PARAMETRIZATIONS AND LOCI

TSOGTGEREL GANTUMUR

Abstract. We consider curvilinear coordinate systems, parametrizations of curves and surfaces, and descriptions of curves and surfaces as loci.

Contents

1. Coordinate systems

Thinking of polar coordinates as an example, a *coordinate system* in $\mathbb{R}^n$ is nothing but a map from one region of $\mathbb{R}^n$ into another region of $\mathbb{R}^n$. More precisely, it is a map $\Phi : D \to R$ from $D \subset \mathbb{R}^n$ into $R \subset \mathbb{R}^n$, where D is called the *domain* of Φ, and R is called the *codomain* or the *range* of Φ. Recall that a *map* $\Phi : D \to R$ assigns an element $Q = \Phi(P)$ of R to every element P of D. The words function, map, transformation, transform, and coordinate change are all interchangeable in this context.

- $\Phi(P)$ is called the *image* of P under Φ, or the *value* of Φ at P.
- The *image of a subset* $S \subset D$ under Φ is defined to be
$$\Phi(S) = \{\Phi(P) : P \in S\}. \tag{1}$$
that is, the collection of all the images $\Phi(P)$ for $P \in S$.
- The image $\Phi(D)$ of the entire domain D is called the *image* of Φ.
- If $\Phi(D) = R$, then we say that the map Φ is *onto*, or *surjective*.
- Given $Q \in \Phi(D)$, the set $\{P \in D : \Phi(P) = Q\} \subset D$ is called the *preimage of* Q *under* Φ, and denoted by $\Phi^{-1}(Q)$.
- If $\Phi(P) = \Phi(P')$ implies $P = P'$ for all $P, P' \in D$, that is, if $\Phi^{-1}(Q)$ consists of a single element for all $Q \in \Phi(D)$, then we say that Φ is *one-to-one*, or *injective*.
- A map that is both injective and surjective is said to be *bijective*.
- If $\Phi : D \to R$ is bijective, then for any $Q \in R$, there exists a unique $P \in D$ such that $\Phi(P) = Q$, and we define the *inverse function* $\Phi^{-1} : R \to D$ by $\Phi^{-1}(Q) = P$.

Note that by replacing the range R by a smaller set, one can always ensure surjectivity. Similarly, on can make any map injective by shrinking the domain D.

Date: 2019/09/21.

Example 1.1. (a) Given any set A, the *identity map* $\mathrm{id} : A \to A$ is defined by $\mathrm{id}(a) = a$ for all $a \in A$. Obviously, the identity map is bijective, and it is its own inverse.

(b) Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ be the set of natural numbers, and let $f : \mathbb{N} \to \mathbb{N}$ be a function defined by $f(n) = 2n$ for $n \in \mathbb{N}$. Then the image of f is the set of all positive even integers $f(\mathbb{N}) = \{2n : n \in \mathbb{N}\}$. It is injective, but not surjective, as the image misses the odd integers.

(c) Consider the function $f(t) = t^2$. With the domain $D = [0, 2]$, the function is injective, and its image is $f(D) = [0, 4]$. On the other hand, the function with the domain $D = [-1, 2]$ has the same image $f(D) = [0, 4]$, and becomes non-injective, as, e.g., $f(-1) = f(1)$.

(d) The polar coordinate change $\Phi(r, \theta) = (r \cos\theta, r \sin\theta)$, considered as a map

$$\Phi : [0, \infty) \times (-\pi, \pi] \to \mathbb{R}^2,$$

is surjective, but not injective, because $\Phi(0, \theta) = (0, 0)$ no matter what θ is. However, shrinking the domain to

$$\Phi : (0, \infty) \times (-\pi, \pi] \to \mathbb{R}^2,$$

ensures injectivity, but loses surjectivity, because the origin is no longer in the image.

Remark 1.2 (Coordinate change)**.** Let $U \subset \mathbb{R}^n$ be an open set, and let $\Phi : U \to \Omega$ be a continuously differentiable and invertible map, whose inverse $\Phi^{-1} : \Omega \to U$ is also continuously differentiable. By default, the points in Ω will be denoted by $x = (x_1, \ldots, x_n)$, and the points in U will be denoted by $y = (y_1, \ldots, y_n)$. We can and should think of y as a new coordinate system in Ω, with $y = \Phi^{-1}(x)$ being the y-coordinates of the point $x \in \Omega$. It will often be convenient to write $y = y(x)$ and $x = x(y)$ instead of $y = \Phi^{-1}(x)$ and $x = \Phi(y)$, respectively. Thus a curve $y = y(t)$ in U corresponds to the curve $x = x(y(t))$ in Ω, and

$$x'(t) = D\Phi(y(t))y'(t), \tag{2}$$

which tells us how the components of a vector should transform under change of coordinates:

$$\alpha = D\Phi(y)\beta, \qquad \text{i.e.,} \qquad \alpha_i = \sum_{k=1}^{n} \frac{\partial x_i}{\partial y_k}(y)\beta_k, \quad (i = 1, \ldots, n), \tag{3}$$

where $y \in U$, $x = x(y)$, $\alpha \in \mathbb{R}^n$ is a vector based at x, and $\beta \in \mathbb{R}^n$ is a vector based at y. In fact, we may think of the columns of $D\Phi(y)$ as a *basis* for vectors based at x, and $\beta_1, \ldots, \beta_n$ as the coordinates of α with respect to this basis. Such a construction, consisting of n vector fields, that form a basis for vectors at each point, is called a *frame*. On the other hand, from the point of view of the domain U, we would have

$$\beta = D\Phi^{-1}(x)\alpha, \qquad \text{i.e.,} \qquad \beta_k = \sum_{i=1}^{n} \frac{\partial y_k}{\partial x_i}(y)\alpha_i, \quad (k = 1, \ldots, n), \tag{4}$$

which means that $\alpha_1 \ldots, \alpha_n$ are the coordinates of β when the columns of $D\Phi^{-1}(x)$ are used as a basis.

Example 1.3 (Polar coordinates)**.** Consider the map $\Phi : \Omega \to U$, defined by

$$\Phi(r, \phi) = \begin{pmatrix} x(r, \theta) \\ y(r, \theta) \end{pmatrix} = \begin{pmatrix} r \cos\theta \\ r \sin\theta \end{pmatrix}, \tag{5}$$

where $\Omega = (0, \infty) \times (-\pi, \pi)$ and $U = \mathbb{R}^2 \setminus \{(x, 0) : x \leq 0\}$. If $\Phi(r, \phi) = (x, y) \in U$, then $x^2 + y^2 = r^2$, or $r = \sqrt{x^2 + y^2} > 0$. This yields $\cos\theta = \frac{x}{r} \in [-1, 1]$, and hence with the function $\arccos t \in [0, \pi]$, we have $\arccos \frac{x}{r} = \phi$ or $\arccos \frac{x}{r} = -\phi$, depending on the sign of θ. In other words, knowing x and r determines θ up to a sign. The sign of ϕ can be determined with the help of the conditions $\sin\theta = \frac{y}{r}$ and $-\pi < \theta < \pi$, because these imply that the sign of θ is the same as the sign of y. To conclude, $(x, y) \in U$ determines $(r, \theta) \in \Omega$ uniquely, i.e., the map Φ is invertible. We can compute

$$D\Phi(r,\theta) = \begin{pmatrix} \partial_r x & \partial_\theta x \\ \partial_r y & \partial_\theta y \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}, \tag{6}$$

and

$$\det D\Phi(r,\theta) = r, \tag{7}$$

which is nonsingular everywhere in Ω. In view of the preceding remark, any vector $\alpha \in \mathbb{R}^2$ based at $q = \Phi(p) \in U$ can be written as

$$\alpha = D\Phi(q)\beta, \tag{8}$$

with $\beta \in \mathbb{R}^2$. If we write $\beta = (\beta_r, \beta_\theta)$ in components, then we have

$$\alpha = \beta_r \hat{e}_r + \beta_\theta \hat{e}_\theta, \tag{9}$$

where

$$\hat{e}_r = \partial_r \Phi(q), \qquad \text{and} \qquad \hat{e}_\theta = \partial_\theta \Phi(q), \tag{10}$$

are the columns of $D\Phi(q)$, forming a frame in U. Note that this frame is an *orthogonal frame*, in the sense that $\hat{e}_r \perp \hat{e}_\theta$ at each point in U. A coordinate system whose associated frame is orthogonal is called an *orthogonal coordinate system.*

Example 1.4 (Elliptic coordinates)**.** Consider the map $(x, y) = \Phi(\rho, \phi)$ defining the *elliptic coordinates* given by

$$\begin{cases} x = \cosh\rho\cos\theta, \\ y = \sinh\rho\sin\theta. \end{cases} \tag{11}$$

For $\rho = 0$, this becomes

$$\begin{cases} x = \cos\theta, \\ y = 0, \end{cases} \tag{12}$$

and hence Φ covers the line segment $-1 \leq x \leq 1$ on the x-axis, twice per period of cosine. For various special values of θ, we have

$$\theta = 0: \quad \begin{cases} x = \cosh\rho, \\ y = 0, \end{cases} \qquad\qquad \theta = \frac{\pi}{2}: \quad \begin{cases} x = 0, \\ y = \sinh\rho, \end{cases}$$

$$\theta = \pi: \quad \begin{cases} x = -\cosh\rho, \\ y = 0, \end{cases} \qquad\qquad \theta = \frac{3\pi}{2}: \quad \begin{cases} x = 0, \\ y = -\sinh\rho. \end{cases}$$

Thus for $\theta = 0$, the map covers the ray $x \geq 1$ on the x-axis twice, and for $\theta = \pi$, it covers the ray $x \leq -1$ on the x-axis twice. Similarly, for $\theta = \frac{\pi}{2}$ and for $\theta = \frac{3\pi}{2}$, it covers respectively the rays $y \geq 0$ and $y \leq 0$ on the y-axis twice. In general, finding $\cos\theta$ and $\sin\theta$ from (62), and invoking the fundamental identity $\cos^2\theta + \sin^2\theta = 1$, we get

$$\frac{x^2}{\cosh^2\rho} + \frac{y^2}{\sinh^2\rho} = 1. \tag{13}$$

Therefore, the curves of constant ρ are ellipses with the semimajor axis $a = \cosh\rho$ and the semiminor axis $b = \sinh\rho$. The focal points of this ellipse are at $x = \pm 1$ and $y = 0$. Now finding $\cosh\rho$ and $\sinh\rho$ from (62), and invoking the identity $\cosh^2\rho - \sinh^2\rho = 1$, we get

$$\frac{x^2}{\cos^2\theta} - \frac{y^2}{\sin^2\theta} = 1. \tag{14}$$

This means that the curves of constant θ are hyperbolae with foci at $x = \pm 1$ on the x-axis. Similarly to polar coordinates, the domain of elliptic coordinates is usually taken to be $\rho > 0$ and $-\pi < \theta < \pi$. The Jacobian matrix of Φ is

$$D\Phi(\rho, \theta) = \begin{pmatrix} \sinh\rho \cos\theta & -\cosh\rho \sin\theta \\ \cosh\rho \sin\theta & \sinh\rho \cos\theta \end{pmatrix}, \tag{15}$$

and its determinant is

$$\det D\Phi(\rho, \theta) = \sinh^2\rho \cos^2\theta + \sinh^2\rho \cos^2\theta = \sinh^2\rho + \sin^2\theta. \tag{16}$$

Note that as the columns of (15) are orthogonal to each other, the elliptic coordinate system is an orthogonal coordinate system.

Example 1.5 (Parabolic coordinates)**.** Consider the map $(x, y) = \Phi(u, v)$ defining the *parabolic coordinates* given by

$$\begin{cases} x = uv, \\ y = \frac{1}{2}(v^2 - u^2). \end{cases} \tag{17}$$

When $u = 0$ and $v = 0$, this becomes

$$u = 0: \quad \begin{cases} x = 0, \\ y = \frac{1}{2}v^2, \end{cases} \qquad\qquad v = 0: \quad \begin{cases} x = 0, \\ y = -\frac{1}{2}u^2, \end{cases}$$

Thus for $u = 0$ and for $v = 0$, the map Φ covers respectively the rays $y \geq 0$ and $y \leq 0$ on the y-axis twice. In general, removing u and subsequently v from (17), we get

$$2y = \frac{x^2}{v^2} - v^2, \qquad 2y = u^2 - \frac{x^2}{u^2}. \tag{18}$$

Therefore, the curves of constant v are "right-side-up" parabolas with their vertices below the origin, while the curves of constant u are "upside-down" parabolas with their vertices above the origin. The domain of parabolic coordinates is usually taken to be $-\infty < u < \infty$ and $v > 0$. The Jacobian matrix of Φ is

$$D\Phi(u, v) = \begin{pmatrix} v & u \\ -u & v \end{pmatrix}, \tag{19}$$

and its determinant is

$$\det D\Phi(u, v) = u^2 + v^2. \tag{20}$$

Since the columns of (19) are orthogonal to each other, the parabolic coordinate system is an orthogonal coordinate system.

Example 1.6 (Cylindrical coordinates)**.** Consider the map $(x, y, z) = \Phi(r, \theta, z)$ defining the *cylindrical coordinates* in 3 dimensions, given by

$$\begin{cases} x = r\cos\theta, \\ y = r\sin\theta, \\ z = z. \end{cases} \tag{21}$$

The Jacobian matrix of Φ is

$$D\Phi(r, \theta, z) = \begin{pmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{22}$$

and its determinant is

$$\det D\Phi(u, v) = r. \tag{23}$$

As the columns of (22) are orthogonal to each other, the cylindrical coordinate system is an orthogonal coordinate system.

Exercise 1.7 (Generalized cylindrical coordinates)**.** One can define a generalized cylindrical coordinate system based on any given coordinate system in the xy-plane. For instance, the *elliptic cylindrical coordinate system* is given by

$$\begin{cases} x = \cosh\rho\cos\theta, \\ y = \cosh\rho\cos\theta, \\ z = z, \end{cases} \tag{24}$$

and the *parabolic cylindrical coordinate system* is given by

$$\begin{cases} x = uv, \\ y = \frac{1}{2}(v^2 - u^2), \\ z = z. \end{cases} \tag{25}$$

Compute the Jacobian matrix and its determinant for each of these coordinate transformations. Are these orthogonal coordinate systems? Depict the coordinate surfaces (i.e., the surfaces with constant ρ, etc.).

Example 1.8 (Spherical coordinates)**.** Consider the map $(x, y, z) = \Phi(r, \theta, \phi)$ defining the *spherical coordinates* in 3 dimensions, given by

$$\begin{cases} x = r\cos\theta\cos\phi, \\ y = r\cos\theta\sin\phi, \\ z = r\sin\theta. \end{cases} \tag{26}$$

The Jacobian matrix of Φ is

$$D\Phi(r, \theta, z) = \begin{pmatrix} \cos\theta\cos\phi & -r\sin\theta\cos\phi & -r\cos\theta\sin\phi \\ \cos\theta\sin\phi & -r\sin\theta\sin\phi & r\cos\theta\cos\phi \\ \sin\theta & r\cos\theta & 0 \end{pmatrix}, \tag{27}$$

and its determinant is

$$\det D\Phi(r, \theta, \phi) = -r^2\cos\theta. \tag{28}$$

Since the columns of (27) are orthogonal to each other, the spherical coordinate system is an orthogonal coordinate system. Note that another convention often used in practice is that one employs $\vartheta = \frac{\pi}{2} - \theta$ instead of θ, and arranges the coordinates as (r, ϕ, ϑ) instead of (r, θ, ϕ). In this case, the transformation is

$$\begin{cases} x = r\cos\phi\sin\vartheta, \\ y = r\sin\phi\sin\vartheta, \\ z = r\cos\vartheta, \end{cases} \tag{29}$$

and the determinant of its Jacobian is

$$\det D\Phi(r, \phi, \vartheta) = r^2\sin\vartheta. \tag{30}$$

Exercise 1.9 (Ellipsoidal and paraboloidal coordinates)**.** Similarly to spherical coordinates, the *ellipsoidal coordinates* (ρ, θ, ϕ) in 3 dimensions is given by

$$\begin{cases} x = \cosh\rho\cos\theta\cos\phi, \\ y = \cosh\rho\cos\theta\sin\phi, \\ z = \sinh\rho\sin\theta, \end{cases} \tag{31}$$

and the *paraboloidal coordinates* (u, v, ϕ) in 3 dimensions is given by

$$\begin{cases} x = uv\cos\phi, \\ y = uv\sin\phi, \\ z = \frac{1}{2}(v^2 - u^2). \end{cases} \tag{32}$$

Depict the coordinate surfaces (i.e., the surfaces with constant u, etc.). What would be convenient ranges of values for the coordinates? Compute the Jacobian matrix and its determinant for each of these coordinate transformations. Are these orthogonal coordinate systems?

2. Smooth curves

Intuitively, a differentiable curves is a curve with the property that the tangent line at each of its points can be defined. To motivate our definition, let us look at some examples.

Example 2.1. (a) Consider the function $\gamma : (0, \pi) \to \mathbb{R}^2$ defined by

$$\gamma(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}. \tag{33}$$

As t varies in the interval $(0, \pi)$, the point $\gamma(t)$ traces out the curve

$$C = [\gamma] \equiv \{\gamma(t) : t \in (0, \pi)\} \subset \mathbb{R}^2, \tag{34}$$

which is a semicircle (without its endpoints). We call γ a *parametrization* of C. If $\gamma(t)$ represents the coordinates of a particle in the plane at time t, then the "instantaneous velocity vector" of the particle at the time moment t is given by

$$\gamma'(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}. \tag{35}$$

Obviously, $\gamma'(t) \neq 0$ for all $t \in (0, \pi)$, and γ is smooth (i.e., infinitely often differentiable). The direction of the velocity vector $\gamma'(t)$ defines the direction of the line tangent to C at the point $p = \gamma(t)$.

(b) Under the substitution $t = s^2$, we obtain a different parametrization of C, given by

$$\eta(s) \equiv \gamma(s^2) = \begin{pmatrix} \cos s^2 \\ \sin s^2 \end{pmatrix}. \tag{36}$$

Note that the parameter s must take values in $(0, \sqrt{\pi})$.

(c) Now consider the function $\xi : [0, \pi] \to \mathbb{R}^2$ defined by

$$\xi(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}. \tag{37}$$

The only difference between the curve $\bar{C} = [\xi]$ and the curve $C = [\gamma]$ from (a) is that $\bar{C}$ contains its endpoints, while C does not. In order to make sense of the velocity vector of ξ at the endpoints of the interval $[0, \pi]$, we may think of ξ as the *restriction* of another function $\tilde{\xi} : (-\varepsilon, \pi + \varepsilon) \to \mathbb{R}^2$ to the interval $[0, \pi]$, where $\varepsilon > 0$ is a small number, and

$$\tilde{\xi}(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \qquad t \in (-\varepsilon, \pi + \varepsilon). \tag{38}$$

We call $\tilde{\xi}$ an *extension* of ξ. With such an extension at hand, the velocity vector of ξ at the endpoints of the interval $[0, \pi]$ can be defined as $\xi'(0) = \tilde{\xi}'(0)$ and $\xi'(\pi) = \tilde{\xi}'(\pi)$.

Exercise 2.2. Let $\delta : \mathbb{R} \to \mathbb{R}^2$ be given by

$$\delta(t) = \begin{pmatrix} (1-\theta(t))t^3 \\ \theta(t)t^3 \end{pmatrix}, \qquad \text{where} \quad \theta(t) = \begin{cases} 1 & \text{for } t > 0, \\ 0 & \text{for } t \leq 0. \end{cases} \tag{39}$$

Show that δ is continuously differentiable in $\mathbb{R}$. Sketch the curve defined by δ. Why there is a "corner" at the origin?

The preceding discussions motivate us to state the following points.

- A curve is a set that admits a parametrization γ.
- In order to have a tangent line at every point of the curve, we require that γ is differentiable and $\gamma' \neq 0$ everywhere.
- By introducing an extension if necessary, we can always assume that γ is defined on some open interval (a, b).

In addition, we require that the tangent lines vary continuously as we traverse along the curve, i.e., we want velocity vector $\gamma'(t)$ to depend continuously on t. This gets rid of the pathological curves such as the graph of the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2 \sin \frac{1}{x}$ for $x \neq 0$ and $f(0) = 0$.

Definition 2.3. A set $L \subset \mathbb{R}^n$ is called an *open curve* if there exists a continuously differentiable function $\gamma : (a, b) \to \mathbb{R}^n$ with $-\infty \leq a < b \leq \infty$, such that $L = \{\gamma(t) : t \in (a, b)\}$ and $\gamma'(t) \neq 0$ for all $t \in (a, b)$. In this setting, γ is called a *parametrization* of L.

Remark 2.4. Strictly speaking, the preceding definition is that of *differentiable* open curves. However, all curves in these notes will be assumed to be differentiable, and we will simply omit the adjective "differentiable."

Example 2.5. (a) Consider the function $\gamma : \mathbb{R} \to \mathbb{R}^2$ defined by

$$\gamma(t) = \begin{pmatrix} t^2 - 1 \\ t(t^2 - 1) \end{pmatrix}. \tag{40}$$

We have

$$\gamma'(t) = \begin{pmatrix} 2t \\ 3t^2 - 1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \text{for} \quad t \in \mathbb{R}, \tag{41}$$

and thus γ defines an open curve in $\mathbb{R}^2$. However, we have $\gamma(-1) = \gamma(1)$, indicating that the curve intersects with itself. At the self-intersection point, we have two possible tangent directions $\gamma'(1) = (2, 2)$ and $\gamma'(-1) = (-2, 2)$. This is not a particularly serious problem, but it is useful to introduce a concept that rules out self-intersecting curves. An idea would be to require injectivity of the parametrization, that is, to require that $\gamma(s) = \gamma(t)$ implies $s = t$.

(b) Consider the unit circle. We may try to parametrize it by

$$\xi(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \qquad t \in (-\varepsilon, 2\pi), \tag{42}$$

but this is not injective, as $\xi(t) = \xi(t + 2\pi)$ for $t \in (-\varepsilon, 0)$. This cannot be avoided if we want a parametrization with an open interval as its domain. A way out would be to "cover" the circle by using multiple parametrizations, meaning that we consider the circle as multiple arcs glued together "nicely."

We will use the notation

$$Q_r(y) = (y_1 - r, y_1 + r) \times \ldots \times (y_n - r, y_n + r), \tag{43}$$

for the n-dimensional cube centred at $y \in \mathbb{R}^n$, with the side length $2r$. This is called an *open* cube, as opposed the the *closed* cube $\bar{Q}_r(y)$. Note that we always have $Q_r(y) \subset \bar{Q}_r(y)$.

Definition 2.6. A set $L \subset \mathbb{R}^n$ is called an *(embedded) curve* if for each $p \in L$, there exists $\delta > 0$ such that $L \cap Q_\delta(p)$ is an open curve admitting an injective parametrization. Recall that a function γ is called injective if $\gamma(s) = \gamma(t)$ implies $s = t$.

Example 2.7. Let us show that the unit circle $C = \{(x, y) \in \mathbb{R}^2 : x^2+y^2 = 1\}$ is a curve in the sense of the preceding definition. Pick an arbitrary $p = (x_*, y_*) \in C$. We consider a few cases. First, assume $y_* > 0$. In this case, we choose $\delta > 0$ so small that $Q_\delta(p) \subset \{(x, y) : -1 < x < 1,\ y > 0\}$, and use the parametrization $\gamma : (x_* - \delta, x_* + \delta) \to \mathbb{R}^2$ defined by $\gamma(t) = (t, \sqrt{1-t^2})$. We can check that this is an injective parametrization of $Q_\delta(p) \cap C$. The second case, where we assume $y_* < 0$, can be treated similarly, by using the parametrization $\gamma(t) = (t, -\sqrt{1-t^2})$. The remaining case is $y = 0$, which can be separated into two subcases: $x_* = 1$ and $x_* = -1$. For $x_* = 1$, we use $\gamma(t) = (\sqrt{1-t^2}, t)$ for $t \in (-1, 1)$, which parametrizes $Q_1(p) \cap C$ injectively. Similarly, for $x_* = -1$, we can use the parametrization $\gamma(t) = (-\sqrt{1-t^2}, t)$.

Definition 2.8. Given a parametrization $\gamma : (a, b) \to \mathbb{R}^n$ of a curve, the *velocity vector* of γ at the point $p = \gamma(t)$ is $\gamma'(t) \in \mathbb{R}^n$.

Remark 2.9. Let $\gamma : (a, b) \to \mathbb{R}^n$ be a parametrization of a curve L, and let $\bar\gamma(s) = \gamma(\phi(s))$ be another parametrization of L, where $\phi : (\bar a, \bar b) \to (a, b)$ is continuously differentiable. One can think of ϕ as a *reparametrization* or a *coordinate change* on the curve. Under this reparametrization, the velocity vector at $p = \bar\gamma(s)$ transforms as

$$\bar\gamma'(s) = \gamma'(\phi(s))\phi'(s). \tag{44}$$

Since $\phi'(s) \in \mathbb{R}$, we see that even though the velocity vector may change during reparametrization, its *direction* stays the same. This direction defines the *tangent line* of L at p, which is an intrinsic property of the curve L independent of parametrization.

3. Hypersurfaces

In this section, we will extend the concept of curves to surfaces, and their higher dimensional generalization, hypersurfaces. We will only be concerned with smooth (or differentiable) surfaces, but in practice, non-smooth objects such as the surface of a cube do not cause much trouble because they can be treated as consisting of a number of smooth pieces.

Example 3.1. The defining characteristic of a curve is that near any of its points, it can be parametrized "nicely" by a single parameter. Intuitively, to parametrize a surface, we need to use two parameters. Let $\Omega = (-1, 1)^2 \subset \mathbb{R}^2$, and let $\Psi : \Omega \to \mathbb{R}^3$ be continuously differentiable in Ω. We imagine that the set $S = \{\Psi(x) : x \in \Omega\}$ is a piece of a surface in $\mathbb{R}^3$, so that Ψ is its parametrization. Consider the 2-dimensional curve $\gamma_\alpha(t) = \alpha t$, $t \in (-1, 1)$, where $\alpha \in \mathbb{R}^2$ is a fixed vector. Under the parametrization Ψ, this curve becomes the 3-dimensional curve $\eta_\alpha(t) = \Psi(\alpha t)$, which is contained in the surface S. The velocity vector of η_α at $\eta_\alpha(0) \in S$ is

$$\eta'_\alpha(0) = D\Psi(0)\gamma'_\alpha(0) = D\Psi(0)\alpha = \frac{\partial \Psi}{\partial x_1}(0)\alpha_1 + \frac{\partial \Psi}{\partial x_2}(0)\alpha_2, \tag{45}$$

where $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2$. If S is a smooth surface, then we expect that the velocity vectors $\eta'_\alpha(0)$ with different $\alpha \in \mathbb{R}^2$ are *not* all aligned to each other. In light of the preceding formula, this means that the columns of $D\Psi(0)$ are expected to be *linearly independent*.

The linear independence condition discussed in the preceding example will appear in the definition of surfaces. Before that, we need to introduce the concept of open sets.

Definition 3.2. A set $\Omega \subset \mathbb{R}^n$ is called *open* if for any $p \in \Omega$, there is $\delta > 0$ such that $Q_\delta(p) \subset \Omega$.

Example 3.3. (a) The square $\Omega = (0,1)^2$ is open, because given any $(x,y) \in \Omega$, we have $(x-\delta, x+\delta) \times (y-\delta, y+\delta) \subset \Omega$ for $\delta = \min\{x, 1-x, y, 1-y\}$.
(b) $\Omega = [0,1)^2$ is *not* open, because taking $p = (0,0) \in \Omega$, there is no $\delta > 0$ with $Q_\delta(p) \subset \Omega$.
(c) The disk $\Omega = \{(x,y) : x^2 + y^2 < 1\}$ is open, because given any $(x,y) \in \Omega$, we have $(x-\delta, x+\delta) \times (y-\delta, y+\delta) \subset \Omega$ for $\delta = \sqrt{1-x^2-y^2}/\sqrt{2}$.

Definition 3.4. A set $M \subset \mathbb{R}^{n+1}$ is called a *hypersurface* if for each $p \in M$, there exist open sets $U \subset \mathbb{R}^n$, $\Omega \subset \mathbb{R}^n$, and a map $\Psi : \Omega \to \mathbb{R}^{n+1}$ such that

(i) $U \cap M = \Psi(\Omega)$ and $p \in U \cap M$.
(ii) Ψ is injective, and continuously differentiable.
(iii) For each $x \in \Omega$, the columns of $D\Psi(x)$ are linearly independent.

If $n = 2$, M is called a *surface*. In this setting, Ψ is called a *local parametrization*, and the triple $(\Psi, \Omega, U \cap M)$ is called a *coordinate chart*. Since Ψ is injective, the inverse $\Psi^{-1} : U \cap M \to \Omega$ exists, and it is called a *local coordinate system* on M.

Example 3.5. Let us introduce local parametrization for the 2-sphere $S^2 = \{y \in \mathbb{R}^3 : y_1^2 + y_2^2 + y_3^2 = 1\}$. Pick an arbitrary point $y^* \in S^2$. We will consider 6 different cases, corresponding to 6 coordinate charts covering S^2. The first case is $y_3^* > 0$. In this case, we set $U = \{y \in \mathbb{R}^3 : y_3 > 0\}$, $\Omega = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 < 1\}$, and $\Psi(x) = (x_1, x_2, \sqrt{1-x_1^2-x_2^2})$. It is easy to see that $U \cap S^2 = \Psi(\Omega)$, and Ψ is injective. Moreover, we have

$$D\Psi(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -x_1/y_3 & -x_2/y_3 \end{pmatrix}, \tag{46}$$

where $y_3 = \sqrt{1-x_1^2-x_2^2}$, which shows that Ψ is continuously differentiable in Ω, and that the columns of $\Psi(x)$ are linearly independent for each $x \in \Omega$. The remaining 5 cases are (ii) $y_3^* < 0$, (iii) $y_3^* = 0$ and $y_2^* > 0$, (iv) $y_3^* = 0$ and $y_2^* < 0$, (v) $y_3^* = y_2^* = 0$ and $y_1^* > 0$, and finally, (vi) $y_3^* = y_2^* = 0$ and $y_1^* < 0$. All these cases can be handled similarly to the first case, with each case corresponding to the positive or the negative half of a coordinate axis, and its associated hemisphere.

Definition 3.6. Given a hypersurface $M \subset \mathbb{R}^{n+1}$ and its point $p \in M$, the *tangent space* of M at p is defined as

$$T_pM = \{\gamma'(0) : \gamma \text{ is a curve on } M \text{ with } \gamma(0) = p\}. \tag{47}$$

Example 3.7. Let us identify the tangent space T_pS^2, for $p = (x,y,z)$, $z > 0$. Consider an arbitrary curve γ on S^2 with $\gamma(0) = p$. Taking the derivative of the relation $\gamma_1(t)^2 + \gamma_2(t)^2 + \gamma_3(t)^2 = 1$ with respect to t, we get

$$\gamma_1(t)\gamma_1'(t) + \gamma_2(t)\gamma_2'(t) + \gamma_3(t)\gamma_3'(t) = 0, \tag{48}$$

and therefore

$$x\gamma_1'(0) + y\gamma_2'(0) + z\gamma_3'(0) = 0. \tag{49}$$

This shows that $T_pS^2 \subset X$, where $X = \{V \in \mathbb{R}^3 : V^\top p = 0\}$. Geometrically, X is the space perpendicular to the vector p. Now let $V = (a,b,c)$ be an arbitrary element of X, meaning that $ax + by + cz = 0$, and let

$$\gamma(t) = \begin{pmatrix} x + at \\ y + bt \\ \sqrt{1-(x+at)^2-(y+bt)^2} \end{pmatrix}. \tag{50}$$

By construction, we have $\gamma(0) = p$. We also have

$$\left.\frac{\mathrm{d}\sqrt{1-(x+at)^2-(y+bt)^2}}{\mathrm{d}t}\right|_{t=0} = \left.\frac{-a(x+at)-b(y+bt)}{\sqrt{1-(x+at)^2-(y+bt)^2}}\right|_{t=0} = \frac{-ax-by}{\sqrt{1-x^2-y^2}}, \qquad (51)$$

and hence

$$\gamma'(0) = \begin{pmatrix} a \\ b \\ -(ax+by)/z \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = V, \qquad (52)$$

where we have used the fact that $ax + by + cz = 0$. This implies $V \in T_pS^2$, and as V is an arbitrary element of X, we have $X \subset T_pS^2$. Since we already established $T_pS^2 \subset X$, we conclude that $T_pS^2 = X \equiv \{V \in \mathbb{R}^3 : V^\top p = 0\}$.

Remark 3.8. Let $M \subset \mathbb{R}^{n+1}$ be a hypersurface, and let $\Psi : \Omega \to M$ be a local parametrization. Then any curve γ on M passing through the point $p \in \Psi(\Omega)$ can be written as

$$\gamma(t) = \Psi(\eta(t)), \qquad (53)$$

with some curve η in Ω. Without loss of generality, assuming that $p = \Psi(q)$ with $q = \eta(0)$, hence that $p = \gamma(0)$, we have

$$\gamma'(0) = D\Psi(q)\eta'(0). \qquad (54)$$

As $\eta'(0) \in \mathbb{R}^n$ can take arbitrary values, we conclude that the tangent space T_pM is spanned by the columns of $D\Psi(q)$:

$$T_pM = \{D\Psi(q)V : V \in \mathbb{R}^n\}. \qquad (55)$$

In fact, the columns of $D\Psi(q)$ form a basis of T_pM, since they are linearly independent.

Example 3.9. The unit 2-sphere S^2 can be described locally (at least for $\theta \neq 0$) by the parametrization

$$\Psi(\theta, \phi) = \begin{pmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ \sin\theta \end{pmatrix}. \qquad (56)$$

Therefore the tangent space T_pS^2 at $p = \Psi(\theta, \phi)$ has the columns of

$$D\Psi(\theta, \phi) = \begin{pmatrix} -\sin\theta\cos\phi & -\cos\theta\sin\phi \\ -\sin\theta\sin\phi & \cos\theta\cos\phi \\ \cos\theta & 0 \end{pmatrix}, \qquad (57)$$

as a basis.

4. Inverse functions: Univariate case

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ and $\alpha \in \mathbb{R}^n$, and consider the equation $f(x) = \alpha$ for the unknown $x \in \mathbb{R}^n$. If this equation has a unique solution for all $\alpha \in U$, where $U \subset \mathbb{R}^n$ is some set, the correspondence $\alpha \mapsto x$ defines the inverse function $f^{-1} : U \to \mathbb{R}^n$ of f on U.

Definition 4.1. Let $K \subset \mathbb{R}^n$ and $U \subset \mathbb{R}^n$, and let $f : K \to U$ be bijective, i.e., for each $\alpha \in U$ there is a unique $x \in K$ such that $f(x) = \alpha$. Then the function $g : U \to K$ defined by $g(f(x)) = x$ for $x \in K$ is called the *inverse function* of f, and denoted by $f^{-1} = g$.

The *inverse function theorem* is the answer to the invertibility question from a differentiable point of view. Suppose that $f : K \to \mathbb{R}^n$ is differentiable at $x^* \in K$, that is, there is $\Lambda \in \mathbb{R}^{n\times n}$ such that

$$f(x) \approx f(x^*) + \Lambda(x - x^*) \qquad \text{as} \quad x \to x^*. \qquad (58)$$

If we define $g : \mathbb{R}^n \to \mathbb{R}^n$ by $g(x) = f(x^*) + \Lambda(x - x^*)$, then $f(x) \approx g(x)$ when x is close to x^*. Moreover, g is invertible if and only if Λ is an invertible matrix. Now, the question is given that g is invertible, can we conclude that f is invertible in a "small" region containing x^*?

Let us consider the case $n = 1$ first. Suppose that $f : (a, b) \to \mathbb{R}$ is differentiable at x^* with $f'(x^*) \neq 0$ for some $x^* \in (a, b)$. Is it true that f is invertible in $(x^* - r, x^* + r)$ for some $r > 0$? However, the following exercise shows that the answer is negative even if f is differentiable everywhere.

Exercise 4.2. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \frac{1}{2}x + x^2 \sin \frac{1}{x} & \text{for } x \neq 0, \\ 0 & \text{for } x = 0. \end{cases} \tag{59}$$

Show that f is differentiable everywhere, and $f'(0) \neq 0$, but f is not invertible in $(-r, r)$ for any $r > 0$. Is f' continuous at 0?

Thus, we need a stronger assumption, and our updated assumption is that $f : (a, b) \to \mathbb{R}$ is *continuously differentiable* in (a, b), and that $f'(x^*) \neq 0$ for some $x^* \in (a, b)$. Under this assumption, it *is* true that f is invertible in $(x^* - r, x^* + r)$ for some $r > 0$.

Theorem 4.3 (Univariate inverse function theorem)**.** *Let $f : (a, b) \to \mathbb{R}$ be continuously differentiable in (a, b), and let $f'(x^*) \neq 0$ for some $x^* \in (a, b)$. Then there exists $r > 0$ such that f is invertible in $I = (x^* - r, x^* + r)$, and for $x \in I$, the inverse function is differentiable at $f(x)$ whenever $f'(x) \neq 0$, with*

$$(f^{-1})'(f(x)) = \frac{1}{f'(x)}. \tag{60}$$

In particular, f^{-1} and $(f^{-1})'$ are continuous at $f(x)$ with $x \in I$ whenever $f'(x) \neq 0$.

Exercise 4.4. The function $f(x) = x^3$ has the inverse $f^{-1}(y) = \sqrt[3]{y}$ for all $y \in \mathbb{R}$, but $f'(0) = 0$. How is this compatible with the inverse function theorem?

Exercise 4.5. Let $f : (a, b) \to (c, d)$ be a continuously differentiable function, whose inverse $f^{-1} : (c, d) \to (a, b)$ is also continuously differentiable. Show that $f'(x) \neq 0$ for all $x \in (a, b)$.

5. The inverse function theorem

The purpose of this section is to extend the inverse function theorem to n-dimensions. Let $Q = (a_1, b_1) \times \ldots \times (a_n, b_n) \subset \mathbb{R}^n$ be a rectangular domain, and let $f : Q \to \mathbb{R}^n$ be differentiable in Q. Then the derivative Df can be considered as a function sending Q to $\mathbb{R}^{n \times n}$, and we will assume that this function is continuous. In other words, we assume that $f : Q \to \mathbb{R}^n$ is *continuously differentiable in* Q. Furthermore, we consider some point $x^* \in Q$, and suppose that the matrix $Df(x^*) \in \mathbb{R}^{n \times n}$ is invertible. Recall that a matrix is invertible (or nonsingular) if and only if its determinant is nonzero.

Theorem 5.1 (Inverse function theorem)**.** *Let $Q \subset \mathbb{R}^n$ be a rectangular domain, and let $f : Q \to \mathbb{R}^n$ be continuously differentiable in Q. Suppose that $Df(x^*)$ is invertible for some $x^* \in Q$. Then there exists $r > 0$ such that f is invertible in $Q_r(x^*)$, and the inverse function is differentiable in $f(Q_r(x^*))$, with*

$$Df^{-1}(f(x)) = (Df(x))^{-1}, \qquad x \in Q_r(x^*). \tag{61}$$

In particular, f^{-1} and Df^{-1} are continuous in $f(Q_r(x^*))$, and moreover, there is $\varepsilon > 0$ such that $Q_\varepsilon(y^*) \subset f(Q_r(x^*))$.

Example 5.2. Consider the map $(u, v) = F(x, y)$ given by

$$\begin{cases} u = x^2 + y^2, \\ v = (x-1)^2 + y^2. \end{cases} \tag{62}$$

Its derivative is

$$DF(x, y) = \begin{pmatrix} 2x & 2y \\ 2(x-1) & 2y \end{pmatrix}. \tag{63}$$

We can also compute

$$\det DF(x, y) = 4y, \tag{64}$$

and for $y \neq 0$,

$$[DF(x, y)]^{-1} = \frac{1}{2y} \begin{pmatrix} y & -y \\ 1-x & x \end{pmatrix}. \tag{65}$$

We want to invert F near $(x^*, y^*) = (0, 1)$. Since $DF(0, 1)$ is invertible, by the inverse function theorem, there is $r > 0$ such that F is invertible in $Q_r(0, 1) = (-r, r) \times (1-r, 1+r)$. Moreover, the derivative of the inverse is given by

$$DF^{-1}(u, v) = DF^{-1}(F(x, y)) = \frac{1}{2y} \begin{pmatrix} y & -y \\ 1-x & x \end{pmatrix}. \tag{66}$$

Example 5.3. Consider the map $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$, defined by $\Psi(r, \phi) = (r \cos \phi, r \sin \phi)$. With $x = x(r, \phi) = r \cos \phi$ and $y = y(r, \phi) = r \sin \phi$ denoting the components of Ψ, the Jacobian of Ψ and its determinant are given by

$$J = \begin{pmatrix} \partial_r x & \partial_\phi x \\ \partial_r y & \partial_\phi y \end{pmatrix} = \begin{pmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{pmatrix}, \qquad \text{and} \qquad \det J = r. \tag{67}$$

Since J is a continuous function of $(r, \phi) \in \mathbb{R}^2$, the map Ψ is differentiable in $\mathbb{R}^2$, with $D\Psi = J$. Moreover, $D\Psi(r, \phi)$ is invertible whenever $r \neq 0$, and

$$(D\Psi)^{-1} = \frac{1}{r} \begin{pmatrix} r \cos \phi & r \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}. \tag{68}$$

By the inverse function theorem, for any $(r^*, \phi^*) \in \mathbb{R}^2$ with $r^* \neq 0$, there exists $\delta > 0$ such that Ψ is invertible in $(r^* - \delta, r^* + \delta) \times (\phi^* - \delta, \phi^* + \delta)$, with

$$D\Psi^{-1}(x, y) = \begin{pmatrix} \partial_x r & \partial_y r \\ \partial_x \phi & \partial_y \phi \end{pmatrix} = \frac{1}{r} \begin{pmatrix} r \cos \phi & r \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}, \tag{69}$$

where $r = r(x, y)$ and $\phi = \phi(x, y)$ are now understood to be the components of Ψ^{-1}. Note that Ψ^{-1} is guaranteed to satisfy $\Psi^{-1}(\Psi(r, \phi)) = (r, \phi)$ for all $(r, \phi) \in (r^*-\delta, r^*+\delta) \times (\phi^*-\delta, \phi^*+\delta)$, and nothing more, so that we would have a potentially different inverse function Ψ^{-1} to Ψ if we change the centre $(r^*, \phi^*) \in \mathbb{R}^2$ and apply the inverse function theorem again. In practice, it does not cause much trouble because we usually work in one such region at a time.

6. The implicit function theorem

In this section, we want to investigate if the equation $g(x, y) = 0$ can be solved as $y = y(x)$. The results will be applied in the next section to derive a convenient criterion to recognize if a set of the form $\{z : \phi(z) = 0\}$ is a smooth curve or a surface. Our approach is based on differentiability, meaning that we fix some point (x_*, y_*), and approximate g as

$$g(x, y) \approx g(x_*, y_*) + \partial_x g(x_*, y_*)(x - x_*) + \partial_y g(x_*, y_*)(y - y_*), \tag{70}$$

for $y \approx y_*$ and $x \approx x_*$. If $\partial_y g(x_*, y_*) \neq 0$, this approximate equation can be solved for y:

$$y - y_* \approx \frac{g(x, y) - g(x_*, y_*) - \partial_x g(x_*, y_*)(x - x_*)}{\partial_y g(x_*, y_*)}. \tag{71}$$

Solving this equation for $g(x,y) \neq g(x_*,y_*)$ would not yield a good approximation, because then $x \approx x_*$ would *not* imply $y \approx y_*$. Thus we put $g(x,y) = g(x_*,y_*) = 0$, and get

$$y - y_* \approx -\frac{\partial_x g(x_*,y_*)}{\partial_y g(x_*,y_*)}(x - x_*). \tag{72}$$

This suggests that the conditions $g(x_*,y_*) = 0$ and $\partial_y g(x_*,y_*) \neq 0$ might be sufficient to solve $g(x,y) = 0$ for a function $y = y(x)$, at least when x is in a small interval containing x_*. In the following remark, we will justify this expectation in full detail.

Remark 6.1. Let $Q_a = (-a,a)^2 \subset \mathbb{R}^2$ be an open square, with $a > 0$, and let $g : Q_a \to \mathbb{R}$ be a continuously differentiable function, satisfying $g(0,0) = 0$ and $\partial_y g(0,0) \neq 0$. We want to find a function $y = h(x)$, defined for $x \in (-\delta,\delta)$ with some $\delta > 0$, such that $g(x,h(x)) = 0$ for all $x \in (-\delta,\delta)$. Note that the point (x_*,y_*) from the previous discussion is now the origin. This is no loss of generality, since we may think of $g(x,y)$ as $\tilde{g}(x_* + x, y_* + y)$ for some function $\tilde{g}$. To proceed further, we introduce the auxiliary map $f : Q_a \to \mathbb{R}^2$, given by $f(x,y) = (x, g(x,y))$ for $(x,y) \in Q_a$. The motivation for considering such a map is that if we can solve $f(x,y) = (\alpha, 0)$ for (x,y) depending on α, then we would have $x = \alpha$ and $g(\alpha, y(\alpha)) = 0$. In order to invert f near the origin, we shall invoke the inverse function theorem. The Jacobian of f is

$$J(x,y) = \begin{pmatrix} 1 & 0 \\ \partial_x g(x,y) & \partial_y g(x,y) \end{pmatrix}, \tag{73}$$

and since g is continuously differentiable, J is continuous in Q_a, and hence we conclude that f is continuously differentiable in Q_a with $Df = J$. At the origin, Df is invertible, and

$$(Df)^{-1} = \begin{pmatrix} 1 & 0 \\ -\partial_x g/\partial_y g & 1/\partial_y g \end{pmatrix}, \tag{74}$$

where all functions are evaluated at the origin $0 \in \mathbb{R}^2$. Now the inverse function theorem guarantees that there exist of $r > 0$ and $f^{-1} : f(Q_r) \to \mathbb{R}^2$, satisfying $f^{-1}(f(x,y)) = (x,y)$ for all $(x,y) \in Q_r$. Note that $f^{-1}(0,0) = (0,0)$. Moreover, $Df(x,y)$ is nonsingular for each $(x,y) \in Q_r$, and f^{-1} is continuously differentiable with $Df^{-1} \circ f = (Df)^{-1}$ in Q_r. If we let $f^{-1}(\alpha,\beta) = (x(\alpha,\beta), y(\alpha,\beta))$, then from $f(f^{-1}(\alpha,\beta)) = (\alpha,\beta)$, we infer that $x(\alpha,\beta) = \alpha$ and $g(\alpha, y(\alpha,\beta)) = \beta$ for $(\alpha,\beta) \in f(Q_r)$. In addition to what we have already mentioned, the inverse function theorem tells us that there is $\delta > 0$ such that $Q_\delta \in f(Q_r)$, implying that we have $g(\alpha, y(\alpha,\beta)) = \beta$ for all $(\alpha,\beta) \in Q_\delta$. In particular, setting $h(\alpha) = y(\alpha, 0)$, we get $g(\alpha, h(\alpha)) = 0$ for all $\alpha \in (-\delta,\delta)$. From $f^{-1}(0,0) = (0,0)$, we get $h(0) = 0$.

The function h we found in the preceding paragraph in fact solves our problem, but our assumptions are strong enough to yield additional results. As a component of f^{-1}, the function $y = y(\alpha,\beta)$ is *continuously differentiable* in Q_δ, and we have

$$Df^{-1} = \begin{pmatrix} 1 & 0 \\ \partial_\alpha y & \partial_\beta y \end{pmatrix}. \tag{75}$$

Comparing this with (74), we get $\partial_\alpha y \circ f = -\partial_x g/\partial_y g$ and $\partial_\beta y \circ f = 1/\partial_y g$. In particular, taking into account that $h'(\alpha) = \partial_\alpha y(\alpha, 0)$, we conclude that

$$h'(x) = -\frac{\partial_x g(x,h(x))}{\partial_y g(x,h(x))}, \qquad \text{for} \quad x \in (-\delta,\delta). \tag{76}$$

Before closing this remark, we make one crucial observation. Fix $x \in (-\delta,\delta)$, and consider $I = \{(x,y) : y \in (-r,r)\}$. The map f sends I to $f(I) = \{(x, g(x,y)) : y \in (-r,r)\} \subset f(Q_r)$. Since f is invertible in Q_r, the only point $(x,y) \in I$ with $g(x,y) = 0$ is $(x, h(x))$. In other words, apart from the curve $\{(x,h(x)) : x \in (-\delta,\delta)\}$, there are *no other* points (x,y) exist in the rectangle $(-\delta,\delta) \times (-r,r)$ satisfying $g(x,y) = 0$.

The preceding remark is the *implicit function theorem* in two dimensions.

Example 6.2. (a) Let us apply the implicit function theorem to the equation $x^2 + y^2 = 1$. Thus we set $g(x, y) = x^2 + y^2 - 1$, and compute $\partial_y g(x, y) = 2y$. This means that as long as (x_*, y_*) satisfies $g(x_*, y_*) = 0$ and $y_* \neq 0$, we can apply the result at the point (x_*, y_*), and infer the existence of $\delta > 0$ and $h : (x_* - \delta, x_* + \delta) \to \mathbb{R}$ such that $g(x, h(x)) = 0$ for all $x \in (x_* - \delta, x_* + \delta)$. We can also compute the derivative of h as

$$h'(x) = -\frac{\partial_x g(x, y)}{\partial_y g(x, y)} = -\frac{2x}{2y} = -\frac{x}{h(x)}, \qquad \text{for} \quad x \in (x_* - \delta, x_* + \delta). \tag{77}$$

The intuitive reason why the case $y_* = 0$ must be excluded is the fact that then the derivative $h'(x_*)$ would have to become infinity.

(b) Let $g(x, y) = y^3 - x$, and let us try to solve $g(x, y) = 0$ for $y = y(x)$ near $(x, y) = (0, 0)$. We have $g(0, 0) = 0$, but $\partial_y g(0, 0) = (3y^2)|_{y=0} = 0$. Therefore the implicit function theorem *cannot* be applied, even though we can explicitly solve the equation as $y(x) = \sqrt[3]{x}$. This has of course to do with the fact that $\sqrt[3]{x}$ is *not* differentiable at $x = 0$.

(c) Let $g(x, y) = x^2 - y^2$, and let us try to solve $g(x, y) = 0$ for $y = y(x)$ near $(x, y) = (0, 0)$. We have $g(0, 0) = 0$, but $\partial_y g(0, 0) = (-2y)|_{y=0} = 0$, and hence the implicit function theorem *cannot* be applied. A close inspection reveals that the solution of $g(x, y) = 0$ is $y = \pm x$, which *cannot* be written as a function $y = y(x)$ near $(x, y) = (0, 0)$.

Let $\Omega \subset \mathbb{R}^n$ and $\Sigma \subset \mathbb{R}^m$ be open sets. Then their *product* $\Omega \times \Sigma \subset \mathbb{R}^{n+m}$ is given by

$$\Omega \times \Sigma = \{(x, y) : x \in \Omega,\ y \in \Sigma\}, \tag{78}$$

where $(x, y) = (x_1, \ldots, x_n, y_1, \ldots, y_m) \in \mathbb{R}^{n+m}$. Let $g : \Omega \times \Sigma \to \mathbb{R}^m$ be a differentiable function. The value of g at $(x, y) \in \Omega \times \Sigma$ is denoted by $g(x, y) \in \mathbb{R}^m$. For any fixed $x \in \Omega$, the correspondence $y \mapsto g(x, y)$ is a function of $y \in \Sigma$, and its derivative will be denoted by $D_y g$ Similarly, we can introduce $D_x g$. In the following, sometimes it will be convenient to specify the dimension of a cube in the notation, as in $Q_r^n(a) = (a - r, a + r)^n \subset \mathbb{R}^n$.

Theorem 6.3. *Let $\Omega \subset \mathbb{R}^n$ and $\Sigma \subset \mathbb{R}^m$ be open sets, and let $g : \Omega \times \Sigma \to \mathbb{R}^m$ be continuously differentiable. Suppose that $(a, b) \in \Omega \times \Sigma$ satisfies $g(a, b) = 0$, and that $D_y g(a, b)$ is nonsingular. Then there exist $\delta > 0$ and $h : Q_\delta^n(a) \to \mathbb{R}^m$ with $h(a) = b$, such that $g(x, h(x)) = 0$ for all $x \in Q_\delta^n(a)$. Moreover, h is continuously differentiable in $Q_\delta^n(a)$, with*

$$Dh(x) = -(D_y g(x, h(x)))^{-1} D_x g(x, h(x)), \qquad x \in Q_\delta^n(a), \tag{79}$$

and we have $\{(x, h(x)) : x \in Q_\delta^n(a)\} = \{(x, y) \in Q_\delta^n(a) \times Q_r^m(b) : g(x, y) = 0\}$ for some $r > 0$.

Example 6.4. (a) Consider the equation

$$g(x, y, z) \equiv \sin(xy + z) + \log(yz^2) = 0. \tag{80}$$

The triple $p = (x, y, z) = (1, 1, -1)$ is a solution: $g(1, 1, -1) = 0$, and g is continuously differentiable in the open set $\{(x, y, z) : x \in \mathbb{R},\ y > 0,\ z < 0\}$. Can we express z as a function of x and y near p? This is exactly the kind of question that could be answered by the implicit function theorem. We have

$$\partial_z g(x, y, z) = \cos(xy + z) + \frac{2z}{yz^2} = \cos(xy + z) + \frac{2}{yz}, \tag{81}$$

and hence

$$\partial_z g(1, 1, -1) = \cos 0 - 2 = -1 \neq 0. \tag{82}$$

Thus there exist $\delta > 0$ and a continuously differentiable function $h : Q_\delta^2(1, 1) \to \mathbb{R}$ such that $g(x, y, h(x, y)) = 0$ for all $(x, y) \in Q_\delta^2(1, 1)$.

(b) Can we solve

$$\begin{aligned} xu^2 + yzv + x^2z &= 3, \\ yv^5 + zu^2 - xv &= 1, \end{aligned} \tag{83}$$

for (u, v) near $(1, 1)$ as a function of (x, y, z) near $(1, 1, 1)$? We can formulate the problem as solving $g(\alpha, \beta) = 0$ for $\beta = \beta(\alpha)$, where $\alpha = (x, y, z)$, $\beta = (u, v)$, and

$$g(\alpha, \beta) = g(x, y, z, u, v) = \begin{pmatrix} xu^2 + yxv + x^2z - 3 \\ yv^5 + 2zu - v^2 - 2 \end{pmatrix}. \tag{84}$$

Obviously, g is continuously differentiable in $\mathbb{R}^5$, and $g(1, 1, 1, 1, 1) = 0$. We can compute the relevant derivative as

$$D_\beta g(\alpha, \beta) = \begin{pmatrix} 2xu & yz \\ 2zu & 5yv^4 - x \end{pmatrix}. \tag{85}$$

so that the matrix

$$D_\beta g(1, 1, 1, 1, 1) = \begin{pmatrix} 2 & 1 \\ 2 & 4 \end{pmatrix}, \tag{86}$$

is invertible. Thus there exist $\delta > 0$ and $h : Q^3_\delta(1, 1, 1) \to \mathbb{R}^2$ continuously differentiable, such that $g(\alpha, h(\alpha)) = 0$ for all $\alpha \in Q^3_\delta(1, 1, 1)$.

7. Level curves and surfaces

With the implicit function theorem at hand, we are now ready to answer the question when the equation $\phi(x) = 0$ defines a smooth curve or surface. We discuss the two dimensional case first, as it involves most of the main ideas. Thus let $g : \mathbb{R}^2 \to \mathbb{R}$ be a continuously differentiable function, and let $L = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 0\}$. We assume that $Dg(x, y) \neq 0$ for all $(x, y) \in L$, that is, at least one component of $Dg(x, y) \in \mathbb{R}^{1\times 2}$ is nonzero whenever (x, y) satisfies $g(x, y) = 0$. Under these assumptions, we want to show that L is a differentiable curve.

Pick an arbitrary point $(\bar{x}, \bar{y}) \in L$, and we shall build a parametrization of L near this point. We will consider the cases $\partial_y g(\bar{x}, \bar{y}) \neq 0$ and $\partial_y g(\bar{x}, \bar{y}) = 0$ separately.

Case 1. We assume that $\partial_y g(\bar{x}, \bar{y}) \neq 0$. Then the implicit function theorem guarantees that we can write y in terms of x at least when x is near $\bar{x}$. Namely, there exist $\delta > 0$ and a continuously differentiable function $h : I \to \mathbb{R}$ such that $g(x, h(x)) = 0$ for all $x \in I$, with $I = (\bar{x} - \delta, \bar{x} + \delta)$. Moreover, apart from the curve $\{(x, h(x)) : x \in I\}$, there are *no other* points (x, y) exist in the rectangle $U = I \times (\bar{y} - r, \bar{y} + r)$ satisfying $g(x, y) = 0$, where $r > 0$ is some constant. Therefore, with $\Psi : I \to \mathbb{R}^2$ defined by $\Psi(t) = (t, h(t))$, conditions (i) and (ii) are satisfied. Moreover, we have $\Psi'(t) = (1, h'(t)) \neq 0$.

Case 2. We assume that $\partial_y g(\bar{x}, \bar{y}) = 0$. In this case, we must have $\partial_x g(\bar{x}, \bar{y}) \neq 0$, because $Dg(\bar{x}, \bar{y}) \neq 0$, and hence the preceding arguments apply with the roles of x and y switched.

We have proved the following result.

Theorem 7.1 (Level curve theorem)**.** *Let $A \subset \mathbb{R}^2$ be an open set, and let $g : A \to \mathbb{R}$ be a continuously differentiable function. Suppose that $\nabla g(x, y) \neq 0$ whenever $(x, y) \in A$ satisfies $g(x, y) = 0$. Then the set $L = \{(x, y) \in A : g(x, y) = 0\}$ is a differentiable curve.*

Example 7.2. Consider $g(x, y) = x^2 + y^2 - \rho$, where $\rho \in \mathbb{R}$ is some constant. This function is continuously differentiable in $\mathbb{R}^2$, with $\nabla g(x, y) = (2x, 2y) \in \mathbb{R}^{1\times 2}$. We see that $\nabla g(x, y) = 0$ if and only if $x = y = 0$. Now let $L = \{(x, y) : g(x, y) = 0\}$.

- If $\rho = 0$, then L is a single point $\{(0, 0)\}$, and $\nabla g = 0$ there. Hence the level curve theorem cannot be applied.

- If $\rho < 0$, then $L = \varnothing$. Since $(0,0) \notin L$, we have $\nabla g(x,y) \neq 0$ for all $(x,y) \in L$. Thus the level curve theorem *can* be applied, to conclude that L is a curve. This example suggests that it is always a good idea to explicitly check if the equations define a nonempty set, in order not to waste efforts working with an empty set.
- If $\rho > 0$, then L is nonempty, because, for example, we have $(0, \sqrt{\rho}) \in L$. Moreover, we have $(0,0) \notin L$, implying that $\nabla g(x,y) \neq 0$ for all $(x,y) \in L$. This means that L is a differentiable curve.

Example 7.3. Consider the equation

$$(x^2 + y^2)^2 = x^2 - y^2, \tag{87}$$

for Bernoulli's lemniscate. We may think of it as the 0-level set of the function

$$\phi(x,y) = (x^2 + y^2)^2 - x^2 + y^2, \tag{88}$$

whose gradient is

$$\nabla\phi(x,y) = (4x(x^2 + y^2) - 2x, 4x(x^2 + y^2) + 2y). \tag{89}$$

Let us identify the points at which this gradient vector vanishes. First, for $x = 0$, we have $\nabla\phi(0,y) = (0, 2y)$, and so $y = 0$ is the only point on the y-axis with $\nabla\phi = 0$. Next, assuming $x \neq 0$, from the condition $4x(x^2 + y^2) - 2x = 0$ for the first component, we get

$$x^2 + y^2 = \tfrac{1}{2}, \tag{90}$$

and substituting this into the second component yields the condition

$$2x + 2y = 0. \tag{91}$$

Thus, the points at which $\nabla\phi = 0$ are $(0,0)$, $(\frac{1}{2}, -\frac{1}{2})$, and $(-\frac{1}{2}, \frac{1}{2})$. Among these, only $(0,0)$ is on the lemniscate, because

$$\phi(\pm\tfrac{1}{2}, \mp\tfrac{1}{2}) = \tfrac{1}{4} \neq 0. \tag{92}$$

To conclude, Bernoulli's lemniscate can be described by a smooth parametrization except near the point $(0,0)$. The reason for this is that $(0,0)$ is a self-intersection point.

Theorem 7.4 (Level surface theorem)**.** *Let $A \subset \mathbb{R}^n$ be an open set, and let $\phi : A \to \mathbb{R}$ be a continuously differentiable function. Suppose that $\nabla\phi(x) \neq 0$ whenever $x \in A$ satisfies $\phi(x) = 0$. Then the set $M = \{x \in A : \phi(x) = 0\}$ is a hypersurface in $\mathbb{R}^n$.*

Example 7.5. Let $a \in \mathbb{R}^n$ be a nonzero vector, and let

$$M = \{x \in \mathbb{R}^n : a_1x_1^2 + a_2x_2^2 + \ldots + a_nx_n^2 = 1\}. \tag{93}$$

We would like to show that M is a hypersurface. Thus we let

$$\phi(x) = a_1x_1^2 + a_2x_2^2 + \ldots + a_nx_n^2 - 1, \tag{94}$$

so that $M = \{\phi = 0\}$, and compute

$$\nabla\phi(x) = (2a_1x_1, 2a_2x_2, \ldots, 2a_nx_n). \tag{95}$$

Since a is a nonzero vector, $\nabla\phi(x) = 0$ if and only if $x = 0$. We know that $0 \notin M$, because $\phi(0) = -1$, and hence $\nabla\phi(x) \neq 0$ for all $x \in M$. Then the level surface theorem implies that M is a hypersurface in $\mathbb{R}^n$.

Remark 7.6. Let $M \subset \mathbb{R}^{n+1}$ be a hypersurface, described by the equation $\phi(x) = 0$, where $\phi : \mathbb{R}^{n+1} \to \mathbb{R}$ is a continuously differentiable function with $\nabla\phi(x) \neq 0$ for all $x \in M$. Then for any curve γ on M, we have $\phi(\gamma(t)) = 0$ for all t, Putting $p = \gamma(0)$, differentiation gives

$$\nabla\phi(p)\gamma'(0) = 0. \tag{96}$$

By definition, the vector $\gamma'(0)$ represents an arbitrary vector in the tangent space T_pM, and hence we conclude that

$$T_pM \subset \ker \nabla\phi(p). \tag{97}$$

Since $\dim\ker\nabla\phi(p) \leq n-1$ and $\dim T_pM = n-1$, we conclude that

$$T_pM = \ker \nabla\phi(p). \tag{98}$$

Example 7.7. Let $S^{n-1} = \{x \in \mathbb{R}^n : x^\top x = 1\}$, which can be written as $\{\phi(x) = 0\}$ with $\phi(x) = x^\top x - 1$. Then $\nabla\phi(x) = 2x^\top$, and hence

$$T_xS^{n-1} = \ker\nabla\phi(x) = \{V \in \mathbb{R}^n : x^\top V = 0\}. \tag{99}$$

8. Locus curves

If ϕ is a scalar function, then $\phi(x) = 0$ is just "one condition on x," so that the equation $\phi(x) = 0$ cannot "reduce the dimension of the underlying space by more than one." That is to say, the equation $\phi(x) = 0$ might be able to describe a surface in $\mathbb{R}^3$, but never a curve in $\mathbb{R}^3$, because in the latter case we would need to "reduce the dimension of $\mathbb{R}^3$ by 2" to get a curve, which is a 1-dimensional object. To describe a curve in $\mathbb{R}^3$, we need a vector valued function of the form $\phi : \mathbb{R}^3 \to \mathbb{R}^2$.

Theorem 8.1 (Locus curve theorem)**.** *Let $A \subset \mathbb{R}^3$ be an open set, and let $\phi : A \to \mathbb{R}^2$ be a continuously differentiable function. Suppose that for each $x \in A$ satisfying $\phi(x) = 0$, there is a 2×2 submatrix of $D\phi(x)$ that is nonsingular. Then the set $M = \{x \in A : \phi(x) = 0\}$ is a differentiable curve.*

Example 8.2. Consider the intersection L between the unit sphere

$$x^2 + y^2 + z^2 = 1, \tag{100}$$

and the plane

$$x + y + z = 0. \tag{101}$$

This can be written as the 0-locus of the map $\phi : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$\phi(x, y, z) = \begin{pmatrix} x^2 + y^2 + z^2 - 1 \\ x + y + z \end{pmatrix}. \tag{102}$$

Its derivative is

$$D\phi(x, y, z) = \begin{pmatrix} 2x & 2y & 2z \\ 1 & 1 & 1 \end{pmatrix}, \tag{103}$$

and the determinants of two of the 2×2 submatrices are

$$\det\begin{pmatrix} 2x & 2y \\ 1 & 1 \end{pmatrix} = 2(x - y), \qquad \det\begin{pmatrix} 2y & 2z \\ 1 & 1 \end{pmatrix} = 2(y - z). \tag{104}$$

There is no point on L satisfying $x = y = z$, meaning that if $x = y$ then $y \neq z$. Hence at least one of the aforementioned determinants is nonzero, and so L is a smooth curve. The tangent space of L can be found as the intersection of two tangent spaces, that is, as the collection of vectors orthogonal to both the gradient of $x^2 + y^2 + z^2 - 1$ and the gradient of $x + y + z$.