Chapter 2

Derivatives, Series, and Vector Fields

2.1 The Derivative of a Function Between Normed Vector Spaces

In most cases, $E = \mathbb{R}^n$ and $F = \mathbb{R}^m$. However, it is sometimes necessary to allow E and F to be infinite dimensional.

Let E and F be two *normed vector spaces*, let $A \subseteq E$ be some open subset of E, and let $a \in A$ be some element of A. Even though a is a vector, we may also call it a point.

The idea behind the derivative of the function f at a is that it is a *linear approximation* of f in a small open set around a.

The difficulty is to make sense of the quotient

$$\frac{f(a+h)-f(a)}{h}$$

where h is a vector.

We circumvent this difficulty in two stages.

A first possibility is to consider the *directional derivative* with respect to a vector $u \neq 0$ in E.

We can consider the vector $f(a+tu)-f(a)$, where $t \in \mathbb{R}$ (or $t \in \mathbb{C}$). Now,

$$\frac{f(a+tu)-f(a)}{t}$$

makes sense.

The idea is that the map from (r, s) to F given by

$$t \mapsto f(a + tu)$$

defines a curve (segment) in F, and the directional derivative $\mathrm{D}_u f(a)$ defines the direction of the tangent line at a to this curve. See Figure 2.1.

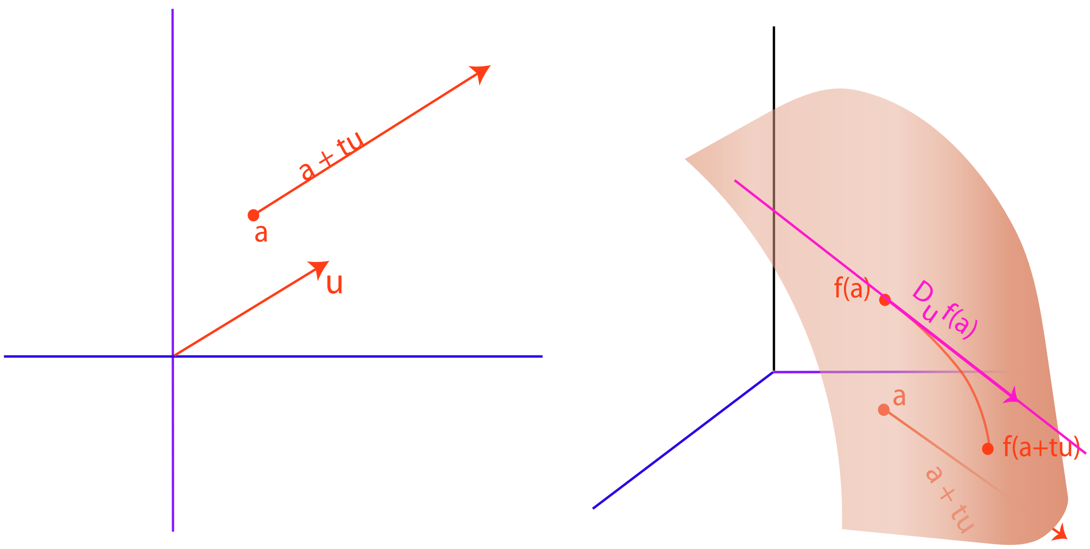

Figure 2.1: Let $f\colon \mathbb{R}^2 \to \mathbb{R}$. The graph of f is the peach surface in $\mathbb{R}^3$, and $t \mapsto f(a + tu)$ is the embedded orange curve connecting $f(a)$ to $f(a + tu)$. Then $\mathrm{D}_u f(a)$ is the slope of the pink tangent line in the direction of u.

Definition 2.1. Let E and F be two normed spaces, let A be a nonempty open subset of E, and let $f\colon A \to F$ be any function. For any $a \in A$, for any $u \neq 0$ in E, the *directional derivative of f at a w.r.t. the vector u*, denoted by $\mathrm{D}_u f(a)$, is the limit (if it exists)

$$\lim_{t \to 0,\, t \in U} \frac{f(a + tu) - f(a)}{t},$$

where $U = \{t \in \mathbb{R} \mid a + tu \in A,\, t \neq 0\}$
(or $U = \{t \in \mathbb{C} \mid a + tu \in A,\, t \neq 0\}$).

Since the map $t \mapsto a + tu$ is continuous, and since $A - \{a\}$ is open, the inverse image U of $A - \{a\}$ under the above map is open, and the definition of the limit in Definition 2.1 makes sense.

The directional derivative is sometimes called the *Gâteaux derivative*.

In the special case where $E = \mathbb{R}$ and $F = \mathbb{R}$, and we let $u = 1$ (i.e., the real number 1, viewed as a vector), it is immediately verified that $\mathrm{D}_1 f(a) = f'(a)$.

When $E = \mathbb{R}$ (or $E = \mathbb{C}$) and F is any normed vector space, the derivative $\mathrm{D}_1 f(a)$, also denoted by $f'(a)$, provides a suitable generalization of the notion of derivative.

However, when E has dimension ≥ 2, directional derivatives present a serious problem, which is that their definition is not sufficiently uniform.

A function can have all directional derivatives at a, and yet not be continuous at a. Two functions may have all directional derivatives in some open sets, and yet their composition may not.

Thus, we introduce a more uniform notion.

Given two normed vector spaces E and F, recall that a linear map $f\colon E \to F$ is *continuous* iff there is some constant $C \geq 0$ such that

$$\|f(u)\| \leq C \|u\| \quad \text{for all } u \in E.$$

The set of *continuous* linear maps from E to F is a vector space denoted $\mathcal{L}(E; F)$, and the set of *all* linear maps from E to F is a vector space denoted by $\mathrm{Hom}(E, F)$.

If E is finite-dimensional, then $\mathcal{L}(E; F) = \mathrm{Hom}(E, F)$, but if E is infinite-dimensional, then there may be linear maps that are *not* continuous, and in general the space $\mathcal{L}(E; F)$ is a proper subspace of $\mathrm{Hom}(E, F)$.

Definition 2.2. Let E and F be two normed spaces, let A be a nonempty open subset of E, and let $f\colon A \to F$ be any function. For any $a \in A$, we say that f is *differentiable at $a \in A$* if there is a *linear continuous* map, $L\colon E \to F$, and a function, $\epsilon(h)$, such that

$$f(a + h) = f(a) + L(h) + \epsilon(h)\|h\|$$

for every $a + h \in A$, where

$$\lim_{h\to 0,\; h\in U} \epsilon(h) = 0,$$

with $U = \{h \in E \mid a + h \in A,\; h \neq 0\}$.

The linear map L is denoted by $\mathrm{D}f(a)$, or $\mathrm{D}f_a$, or $df(a)$, or df_a, or $f'(a)$, and it is called the *Fréchet derivative*, or *derivative*, or *total derivative*, or *total differential*, or *differential*, of f at a. See Figure 2.2.

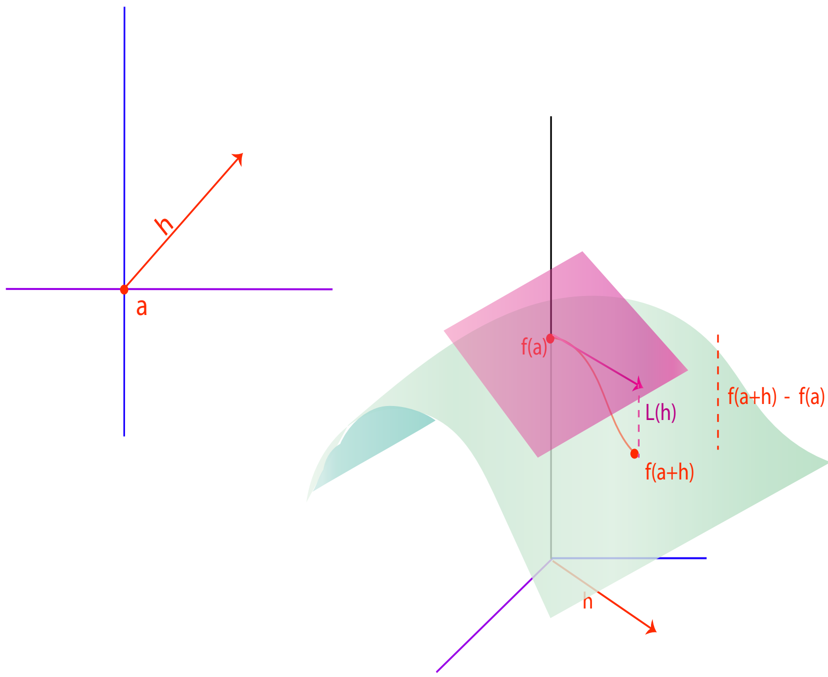

Figure 2.2: Let $f\colon \mathbb{R}^2 \to \mathbb{R}$. The graph of f is the green surface in $\mathbb{R}^3$. The linear map $L = \mathrm{D}f(a)$ is the pink tangent plane. For any vector $h \in \mathbb{R}^2$, $L(h)$ is approximately equal to $f(a+h) - f(a)$. Note that $L(h)$ is also the direction tangent to the curve $t \mapsto f(a+tu)$.

Since the map $h \mapsto a + h$ from E to E is continuous, and since A is open in E, the inverse image U of $A - \{a\}$ under the above map is open in E, and it makes sense to say that

$$\lim_{h \to 0,\, h \in U} \epsilon(h) = 0.$$

Note that for every $h \in U$, since $h \neq 0$, $\epsilon(h)$ is uniquely determined since

$$\epsilon(h) = \frac{f(a + h) - f(a) - L(h)}{\|h\|},$$

and the value $\epsilon(0)$ plays absolutely no role in this definition.

It does no harm to assume that $\epsilon(0) = 0$, and we will assume this from now on.

Note that *the continuous linear map L is unique*, if it exists.

The following proposition shows that our new definition is consistent with the definition of the directional derivative.

Proposition 2.1. *Let E and F be two normed spaces, let A be a nonempty open subset of E, and let $f\colon A \to F$ be any function. For any $a \in A$, if $\mathrm{D}f(a)$ is defined, then f is continuous at a and f has a directional derivative $\mathrm{D}_u f(a)$ for every $u \neq 0$ in E. Furthermore,*

$$\mathrm{D}_u f(a) = \mathrm{D}f(a)(u).$$

The uniqueness of L follows from Proposition 2.1.

Also, when E is of finite dimension, it is easily shown that every linear map is continuous, and this assumption is then redundant.

Example 2.1. Consider the map $f \colon \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ given by

$$f(A) = A^\top A - I,$$

where $\mathrm{M}_n(\mathbb{R})$ denotes the vector space of all $n \times n$ matrices with real entries equipped with any matrix norm, since they are all equivalent;

for example, pick the Frobenius norm $\|A\|_F = \sqrt{\mathrm{tr}(A^\top A)}$.

We claim that

$$Df(A)(H) = A^\top H + H^\top A, \quad \text{for all } A \text{ and } H \text{ in } \mathrm{M}_n(\mathbb{R}).$$

We have

$$\begin{aligned} & f(A+H) - f(A) - (A^\top H + H^\top A) \\ & = (A+H)^\top (A+H) - I - (A^\top A - I) - A^\top H - H^\top A \\ & = A^\top A + A^\top H + H^\top A + H^\top H - A^\top A - A^\top H - H^\top A \\ & = H^\top H. \end{aligned}$$

It follows that

$$\epsilon(H) = \frac{f(A+H) - f(A) - (A^\top H + H^\top A)}{\|H\|} = \frac{H^\top H}{\|H\|},$$

and since our norm is the Frobenius norm,

$$\|\epsilon(H)\| = \left\| \frac{H^\top H}{\|H\|} \right\| \leq \frac{\left\|H^\top\right\| \|H\|}{\|H\|} = \left\|H^\top\right\| = \|H\|,$$

so

$$\lim_{H \mapsto 0} \epsilon(H) = 0,$$

and we conclude that

$$Df(A)(H) = A^\top H + H^\top A.$$

If $\mathrm{D}f(a)$ exists for every $a \in A$, we get a map

$$\mathrm{D}f\colon A \to \mathcal{L}(E; F),$$

called the *derivative of f on A*, and also denoted by df.

Here, $\mathcal{L}(E; F)$ denotes the vector space of continuous linear maps from E to F.

We now consider a number of standard results about derivatives.

A function $f\colon E \to F$ is said to be *affine* if there is some linear map $\overrightarrow{f} : E \to F$ and some fixed vector $c \in F$, such that

$$f(u) = \overrightarrow{f}(u) + c$$

for all $u \in E$.

We call $\overrightarrow{f}$ the *linear map associated with f*.

Proposition 2.2. *Given two normed spaces E and F, if $f\colon E \to F$ is a constant function, then*

$$\mathrm{D}f(a) = 0, \quad \textit{for every } a \in E.$$

If $f\colon E \to F$ is a continuous affine map, then

$$\mathrm{D}f(a) = \overrightarrow{f}, \quad \textit{for every } a \in E,$$

where $\overrightarrow{f}$ is the linear map associated with f.

In particular, if $f\colon E \to F$ is a continuous linear map, then

$$Df(a) = f, \quad \textit{for every } a \in E.$$

Proposition 2.3. *Given a normed space E and a normed vector space F, for any two functions $f, g \colon E \to F$, for every $a \in E$, if $\mathrm{D}f(a)$ and $\mathrm{D}g(a)$ exist, then $\mathrm{D}(f+g)(a)$ and $\mathrm{D}(\lambda f)(a)$ exist, and*

$$\begin{aligned} \mathrm{D}(f+g)(a) &= \mathrm{D}f(a) + \mathrm{D}g(a), \\ \mathrm{D}(\lambda f)(a) &= \lambda \mathrm{D}f(a). \end{aligned}$$

Given two normed vector spaces $(E_1, \| \|_1)$ and $(E_2, \| \|_2)$, there are three natural and equivalent norms that can be used to make $E_1 \times E_2$ into a normed vector space:

1. $\|(u_1, u_2)\|_1 = \|u_1\|_1 + \|u_2\|_2$.
2. $\|(u_1, u_2)\|_2 = (\|u_1\|_1^2 + \|u_2\|_2^2)^{1/2}$.
3. $\|(u_1, u_2)\|_\infty = \max(\|u_1\|_1, \|u_2\|_2)$.

We usually pick the first norm.

If E_1, E_2, and F are three normed vector spaces, recall that a bilinear map $f \colon E_1 \times E_2 \to F$ is *continuous* iff there is some constant $C \geq 0$ such that

$$\|f(u_1, u_2)\| \leq C \|u_1\|_1 \|u_2\|_2 \text{ for all } u_1 \in E_1 \text{ and } u_2 \in E_2.$$

Proposition 2.4. *Given three normed vector spaces E_1, E_2, and F, for any continuous bilinear map $f\colon E_1 \times E_2 \to F$, for every $(a, b) \in E_1 \times E_2$, $\mathrm{D}f(a, b)$ exists, and for every $u \in E_1$ and $v \in E_2$,*

$$\mathrm{D}f(a, b)(u, v) = f(u, b) + f(a, v).$$

We now state the very useful *chain rule*.

Theorem 2.5. *Given three normed spaces E, F, and G, let A be an open set in E, and let B an open set in F. For any functions $f\colon A \to F$ and $g\colon B \to G$, such that $f(A) \subseteq B$, for any $a \in A$, if $\mathrm{D}f(a)$ exists and $\mathrm{D}g(f(a))$ exists, then $\mathrm{D}(g \circ f)(a)$ exists, and*

$$\mathrm{D}(g \circ f)(a) = \mathrm{D}g(f(a)) \circ \mathrm{D}f(a).$$

Theorem 2.5 has many interesting consequences. We mention one corollary.

Proposition 2.6. *Given two normed spaces E and F, let A be some open subset in E, let B be some open subset in F, let $f\colon A \to B$ be a bijection from A to B, and assume that $\mathrm{D}f$ exists on A and that $\mathrm{D}f^{-1}$ exists on B. Then, for every $a \in A$,*

$$\mathrm{D}f^{-1}(f(a)) = (\mathrm{D}f(a))^{-1}.$$

Proposition 2.6 has the remarkable consequence that the two vector spaces E and F have the same dimension.

In other words, a *local property*, the existence of a bijection f between an open set A of E and an open set B of F, such that f is differentiable on A and f^{-1} is differentiable on B, implies a *global property*, that the two vector spaces E and F have the same dimension.

Let us mention two more rules about derivatives that are used all the time.

Let $\iota\colon \mathbf{GL}(n, \mathbb{C}) \to \mathrm{M}_n(\mathbb{C})$ be the function (inversion) defined on invertible $n \times n$ matrices by $\iota(A) = A^{-1}$. Then we have

$$d\iota_A(H) = -A^{-1}HA^{-1},$$

for all $A \in \mathbf{GL}(n, \mathbb{C})$ and for all $H \in \mathrm{M}_n(\mathbb{C})$.

Next, if $f\colon \mathrm{M}_n(\mathbb{C}) \to \mathrm{M}_n(\mathbb{C})$ and $g\colon \mathrm{M}_n(\mathbb{C}) \to \mathrm{M}_n(\mathbb{C})$ are differentiable matrix functions, then

$$d(fg)_A(B) = df_A(B)g(A) + f(A)dg_A(B),$$

for all $A, B \in \mathrm{M}_n(\mathbb{C})$. This is known as the *product rule*.

When E is of finite dimension n, for any basis $(u_1, \ldots, u_n)$ of E, we can define the directional derivatives with respect to the vectors in the basis $(u_1, \ldots, u_n)$

This way, we obtain the definition of partial derivatives, as follows.

Definition 2.3. For any two normed spaces E and F, if E is of finite dimension n, for every basis $(u_1, \ldots, u_n)$ for E, for every $a \in E$, for every function $f \colon E \to F$, the directional derivatives $\mathrm{D}_{u_j} f(a)$ (if they exist) are called the *partial derivatives of f with respect to the basis $(u_1, \ldots, u_n)$*. The partial derivative $\mathrm{D}_{u_j} f(a)$ is also denoted by $\partial_j f(a)$, or $\dfrac{\partial f}{\partial x_j}(a)$.

The notation $\dfrac{\partial f}{\partial x_j}(a)$ for a partial derivative, although customary and going back to Leibniz, is a "logical obscenity."

Indeed, the variable x_j really has nothing to do with the formal definition.

This is just another of these situations where tradition is just too hard to overthrow!

If both E and F are of finite dimension, for any basis $(u_1, \ldots, u_n)$ of E and any basis $(v_1, \ldots, v_m)$ of F, every function $f\colon E \to F$ is determined by m functions $f_i\colon E \to \mathbb{R}$ (or $f_i\colon E \to \mathbb{C}$), where

$$f(x) = f_1(x)v_1 + \cdots + f_m(x)v_m,$$

for every $x \in E$.

Then, we get

$$\mathrm{D}f(a)(u_j) = \\ \mathrm{D}f_1(a)(u_j)v_1 + \cdots + \mathrm{D}f_i(a)(u_j)v_i + \cdots + \mathrm{D}f_m(a)(u_j)v_m,$$

that is,

$$\mathrm{D}f(a)(u_j) = \partial_j f_1(a)v_1 + \cdots + \partial_j f_i(a)v_i + \cdots + \partial_j f_m(a)v_m.$$

The linear map $\mathrm{D}f(a)$ is determined by the $m \times n$-matrix

$J(f)(a) = (\partial_j f_i(a))$, or

$J(f)(a) = \left(\dfrac{\partial f_i}{\partial x_j}(a)\right)$:

$$J(f)(a) = \begin{pmatrix} \partial_1 f_1(a) & \partial_2 f_1(a) & \dots & \partial_n f_1(a) \\ \partial_1 f_2(a) & \partial_2 f_2(a) & \dots & \partial_n f_2(a) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_m(a) & \partial_2 f_m(a) & \dots & \partial_n f_m(a) \end{pmatrix}$$

or

$$J(f)(a) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \dfrac{\partial f_1}{\partial x_2}(a) & \dots & \dfrac{\partial f_1}{\partial x_n}(a) \\ \dfrac{\partial f_2}{\partial x_1}(a) & \dfrac{\partial f_2}{\partial x_2}(a) & \dots & \dfrac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(a) & \dfrac{\partial f_m}{\partial x_2}(a) & \dots & \dfrac{\partial f_m}{\partial x_n}(a) \end{pmatrix}.$$

This matrix is called the *Jacobian matrix* of $\mathrm{D}f$ at a.

When $m = n$, the determinant, $\det(J(f)(a))$, of $J(f)(a)$ is called the *Jacobian* of $\mathrm{D}f(a)$.

We know that this determinant only depends on $\mathrm{D}f(a)$, and not on specific bases. However, partial derivatives give a means for computing it.

When $E = \mathbb{R}^n$ and $F = \mathbb{R}^m$, for any function $f\colon \mathbb{R}^n \to \mathbb{R}^m$, it is easy to compute the partial derivatives $\dfrac{\partial f_i}{\partial x_j}(a)$.

We simply treat the function $f_i\colon \mathbb{R}^n \to \mathbb{R}$ as a function of its j-th argument, leaving the others fixed, and compute the derivative as the usual derivative.

Example 2.2. For example, consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$, defined by

$$f(r,\theta) = (r\cos\theta, r\sin\theta).$$

Then, we have

$$J(f)(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and the Jacobian (determinant) has value

$\det(J(f)(r,\theta)) = r.$

In the case where $E = \mathbb{R}$ (or $E = \mathbb{C}$), for any function $f\colon \mathbb{R} \to F$ (or $f\colon \mathbb{C} \to F$), the Jacobian matrix of $\mathrm{D}f(a)$ is a column vector. In fact, this column vector is just $\mathrm{D}_1 f(a)$. Then, for every $\lambda \in \mathbb{R}$ (or $\lambda \in \mathbb{C}$),

$$\mathrm{D}f(a)(\lambda) = \lambda \mathrm{D}_1 f(a).$$

Definition 2.4. Given a function $f\colon \mathbb{R} \to F$ (or $f\colon \mathbb{C} \to F$), where F is a normed space, the vector

$$\mathrm{D}f(a)(1) = \mathrm{D}_1 f(a)$$

is called the *vector derivative or velocity vector (in the real case)* at a. We usually identify $\mathrm{D}f(a)$ with its Jacobian matrix $\mathrm{D}_1 f(a)$, which is the column vector corresponding to $\mathrm{D}_1 f(a)$.

By abuse of notation, we also let $\mathrm{D}f(a)$ denote the vector $\mathrm{D}f(a)(1) = \mathrm{D}_1 f(a)$.

When $E = \mathbb{R}$, the physical interpretation is that f defines a (parametric) curve that is the trajectory of some particle moving in $\mathbb{R}^m$ as a function of time, and the vector $\mathrm{D}_1 f(a)$ is the *velocity* of the moving particle $f(t)$ at $t = a$. See Figure 2.3.

Example 2.3.

1. When $A = (0, 1)$ and $F = \mathbb{R}^3$, a function $f\colon (0, 1) \to \mathbb{R}^3$ defines a (parametric) curve in $\mathbb{R}^3$.

   If $f = (f_1, f_2, f_3)$, its Jacobian matrix at $a \in \mathbb{R}$ is

$$J(f)(a) = \begin{pmatrix} \dfrac{\partial f_1}{\partial t}(a) \\ \dfrac{\partial f_2}{\partial t}(a) \\ \dfrac{\partial f_3}{\partial t}(a) \end{pmatrix}.$$

   See Figure 2.3.

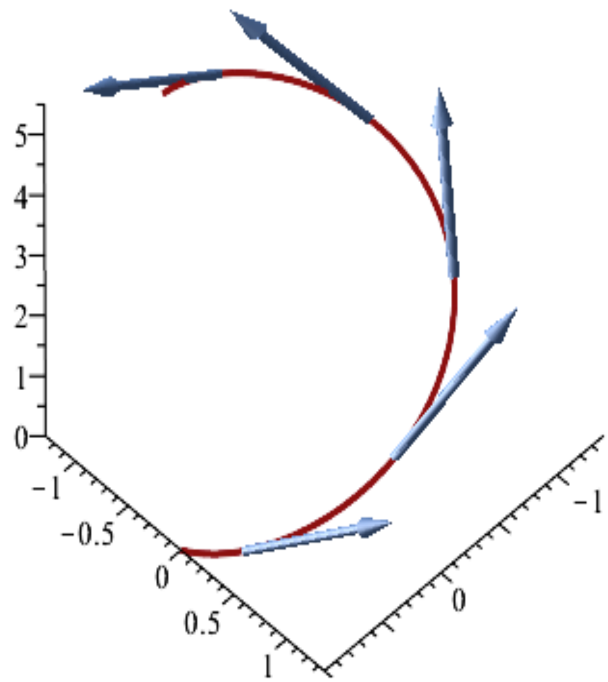

Figure 2.3: The red space curve $f(t) = (\cos(t), \sin(t), t)$.

The velocity vectors $J(f)(a) = \begin{pmatrix} -\sin(t) \\ \cos(t) \\ 1 \end{pmatrix}$ are represented by the blue arrows.

2. When $E = \mathbb{R}^2$ and $F = \mathbb{R}^3$, a function $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^3$ defines a parametric surface.

   Letting $\varphi = (f, g, h)$, its Jacobian matrix at $a \in \mathbb{R}^2$ is

$$J(\varphi)(a) = \begin{pmatrix} \frac{\partial f}{\partial u}(a) & \frac{\partial f}{\partial v}(a) \\ \frac{\partial g}{\partial u}(a) & \frac{\partial g}{\partial v}(a) \\ \frac{\partial h}{\partial u}(a) & \frac{\partial h}{\partial v}(a) \end{pmatrix}.$$

   See Figure 2.4.

   The Jacobian matrix is

$$J(f)(a) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2u & 2v \end{pmatrix}.$$

   The first column is the vector tangent to the pink u-direction curve, while the second column is the vector tangent to the blue v-direction curve.

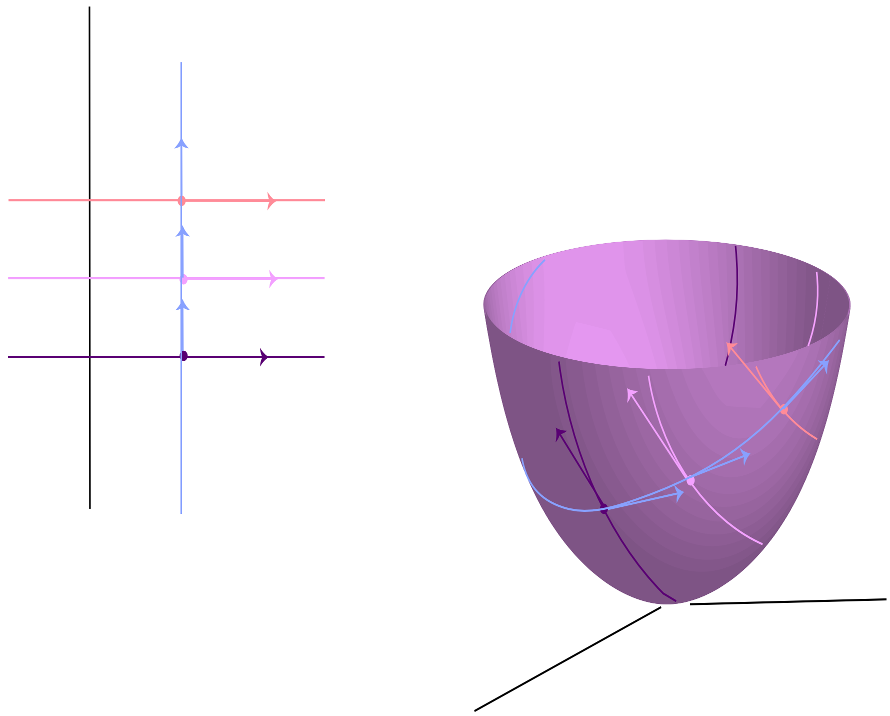

Figure 2.4: The parametric surface $x = u, y = v, z = u^2 + v^2$.

3. When $E = \mathbb{R}^3$ and $F = \mathbb{R}$, for a function $f \colon \mathbb{R}^3 \to \mathbb{R}$, the Jacobian matrix at $a \in \mathbb{R}^3$ is

$$J(f)(a) = \left(\frac{\partial f}{\partial x}(a) \ \ \frac{\partial f}{\partial y}(a) \ \ \frac{\partial f}{\partial z}(a) \right).$$

More generally, when $f\colon \mathbb{R}^n \to \mathbb{R}$, the Jacobian matrix at $a \in \mathbb{R}^n$ is the row vector

$$J(f)(a) = \left(\frac{\partial f}{\partial x_1}(a) \;\cdots\; \frac{\partial f}{\partial x_n}(a)\right).$$

Its transpose is a column vector called the *gradient* of f at a, denoted by $\mathrm{grad} f(a)$ or $\nabla f(a)$.

Then, given any $v \in \mathbb{R}^n$, note that

$$\mathrm{D}f(a)(v) = \frac{\partial f}{\partial x_1}(a)\, v_1 + \cdots + \frac{\partial f}{\partial x_n}(a)\, v_n = \mathrm{grad} f(a) \cdot v,$$

the scalar product of $\mathrm{grad} f(a)$ and v.

Example 2.4. Consider the quadratic function $f\colon \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = x^\top A x, \quad x \in \mathbb{R}^n,$$

where A is a real $n \times n$ symmetric matrix.

We claim that

$$df_u(h) = 2u^\top A h \quad \text{for all } u, h \in \mathbb{R}^n.$$

Since A is symmetric, we have

$$\begin{aligned} f(u+h) &= (u^\top + h^\top)A(u+h) \\ &= u^\top A u + u^\top A h + h^\top A u + h^\top A h \\ &= u^\top A u + 2u^\top A h + h^\top A h, \end{aligned}$$

so we have

$$f(u+h) - f(u) - 2u^\top A h = h^\top A h.$$

If we write

$$\epsilon(h) = \frac{h^\top A h}{\|h\|}$$

for $h \notin 0$ where $\| \, \|$ is the 2-norm, by Cauchy–Schwarz we have

$$|\epsilon(h)| \leq \frac{\|h\| \, \|Ah\|}{\|h\|} \leq \frac{\|h\|^2 \, \|A\|}{\|h\|} = \|h\| \, \|A\| \, ,$$

which shows that $\lim_{h \mapsto 0} \epsilon(h) = 0$. Therefore,

$$df_u(h) = 2u^\top A h \quad \text{for all } u, h \in \mathbb{R}^n,$$

as claimed.

This formula shows that the gradient ∇f_u of f at u is given by

$$\nabla f_u = 2Au.$$

As a first corollary we obtain the gradient of a function of the form

$$f(x) = \frac{1}{2} x^\top A x - b^\top x,$$

where A is a symmetric $n \times n$ matrix and b is some vector $b \in \mathbb{R}^n$.

Since the derivative of a linear function is itself, we obtain

$$df_u(h) = u^\top A h - b^\top h,$$

and the gradient of $f(x) = \frac{1}{2} x^\top A x - b^\top x$, is given by

$$\nabla f_u = Au - b.$$

As a second corollary we obtain the gradient of the function

$$f(x) = \|Ax - b\|_2^2 = (Ax - b)^\top (Ax - b),$$

which is the function to minimize in a least squares problem, where A is an $m \times n$ matrix. We obtain

$$df_u(h) = 2u^\top A^\top A h - 2b^\top A h.$$

Consequently, the gradient of $f(x) = \|Ax - b\|_2^2$ is given by

$$\nabla f_u = 2A^\top A u - 2A^\top b.$$

These two results will be heavily used in quadratic optimization.

When E, F, and G have finite dimensions, if A is an open subset of E, B is an open subset of F, for any functions $f\colon A \to F$ and $g\colon B \to G$, such that $f(A) \subseteq B$, for any $a \in A$, letting $b = f(a)$, and $h = g \circ f$, if $\mathrm{D}f(a)$ exists and $\mathrm{D}g(b)$ exists, by Theorem 2.5, the Jacobian matrix $J(h)(a) = J(g \circ f)(a)$ is given by

$$J(h)(a) = J(g)(b)J(f)(a) =$$
$$\begin{pmatrix} \frac{\partial g_1}{\partial y_1}(b) & \frac{\partial g_1}{\partial y_2}(b) & \dots & \frac{\partial g_1}{\partial y_n}(b) \\ \frac{\partial g_2}{\partial y_1}(b) & \frac{\partial g_2}{\partial y_2}(b) & \dots & \frac{\partial g_2}{\partial y_n}(b) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial y_1}(b) & \frac{\partial g_m}{\partial y_2}(b) & \dots & \frac{\partial g_m}{\partial y_n}(b) \end{pmatrix} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \dots & \frac{\partial f_1}{\partial x_p}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \dots & \frac{\partial f_2}{\partial x_p}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(a) & \frac{\partial f_n}{\partial x_2}(a) & \dots & \frac{\partial f_n}{\partial x_p}(a) \end{pmatrix}$$

Thus, we have the familiar formula

$$\frac{\partial h_i}{\partial x_j}(a) = \sum_{k=1}^{k=n} \frac{\partial g_i}{\partial y_k}(b) \frac{\partial f_k}{\partial x_j}(a).$$

Given two normed spaces E and F of finite dimension, given an open subset A of E, if a function $f\colon A \to F$ is differentiable at $a \in A$, then its Jacobian matrix is well defined.

One should be warned that the converse is false. There are functions such that all the partial derivatives exist at some $a \in A$, but yet, the function is not differentiable at a, and not even continuous at a.

However, there are sufficient conditions on the partial derivatives for $\mathrm{D}f(a)$ to exist, namely, continuity of the partial derivatives.

If f is differentiable on A, then f defines a function $\mathrm{D}f\colon A \to \mathcal{L}(E; F)$.

It turns out that the continuity of the partial derivatives on A is a necessary and sufficient condition for $\mathrm{D}f$ to exist and to be continuous on A.

To prove this, we need an important result known as the mean value theorem.

If E is a vector space (over $\mathbb{R}$ or $\mathbb{C}$), given any two points $a, b \in E$, the *closed segment* $[a, b]$ is the set of all points $a + \lambda(b - a)$, where $0 \leq \lambda \leq 1$, $\lambda \in \mathbb{R}$, and the *open segment* (a, b) is the set of all points $a + \lambda(b - a)$, where $0 < \lambda < 1$, $\lambda \in \mathbb{R}$.

The following result is known as the *mean value theorem*.

Proposition 2.7. *Let E and F be two normed vector spaces, let A be an open subset of E, and let $f\colon A \to F$ be a continuous function on A. Given any $a \in A$ and any $h \neq 0$ in E, if the closed segment $[a, a+h]$ is contained in A, if $f\colon A \to F$ is differentiable at every point of the open segment $(a, a+h)$, and if*

$$\sup_{x \in (a, a+h)} \|\mathrm{D}f(x)\| \leq M$$

for some $M \geq 0$, then

$$\|f(a+h) - f(a)\| \leq M\|h\|.$$

As a corollary, if $L\colon E \to F$ is a continuous linear map, then

$$\|f(a+h) - f(a) - L(h)\| \leq M\|h\|,$$

where $M = \sup_{x \in (a, a+h)} \|\mathrm{D}f(x) - L\|$.

Recall that an open subset A of a topological space E is *connected* if it is not the union of two disjoint nonempty open subsets.

More generally, a subset A of a topological space E is *connected* if it is connected with respect to the subspace topology induced on A by E.

Proposition 2.8. *Let $f\colon A \to F$ be any function between two normed vector spaces E and F, where A is an open subset of E. If A is connected and if $\mathrm{D}f(a) = 0$ for all $a \in A$, then f is a constant function on A.*

The mean value theorem also implies the following important result.

Theorem 2.9. *Given two normed vector spaces E and F, where E is of finite dimension n and where $(u_1, \ldots, u_n)$ is a basis of E, given any open subset A of E, given any function $f\colon A \to F$, the derivative $\mathrm{D}f\colon A \to \mathcal{L}(E; F)$ is defined and continuous on A iff every partial derivative $\partial_j f$ (or $\dfrac{\partial f}{\partial x_j}$) is defined and continuous on A, for all j, $1 \leq j \leq n$.*

As a corollary, if F is of finite dimension m, and $(v_1, \ldots, v_m)$ is a basis of F, the derivative $\mathrm{D}f\colon A \to \mathcal{L}(E; F)$ is defined and continuous on A iff every partial derivative $\partial_j f_i$ $\left(\text{or } \dfrac{\partial f_i}{\partial x_j}\right)$ is defined and continuous on A, for all i, j, $1 \leq i \leq m$, $1 \leq j \leq n$.

Definition 2.5. Given two normed vector spaces E and F, and an open subset A of E, we say that a function $f\colon A \to F$ is a *C^0-function on A* if f is continuous on A. We say that $f\colon A \to F$ is a *C^1-function on A* if $\mathrm{D}f$ exists and is continuous on A.

Let E and F be two normed vector spaces, let $U \subseteq E$ be an open subset of E and let $f\colon E \to F$ be a function such that $Df(a)$ exists for all $a \in U$.

If $Df(a)$ is injective for all $a \in U$, we say that f is an *immersion* (on U) and if $Df(a)$ is surjective for all $a \in U$, we say that f is a *submersion* (on U).

When E and F are finite dimensional with $\dim(E) = n$ and $\dim(F) = m$, if $m \geq n$, then f is an immersion iff the Jacobian matrix $J(f)(a)$, has full rank (n) for all $a \in E$ and if $n \geq m$, then f is a submersion iff the Jacobian matrix $J(f)(a)$, has full rank (m) for all $a \in E$.

For example, $f\colon \mathbb{R} \to \mathbb{R}^2$ defined by $f(t) = (\cos(t), \sin(t))$ is an immersion since

$$J(f)(t) = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix}$$

has rank 1 for all t.

On the other hand, $f\colon \mathbb{R} \to \mathbb{R}^2$ defined by $f(t) = (t^2, t^2)$ is not an immersion since

$$J(f)(t) = \begin{pmatrix} 2t \\ 2t \end{pmatrix}$$

vanishes at $t = 0$. See Figure 2.5.

An example of a submersion is given by the projection map $f\colon \mathbb{R}^2 \to \mathbb{R}$, where $f(x, y) = x$, since $J(f)(x, y) = \begin{pmatrix} 1 & 0 \end{pmatrix}$.

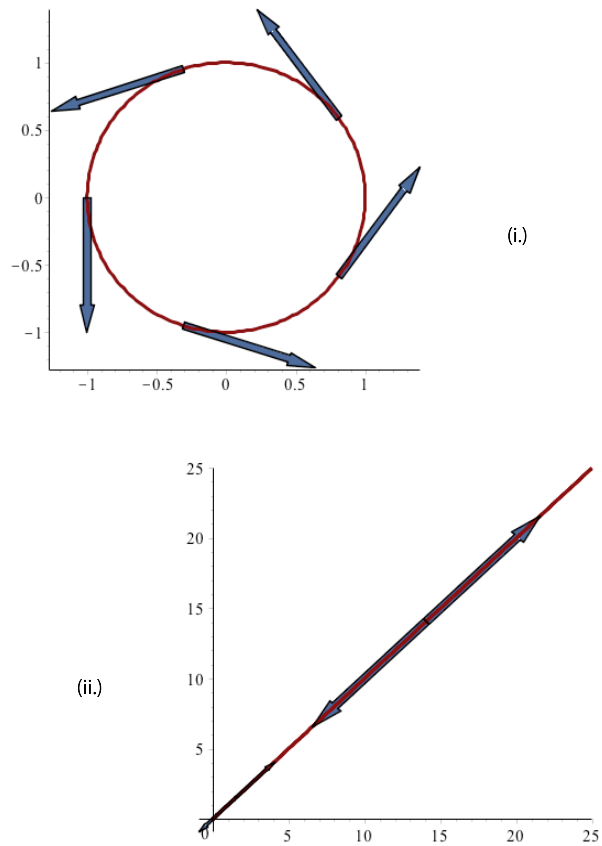

Figure 2.5: Figure $(i.)$ is the immersion of $\mathbb{R}$ into $\mathbb{R}^2$ given by $f(t) = (\cos(t), \sin(t))$. Figure $(ii.)$, the parametric curve $f(t) = (t^2, t^2)$, is not an immersion since the tangent vanishes at the origin.

A very important theorem is the inverse function theorem. In order for this theorem to hold for infinite dimensional spaces, it is necessary to assume that our normed spaces are complete.

Given a normed vector space, E, we say that a sequence, $(u_n)_n$, with $u_n \in E$, is a *Cauchy sequence* iff for every $\epsilon > 0$, there is some $N > 0$ so that for all $m, n \geq N$,

$$\|u_n - u_m\| < \epsilon.$$

A normed vector space, E, is *complete* iff every Cauchy sequence converges.

A complete normed vector space is also called a *Banach space*, after Stefan Banach (1892-1945).

Fortunately, $\mathbb{R}, \mathbb{C}$, and every finite dimensional (real or complex) normed vector space is complete.

A real (resp. complex) vector space, E, is a real (resp. complex) *Hilbert space* if it is complete as a normed space with the norm $\|u\| = \sqrt{\langle u, u \rangle}$ induced by its Euclidean (resp. Hermitian) inner product (of course, positive, definite).

Definition 2.6. Given two topological spaces E and F and an open subset A of E, we say that a function $f\colon A \to F$ is a *local homeomorphism from A to F* if for every $a \in A$, there is an open set $U \subseteq A$ containing a and an open set V containing $f(a)$ such that f is a homeomorphism from U to $V = f(U)$.

If B is an open subset of F, we say that $f\colon A \to F$ is a *(global) homeomorphism from A onto B* if f is a homeomorphism from A to $B = f(A)$.

If E and F are normed spaces, we say that $f\colon A \to F$ is a *local diffeomorphism from A to F* if for every $a \in A$, there is an open set $U \subseteq A$ containing a and an open set V containing $f(a)$ such that f is a bijection from U to V, f is a C^1-function on U, and f^{-1} is a C^1-function on $V = f(U)$.

We say that $f\colon A \to F$ is a *(global) diffeomorphism from A to B* if f is a homeomorphism from A to $B = f(A)$, f is a C^1-function on A, and f^{-1} is a C^1-function on B.

Note that a local diffeomorphism is a local homeomorphism.

Also, as a consequence of Proposition 2.6, if f is a diffeomorphism on A, then $\mathrm{D}f(a)$ is a linear isomorphism for every $a \in A$.

Theorem 2.10. *(Inverse Function Theorem) Let E and F be complete normed spaces, let A be an open subset of E, and let $f\colon A \to F$ be a C^1-function on A. The following properties hold:*

(1) For every $a \in A$, if $\mathrm{D}f(a)$ is a linear isomorphism (which means that both $\mathrm{D}f(a)$ and $(\mathrm{D}f(a))^{-1}$ are linear and continuous),[1] then there exist some open subset $U \subseteq A$ containing a, and some open subset V of F containing $f(a)$, such that f is a diffeomorphism from U to $V = f(U)$. Furthermore,

$$\mathrm{D}f^{-1}(f(a)) = (\mathrm{D}f(a))^{-1}.$$

For every neighborhood N of a, the image $f(N)$ of N is a neighborhood of $f(a)$, and for every open ball $U \subseteq A$ of center a, the image $f(U)$ of U contains some open ball of center $f(a)$.

[1] Actually, since E and F are Banach spaces, by the Open Mapping Theorem, it is sufficient to assume that $\mathrm{D}f(a)$ is continuous and bijective; see Lang [29].

(2) If $\mathrm{D}f(a)$ *is invertible for every* $a \in A$*, then* $B = f(A)$ *is an open subset of* F*, and* f *is a local diffeomorphism from* A *to* B*. Furthermore, if* f *is injective, then* f *is a diffeomorphism from* A *to* B*.*

Part (1) of Theorem 2.10 is often referred to as the "(local) inverse function theorem." It plays an important role in the study of manifolds and (ordinary) differential equations.

If E and F are both of finite dimension, the case where $\mathrm{D}f(a)$ is just injective or just surjective is also important for defining manifolds, using implicit definitions.

We now briefly consider second-order and higher-order derivatives.

2.2 Second-Order and Higher-Order Derivatives

Given two normed vector spaces E and F, and some open subset A of E, if $\mathrm{D}f(a)$ is defined for every $a \in A$, then we have a mapping $\mathrm{D}f \colon A \to \mathcal{L}(E; F)$.

Since $\mathcal{L}(E; F)$ is a normed vector space, if $\mathrm{D}f$ exists on an open subset U of A containing a, we can consider taking the derivative of $\mathrm{D}f$ at some $a \in A$.

Definition 2.7. Given a function $f \colon A \to F$ defined on some open subset A of E such that $\mathrm{D}f(a)$ is defined for every $a \in A$, if $\mathrm{D}(\mathrm{D}f)(a)$ exists for every $a \in A$, we get a mapping $\mathrm{D}^2 f \colon A \to \mathcal{L}(E; \mathcal{L}(E; F))$ called the *second derivative of f on A*, where $\mathrm{D}^2 f(a) = \mathrm{D}(\mathrm{D}f)(a)$, for every $a \in A$.

Proposition 2.11. *If* $\mathrm{D}^2 f(a)$ *exists, then* $\mathrm{D}_u(\mathrm{D}_v f)(a)$ *exists and*

$$\mathrm{D}^2 f(a)(u)(v) = \mathrm{D}_u(\mathrm{D}_v f)(a), \quad \textit{for all } u, v \in E.$$

Definition 2.8. We denote $\mathrm{D}_u(\mathrm{D}_v f)(a)$ by $\mathrm{D}^2_{u,v} f(a)$ (or $\mathrm{D}_u \mathrm{D}_v f(a)$).

Recall from Proposition **??**, that the map from $\mathcal{L}_2(E, E; F)$ to $\mathcal{L}(E; \mathcal{L}(E; F))$ defined such that $g \mapsto \varphi$ iff for every $g \in \mathcal{L}_2(E, E; F)$,

$$\varphi(u)(v) = g(u, v),$$

is an isomorphism of vector spaces.

Thus, we will consider $\mathrm{D}^2 f(a) \in \mathcal{L}(E; \mathcal{L}(E; F))$ *as a continuous bilinear map in* $\mathcal{L}_2(E, E; F)$*, and we write* $\mathrm{D}^2 f(a)(u, v)$*, instead of* $\mathrm{D}^2 f(a)(u)(v)$*.*

Then the above discussion can be summarized by saying that when $\mathrm{D}^2 f(a)$ is defined, we have

$$\mathrm{D}^2 f(a)(u, v) = \mathrm{D}_u \mathrm{D}_v f(a).$$

Definition 2.9. When E has finite dimension and $(e_1, \ldots, e_n)$ is a basis for E, we denote $\mathrm{D}_{e_j} \mathrm{D}_{e_i} f(a)$ by $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(a)$, when $i \neq j$, and we denote $\mathrm{D}_{e_i} \mathrm{D}_{e_i} f(a)$ by $\dfrac{\partial^2 f}{\partial x_i^2}(a)$.

Proposition 2.12. *(Schwarz's lemma) Given two normed vector spaces E and F, given any open subset A of E, given any $f \colon A \to F$, for every $a \in A$, if $\mathrm{D}^2 f(a)$ exists, then $\mathrm{D}^2 f(a) \in \mathcal{L}_2(E, E; F)$ is a continuous symmetric bilinear map. As a corollary, if E is of finite dimension n, and $(e_1, \ldots, e_n)$ is a basis for E, we have*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a).$$

When $E = \mathbb{R}^2$, the existence of $\dfrac{\partial^2 f}{\partial x \partial y}(a)$ and $\dfrac{\partial^2 f}{\partial y \partial x}(a)$ is not sufficient to insure the existence of $\mathrm{D}^2 f(a)$.

Therefore, for every $a \in A$, where it exists, $\mathrm{D}^2 f(a)$ belongs to the space $\mathcal{S}\mathrm{ym}_2(E^2; F)$ of continuous symmetric bilinear maps from E^2 to F.

If E has finite dimension n and $F = \mathbb{R}$, with respect to any basis $(e_1, \ldots, e_n)$ of E,

$$\mathrm{D}^2 f(a)(u, v) = u^\top \mathrm{Hess} f(a) v,$$

where

$$\mathrm{Hess} f(a) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)$$

is the *Hessian matrix* of f at a.

Example 2.5. Consider the function f defined on real invertible 2×2 matrices such that $ad - bc > 0$ given by

$$f(a, b, c, d) = \log(ad - bc).$$

We immediately verify that the Jacobian matrix of f is given by

$$df_{a,b,c,d} = \frac{1}{ad - bc} \begin{pmatrix} d & -c & -b & a \end{pmatrix}.$$

It is easily checked that if

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad X = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix},$$

then

$$df_A(X) = \mathrm{tr}(A^{-1}X) = \frac{1}{ad - bc} \mathrm{tr}\left(\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix} \right).$$

Computing second-order derivatives, we find that the Hessian matrix of f is given by

$$Hf(A) = \frac{1}{(ad-bc)^2}\begin{pmatrix} -d^2 & cd & bd & -bc \\ cd & -c^2 & -ad & ac \\ bd & -ad & -b^2 & ab \\ -bc & ac & ab & -a^2 \end{pmatrix}.$$

Using the formula for the derivative of the inversion map and the chain rule we can show that

$$\mathrm{D}^2 f(A)(X_1, X_2) = -\mathrm{tr}(A^{-1}X_1A^{-1}X_2),$$

and so

$$Hf(A)(X_1, X_2) = -\mathrm{tr}(A^{-1}X_1A^{-1}X_2),$$

a formula which is far from obvious.

The function f can be generalized to matrices $A \in \mathbf{GL}^+(n, \mathbb{R})$, that is, $n \times n$ real invertible matrices of positive determinants, as

$$f(A) = \log\det(A).$$

It can be shown that the formulae

$$\begin{aligned} df_A(X) &= \mathrm{tr}(A^{-1}X) \\ \mathrm{D}^2 f(A)(X_1, X_2) &= -\mathrm{tr}(A^{-1}X_1A^{-1}X_2) \end{aligned}$$

also hold.

By induction, if $\mathrm{D}^m f\colon A \to \mathcal{S}\mathrm{ym}_m(E^m; F)$ exists for $m \geq 1$, where $\mathcal{S}\mathrm{ym}_m(E^m; F)$ denotes the vector space of continuous symmetric multilinear maps from E^m to F, and if $\mathrm{D}\mathrm{D}^m f(a)$ exists for all $a \in A$, we obtain the $(m+1)$th derivative

$$\mathrm{D}^{m+1} f = \mathrm{D}\mathrm{D}^m f(a)$$

of f, and $\mathrm{D}^{m+1} f \in \mathcal{S}\mathrm{ym}_{m+1}(E^{m+1}; F)$, where $\mathcal{S}\mathrm{ym}_{m+1}(E^{m+1}; F)$ is the vector space of continuous symmetric multilinear maps from E^{m+1} to F.

For any $m \geq 1$, we say that the map $f\colon A \to F$ is a *C^m function* (or simply that *f is C^m*) if $\mathrm{D}f, \mathrm{D}^2 f, \ldots, \mathrm{D}^m f$ exist and are continuous on A.

We say that f is C^∞ or *smooth* if $\mathrm{D}^m f$ exists and is continuous on A for all $m \geq 1$. If E has finite dimension n, it can be shown that f is smooth iff all of its partial derivatives

$$\frac{\partial^m f}{\partial x_{i_1} \cdots \partial x_{i_m}}(a)$$

are defined and continuous for all $a \in A$, all $m \geq 1$, and all $i_1, \ldots, i_m \in \{1, \ldots, n\}$.

The function $f\colon A \to F$ is a *C^m diffeomorphism* between A and $B = f(A)$ if f is a bijection from A to B and if f and f^{-1} are C^m.

Similarly, f is a *smooth diffeomorphism* between A and $B = f(A)$ if f is a bijection from A to B and if f and f^{-1} are smooth.

Several variants of Taylor's formula exist but we will not cover these here. The reader is referred to Section 3.6 of our book, Vol II.

2.3 Series and Power Series of Matrices

Since a number of important functions on matrices are defined by power series, in particular the exponential, we review quickly some basic notions about series in a complete normed vector space.

Given a normed vector space $(E, \| \|)$, a *series* is an infinite sum $\sum_{k=0}^{\infty} a_k$ of elements $a_k \in E$.

We denote by S_n the partial sum of the first $n+1$ elements,

$$S_n = \sum_{k=0}^{n} a_k.$$

Definition 2.10. We say that the series $\sum_{k=0}^{\infty} a_k$ *converges* to the limit $a \in E$ if the sequence (S_n) converges to a. In this case, we say that the series is *convergent*. We say that the series $\sum_{k=0}^{\infty} a_k$ *converges absolutely* if the series of norms $\sum_{k=0}^{\infty} \|a_k\|$ is convergent.

There are series that are convergent but not absolutely convergent; for example, the series

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}.$$

However, if E is complete (which means that every Cauchy sequence converges), the converse is an *enormously useful result*.

Proposition 2.13. *Let $(E, \| \, \|)$ be a complete normed vector space. If a series $\sum_{k=0}^{\infty} a_k$ is absolutely convergent, then it is convergent.*

Remark: It can be shown that if $(E, \| \, \|)$ is a normed vector space such that every absolutely convergent series is also convergent, then E must be complete.

If $E = \mathbb{C}$, then there are several conditions that imply the absolute convergence of a series.

The *ratio test* is the following test. Suppose there is some $N > 0$ such that $a_n \neq 0$ for all $n \geq N$, and either

$$r = \lim_{n \mapsto \infty} \left| \frac{a_{n+1}}{a_n} \right|$$

exists, or the sequence of ratios diverges to infinity, in which case we write $r = \infty$. Then, if $0 \leq r < 1$, the series $\sum_{k=0}^{n} a_k$ converges absolutely, else if $1 < r \leq \infty$, the series diverges.

If (r_n) is a sequence of real numbers, recall that

$$\limsup_{n \mapsto \infty} r_n = \lim_{n \mapsto \infty} \sup_{k \geq n} \{r_k\}.$$

If $r_n \geq 0$ for all n, then there are two cases:

(1) The sequence (r_n) is unbounded, in which case $\limsup_{n \mapsto \infty} r_n = +\infty$.

(2) The sequence (r_n) is bounded, in which case $r = \limsup_{n \mapsto \infty} r_n$ is finite and is characterized as follows. For every $\epsilon > 0$, there is some $N \in \mathbb{N}$ such that $r_n < r + \epsilon$ for all $n \geq N$, and $r_n > r - \epsilon$ for infinitely many n.

Then, the *root test* is this. Let

$$r = \limsup_{n \mapsto \infty} |a_n|^{1/n}$$

if the limit exists (is finite), else write $r = \infty$. Then, if $0 \leq r < 1$, the series $\sum_{k=0}^{n} a_k$ converges absolutely, else if $1 < r \leq \infty$, the series diverges.

The root test also applies if $(E, \| \|)$ is a complete normed vector space by replacing $|a_n|$ by $\|a_n\|$.

Let $\sum_{k \geq 0}^{\infty} a_k$ be a series of elements $a_k \in E$ and let

$$r = \limsup_{n \mapsto \infty} \|a_n\|^{1/n}$$

if the limit exists (is finite), else write $r = \infty$. Then, if $0 \leq r < 1$, the series $\sum_{k=0}^{n} a_k$ converges absolutely, else if $1 < r \leq \infty$, the series diverges.

A *power series* with coefficients $a_k \in \mathbb{C}$ in the indeterminate z is a formal expression $f(z)$ of the form

$$f(z) = \sum_{k=0}^{\infty} a_k z^k,$$

For any fixed value $z \in \mathbb{C}$, the series $f(z)$ may or may not converge. It always converges for $z = 0$, since$f(0) = a_0$.

A fundamental fact about power series is that they have a *radius of convergence*.

Proposition 2.14. *Given any power series*

$$f(z) = \sum_{k=0}^{\infty} a_k z^k,$$

there is a nonnegative real R, possibly infinite, called the **radius of convergence** *of the power series, such that if $|z| < R$, then $f(z)$ converges absolutely, else if $|z| > R$, then $f(z)$ diverges. Moreover (Hadamard), we have*

$$R = \frac{1}{\limsup_{n \mapsto \infty} |a_n|^{1/n}}.$$

Note that Proposition 2.14 does not say anything about the behavior of the power series for boundary values, that is, values of z such that $|z| = R$.

Proposition 2.15. *Let $f(z) = \sum_{k=0}^{\infty} a_k z^k$ be a power series with coefficients $a_k \in \mathbb{C}$. Suppose there is some $N > 0$ such that $a_n \neq 0$ for all $n \geq N$, and either*

$$R = \lim_{n \mapsto \infty} \left| \frac{a_n}{a_{n+1}} \right|$$

exists, or the sequence on the righthand side diverges to infinity, in which case we write $R = \infty$. Then, the power series $\sum_{k=0}^{\infty} a_k z^k$ has radius of convergence R.

Power series behave very well with respect to derivatives.

Proposition 2.16. *Suppose the power series* $f(z) = \sum_{k=0}^{\infty} a_k z^k$ *(with real coefficients) has radius of convergence* $R > 0$*. Then,* $f'(z)$ *exists if* $|z| < R$*, the power series* $\sum_{k=1}^{\infty} k a_k z^{k-1}$ *has radius of convergence* R*, and*

$$f'(z) = \sum_{k=1}^{\infty} k a_k z^{k-1}.$$

Let us now assume that $f(z) = \sum_{k=0}^{\infty} a_k z^k$ is a power series with coefficients $a_k \in \mathbb{C}$, and that its radius of convergence is R.

Given any matrix $A \in \mathrm{M}_n(\mathbb{C})$ we can form the power series obtained by substituting A for z,

$$f(A) = \sum_{k=0}^{\infty} a_k A^k.$$

Let $\| \ \|$ be any matrix norm on $\mathrm{M}_n(\mathbb{C})$.

Proposition 2.17. *Let $f(z) = \sum_{k=1}^{\infty} a_k z^k$ be a power series with complex coefficients, write R for its radius of convergence, and assume that $R > 0$. For every ρ such that $0 < \rho < R$, the series $f(A) = \sum_{k=1}^{\infty} a_k A^k$ is absolutely convergent for all $A \in \mathrm{M}_n(\mathbb{C})$ such that $\|A\| \leq \rho$. Furthermore, f is continuous on the open ball $B(R) = \{A \in \mathrm{M}_n(\mathbb{C}) \mid \|A\| < R\}$.*

Note that unlike the case where $A \in \mathbb{C}$, if $\|A\| > R$, we cannot claim that the series $f(A)$ diverges.

This has to do with the fact that even for the operator norm we may have $\|A^n\| < \|A\|^n$. We leave it as an exercise to find an example of a series and a matrix A with $\|A\| > R$, and yet $f(A)$ converges.

As an application of Proposition 2.17, the exponential power series

$$e^A = \exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

is absolutely convergent for all $A \in \mathrm{M}_n(\mathbb{C})$, and continuous everywhere.

Proposition 2.17 also implies that the series

$$\log(I + A) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{A^k}{k}$$

is absolutely convergent if $\|A\| < 1$.

Now, it is known (see Cartan [10]) that the formal power series

$$E(A) = \sum_{k=1}^{\infty} \frac{A^k}{k!}$$

and

$$L(A) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{A^k}{k}$$

are mutual inverses; that is,

$$E(L(A)) = A, \quad L(E(A)) = A, \quad \text{for all } A.$$

Observe that $E(A) = e^A - I = \exp(A) - I$ and $L(A) = \log(I + A)$. It follows that

$$\begin{aligned} \log(\exp(A)) &= A \quad \text{for all } A \text{ with } \|A\| < \log(2) \\ \exp(\log(I + A)) &= I + A \quad \text{for all } A \text{ with } \|A\| < 1. \end{aligned}$$

Finally, let us consider the generalization of the notion of a power series $f(t) = \sum_{k=1}^{\infty} a_k t^k$ of a real variable t, where the coefficients a_k belong to a complete normed vector space $(F, \| \|)$.

Proposition 2.18. *Let $(F, \| \|)$ be a complete normed vector space. Given any power series*

$$f(t) = \sum_{k=0}^{\infty} a_k t^k,$$

with $t \in \mathbb{R}$ and $a_k \in F$, there is a nonnegative real R, possibly infinite, called the **radius of convergence** *of the power series, such that if $|t| < R$, then $f(t)$ converges absolutely, else if $|t| > R$, then $f(t)$ diverges. Moreover, we have*

$$R = \frac{1}{\limsup_{n \mapsto \infty} \|a_n\|^{1/n}}.$$

Proposition 2.19. *Let* $(F, \| \|)$ *be a complete normed vector space. Suppose the power series* $f(t) = \sum_{k=0}^{\infty} a_k t^k$ *(with coefficients* $a_k \in F$*) has radius of convergence* R*. Then,* $f'(t)$ *exists if* $|t| < R$*, the power series* $\sum_{k=1}^{\infty} k a_k t^{k-1}$ *has radius of convergence* R*, and*

$$f'(t) = \sum_{k=1}^{\infty} k a_k t^{k-1}.$$

2.4 Linear Vector Fields and the Exponential

We can apply Propositions 2.18 and 2.19 to the map $f\colon t \mapsto e^{tA}$, where A is any matrix $A \in \mathrm{M}_n(\mathbb{C})$.

This power series has a infinite radius of convergence, and we have

$$f'(t) = \sum_{k=1}^{\infty} k\frac{t^{k-1}A^k}{k!} = A\sum_{k=1}^{\infty} \frac{t^{k-1}A^{k-1}}{(k-1)!} = Ae^{tA}.$$

Note that

$$Ae^{tA} = e^{tA}A.$$

Definition 2.11. Given some open subset A of $\mathbb{R}^n$, a *vector field* X on A is a function $X\colon A \to \mathbb{R}^n$, which assigns to every point $p \in A$ a vector $X(p) \in \mathbb{R}^n$.

Usually, we assume that X is at least C^1 function on A.

For example, if $f\colon \mathbb{R}^2 \to \mathbb{R}$ is

$$f(x,y) = \cos(xy^2),$$

the gradient vector field X is

$$(-y^2\sin(xy^2), -2xy\sin(xy^2)) = (X_1, X_2).$$

Note that

$$\frac{\partial X_1}{\partial y} = -2y\sin(xy^2) - 2xy^3\cos(xy^2) = \frac{\partial X_2}{\partial x}.$$

This example is easily generalized to $\mathbb{R}^n$.

If $f\colon A \to \mathbb{R}$ is a C^1 function, then its *gradient* defines a vector field X; namely, $p \mapsto \operatorname{grad} f(p)$.

If f is C^2, then its second partials commute; that is,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(p) = \frac{\partial^2 f}{\partial x_j \partial x_i}(p), \quad 1 \leq i, j \leq n,$$

so this vector field $X = (X_1, \ldots, X_n)$ has a very special property:

$$\frac{\partial X_i}{\partial x_j} = \frac{\partial X_j}{\partial x_i}, \quad 1 \leq i, j \leq n.$$

This is a necessary condition for a vector field to be the gradient of some function, but not a sufficient condition in general.

The existence of such a function depends on the topological shape of the domain A.

Understanding what are sufficient conditions to answer the above question led to the development of differential forms and cohomology.

Definition 2.12. Given a vector field $X\colon A \to \mathbb{R}^n$, for any point $p_0 \in A$, a C^1 curve $\gamma\colon (-\epsilon, \epsilon) \to \mathbb{R}^n$ (with $\epsilon > 0$) is an *integral curve for X with initial condition p_0* if $\gamma(0) = p_0$, and

$$\gamma'(t) = X(\gamma(t)) \quad \text{for all } t \in (-\epsilon, \epsilon).$$

An integral curve has the property that for every time $t \in (-\epsilon, \epsilon)$, the tangent vector $\gamma'(t)$ to the curve γ at the point $\gamma(t)$ coincides with the vector $X(\gamma(t))$ given by the vector field at the point $\gamma(t)$. See Figure 2.6.

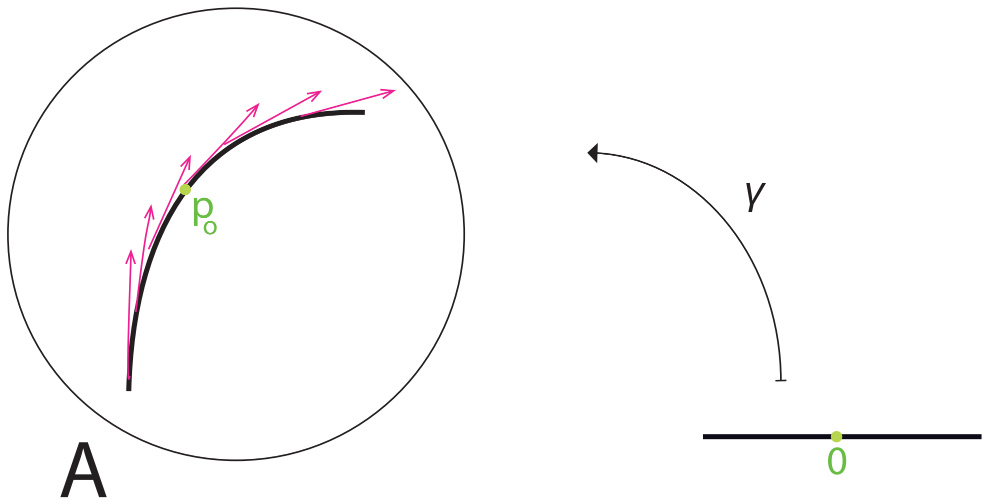

Figure 2.6: An integral curve in $\mathbb{R}^2$.

Definition 2.13. Given a C^1 vector field $X\colon A \to \mathbb{R}^n$, for any point $p_0 \in A$, a *local flow for X at p_0* is a function

$$\varphi\colon J \times U \to \mathbb{R}^n,$$

where $J \subseteq \mathbb{R}$ is an open interval containing 0 and U is an open subset of A containing p_0, so that for every $p \in U$, the curve $t \mapsto \varphi(t, p)$ is an integral curve of X with initial condition p. See Figure 2.7

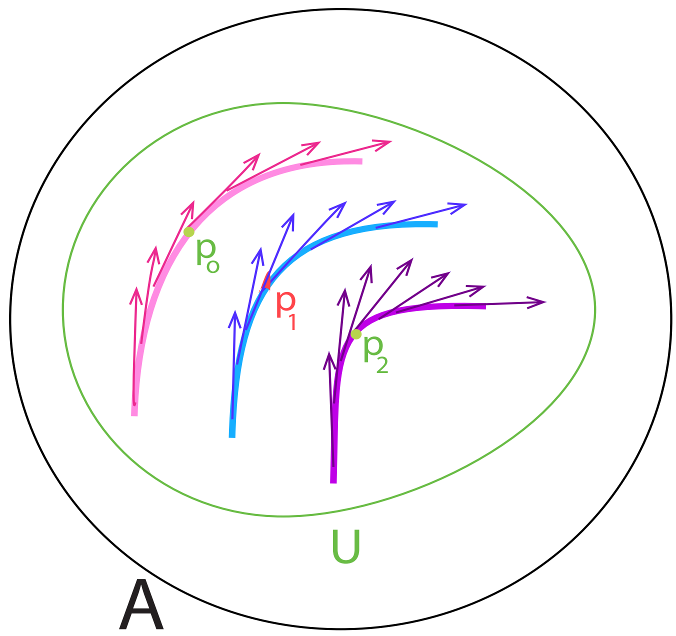

Figure 2.7: A portion of local flow $\varphi\colon J \times U \to \mathbb{R}^2$. If p is fixed and t varies, the flow moves along one of the colored curves. If t is fixed and p varies, p acts as a parameter for the individually colored curves.

The theory of ODE tells us that if X is C^1, then for every $p_0 \in A$, there is a pair (J, U) as above such that there is a *unique* C^1 local flow $\varphi\colon J \times U \to \mathbb{R}^n$ for X at p_0.

Let us now consider the special class of vector fields induced by matrices in $\mathrm{M}_n(\mathbb{R})$.

For any matrix $A \in \mathrm{M}_n(\mathbb{R})$, let X_A be the vector field given by

$$X_A(p) = Ap \quad \text{for all } p \in \mathbb{R}^n.$$

Such vector fields are obviously C^1 (in fact, C^∞).

The vector field induced by the matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

is shown in Figure 2.8. Integral curves are circles of center $(0, 0)$.

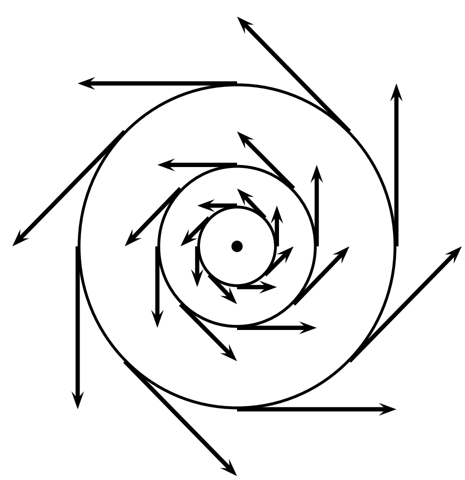

Figure 2.8: A vector field in $\mathbb{R}^2$

It turns out that the local flows of X_A are global, in the sense that $J = \mathbb{R}$ and $U = \mathbb{R}^n$, and that they are given by the matrix exponential.

Proposition 2.20. *For any matrix $A \in \mathrm{M}_n(\mathbb{R})$, for any $p_0 \in \mathbb{R}^n$, there is a unique local flow $\varphi\colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ for the vector field X_A given by*

$$\varphi(t, p) = e^{tA}p,$$

for all $t \in \mathbb{R}$ and all $p \in \mathbb{R}^n$.

For t fixed, the map $\Phi_t\colon p \mapsto e^{tA}p$ is a smooth diffeomorphism of $\mathbb{R}^n$ (with inverse given by e^{-tA}).

We can think of Φ_t as the map which, given any p, moves p along the integral curve γ_p from p to $\gamma_p(t) = e^{tA}p$.

For the vector field of Figure 2.8, each Φ_t is the rotation

$$e^{tA} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

The map $\Phi\colon \mathbb{R} \to \mathrm{Diff}(\mathbb{R}^n)$ is a group homomorphism, because

$$\Phi_s \circ \Phi_t = \Phi_{s+t} \quad \text{for all } s, t \in \mathbb{R}.$$

Observe that $\Phi_t(p) = \varphi(t, p)$.

If we hold p fixed, we obtain the integral curve with initial condition p, which is also called a *flow line* of the local flow.

If we hold t fixed, we obtain a smooth diffeomorphism of $\mathbb{R}^n$. The family $\{\Phi_t\}_{t\in\mathbb{R}}$ is called the *1-parameter group generated by X_A*, and Φ is called the *(global) flow generated by X_A*.

In the case of 2×2 matrices, it is possible to describe explicitly the shape of all integral curves; see Rossmann [42] (Section 1.1).