$ is differentiable with $t(p) = \pm d_m(p)$. Making sure that this neighborhood does not intersect m, we can ensure that the sign of $t(p)$ is constant so that d_m is differentiable in this neighborhood and, in particular, at q. □

Consider the mapping (a local version of which was introduced in the previous proof)

$$\begin{aligned} \varphi_m : \quad [0, L]_* \times (-r, r) &\to \quad \mathbb{R}^2 \\ (s, t) &\mapsto \quad m(s) + tN_m(s) \end{aligned}$$

for some $r < r(\mathcal{R}_m)$. As shown in the proof of Proposition 1.23, φ_m is locally invertible, but because it is also one-to-one, it provides a diffeomorphism from $[0, L]_* \times (-r, r)$ to the set

$$V_m(r) = \{q : d_m(q) < r\}.$$

Consider now the set $V_m^+(r) = \varphi_m([0, L]_* \times (0, r))$. We can write

$$\begin{aligned}\mathrm{Area}(V_m^+(r)) &= \int_0^L \int_0^r \det d\varphi_m(s,t)\, ds\, dt \\ &= \int_0^L \int_0^r (1 - t\kappa_m)\, ds\, dt \\ &= Lr - \frac{r^2}{2}\int_0^L \kappa_m\, ds.\end{aligned}$$

The interesting conclusion is that the area is a second-degree polynomial in r. The first-degree coefficient is the curve's length and the second-degree coefficient is the integral of the curvature, i.e., the rotation index of the curve.

The formula can be localized without difficulty by restricting $V^+(m)$ to points s, t such that $s_0 < s < s_1$, the result being obviously

$$(s_1 - s_0)r - \frac{r^2}{2}\int_{s_0}^{s_1} \kappa_m\, ds = r\int_{s_0}^{s_1} (1 - \kappa_m r/2)\, ds.$$

The "infinitesimal limit" $r(1 - \kappa_m(s)r/2)ds$ provides the infinitesimal area of the set of points that are within distance r to the curve and project on $m(s)$ for some $s \in (s_0, s_1)$. This area is at first order given by the arc length times r, with a corrective term involving the curvature.

This computation is a special case of a very general construction of what are called curvature measures [106]. They can be defined for a large variety of sets, in any dimension. We will see a two-dimensional description of them when discussing surfaces.

Proposition 1.23 needs to be modified to apply to open curves. Consider such a curve, $m : [0, L] \to \mathbb{R}^2$. Then point (i) in the proposition remains true with a proper definition of a normal vector to $\mathcal{R}_m$: one says that N is a unit normal to $\mathcal{R}_m$ (or simply to m) at $m(s)$ if

$$\begin{cases} N = \pm N_m(s) & \text{if } s \in (0, L) \\ N = t_1 N_m(0) + t_2 T_m(0) & \text{if } s = 0 \\ N = t_1 N_m(L) - t_2 T_m(L) & \text{if } s = L \end{cases}$$

with $t_1^2 + t_2^2 = 1$, $t_2 > 0$. Denoting by $\mathcal{N}_m(s)$ the set of unit normals to m at $m(s)$, the first statement in (i) can be replaced by: $p = m(s) + d_m(p)N$ where $N \in \mathcal{N}(s)$. The fact that $\kappa_m(s)d_m(p) \leq 1$ holds for $s \in (0, L)$.

If one replaces the definition of ρ_m by

$$\rho_m = \max\left\{\frac{2\,(m(\tilde{s}) - m(s))^T N}{|m(\tilde{s}) - m(s)|^2} : s, \tilde{s} \in [0, L]_*, s \neq \tilde{s}, N \in \mathcal{N}_m(s)\right\}, \quad (1.24)$$

then (ii) remains true. Note that (1.24) boils down to (1.23) for closed curves, where $\mathcal{N}_m(s) = \{\pm N_m(s)\}$. Finally, (iii) is true.

The reader can try to prove these statements directly, or refer to [106], where these statements are proved for arbitrary closed sets, with a proper definition of the set of unit normal vectors, and without the finiteness of ρ_m, which does not hold in general. (It does not hold, for example, for polygonal curves.)

1.14 Implicit Representation

1.14.1 Introduction

Implicit representations can provide simple descriptions of relatively complex shapes and can in many cases be a good choice when designing stable shape processing algorithms. The zero level set of a function $f : \mathbb{R}^2 \to \mathbb{R}$ is the set C_f of all $p \in \mathbb{R}^2$ such that $f(p) = 0$ (cf. Fig. 1.2). One says that f is regular if its derivative never vanishes on C_f, that is,

$$f(p) = 0 \Rightarrow \nabla f(p) \neq 0. \tag{1.25}$$

The set C_f can have several connected components, each of them being the image of a curve (level sets can therefore be used to represent multiple curves). Our first goal is to show how local properties of curves can be computed directly from the function f. We will always assume, in this chapter, that the function f tends to infinity as p tends to infinity. This implies that the zero level sets are bounded.

The implicit function theorem implies that, in a neighborhood of any regular point of f (such that $\nabla f(m) \neq 0$), the set C_f can be locally parametrized as a regular curve, for example by expressing one of the coordinates (x, y) as a function of the other. This fact and Proposition 1.11 implies that, if f is regular, each connected component of C_f can be parametrized as a regular curve. The existence of higher derivatives in f implies the same regularity for the parametrization.

Fix a connected component and assume that such a parametrization has been chosen. This results in a curve $m : I \to \mathbb{R}^2$ such that $m(0) = m_0$ and $f(m(u)) = 0$ for $u \in I$ ($\mathcal{R}_m$ coincides with the chosen connected component). From the chain rule, we have:

$$\nabla f(m)^T \partial_u m = 0.$$

This implies that $\nabla f(m)$ is normal to m.

Orientation. *We will say that f is positively oriented if $f < 0$ in the bounded connected components of $\mathbb{R}^2 \setminus C_f$ and $f > 0$ otherwise. If m is also positively oriented, then $\nabla f(m)$ points outward while the normal N to m points inward, so that $\nabla f(m) = -|\nabla f(m)|N$ (recall that (T, N) must have determinant 1, with $T = \dot{m}/|\dot{m}|$).*

Assuming positive orientation, we obtain

$$T = \frac{1}{|\nabla f|}\big(-\partial_2 f, \partial_1 f\big)$$

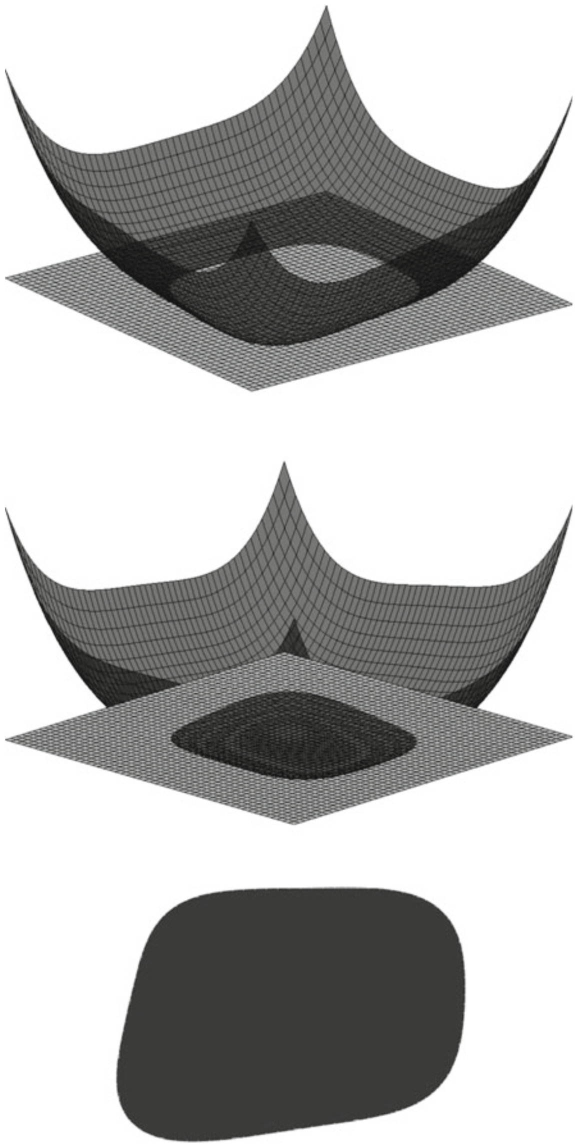

Fig. 1.2 Implicit representation: The first two panels provide two views of the graph of a function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ intersecting the plane $z = 0$. The third panel is the corresponding level set

Assume that f is twice differentiable. From the second derivative of the equation $f(m(u)) = 0$, we have

$$\dot{m}^T d^2 f(m)\dot{m} + \nabla f(m)^T \ddot{m} = 0\,.$$

(recall that the second derivative of f is a 2 by 2 matrix).

Since $\nabla f(m) = -|\nabla f(m)|N$ and $\ddot{m}^T N = \kappa|\dot{m}|^2$, the previous equation yields (after division by $|\dot{m}|^2$),

$$T^T d^2 f(m)\, T - \kappa|\nabla f(m)| = 0\,.$$

so that

$$\kappa = \frac{T^T d^2 f\, T}{|\nabla f|} = \frac{\partial_1^2 f\, \partial_2 f^2 - 2\partial_1\partial_2 f\, \partial_1 f\, \partial_2 f + \partial_2^2 f\, \partial_1 f^2}{(\partial_1 f^2 + \partial_2 f^2)^{3/2}}.$$

This can also be written as (the computation being left to the reader)

$$\kappa = \operatorname{div}\frac{\nabla f}{|\nabla f|}. \tag{1.26}$$

1.14.2 Example: Implicit Polynomials

A large variety of shapes can be obtained by restricting the function f to be a polynomial of small degree [169], therefore involving a dependency on a small number of parameters. A polynomial in two variables and total degree less than n is given by the general formula

$$f(x, y) = \sum_{p+q\leq n} a_{pq} x^p y^q\,.$$

The zero level set of f, $C_f = \{z = (x, y),\ f(x, y) = 0\}$, is called an algebraic curve. It can be a complicated object, with branches at infinity, self-intersections, or multiple loops.

The principal part of f is the homogeneous polynomial

$$g(x, y) = \sum_{k=0}^{n} a_{k,n-k} x^k y^{n-k}.$$

A sufficient condition for the compactness of C_f is that g has no non-trivial zeros, i.e., $g(x, y) = 0 \Rightarrow x = y = 0$. Adding our usual regularity condition, $f = 0 \Rightarrow \nabla f \neq 0$, ensures that C_f is a union of Jordan curves.

Figure 1.3 provides a few examples of zero level sets of implicit polynomials.

Fig. 1.3 Shapes generated by implicit polynomials of degree 4. The first curve is the level set of the polynomial $f(x, y) = x^4 + y^4 - xy - 0.1$. The other three are generated by adding a small noise to the coefficients (including zeros) of f

1.15 Invariance for Affine and Projective Transformations

Invariance, which searches for quantities that remain unchanged under certain classes of transformations, is a fundamental concept when dealing with shapes. So far, we have discussed two classes of transformations: parameter change and Euclidean motion (rotations, translations). We found in particular that Euclidean curvature was an invariant for these two classes together. We now consider additional invariants to complement these two.

We will start with transformation by scaling. This corresponds to replacing the curve m by $\tilde{m} = \lambda m$ where λ is a positive number. Visually, this corresponds to viewing the shape from a location that is closer or further away. Because of the renormalization, the unit tangent, normal and the angles θ_m are invariant. However, the length and arc length are multiplied by the constant factor λ. Finally, since the curvature is the rate of change of the angle as a function of arc length, it is divided by the same constant, $\kappa_{\tilde{m}} = \kappa_m/\lambda$.

It will also be interesting to consider invariants of affine transformations $m \mapsto Am + b$ where A is a 2 by 2 invertible matrix (a general affine transformation). Arc length and curvature are not conserved by such transformations, and there is no simple formula to compute their new value. This section describes how new quantities, which will be called affine arc length and affine curvature, can be introduced to obtain the same type of invariance.

However, a comprehensive study of the theory of differential invariants of curves [224] lies beyond the scope of this book. Here, we content ourselves with the computation in some particular cases. Although this repeats what we have already done with arc length and curvature, it will be easier to start with the simple case of rotation invariance. We know that s_m and κ_m are invariant under translation and rotation, and we now show how this can be obtained with a systematic approach that will in turn be applied to more general cases.

1.15.1 Euclidean Invariance

The generic approach to defining generalized notions of length and arc length is to look for a function Q which depends only on the derivatives of a curve at a given point, such that $Q(\dot{m}, \ddot{m}, \ldots)du$ provides the length of an element of the curve between u and $u + du$.

An arc length is then defined by

$$\sigma_m(u) = \int_0^u Q(\dot{m}, \ddot{m}, \ldots)dv.$$

The function Q will be designed to meet invariance properties. We will always require parametrization invariance, ensuring that $m = \tilde{m} \circ \varphi$ implies $\sigma_m = \sigma_{\tilde{m}} \circ \varphi$. Computing the derivative of this identity yields, in terms of Q:

$$Q(\dot{m}, \ddot{m}, \ldots) = \dot{\varphi} Q(\dot{\tilde{m}} \circ \varphi, \ddot{\tilde{m}} \circ \varphi, \ldots). \tag{1.27}$$

Now, for $m = \tilde{m} \circ \varphi$, we have

$$\begin{aligned} \dot{m} &= \dot{\varphi}\, \dot{\tilde{m}} \circ \varphi, \\ \ddot{m} &= \ddot{\varphi}\, \dot{\tilde{m}} \circ \varphi + \dot{\varphi}^2 \ddot{\tilde{m}} \circ \varphi, \end{aligned}$$

and so on for higher derivatives.

As a consequence, if Q only depends on the first derivative, we must have

$$Q(\dot{\varphi}\, \dot{\tilde{m}} \circ \varphi) = \dot{\varphi}\, Q(\dot{\tilde{m}} \circ \varphi).$$

This is true in particular when, for all $z_1 \in \mathbb{R}^2$, $\lambda_1 > 0$:

$$Q(\lambda_1 z_1) = \lambda_1 Q(z_1).$$

This is the order 1 condition for Q. It is sufficient by the discussion above, but one can show that it is also necessary. Similarly, the order 2 condition is that, for all $z_1, z_2 \in \mathbb{R}^2$, for all $\lambda_1 > 0$, $\lambda_2 \in \mathbb{R}$:

$$Q(\lambda_1 z_1, \lambda_2 z_1 + \lambda_1^2 z_2) = \lambda_1 Q(z_1, z_2).$$

This argument can be applied to any number of derivatives. The general expression (based on the *Faà di Bruno formula*) is quite heavy, and we will not need it for this discussion, but the trick for deriving new terms is quite simple. Think in terms of derivatives: the derivative of λ_k is λ_{k+1} and the derivative of z_k is $\lambda_1 z_{k+1}$; then apply the product rule. For example, the second term is the derivative of the first term, $\lambda_1 z_1$, and therefore:

$$(\lambda_1 z_1)' = (\lambda_1)' z_1 + \lambda_1 (z_1)'$$
$$= \lambda_2 z_1 + \lambda_1^2 z_2,$$

which is what we found by direct computation. The constraint with three derivatives would be

$$Q(\lambda_1 z_1, \lambda_2 z_1 + \lambda_1^2 z_2, \lambda_3 z_1 + 3\lambda_2\lambda_1 z_2 + \lambda_1^3 z_3) = \lambda_1 Q(z_1, z_2, z_3).$$

The second type of constraint which is required for Q is invariance under some class of transformations of the plane. If A is such a transformation, and $\tilde{m} = Am$, the requirement is $\sigma_{\tilde{m}} = \sigma_m$, or

$$Q(\dot{m}, \ddot{m}, \ldots) = Q(\partial(Am), \partial^2(Am), \ldots). \tag{1.28}$$

We consider affine transformations (the results will be extended to projective transformations at the end of this section). The equality is always true for translations $Am = m + b$, since Q only depends on the derivatives of m, and therefore we can assume that A is purely linear. Equality (1.28) therefore becomes: for all $z_1, z_2, \ldots \in \mathbb{R}^2$,

$$Q(z_1, z_2, \ldots) = Q(Az_1, Az_2, \ldots).$$

We now specialize to rotations. We will favor the lowest complexity for Q, and therefore first study whether a solution involving only one derivative exists. In this case, Q must satisfy: for all $\lambda_1 > 0$, for all $z_1 \in \mathbb{R}^2$ and for any rotation A,

$$Q(Az_1) = Q(z_1) \text{ and } Q(\lambda_1 z_1) = \lambda_1 Q(z_1)\,.$$

Let $e_1 = (1, 0)$ be the unit vector in the x-axis. Since one can always use a rotation to transform any vector z_1 into $|z_1|e_1$, the first condition implies that $Q(z_1) = Q(|z_1|e_1)$, which is equal to $|z_1|Q(e_1)$ from the second condition. We therefore find that $Q(z_1) = c|z_1|$ for some constant c, yielding $Q(\dot{m}) = c|\dot{m}| = c\sqrt{\dot{x}^2 + \dot{y}^2}$. We therefore retrieve the previously defined arc length up to a multiplicative constant c. The choice $c = 1$ is quite arbitrary, and corresponds to the condition that e_1 provides a unit speed: $Q(e_1) = 1$. We will refer to this σ_m as the Euclidean arc length, since we now consider other choices to obtain more invariants.

1.15.2 Scale Invariance

Let us now add scale to translation and rotation. Since it is always possible to transform any vector z_1 into e_1 with a rotation and scaling, considering only one derivative is not enough anymore.[2] We need at least two derivatives and therefore consider z_1 and z_2 with the constraints

$$Q(Az_1, Az_2) = Q(z_1, z_2) \text{ and } Q(\lambda_1 z_1, \lambda_2 z_1 + \lambda_1^2 z_2) = \lambda_1 Q(z_1, z_2)\,.$$

Similar to rotations, the first step is to use the first condition to place z_1 and z_2 into a canonical position. Consider the combination of rotation and scaling which maps e_1 to z_1. The first column of its matrix must therefore be z_1, but, because combinations of rotation and scaling have matrices of the form $S = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, we see that, letting $z_1 = (x_1, y_1)$, the obtained matrix is

$$S_{z_1} = \begin{pmatrix} x_1 & -y_1 \\ y_1 & x_1 \end{pmatrix}.$$

Now take $A = S_{z_1}^{-1}$ to obtain, from the first condition:

$$Q(z_1, z_2) = Q(e_1, S_{z_1}^{-1} z_2)\,.$$

A direct computation yields

$$S_{z_1}^{-1} z_2 = \frac{1}{x_1^2 + y_1^2} \begin{pmatrix} x_1 x_2 + y_1 y_2 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}.$$

So far, we have obtained the fact that Q must be a function F of the quantities $a = (z_1^T z_2)/|z_1|^2$ and $b = \det(z_1, z_2)/|z_1|^2$.

Now consider the second condition. The transformation $z_1 \to \lambda_1 z_1$ and $z_2 \to \lambda_1^2 z_2 + \lambda_2 z_1$ takes a to $\lambda_1 a + \lambda_2/\lambda_1$ and b to $\lambda_1 b$. Thus, if $Q(z_1, z_2) = F(a, b)$, we must have

$$F(\lambda_1 a + \lambda_2/\lambda_1, \lambda_1 b) = \lambda_1 F(a, b)$$

for all real numbers a, b, λ_2 and $\lambda_1 > 0$. Given a, b we can take $\lambda_2 = -\lambda_1^2 a$ and $\lambda_1 = 1/|b|$, at least when $b \neq 0$. This yields, for $b \neq 0$:

$$F(a, b) = |b| F(0, \mathrm{sign}(b)).$$

For $b = 0$, we can take the same value for λ_2 to obtain $F(0, 0) = \lambda_1 F(a, 0)$ for every λ_1 and a, which is only possible if $F(a, 0) = 0$ for all a. Thus, in full generality, the function Q must take the form

[2]This would give $Q(z_1) = Q(e_1) = \text{const}$ and $Q(\lambda_1 z_1) = \lambda_1 Q(z_1) = Q(z_1)$ for all $\lambda_1 > 0$, yielding $Q = 0$.

$$Q(z_1, z_2) = \begin{cases} c_+ |\det(z_1, z_2)|/|z_1|^2 \text{ if } \det(z_1, z_2) > 0, \\ 0 \text{ if } \det(z_1, z_2) = 0, \\ c_- |\det(z_1, z_2)|/|z_1|^2 \text{ if } \det(z_1, z_2) < 0, \end{cases}$$

where c_0, c_+, c_- are positive constants. To ensure invariance by a change of orientation, however, it is natural to choose $c_+ = c_-$. Taking this value equal to 1 yields

$$Q(z_1, z_2) = |\det(z_1, z_2)|/|z_1|^2.$$

We obtain the definition of the arc length for similitudes[3]:

$$d\sigma^{sim} = \frac{|\dot{x}\ddot{y} - \ddot{x}\dot{y}|}{\dot{x}^2 + \dot{y}^2} du\,. \tag{1.29}$$

1.15.3 Special Affine Transformations

We now consider the case of area-preserving, or special affine transformations. These are affine transformations A such that $\det(A) = 1$. As before, we need two derivatives, and the first step is again to normalize $[z_1, z_2]$ using a suitably chosen matrix A. Here, the choice is natural and simple, at least when z_1 and z_2 are independent: take A to be the inverse of $[z_1, z_2]$, normalized to have determinant 1, namely

$$A = \begin{cases} \sqrt{\det(z_1, z_2)}[z_1, z_2]^{-1} \text{ if } \det(z_1, z_2) > 0, \\ \sqrt{\det(z_2, z_1)}[z_2, z_1]^{-1} \text{ if } \det(z_1, z_2) < 0. \end{cases}$$

When $\det(z_1, z_2) > 0$, this yields

$$Q(z_1, z_2) = Q(\sqrt{\det(z_1, z_2)}e_1, \sqrt{\det(z_1, z_2)}e_2)$$

so that Q must be a function F of $\sqrt{\det(z_1, z_2)}$. Applying the parametrization invariance condition, we find

$$F(\lambda_1^{3/2}\sqrt{\det(z_1, z_2)}) = \lambda_1 F(\sqrt{\det(z_1, z_2)}),$$

which implies, taking $\lambda_1 = (\det(z_1, z_2))^{-1/3}$, that

$$Q(z_1, z_2) = F(1)(\det(z_1, z_2))^{1/3}.$$

The same result is true for $\det(z_1, z_2) < 0$, yielding

[3] To complete the argument, one needs to check that the required conditions are satisfied for the obtained Q; this is indeed the case, although we skip the computation.

$$Q(z_1, z_2) = \tilde{F}(1)(\det(z_2, z_1))^{1/3}.$$

with a possibly different constant $\tilde{F}(1)$. Again, for orientation invariance, it is natural to define the area-preserving arc length by

$$d\sigma^{s.aff} = |\ddot{x}\dot{y} - \ddot{y}\dot{x}|^{1/3} du.$$

We have left aside the case $\det(z_1, z_2) = 0$. In this case, assume that $z_2 = \alpha z_1$. The second condition implies, taking $\lambda_2 = -\lambda_1^2\alpha$:

$$\lambda_1 Q(z_1, \alpha z_1) = Q(\lambda_1 z_1, \lambda_1^2 \alpha z_1 + \lambda_2 z_1) = Q(\lambda_1 z_1, 0),$$

but we can always find an area-preserving transformation which maps $\lambda_1 z_1$ to e_1 so that $\lambda_1 Q(z_1, \alpha z_1) = Q(e_1, 0)$ is true for every $\lambda_1 > 0$ only if $Q(z_1, \alpha z_1) = 0$. This is consistent with the formula obtained for $\det(z_1, z_2) \neq 0$.

Computations are also possible for the full affine group and also for the projective group, but they require us to deal with four and more derivatives and are quite lengthy. They will be provided at the end of this section. The reader may refer to Sect. B.4 for a quick introduction to groups of linear transformations and their actions.

1.15.4 Generalized Curvature

In addition to arc length, new definitions of curvature can be adapted to more invariance constraints. One way to understand the definition is to return to the rotation case, and our original definition of curvature.

We have interpreted the curvature as the speed of rotation of the tangent with respect to arc length. Consider the matrix $P_m = [T_m, N_m]$ associated to the tangent and normal to m. Because (T_m, N_m) is an orthonormal system, this matrix is a rotation, called a moving frame [55, 104, 108, 109], along the curve. The rate of variation of this matrix is defined by

$$W_m = P_m^{-1} \partial_s P_m.$$

In the Euclidean case, it is

$$W_m = \partial_s \theta_m \begin{pmatrix} \cos\theta_m & \sin\theta_m \\ -\sin\theta_m & \cos\theta_m \end{pmatrix} \begin{pmatrix} -\sin\theta_m & -\cos\theta_m \\ \cos\theta_m & -\sin\theta_m \end{pmatrix} = \kappa_m(s) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

This illustrates the moving frame method, which provides here the Euclidean curvature. It can be shown to always provide a function which is invariant under the considered transformations and change of parametrization. More precisely, we have the following definition. For a group G with associated arc length $d\sigma = Q du$, we will use the notation

$$\partial_\sigma = \partial_u / Q,$$

which generalizes the arc-length derivative defined in the Euclidean case. The following discussion concerns curves such that $Q \neq 0$, which generalizes the notion of regular curves.

Let $J_k(G)$ be the set of vectors $(z_0, z_1, \ldots, z_k) \in (\mathbb{R}^2)^{k+1}$ such that there exists a curve m such that $z_k = \partial_\sigma^k m$. That this condition induces restrictions on $z_1, \ldots, z_k$ is already clear in the case of rotations, for which one must have $|z_1| = 1$.

Definition 1.24 Let G be a group acting on $\mathbb{R}^2$ (e.g., a subgroup of $\mathrm{GL}_2(\mathbb{R})$). A G-moving frame of order k is a one-to-one function $P_0 : J_k(G) \to G$ with the following property. For all curves $m : I \to \mathbb{R}^2$ with $Q \neq 0$ on m, define $P_m : I \to G$ by

$$P_m = P_0(m, \partial_\sigma m, \ldots, \partial_\sigma^k m).$$

Then, one must have $P_{gm} = g P_m$ for all $g \in G$.

We now consider affine transformations, with group G a subgroup of $\mathrm{GL}_2(\mathbb{R}) \ltimes \mathbb{R}^2$ (cf. Sect. B.4.3). An element of G is represented by a pair (A, b) for a linear map A and $b \in \mathbb{R}^2$. We will therefore write $P_0 = (A_0, b_0)$, $P_m = (A_m, b_m)$. We denote by G_0 the linear part of G, i.e., $(A, b) \in G \Rightarrow A \in G_0$. The invariance condition in Definition 1.24 yields, for all $U \in G_0$, $h \in \mathbb{R}^2$,

$$\begin{aligned} A_0(Uz_0 + h, Uz_1, Uz_2, \ldots, Uz_k) &= UA_0(z_0, z_1, z_2, \ldots, z_k), \\ b_0(Uz_0 + h, Uz_1, Uz_2, \ldots, Uz_k) &= Ub_0(z_0, z_1, z_2, \ldots, z_k) + h. \end{aligned} \tag{1.30}$$

We have the following result, which generalizes Theorem 1.13. We here use the same notation as in Sect. B.5.

Theorem 1.25 (Moving Frame: affine case) *Let $G = G_0 \ltimes \mathbb{R}^2$ be a subgroup of* $\mathrm{GL}_2(\mathbb{R}) \ltimes \mathbb{R}^2$. *If $P_0 = (A_0, b_0)$ is a G-moving frame, then, for any plane curve m*

$$\bar{W}_m = A_m^{-1} \partial_\sigma P_m = (A_m^{-1} \partial_\sigma A_m, A_m^{-1} \partial_\sigma b_m)$$

is invariant under change of parametrization and under the action of G. It moreover characterizes the curve up to the action of G: if $\bar{W}_{m^*} = \bar{W}_{\tilde{m}^*}$, where m^* and $\tilde{m}^*$ are respectively the arc-length reparametrization of m and $\tilde{m}$, then $\tilde{m} = gm \circ \psi$ for some $g \in G$ and a change of parameter ψ.

Proof Invariance by change of parametrization relies on the fact the arc length is, by construction, invariant and the details are left to the reader. If $\tilde{m} = Um + h$, then $P_{\tilde{m}} = (UA_m, Ub_m + h)$ and

$$\bar{W}_{\tilde{m}} = A_m^{-1} U^{-1} (U \partial_\sigma A_m, U \partial_\sigma b_m) = P_m^{-1} \partial_\sigma P_m = \bar{W}_m,$$

which proves G-invariance.

Conversely, assume that $\bar{W}_{\tilde{m}} = \bar{W}_m = W$, and assume, without loss of generality, that they are both parametrized by arc length. Let $g = (U, h) = P_{\tilde{m}}(0)P_m(0)^{-1}$. The proof that $\tilde{m} = gm$ for some g derives from the uniqueness theorem for ordinary differential equations (cf. Appendix C); $P_m = (A_m, b_m)$ and $P_{\tilde{m}} = (A_{\tilde{m}}, b_{\tilde{m}})$ are both solutions of the equation $\partial_\sigma(A, b) = AW$, and gP_m is another solution, as can easily be checked. Since $gP_m(0) = P_{\tilde{m}}(0)$ by definition of g, we have

$$P_0(\tilde{m}, \dot{\tilde{m}}, \ldots, \tilde{m}^{(k)}) = gP_0(\dot{m}, \ldots, m^{(k)}) = P_0(gm, U\dot{m}, \ldots, Um^{(k)}).$$

Because P_0 is assumed to be one-to-one, we have $\tilde{m} = gm$, which proves the theorem. □

For affine groups, we select a moving frame P_0 of the form $P_0(z_0, z_1, \ldots, z_k) = (A_0(z_1, \ldots, z_k), z_0)$. This implies that

$$\bar{W}_m = \left(A_m^{-1}\partial_\sigma A_m,\ A_m^{-1}\partial_\sigma m\right).$$

We will mainly focus on the first term, which we denote by

$$W_m = A_m^{-1}\partial_\sigma A_m.$$

The choice made for rotations corresponds to $A_0(z_1) = [z_1, Rz_1]$, R being the $(\pi/2)$-rotation. It is obviously one-to-one and satisfies the invariance requirements. The second term in $\bar{W}_m$ is constant, namely $A_m^{-1}\partial_\sigma m = (1, 0)$.

It can be shown that W_m can lead to only one, "fundamental", scalar invariant. All other coefficients are either constant, or can be deduced from this fundamental invariant. This invariant will be called the curvature associated to the group.

Consider this approach applied to similitudes. Assume that the curve is parametrized by the related arc length, σ. The frame, here, must be a similitude, A_m, and, as above, we take

$$A_m = \begin{pmatrix} \dot{x} & -\dot{y} \\ \dot{y} & \dot{x} \end{pmatrix}.$$

Define $W_m = A_m^{-1}\partial_\sigma A_m$, so that

$$\begin{aligned} W_m &= \frac{1}{\dot{x}^2 + \dot{y}^2}\begin{pmatrix} \dot{x} & \dot{y} \\ -\dot{y} & \dot{x} \end{pmatrix}\begin{pmatrix} \ddot{x} & -\ddot{y} \\ \ddot{y} & \ddot{x} \end{pmatrix} \\ &= \frac{1}{\dot{x}^2 + \dot{y}^2}\begin{pmatrix} \ddot{x}\dot{x} + \ddot{y}\dot{y} & \ddot{x}\dot{y} - \ddot{y}\dot{x} \\ -\ddot{x}\dot{y} + \ddot{y}\dot{x} & \ddot{x}\dot{x} + \ddot{y}\dot{y} \end{pmatrix}. \end{aligned}$$

When the curve is parametrized by arc length, we have

$$\frac{|\dot{x}\ddot{y} - \ddot{x}\dot{y}|}{\dot{x}^2 + \dot{y}^2} = 1$$

along the curve. Therefore

$$W_m(\sigma) = \begin{pmatrix} \frac{\ddot{x}\dot{x}+\ddot{y}\dot{y}}{\dot{x}^2+\dot{y}^2} & \mp 1 \\ \pm 1 & \frac{\ddot{x}\dot{x}+\ddot{y}\dot{y}}{\dot{x}^2+\dot{y}^2} \end{pmatrix}.$$

(σ being the similitude arc length). The computation exhibits a new quantity, which is

$$K = \frac{\ddot{x}\dot{x} + \ddot{y}\dot{y}}{\dot{x}^2 + \dot{y}^2}. \tag{1.31}$$

This is the curvature for the group of similitudes: it is invariant under translation, rotation and scaling, and characterizes curves up to similitudes.

We now consider special affine transformations (affine with determinant 1). Assume that the curve is parametrized by the corresponding arc length, σ, i.e.,

$$|\ddot{x}\dot{y} - \ddot{y}\dot{x}|^{1/3} = 1.$$

One can choose $A_m = \begin{pmatrix} \dot{x} & \ddot{x} \\ \dot{y} & \ddot{y} \end{pmatrix}$, which has determinant 1. Since $A_m(1,0)^T = \dot{m}$, the term $A_m^{-1}\dot{m}$ is trivial. We have

$$A_m^{-1}\partial A_m = \begin{pmatrix} \ddot{y} & -\ddot{x} \\ -\dot{y} & \dot{x} \end{pmatrix} \begin{pmatrix} \ddot{x} & x^{(3)} \\ \ddot{y} & y^{(3)} \end{pmatrix} = \begin{pmatrix} 0 & \ddot{y}x^{(3)} - \ddot{x}y^{(3)} \\ 1 & -\dot{y}x^{(3)} + \dot{x}y^{(3)} \end{pmatrix}.$$

Since $\partial(\ddot{x}\dot{y} - \ddot{y}\dot{x}) = \dot{y}x^{(3)} - \dot{x}y^{(3)} = 0$, the only non-trivial coefficient is $\ddot{y}x^{(3)} - \ddot{x}y^{(3)}$, which can be taken (up to a sign change) as a definition of the *special affine curvature*:

$$K = \det(\ddot{m}, m^{(3)}). \tag{1.32}$$

Again, this is expressed as a function of the affine arc length and is invariant under the action of special affine transformations.

The local invariants with respect to rotation, similitude and the special affine group probably reach the limits of numerical feasibility, based on the number of derivatives they require. Going further involves even higher derivatives, and has only theoretical interest. However, we include here, for completeness, the definition of the affine and projective arc lengths and curvatures. This section can be safely skipped. In discussing the projective arc lengths, we will use a few notions that are related to Lie groups and manifolds. The reader can refer to Appendix B for more details.

1.15.5 Affine Arc Length

We first introduce new parameters which depend on the sequence $z_1, \ldots, z_n$ that describes the first derivatives. We assume that $\det(z_1, z_2) \neq 0$ and let

$$\alpha_k = \alpha_k(z_1, \ldots, z_n) = \frac{\det(z_k, z_2)}{\det(z_1, z_2)}$$
$$\text{and } \beta_k = \beta_k(z_1, \ldots, z_n) = \frac{\det(z_1, z_k)}{\det(z_1, z_2)}.$$

These are defined so that

$$z_k = \alpha_k z_1 + \beta_k z_2, \tag{1.33}$$

which also yields

$$\begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix} = [z_1, z_2]^{-1} z_k.$$

In particular, we have $\alpha_1 = \beta_2 = 1$, $\alpha_2 = \beta_1 = 0$.

Assuming affine invariance, we must have

$$Q(z_1, \ldots, z_n) = Q([z_1, z_2]^{-1} z_1, \ldots, [z_1, z_2]^{-1} z_n),$$

which implies that Q must be a function of the α_k's and β_k's. We see also that we must have at least $n = 3$ to ensure a non-trivial solution. In fact, we need to go to $n = 4$, as will be shown by the following computation.

For $n = 4$, the parametric invariance constraint yields: for all $\lambda_1 > 0, \lambda_2, \lambda_3, \lambda_4$,

$$Q(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3, \tilde{z}_4) = \lambda_1 Q(z_1, z_2, z_3, z_4)$$

with $\tilde{z}_1 = \lambda_1 z_1$, $\tilde{z}_2 = \lambda_2 z_1 + \lambda_1^2 z_2$, $\tilde{z}_3 = \lambda_3 z_1 + 3\lambda_2\lambda_1 z_2 + \lambda_1^3 z_3$ and

$$\tilde{z}_4 = \lambda_4 z_1 + (3\lambda_2^2 + 4\lambda_3\lambda_1) z_2 + 6\lambda_1^2\lambda_2 z_3 + \lambda_1^4 z_4.$$

We now make specific choices for $\lambda_1, \lambda_2, \lambda_3$ and λ_4 to progressively reduce the functional form of Q. We will abuse the notation by keeping the letter Q to design the function at each step. Our starting point is $Q = Q(\alpha_3, \beta_3, \alpha_4, \beta_4)$.

(i) We start by taking $\lambda_1 = 1$, $\lambda_2 = \lambda_3 = 0$, yielding $\tilde{z}_1 = z_1$, $\tilde{z}_2 = z_2$, $\tilde{z}_3 = z_3$ and $\tilde{z}_4 = z_4 + \lambda_4 z_1$. Denote by $\tilde{\alpha}_k$, $\tilde{\beta}_k$ the α_k, β_k coefficients associated to the $\tilde{z}$'s. For the considered variation, the only coefficient that changes is α_4, which becomes $\tilde{\alpha}_4 = \alpha_4 + \lambda_4$. This implies that

$$Q(\alpha_3, \beta_3, \alpha_4, \beta_4) = Q(\alpha_3, \beta_3, \alpha_4 + \lambda_4, \beta_4).$$

Taking $\lambda_4 = -\alpha_4$, we see that Q does not depend on α_4, yielding the new functional form $Q = Q(\alpha_3, \beta_3, \beta_4)$.

(ii) Let's now consider $\lambda_1 = 1, \lambda_2 = \lambda_4 = 0$. In this case, z_1, z_2 remain unchanged, and z_3 and z_4 become $\tilde{z}_3 = z_3 + \lambda_3 z_1$, $\tilde{z}_4 = z_4 + 4\lambda_3 z_2$. This implies $\tilde{\alpha}_3 = \alpha_3 + \lambda_3$, $\tilde{\beta}_3 = \beta_3$ and $\tilde{\beta}_4 = \beta_4 + 4\lambda_3$. Taking $\lambda_3 = -\alpha_3$ yields the new functional form $Q = Q(\beta_3, \beta_4 - 4\alpha_3)$.

(iii) Now, take $\lambda_1 = 1$, $\lambda_2 = \lambda_4 = 0$, yielding $\tilde{z}_1 = z_1$, $\tilde{z}_2 = z_2 + \lambda_2 z_1$, $\tilde{z}_3 = z_3 + 3\lambda_2 z_2$ and $\tilde{z}_4 = z_4 + 6\lambda_2 z_3 + 3\lambda_2^2 z_2$, so that $\tilde{\beta}_3 = \beta_3 + 3\lambda_2$, $\tilde{\alpha}_3 = \alpha_3 - 3\lambda_2^2 - \lambda_2\beta_3$ and $\tilde{\beta}_4 = \beta_4 + 3\lambda_2^2 + 6\lambda_2\beta_3$. In particular,

$$\tilde{\beta}_4 - 3\tilde{\alpha}_3 = \beta_4 - 4\alpha_3 + 15\lambda_2^2 + 10\lambda_2\beta_3.$$

Taking $\lambda_2 = -\beta_3/3$ yields $Q = Q(\beta_4 - 4\alpha_3 - 5\beta_3^2/3)$.

(iv) Finally, take $\lambda_2 = \lambda_2 = \lambda_4 = 0$ yielding $\tilde{\beta}_3 = \lambda_1\beta_3$ $\tilde{\beta}_4 = \lambda_1^2\beta_4$ and $\tilde{\alpha}_3 = \lambda_1^2\alpha_3$. This gives

$$Q(\lambda_1^2(\beta_4 - 4\alpha_3 - 5\beta_3^2/3)) = \lambda_1 Q(\beta_4 - 4\alpha_3 - 5\beta_3^2/3).$$

Taking $\lambda_1 = 1/\sqrt{|\beta_4 - 4\alpha_3 - 5\beta_3/3|}$, assuming this expression does not vanish, yields

$$Q(\beta_4 - 4\alpha_3 - 5\beta_3^2/3) = \begin{cases} Q(1)\sqrt{|\beta_4 - 4\alpha_3 - 5\beta_3^2/3|} \text{ if } \beta_4 - 4\alpha_3 - 5\beta_3^2/3 > 0, \\ Q(-1)\sqrt{|\beta_4 - 4\alpha_3 - 5\beta_3^2/3|} \text{ if } \beta_4 - 4\alpha_3 - 5\beta_3^2/3 < 0. \end{cases}$$

Here again, it is natural to ensure an invariance by a change of orientation and let $Q(1) = Q(-1) = 1$ so that

$$Q(z_1, z_2, z_3, z_4) = \sqrt{|\beta_4 - 4\alpha_3 - 5\beta_3^2/3|}.$$

This provides the affine-invariant arc length.

We can take the formal derivative in (1.33), yielding

$$z_{k+1} = \alpha_k' z_1 + \alpha_k z_2 + \beta_k' z_2 + \beta_k z_3 = (\alpha_k' + \beta_k\alpha_3)z_1 + (\alpha_k + \beta_k' + \beta_k\beta_2)z_2,$$

so that $\alpha_{k+1} = \alpha_k' + \beta_k\alpha_3$ and $\beta_{k+1} = \beta_k' + \alpha_k + \beta_k\beta_3$. This implies that higher-order coefficients can always be expressed in terms of α_3, β_3 and their (formal) derivatives, which are represented using prime exponents. In particular, using $\beta_4 = \beta_3' + \alpha_3 + \beta_3^2$, we get

$$Q(z_1, z_2, z_3, z_4) = \sqrt{|\beta_3' - 3\alpha_3 - 2\beta_3^2/3|}. \tag{1.34}$$

Returning to parametrized curves, let $\alpha_{m,k}$ and $\beta_{m,k}$ be the coefficients α_k, β_k in which $(z_1, z_2, \ldots)$ are replaced by their corresponding derivatives $(\dot{m}_u, \ddot{m}_{uu}, \ldots)$, so that

$$m^{(k)} = \alpha_{m,k}\dot{m} + \beta_{m,k}\ddot{m}.$$

We want to express the affine arc length in terms of the Euclidean curvature. Assuming that m is parametrized by Euclidean arc length, we have $\ddot{m} = \kappa R\dot{m}$, where R is the $\pi/2$ rotation. Taking one derivative yields (using $R^2 = -\mathrm{Id}$)

$$m^{(3)} = \kappa R\ddot{m} + \dot{\kappa}R\dot{m} = -\kappa^2\dot{m} + (\dot{\kappa}/\kappa)\ddot{m}.$$

This implies that $\alpha_{m,3} = -\kappa^2$ and $\beta_{m,3} = \dot{\kappa}/\kappa$; thus, (1.34) implies that the affine arc length, σ, and the Euclidean arc length are related by

$$d\sigma = \sqrt{|\partial(\dot{\kappa}/\kappa) + 3\kappa^2 - 2(\dot{\kappa}/\kappa)^2/3|}ds.$$

1.15.6 Projective Arc Length

The problem is harder to address for the projective group (see Sect. B.4.3 for a definition) because of the non-linearity of the transformations. We keep the same notation for α_k and β_k as in the affine case (since the projective group includes the affine group, we know that the function Q will have to depend on these reduced coordinates).

Before the computation, we need to express the effects that a projective transformation has on the derivative of the curve. We still let the symbol z_k hold for the kth derivative. A projective transformation applied to a point $z \in \mathbb{R}^2$ takes the form $g : z \mapsto (Uz + b)/(w^T z + 1)$ for a 2 by 2 matrix U, and vectors $b, w \in \mathbb{R}^2$. Let $\gamma_0 = (w^T z_0 + 1)^{-1}$ so that z_0 is transformed as $\tilde{z}_0 = \gamma_0(Uz_0 + b)$. We need to express the higher derivatives $\tilde{z}_1, \tilde{z}_2, \ldots$ as functions of the initial $z_1, z_2, \ldots$ and the parameters of the transformations. Letting γ_k represent the kth derivative of γ_0, the rule for the derivation of a product (Leibniz's formula) yields

$$\tilde{z}_k = \gamma_k(Uz_0 + b) + \sum_{q=1}^{k} \binom{k}{q} \gamma_{k-q} U z_q. \tag{1.35}$$

This provides a group action, which will be denoted $\tilde{z} = g \star z$. Our goal is to find a function Q such that $Q(z_1, z_2, \ldots, z_k) = Q(\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_k)$, and which is also invariant under the transformations induced by a change of variables. It will be necessary to go to $k = 5$ for the projective group.

We first focus on projective invariance, and make an analysis equivalent to the one that allowed us to remove z_0, z_1 and z_2 in the affine case. More precisely, we show that U, b, and w can be found such that $\tilde{z}_0 = 0$, $\tilde{z}_1 = e_1$, $\tilde{z}_2 = e_2$ and $\tilde{z}_3 = 0$, with $e_1 = (1, 0)$ and $e_2 = (0, 1)$.

First note that $\gamma_1 = -w^T z_1 \gamma_0^2$ and $\gamma_2 = -w^T z_2 \gamma_0^2 + 2(w^T z_1)^2 \gamma_0^3$. Take $b = -Uz_0$ to ensure $\tilde{z}_0 = 0$. We have $\tilde{z}_1 = \gamma_0 U z_1$, $\tilde{z}_2 = 2\gamma_1 U z_1 + \gamma_0 U z_2$ and

$$\tilde{z}_3 = 3\gamma_2 U z_1 + 3\gamma_1 U z_2 + \gamma_0 U z_3.$$

We therefore need

$$U z_1 = e_1/\gamma_0, U z_2 = e_2/\gamma_0 - (2\gamma_1/\gamma_0^2) e_1 = e_2/\gamma_0 + 2w^T z_1 e_1$$

and (after some algebra)

$$\begin{aligned} U z_3 &= -3(\gamma_2/\gamma_0) U z_1 - 3(\gamma_1/\gamma_0) U z_2 \\ &= 3w^T z_2 e_1 + 3w^T z_1 e_2. \end{aligned}$$

Using the decomposition $z_k = \alpha_k z_1 + \beta_k z_2$, we also have $U z_3 = \alpha_3(e_1/\gamma_0) + \beta_3(e_2/\gamma_0 - (2\gamma_1/\gamma_0^2)e_1)$, which yields the identification

$$w^T z_1 = \beta_3/(3\gamma_0) \text{ and } w^T z_2 = (3\alpha_3 + 2\beta_3^2)/9.$$

Using the definition of γ_0, this can be written as

$$\begin{cases} w^T(z_1 - \beta_3/3 z_0) = \beta_3/3 \\ w^T(z_2 - (\alpha_3/3 + 2\beta_3/9) z_0) = (\alpha_3/3 + 2\beta_3/9), \end{cases}$$

which uniquely defines w, under the assumption (which we make here) that z_0, $(3/\beta_3) z_1$, $(9/(3\alpha_3 + 2\beta_3^2)) z_2$ forms an affine frame. Given W, we can compute b and U. We have in particular, using the decomposition of z_k:

$$U z_k = (\alpha_k/\gamma_0 + 2\beta_k \beta_3/(3\gamma_0)) e_1 + (\beta_k/\lambda) e_2.$$

Similarly, we have

$$w^T z_k = \alpha_k \beta_3/(3\gamma_0) + \beta_k (3\alpha_3 + 2\beta_3^2)/9.$$

With this choice of U, w and b, the resulting expressions of $\tilde{z}_3$, $\tilde{z}_4$ and $\tilde{z}_5$ can be obtained. This is a heavy computation for which the use of a mathematical software is helpful; the result is that the projective invariance implies that the function Q must be a function of the following four expressions:

$$A = \alpha_4 - \frac{8}{3}\alpha_3 \beta_3 - \frac{8\beta_3^3}{9} + \frac{2}{3}\beta_3 \beta_4$$

$$
\begin{aligned}
B &= \alpha_5 - \frac{10}{3}\alpha_4\beta_3 + \frac{40}{9}\alpha_3\beta_3^3 + \frac{40\beta_3^4}{27} - \\
&\quad \frac{5}{3}\alpha_3\beta_4 - \frac{20}{9}\beta_3^2\beta_4 + \frac{2}{3}\beta_3\beta_5 \\
C &= -2\alpha_3 - \frac{4\beta_3^2}{3} + \beta_4 \\
D &= -\frac{10}{3}\alpha_3\beta_3 - \frac{5}{3}\beta_3\beta_4 + \beta_5.
\end{aligned}
$$

Given this, it remains to carry out the reductions associated to the invariance by change of parameter. This is done as in the affine case, progressively selecting the coefficients λ_i to eliminate one of the expressions and modify the others, the difference being that there is one extra constraint here associated to the fifth derivative. Note that with five constraints, we would normally be short of one expression, but one of the invariances is (magically) satisfied in the reduction process, which would otherwise have required using six derivatives. We spare the reader the details, and directly provide the final expression for Q, which is

$$
Q = \left| \frac{40\beta_3^3}{9} + \beta_5 - 5\beta_3(\beta_4 - 2\alpha_3) - 5\alpha_4 \right|^{1/3}.
$$

As before, this can be expressed in terms of the formal derivatives of α_3 and β_3, yielding

$$
Q = \left[\beta_3'' - 3\alpha_3' - 2\beta_3\beta_3' + 2\beta_3\alpha_3 + (4/9)\beta_3^3\right]^{1/3}. \tag{1.36}
$$

1.15.7 Affine Curvature

We can apply the moving frame method described in Sect. 1.15.4 to obtain the affine curvature of a curve m. We assume here that m is parametrized by affine arc length, σ. A moving frame on m is immediately provided by the matrix $A_m = [\dot{m}_\sigma, \ddot{m}_{\sigma\sigma}]$, or, with our z notation, $A_0 = [z_1, z_2]$. By definition of α_3 and β_3, the matrix $W_m = A_m^{-1}\partial_\sigma A_m$ is equal to

$$
W_m = \begin{pmatrix} 0 & \alpha_{m,3} \\ 1 & \beta_{m,3} \end{pmatrix}.
$$

Since the curve is parametrized by affine arc length, we have $Q = 1$, where Q is given by $\sqrt{|\dot{\beta}_{m,3} - 3\alpha_{m,3} - 2\beta_{m,3}^2/3|}$. This implies that $\alpha_{m,3}$ is a function of $\beta_{m,3}$ and $\dot{\beta}_{m,3}$ along the curve; the moving frame therefore only depends on $\beta_{m,3}$ and its derivatives, which indicates that $\beta_{m,3}$ is the affine curvature. Thus, when a curve is parametrized by affine arc-length, σ, its curvature is given by

$$\kappa_m(\sigma) = \frac{\det(\dot{m}, m^{(3)})}{\det(\dot{m}, \ddot{m})}.$$

If the curve now has an arbitrary parametrization, the curvature is obtained by using $d\sigma = Q du$, where Q is given by (1.34). This yields the following expression:

$$\kappa_m(s) = \frac{1}{Q}\frac{\det(\dot{m}, m^{(3)})}{\det(\dot{m}, \ddot{m})} - \frac{3\dot{Q}}{Q}.$$

1.15.8 Projective Curvature

In the projective case, the moving frame method cannot be used exactly as described in Sect. 1.15.4, because of the non-linearity of the transformations. The moving frame is still associated to a one-to-one function $P_0(z_0, \ldots, z_k) \in G = \mathrm{PGL}_2(\mathbb{R})$. The invariance property in this case gives, with the definition of the action $z \mapsto g \star z$ given in (1.35), $P_0(g \star z) = g P_0(z)$. For Theorem 1.25 to make sense, we must use the differential of the left translation $L_g : h \mapsto hg$ on $\mathrm{PGL}_2(\mathbb{R})$, and define

$$\bar{W}_m = dL_{P_m}(\mathrm{Id})^{-1}\partial_\sigma P_m,$$

which belongs to the Lie algebra of $\mathrm{PGL}_2(\mathbb{R})$. This is the general definition of a moving frame on a Lie group [108], and coincides with the definition that has been given for affine groups, for which we had $dL_g = A$ when $g = (A, b)$.

We first need to build the matrix A_0. For this, using as before the notation (e_1, e_2) for the canonical basis of $\mathbb{R}^2$, we define a projective transformation that takes the family $\omega = (0, e_1, e_2, 0)$ to the family $z = (z_0, z_1, z_2, z_3)$, i.e., we want to determine g such that $g \star \omega = z$ (we showed that its inverse exists in Sect. 1.15.6, but we need to compute it explicitly). Since this provides eight equations for eight dimensions, one can expect that a unique such transformation exists; this will be our $A_0(z)$.

Assuming that this existence and uniqueness property is satisfied, such a construction ensures the invariance of the moving frame under the group action. Indeed, letting z be associated to a curve m and $\tilde{z}$ to $\tilde{m} = g(m)$ for some $g \in \mathrm{PGL}_2(\mathbb{R})$, we have $\tilde{z} = g \star z$. Since $A_0(z)$ is defined by $A_0(z) \star \omega = z$, the equality $A_0(\tilde{z}) \star \omega = \tilde{z}$ is achieved by $A_0(\tilde{z}) = g A_0(z)$, which is the required invariance. (Indeed, because $\star$ is a group action, we have $(g A_0(z)) \star \omega = g \star (A_0(z)\omega) = g \star z = \tilde{z}$.)

We now proceed to the computation. The first step is to obtain the expression of $g \star z$ for $z = (z_0, z_1, z_2, z_3)$. We do this in the special case in which g is given by:

$$g(m) = (Um + b)/(1 + w^T m),$$

w and b being two vectors in $\mathbb{R}^2$ and $U \in \mathrm{GL}_2(\mathbb{R})$. Define $g \star (\tilde{z}_0, \tilde{z}_1, \tilde{z}_2, \tilde{z}_3) = (z_0, z_1, z_2, z_3)$. From $(1 + w^T \tilde{z}_0) z_0 = U\tilde{z}_0 + b$, we obtain

$$\begin{cases} (1 + w^T \tilde{z}_0) z_0 = U \tilde{z}_0 + b \\ (1 + w^T \tilde{z}_0) z_1 + w^T \tilde{z}_1 z_0 = U \tilde{z}_1 \\ (1 + w^T \tilde{z}_0) z_2 + 2 w^T \tilde{z}_1 z_1 + w^T \tilde{z}_2 z_0 = U \tilde{z}_2 \\ (1 + w^T \tilde{z}_0) z_3 + 3 w^T \tilde{z}_1 z_2 + 3 w^T \tilde{z}_2 z_1 + w^T \tilde{z}_3 z_0 = U \tilde{z}_3. \end{cases} \tag{1.37}$$

Taking $\tilde{z} = \omega$, we get

$$\begin{cases} z_0 = b \\ z_1 + w_1 z_0 = u_1 \\ z_2 + 2 w_1 z_1 + w_2 z_0 = u_2 \\ z_3 + 3 w_1 z_2 + 3 w_2 z_1 = 0, \end{cases} \tag{1.38}$$

where $w = (w_1, w_2)$, $U e_1 = u_1$ and $U e_2 = u_2$. The third equation yields

$$z_3 = -3 w_2 z_1 - 3 w_1 z_2. \tag{1.39}$$

We will assume that z_1 and z_2 are linearly independent, so that $w = (w_1, w_2)$ is uniquely defined by this equation, and therefore $U = [u_1, u_2]$ by the middle equations of (1.38). Using again the notation $z_3 = \alpha_3 z_1 + \beta_3 z_2$, we get

$$\begin{cases} w_1 = -\beta_3/3 \\ w_2 = -\alpha_3/3. \end{cases}$$

This fully defines our moving frame $A_0(z)$.

Recall that the formal derivative of a quantity M that depends on $z_0, \dots, z_3$ is given, in our notation, by $M' = \sum_{k=0}^{3} (\partial M / \partial z_k) z_{k+1}$. Since $b = z_0$, we have $b' = z_1$; from $u_1 = z_1 + w_1 z_0$, we get

$$u_1' = w_1 z_1 + z_2 + w_1' z_0 \,,$$

and from (1.39) and $u_2 = z_2 + 2 w_1 z_1 + w_2 z_0$,

$$\begin{aligned} u_2' &= z_3 + 2 w_1 z_2 + (2 w_1' + w_2) z_1 + w_2' z_0 \\ &= (-2 w_2 + 2 w_1') z_1 - w_1 z_2 + w_2' z_0. \end{aligned}$$

We have $w_1' = -\beta_3'/3$ and $w_2' = -\alpha_3'/3$, which are therefore directly computable along the curve.

By taking the representation of a projective transformation by the triplet (U, b, w), we have chosen a local chart on $\mathrm{PGL}_2(\mathbb{R})$ which obviously contains the identity represented by $(\mathrm{Id}, 0, 0)$. To be able to compute the differential of the left translation $L_{A(z)}$, we need to express the product in this chart. One way to do this efficiently is to observe that, by definition of the projective group, products in $\mathrm{PGL}_2(\mathbb{R})$ can be deduced from matrix products in $\mathrm{GL}_3(\mathbb{R})$, up to a multiplicative constant. A function g with coordinates (U, b, w) in the chart is identified (up to multiplication by a scalar) with the matrix

$$\begin{pmatrix} U & b \\ w^T & 1 \end{pmatrix}$$

and the product of $\tilde{g} = (\tilde{U}, \tilde{b}, \tilde{w})$ and $\bar{g} = (\bar{U}, \bar{b}, \bar{w})$ is therefore identified with the product of the associated matrices, which is

$$\begin{pmatrix} \bar{U} & \bar{b} \\ \bar{w}^T & 1 \end{pmatrix} \begin{pmatrix} \tilde{U} & \tilde{b} \\ \tilde{w}^T & 1 \end{pmatrix} = \begin{pmatrix} \bar{U}\tilde{U}' + \bar{b}\tilde{w}^T & \bar{U}\tilde{b} + \bar{b} \\ \bar{w}^T\tilde{U} + \tilde{w}^T & \bar{w}^T\tilde{b} + 1 \end{pmatrix},$$

which yields the chart representation for the product

$$\bar{g}\tilde{g} = \Big((\bar{U}\tilde{U} + \bar{b}\tilde{w}^T)/(1 + \bar{w}^T\tilde{b}), \\ (\bar{U}\tilde{b} + \bar{b})/(1 + \bar{w}^T\tilde{b}), (\tilde{U}^T\bar{w} + \tilde{w}')/(1 + \bar{w}^T\tilde{b}) \Big).$$

To compute the differential of the left translation in local coordinates, it suffices to take $\tilde{U} = \mathrm{Id} + \varepsilon H$, $\tilde{b} = \varepsilon\beta$ and $\tilde{w} = \varepsilon\gamma$, and compute the first derivative of the product with respect to ε at $\varepsilon = 0$. This yields

$$d_{\mathrm{Id}} L_{\bar{g}}(H, \beta, \gamma) = (\bar{U}H + \bar{b}\gamma^T - \bar{w}^T\beta\bar{U}, \bar{U}\beta - \bar{w}^T\beta\bar{b}, \gamma + H^T\bar{w} - \bar{w}^T\beta\bar{w}).$$

We need to compute the inverse of this linear transformation, and therefore solve

$$\begin{cases} \bar{U}H + \bar{b}\gamma^T - \bar{w}^T\beta\bar{U} = \tilde{H} \\ \bar{U}\beta - \bar{w}^T\beta\bar{b} = \tilde{\beta} \\ \gamma + H^T\bar{w} - \bar{w}^T\beta\bar{w} = \tilde{\gamma}. \end{cases}$$

The second equation yields $\beta = (\bar{U} - \bar{b}\bar{w}^T)^{-1}\tilde{\beta}$. Substituting γ in the first by its expression in the third yields

$$\tilde{H} = (\bar{U} - \bar{b}\bar{w}^T)H + \bar{b}\tilde{\gamma}^T + (\bar{w}^T\beta)\bar{b}\bar{w}^T - \bar{w}^T\beta\bar{U}$$

so that

$$H = (\bar{U} - \bar{b}\bar{w}^T)^{-1}(\tilde{H} - \bar{b}\tilde{\gamma}^T) + (\bar{w}^T\beta)\mathrm{Id}.$$

Finally, we have

$$\gamma = \tilde{\gamma} - H^T\bar{w} + \bar{w}^T\beta\bar{w}.$$

$\bar{W}$ is obtained by applying these formulae to $\bar{g} = A(z) = (U, b, w)$ and $\tilde{H} = (\theta_1, \theta_2)$ with

$$\begin{cases} \theta_1 = u_1' = w_1 z_1 + z_2 + w_1' z_0 \\ h_2' = u_2' = (2w_1' - 2w_2)z_1 - w_1 z_2 + w_2' z_0 \\ \tilde{\beta} = z_1 \\ \tilde{\gamma} = w'. \end{cases}.$$

Note that, since $(A - bw^T)h = Ah - w^T hb$, the identity $z_1 = u_1 - w_1 b$ implies

$$\beta = (U - bw^T)^{-1} z_1 = e_1 .$$

Similarly, from $u_2 - w_2 b = z_2 + 2w_1 z_1$, we get

$$(U - bw^T)^{-1} z_2 = e_2 - 2w_1 e_1 .$$

We have, using $b = z_0$ and $\tilde{\gamma} = w'$,

$$\tilde{H} - b\tilde{\gamma}^T = \big(w_1 z_1 + z_2,\ (w_1' - 2w_2)z_1 - w_1 z_2\big).$$

We therefore obtain

$$h_1 = (U - bw^T)^{-1}(w_1 z_1 + z_2) + w_1 e_1 = w_1 e_1 + e_2 - 2w_1 e_1 + w_1 e_1 = e_2 ,$$

$$\begin{aligned} h_2 &= (U - bw^T)^{-1}((2w_1' - 2w_2)z_1 - w_1 z_2) + w_1 e_2 \\ &= (2w_1' - 2w_2)e_1 - w_1(e_2 - 2w_1 e_1) + w_1 e_2 \\ &= (2w_1' + 2w_1^2 - 2w_2)e_1 . \end{aligned}$$

With $c = w_1' + w_1^2 - w_2$, we have obtained $W = \begin{pmatrix} 0 & 2c \\ 1 & 0 \end{pmatrix}$. Moreover, we have

$$\gamma = w' - H^T w + w_1 w = \begin{pmatrix} w_1' - w_2 + w_1^2 \\ w_2' - 2cw_1 + w_1 w_2 \end{pmatrix}.$$

Because we assume that $\left[\beta_3'' - 3\alpha_3' - 2\beta_3\beta_3' + 2\beta_3\alpha_3 + (4/9)\beta_3^3\right]^{1/3} = 1$, we see that $w_2' = -\alpha_3'/3$ can be expressed as a function of α_3 and the derivatives of β_3 (up to the second one), while c is equal to $-(\beta_3' - \beta_3^2/3 - \alpha_3)/3$. The invariant of smallest degree can therefore be taken to be $\beta_3' - \beta_3^2/3 - \alpha_3$ (in fact, $w_2' - 2cw_1 + w_1 w_2 = -c'/6$). The projective curvature can therefore be taken as (assuming a curve parametrized by projective arc length)

$$\kappa_m(\sigma) = \partial\left(\frac{\det(\dot{m}, m^{(3)})}{\det(\dot{m}, \ddot{m})}\right) - \frac{\det(m^{(3)}, \ddot{m})}{\det(\dot{m}, \ddot{m})} + \frac{1}{3}\left(\frac{\det(\dot{m}, m^{(3)})}{\det(\dot{m}, \ddot{m})}\right)^2 .$$

The computation of the expression of the curvature for an arbitrary parametrization is left to the reader. It involves the second derivative of the arc length, and therefore the seventh derivative of the curve.

1.16 Non-local Representations

1.16.1 Semi-local Invariants

The invariants that we have defined so far depend on derivatives that can be difficult to estimate in the presence of noisy data (as seen in Fig. 1.1). Semi-local invariants attempt to address this issue by replacing derivatives by estimates depending on nearby, but not coincident, points. They provide new curve "signatures", different from the one associated to the curvature.

A general recipe for building semi-local invariants can be described as follows [48]. For a given integer, k, one needs to provide:

1. An algorithm to select k points on the curve, relative to a single point $m(u)$.
2. A formula to compute a signature based on the k selected points.

We introduce some notation. First, let S_m represent the selection of k points along m. If $p = m(u)$ is a point on m, we let $S_m(p) = (p_1, \ldots, p_k)$. Second, let F be the signature function: it takes $p_1, \ldots, p_k$ as input and returns a real number.

We need to enforce invariance at both steps of the method. Reparametrization invariance is implicitly enforced by the assumption that S_m only depends on $p = m(u)$ (and not on u). Consider now the issue of invariance with respect to a class G of affine transformations. For A in this class, we want that:

1. The point selection process "commutes": if $S_m(p) = (p_1, \ldots, p_k)$, then $S_{Am}(Ap) = (Ap_1, \ldots Ap_k)$.
2. The function F is invariant: $F(Ap_1, \ldots, Ap_k) = F(p_1, \ldots, p_k)$.

Enforcing Point 2 becomes easy if one introduces a transformation A which places the first points in $S_m(p)$ in a generic position, leading to a normalization of the function F. We clarify this operation with examples. Assume that the class of transformations being considered are translations and rotations. Then, there is a unique such transformation that displaces p_1 on O and p_2 on $|p_1 - p_2|e_1$, where e_1 is the unit vector of the horizontal axis. Denote this transformation by A_{p_1,p_2}. Then, we must have

$$\begin{aligned} F(p_1, p_2, \ldots, p_k) &= F(A_{p_1,p_2}p_1, A_{p_1,p_2}p_2, \ldots, A_{p_1,p_2}p_k) \\ &= F(0, |p_1 - p_2|e_1, A_{p_1,p_2}p_3, \ldots, A_{p_1,p_2}p_k). \end{aligned}$$

Conversely, it is clear that any function F of the form

$$F(p_1, p_2, \ldots, p_k) = \tilde{F}(|p_1 - p_2|, A_{p_1,p_2}p_3, \ldots, A_{p_1,p_2}p_k)$$

is invariant under rotation and translation. The transformation $A_{p_1 p_2}$ can be made explicit: skipping the computation, this yields ((x_i, y_i) being the coordinates of p_i)

$$A_{p_1,p_2}p_j = \frac{1}{|p_2 - p_1|}\begin{pmatrix}(x_2 - x_1)(x_j - x_1) + (y_2 - y_1)(y_j - y_1)\\(x_2 - x_1)(y_j - y_1) - (y_2 - y_1)(x_j - x_1)\end{pmatrix}.$$

Thus, with three selected points, the general form of F is

$$F(p_1, p_2, p_3) = \tilde{F}\Big(|p_2 - p_1|, \frac{(p_2 - p_1)^T(p_3 - p_1)}{|p_2 - p_1|}, \frac{\det(p_2 - p_1, p_3 - p_1)}{|p_2 - p_1|}\Big).$$

If scaling is added to the class of transformations, the same argument shows that the only choice with three points is:

$$F(p_1, p_2, p_3) = \tilde{F}\left(\frac{(p_2 - p_1)^T(p_3 - p_1)}{|p_2 - p_1|^2}, \frac{\det(p_2 - p_1, p_3 - p_1)}{|p_2 - p_1|^2}\right).$$

Similar computations hold for larger classes of transformations.

There are several possible choices for point selection (Step 1). One can use the arc length (relative to the class of transformations) that we have defined in the previous sections, and choose $p_1, \dots, p_k$ symmetrically around p, with fixed relative arc lengths $\sigma_m(p_1) - \sigma_m(p), \dots, \sigma_m(p_k) - \sigma_m(p)$. For example, letting $\delta_i = \sigma_m(p_i) - \sigma_m(p)$, and if $k = 2l + 1$, one can take $\delta_1 = -l\varepsilon, \delta_2 = -(l-1)\varepsilon, \dots, \delta_k = l\varepsilon$.

However, the arc length requires using curve derivatives, and this is precisely what we wanted to avoid. Some purely geometric constructions can be used instead. For rotations, for example, we can choose $p_1 = p$, and p_2 and p_3 to be the two intersections of the curve m with a circle of radius ε centered at p (taking the ones closest to p on the curves) with ε small enough. For scale and rotation, consider again circles, but instead of fixing the radius in advance, adjust it so that $|p_2 - p_3|$ becomes smaller that $1 - \varepsilon$ times the radius of the circle. This is always possible, unless the curve is a straight line.

Considering the class of special affine transformations [48], one can choose p_1, p_2, p_3, p_4 such that the line segments (p_1, p_2) and (p_3, p_4) are parallel to the tangent at p, and the areas of the triangles (p_0, p_1, p_2) and (p_0, p_3, p_4) are respectively given by ε and 2ε.

1.16.2 The Shape Context

The shape context [33] represents a shape by a collection of histograms along its outline. Here we give a presentation of this concept in the continuum and do not discuss discretization issues.

Let $s \mapsto m(s)$ be a parametrized curve, defined on some interval I. For $s, t \in I$, let $v(s, t) = m(t) - m(s)$. Fixing t, the function $s \mapsto v(s, t)$ takes values in $\mathbb{R}^2$.

Consider a density kernel, i.e, a function $K : \mathbb{R}^2 \to \mathbb{R}^2$ such that, for fixed x, $K(x, \cdot)$ is a probability density on $\mathbb{R}^2$, usually symmetric around x. The typical example is

$$K(x, y) = \exp(-|x - y|^2/(2\sigma^2))/(2\pi\sigma^2). \tag{1.40}$$

Using this kernel, let, for $s \in I$

$$f^{(m)}(s, y) = \int_I K(y, v(s, t))dt.$$

The density $f^{(m)}(s, \cdot)$ is the shape context of the curve at s and the bivariate function $f^{(m)}$ is the shape context of the whole curve. To discuss some invariance properties of this representation, we assume that the curve is parametrized by arc length (and therefore focus on translation and rotations), and that K is radial, i.e., $K(x, y)$ only depends on $|x - y|$, which is true for (1.40).

A translation applied to the curve has no effect on $v(s, t)$ and therefore leaves the shape context invariant. A rotation R transforms v into Rv, and we have $f^{(Rm)}(s, Ry) = f^{(m)}(s, y)$. The representation is not scale-invariant, but can be made so with an additional normalization (e.g., by forcing the mean distance between different points in the shape to be equal to 1, cf. [33]).

The shape context is a global representation, since it depends for any point on the totality of the curve. To some extent, however, it shares the property of local representations that small variations of the contour will have a small influence on the shape context of other points, by only slightly modifying the density $f(s, \cdot)$.

1.16.3 Conformal Welding

Conformal welding is a complex analysis operation that provides a representation of a curve by a diffeomorphism of the unit circle. While a rigorous description of the method requires advanced mathematical concepts (compared to the rest of this book), the resulting representation is interesting enough to justify the effort.

We will identify $\mathbb{R}^2$ with $\mathbb{C}$, via the usual correspondence $(x, y) \to x + iy$, and add to $\mathbb{C}$ a point at infinity that will confer the structure of a two-dimensional sphere to it. This can be done using the mapping

$$F(re^{i\theta}) = \left(\frac{2r\cos\theta}{r^2+1}, \frac{2r\sin\theta}{r^2+1}, \frac{r^2-1}{r^2+1}\right).$$

This mapping can be interpreted as identifying parallel circles on the sphere with zero-centered circles on the plane; zero is mapped to the south pole, the unit disc is mapped to the equator, and the representation tends to the north pole as $r \to \infty$. With this representation, the interior and the exterior of the unit disc are mapped to hemispheres and therefore play a symmetric role. We will let $\bar{\mathbb{C}}$ denote $\mathbb{C} \cup \infty$.

The complex derivative of a function is defined as the limit of $(f(z+h) - f(z))/h$ as $h \to 0$ in $\mathbb{C}$.

Two domains $\Omega_1, \Omega_2 \subset \bar{\mathbb{C}}$ are said to be conformally equivalent if there exists a function $f : \Omega_1 \to \Omega_2$ such that f is onto and one-to-one and the complex derivative $f'(z)$ exists for all $z \in \Omega_1$, with $f'(z) \neq 0$. Such a function has the property of conserving angles, in the sense that the angle made by two curves passing by z remains unchanged after a transformation by f.

The Riemann mapping theorem [249] states that any simply connected domain (i.e., any domain within which any simple closed curve can be continuously deformed into a point) is conformally equivalent to the unit disc. This domain may or may not include a point at infinity and therefore may or may not be bounded. For example, the transformation $z \mapsto 1/z$ maps the interior of the unit disc to its exterior and vice-versa. This conformal transformation is obviously unique up to any conformal mapping of the unit disc onto itself. It can be shown that the latter transformations must belong to a three-parameter family (a sub-class of the family of Möbius transformations of the plane), containing functions of the form

$$z \mapsto e^{i\alpha} \frac{z^{i\beta} + r}{rze^{i\beta} + 1} \tag{1.41}$$

with $r < 1$. We let M_1 be the set of such transformations (which forms a three-parameter group of diffeomorphisms of the unit disc). A transformation in M_1 can be decomposed into three steps: a rotation $z \mapsto ze^{i\beta}$ followed by the transformation $z \mapsto (z+r)/(zr+1)$, followed again by a rotation $z \mapsto ze^{i\alpha}$.

The Riemann mapping theorem can be applied to the interior and to the exterior of any Jordan curve γ. Letting Ω_γ represent the interior, and $\overline{\Omega}_\gamma^c$ the exterior (the notation holding for the complement of the closure of Ω_γ), and D being the open unit disc, we therefore have two conformal transformations $\Phi_- : \Omega_\gamma \to D$ and $\Phi_+ : \overline{\Omega}_\gamma^c \to D$. These two maps can be extended to the boundary of Ω_γ, i.e., the range R_γ of the curve γ, and the extension remains a homeomorphism. Restricting Φ^+ to R_γ yields a map $\varphi^+ : R_\gamma \to S^1$ (where S^1 is the unit circle) and similarly $\varphi^- : R_\gamma \to S^1$. In particular, the mapping $\varphi = \varphi^- \circ (\varphi^+)^{-1}$ is a homeomorphism of S^1 onto itself. It is almost uniquely defined by γ. In fact Φ^+ and Φ^- are both unique up to composition (on the left) by a Möbius transformation, as given by (1.41), so that φ is unique up to a Möbius transformation applied on the left or on the right. The indeterminacy on the right can be removed by the following normalization; one can constrain Φ^+, which associates two unbounded domains, to transform the point at infinity into itself, and be such that its differential at this point has a positive real part and a vanishing imaginary part. Under this constraint, φ is unique up to the left action of Möbius transformations.

In mathematical terms, we obtain a representation of (smooth) Jordan plane curves by the set of diffeomorphisms of S^1 (denoted $\mathrm{Diff}(S^1)$) modulo the Möbius transformations (denoted $PSL_2(S^1)$), writing

$$\text{2D shapes} \sim \mathrm{Diff}(S^1)/PSL_2(S^1).$$

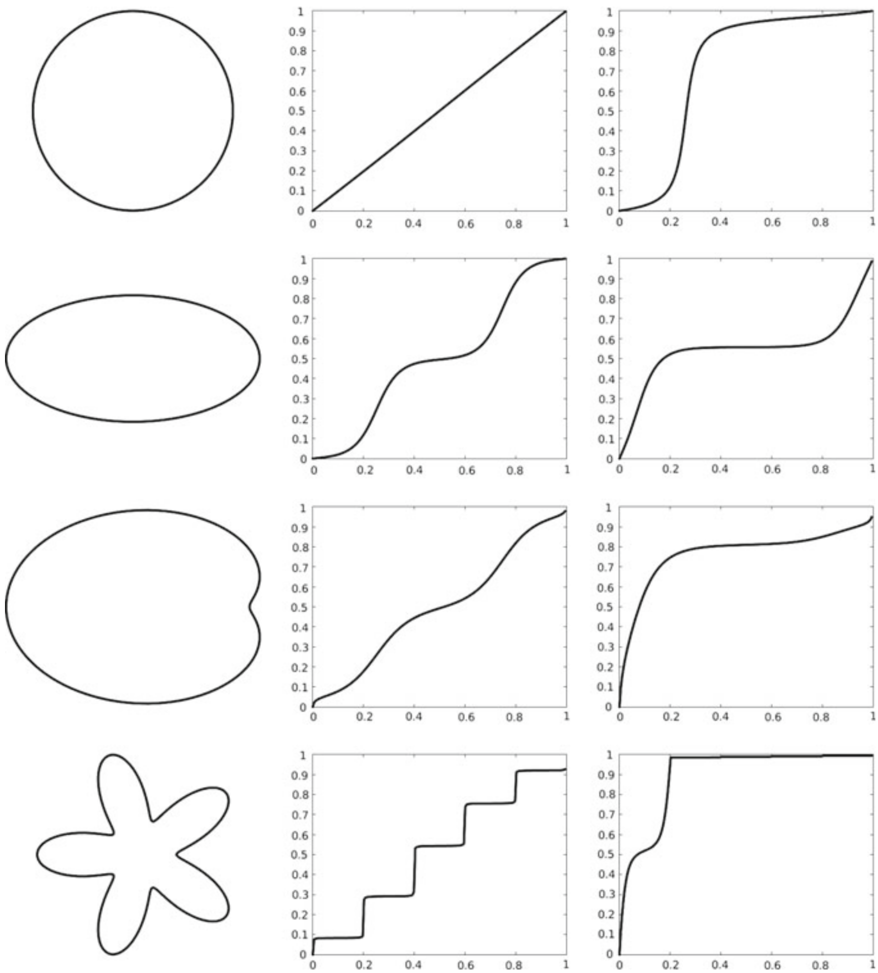

Fig. 1.4 Conformal disparity between the interior and exterior of four planar curves. *First column:* original curves; *second and third columns:* two representations of the curve signature rescaled over the unit interval, related by a Möbius transformation, illustrating the fact that these signatures are equivalent classes of diffeomorphisms of the unit disc

We now describe the two basic operations associated to this equivalence, namely computing this representation from the curve, and retrieving the curve from the representation. The first operation requires computing the trace of the conformal maps of the interior and exterior of the curve. Several algorithms are available to compute conformal maps. The plots provided in Fig. 1.4 were obtained using the Schwarz–Christoffel toolbox developed by T. Driscoll.

The solution to the second problem (going from the representation to the curves) is described in [260, 261] (Fig. 1.5). It is proved in [261] that, if φ is the mapping

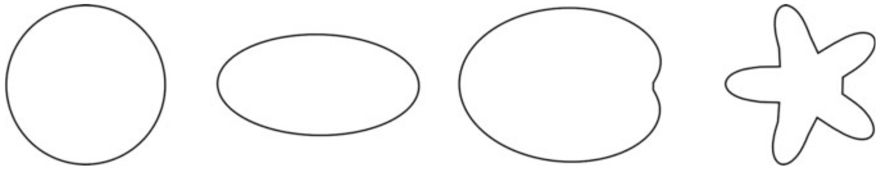

Fig. 1.5 Reconstruction of the curves in Fig. 1.4 from their signatures

above, and $\psi = \varphi^{-1}$, the corresponding shape (defined up to translation, rotation and scaling) can be parametrized as $\theta \mapsto F(\theta) \in \mathbb{C}, \theta \in [0, 2\pi]$, where F is the solution of the integral equation

$$K(F)(\theta) + F(\theta) = e^{i\theta},$$

where $K(F)(\theta) = \int_0^{2\pi} K(\theta, \tilde{\theta}) F(\tilde{\theta}) d\tilde{\theta}$, and the kernel K is given by

$$K(\theta, \tilde{\theta}) = \frac{i}{2} \operatorname{ctn}\left(\frac{\theta - \tilde{\theta}}{2}\right) - \frac{i}{2} \dot{\psi}(\tilde{\theta}) \operatorname{ctn}\left(\frac{\psi(\theta) - \psi(\tilde{\theta})}{2}\right)$$

which has limit $i\ddot{\psi}(\theta)/4\dot{\psi}(\theta)$ as $\theta \to \tilde{\theta}$. The inverse representation can then be computed by solving, after discretization, a linear equation in F. More precisely, assume that $((\theta_k, \varphi_k), i = 0, \ldots, N)$ is a discretization of φ (with $\varphi_N = \varphi_0 + 2\pi$ and $\theta_N = \theta_0 + 2\pi$). Following [261], one then makes the approximation

$$\begin{aligned}\int_0^{2\pi} \operatorname{ctn}\left(\frac{\varphi - \tilde{\varphi}}{2}\right) F(\tilde{\varphi}) d\tilde{\varphi} &\simeq \sum_{k=1}^{N} F_k \int_{\varphi_{k-1}}^{\varphi_k} \operatorname{ctn}\left(\frac{\varphi - \tilde{\varphi}}{2}\right) d\tilde{\varphi} \\ &= 2 \sum_{k=1}^{N} F_k \log \frac{|\sin((\varphi - \varphi_k)/2)|}{|\sin((\varphi - \varphi_{k-1})/2)|},\end{aligned}$$

where we have set $F_k = F((\varphi_k + \varphi_{k-1})/2)$. Similarly, letting $\theta = \psi(\varphi)$,

$$\begin{aligned}\int_0^{2\pi} \operatorname{ctn}\left(\frac{\theta - \psi(\tilde{\varphi})}{2}\right) F(\tilde{\varphi}) \dot{\psi}(\tilde{\varphi}) d\tilde{\varphi} &\simeq \sum_{k=1}^{N} F_k \int_{\theta_{k-1}}^{\theta_k} \operatorname{ctn}\left(\frac{\theta - \tilde{\theta}}{2}\right) d\tilde{\theta} \\ &= 2 \sum_{k=1}^{N} F_k \log \frac{|\sin((\theta - \theta_k)/2)|}{|\sin((\theta - \theta_{k-1})/2)|}.\end{aligned}$$

Letting $\bar{\varphi}_l = (\varphi_l + \varphi_{l-1})/2$ and $\bar{\theta}_l = (\theta_l + \theta_{l-1})/2$, one obtains a discretization $((\bar{\varphi}_l, F_l), l = 1, \ldots, N)$ of F by solving the equation

$$F_l + i\sum_{k=1}^{N} K_{lk}F_k = e^{i\bar{\varphi}_l}, \quad l = 1, \ldots, N$$

with

$$K_{lk} = \log \frac{|\sin((\bar{\varphi}_l - \varphi_k)/2)\,\sin((\bar{\theta}_l - \theta_{k-1})/2)|}{|\sin((\bar{\varphi}_l - \varphi_{k-1})/2)\,\sin((\bar{\theta}_l - \theta_k)/2)|}.$$