Some sources and inspiration for this treatment are the advanced calculus or analysis books by Dieudonné, Loomis & Sternberg, and Spivak, and notes and books by Milnor.

1. The derivative

**Definition.** Let $U \subset \mathbb{R}^m$ be an open set, $a \in U$, and $f : U \rightarrow \mathbb{R}^n$. The map $f$ is differentiable at $a$ if there is a linear map $\lambda \in \text{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ with

$$\lim_{x \to a} \frac{|f(x) - f(a) - \lambda(x - a)|}{|x - a|} = 0.$$

**Lemma.** If there is such a $\lambda$, it is unique.

**Proof.** Let $\lambda$ and $\lambda_1$ both satisfy the definition. Then

$$|(\lambda - \lambda_1)(x - a)| \leq |f(x) - f(a) - \lambda(x - a)| + | - f(x) + f(a) + \lambda_1(x - a)|$$

hence $|(\lambda - \lambda_1)(x - a)|/|x - a| \to 0$ as $x \to a$. For $v \neq 0$, letting $x = a + v \in U$,

$$|(\lambda - \lambda_1)(v)|/|v| = |(\lambda - \lambda_1)(tv)|/|tv| \to 0 \text{ as } t \to 0.$$ 

Therefore $\lambda(v) = \lambda_1(v)$.

When $f$ is differentiable at $a$ this unique linear map is denoted $Df(a)$.

2. The case $m = n = 1$

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ and assume $f'(a)$ exists. Then

$$\frac{|f(x) - f(a) - f'(a)(x - a)|}{|x - a|} = \left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| \rightarrow 0 \text{ as } x \to a$$

so $Df(a)(v) = f'(a)v$. The $1 \times 1$-matrix for the linear map $Df(a)$ has entry $f'(a)$.

3. The case $n = 1$ of real-valued functions, partial derivatives

**Proposition.** If $f : U \rightarrow \mathbb{R}$ is differentiable at $a \in U \subset \mathbb{R}^m$, then the partial derivatives of $f$ exist at $a$ and determine $Df(a)$. 

PROOF. Let $e_1, \ldots, e_m$ be the standard orthonormal basis for $\mathbb{R}^m$. Then

$$ \lim_{t \to 0} \left| \frac{f(a + te_i) - f(a)}{t} - Df(a)(e_i) \right| = \lim_{t \to 0} \left| \frac{f(a + te_i) - f(a) - D(f)(a)(te_i)}{|te_i|} \right| = 0,$$

hence the partial derivative with respect to the $i$th variable exists:

$$ \frac{\partial f}{\partial x_i}(a) = D_1f(a) = Df(a)(e_i) = \lim_{t \to 0} \frac{f(a + te_i) - f(a)}{t}.$$

If $v = \sum_i v_i e_i$, then $Df(a)v = \sum_i D_i f(a)v_i$.

More generally, the directional derivative is defined by

$$ D_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t}. $$

This limit may exist, in some or all directions, even if $f$ is not differentiable at $a$. The gradient of $f$ at $a$ is the vector $\nabla f(a) = \sum_i D_i f(a)e_i$ and, if $f$ is differentiable at $a$,

$$ Df(a)v = D_v f(a) = \nabla f(a) \cdot v. $$

For $f$ to be differentiable at $a$ it is necessary, but not sufficient, for the partial derivatives to exist at $a$. It is even necessary, but not sufficient, for the directional derivative to exist at $a$ for all $v$ and to define a linear function. A sufficient condition for $f$ to be differentiable is given by the following theorem, but this condition is not necessary.

**Theorem.** Let $f : U \to \mathbb{R}$, $U$ open in $\mathbb{R}^m$. Suppose the partial derivatives $D_i f$ are each continuous at $a \in U$. Then $f$ is differentiable at $a$ and $Df(a)v = \sum_i D_i f(a)v_i$.

**Proof.** Given $\varepsilon > 0$ there exists $\delta > 0$ such that

$$ |x - a| < \delta \Rightarrow |D_i f(x) - D_i f(a)| < \varepsilon \text{ for all } i.$$

Let $\xi_i = (x_1, \ldots, x_i, a_{i+1}, \ldots, a_m); \xi_0 = a, \xi_m = x$. Then $|\xi_i - a| < \delta$ and

$$ f(x) - f(a) = \sum_{i=0}^m f(\xi_i) - f(\xi_{i-1}). $$

Let $\varphi_i(t) = f(\xi_{i-1} + te_i)$. Then

$$ f(\xi_i) - f(\xi_{i-1}) = \varphi_i(x_i - a_i) - \varphi_i(0) = \varphi'(t_i)(x_i - a_i) = D_i f(\xi_{i-1} + t_i e_i)(x_i - a_i) $$

for some $t_i$ with $0 < t_i < x_i - a_i$, by the mean value theorem in one variable. Now

$$ \left| f(x) - f(a) - \sum D_i f(a)(x_i - a_i) \right| \leq \sum |f(\xi_i) - f(\xi_{i-1}) - D_i f(a)(x_i - a_i)| $$

$$ \leq \sum |f(\xi_i) - f(\xi_{i-1}) - D_i f(\xi_{i-1} + t_i e_i)(x_i - a_i)| + \sum |(D_i f(\xi_{i-1} + t_i e_i) - D_i f(a))(x_i - a_i)| $$

$$ \leq 0 + n\varepsilon |x - a|. $$

Hence

$$ \frac{|f(x) - f(a) - \lambda(x - a)|}{|x - a|} \to 0 \text{ as } x \to a \text{ where } \lambda \text{ is the linear map defined by } \lambda(v) = \sum D_i f(a)v_i. \text{ Therefore } f \text{ is differentiable at } a.
4. The derivative of linear and bilinear maps

**Lemma.** If \( f \) is a linear map then \( Df(a) = f \).

**Proof.** Since \( f \) is linear, \( f(x) - f(a) - f(x - a) = 0 \).

**Lemma.** If \( U, V, W \) are vector spaces and \( \beta : U \times V \longrightarrow W \) is bilinear, then

\[
D\beta(a, b)(u, v) = \beta(a, v) + \beta(u, b).
\]

**Proof.** Note that the map \( \ell(a, b) \) defined by \( \ell(a, b)(u, v) = \beta(a, v) + \beta(u, b) \) is linear from \( U \times V \longrightarrow W \) and

\[
\beta(a + u, b + v) - \beta(a, b) - \ell(a, b)(u, v) = \beta(u, v).
\]

The norm \(|(u, v)| = \sqrt{|u|^2 + |v|^2} \), and \(|u||v| \leq \max\{|u|^2, |v|^2\} \leq |u|^2 + |v|^2 \), hence

\[
\beta(u, v) = |u||v|\beta(u/|u|, v/|v|) \leq |(u, v)|^2 \beta(u/|u|, v/|v|), \text{ for } u \neq 0, v \neq 0.
\]

Therefore \(|\beta(u, v)| / |(u, v)| \to 0 \text{ as } (u, v) \to (0, 0)\).

**Examples of bilinear maps** \( \beta : \mathbb{R}^k \times \mathbb{R}^m \longrightarrow \mathbb{R}^n \).

\[
\ell = m = n = 1, \quad \beta(r, s) = rs
\]
\[
\ell = 1, \quad m = n, \quad \beta(r, u) = ru,
\]
\[
\ell = m, \quad n = 1, \quad \beta(u, v) = u \cdot v,
\]
\[
\ell = m = n = 3, \quad \beta(u, v) = u \times v.
\]

5. A norm on \( \text{Hom}(\mathbb{R}^m, \mathbb{R}^n) \)

Let \( e_1, \ldots, e_m \) be the standard orthonormal basis for \( \mathbb{R}^m \) and \( \bar{e}_1, \ldots, \bar{e}_n \) be the standard orthonormal basis for \( \mathbb{R}^n \). Let \( x = \sum_i x_i e_i \in \mathbb{R}^m \), so \( x_i = x \cdot e_i \). Let \( \ell \in \text{Hom}(\mathbb{R}^m, \mathbb{R}^n) \) and set \( \ell_i = \ell(e_i) \cdot \bar{e}_j \). Then \( \ell(x) = \sum_i x_i \ell(e_i) = \sum_j \sum_i \ell_i \cdot e_i \cdot \bar{e}_j \).

**Proposition.** If \(|\ell_i|^2 \leq k \) for all \( i, j \), then \(|\ell(x)| \leq \sqrt{mn}k|x| \).

**Proof.** By Cauchy’s inequality, \(|\sum_i \ell_i \cdot x_i| \leq \left\{ \sum_i (\ell_i)^2 \right\}^{1/2} |x| \leq \sqrt{m}k|x| \). Then

\[
|\ell(x)| = \left\{ \sum_j \left( \sum_i \ell_i \cdot x_i \right)^2 \right\}^{1/2} \leq \sqrt{mn}k|x|.
\]

The continuous real-valued function \(|\ell(x)| \) is bounded on the compact unit sphere, \(|x : |x| = 1| \subset \mathbb{R}^m \), and attains its bound.

**Definition.** For a linear map \( \ell \), define \( \|\ell\| = \sup\{|\ell(x)| : |x| = 1\} \).

**Corollary.** (i) \(|\ell(x)| \leq \|\ell\| |x| \) and (ii) \( \|\ell\| \leq \sqrt{mn}k \).
6. Lipschitz continuity of differentiable functions

Proposition. If \( f : U \rightarrow \mathbb{R}^n \) where \( U \) is open in \( \mathbb{R}^m \) and \( f \) is differentiable at \( a \), then there exist \( \delta > 0 \) and \( k > 0 \) such that \( |x - a| < \delta \Rightarrow |f(x) - f(a)| \leq k|x - a| \).

Proof. There is a linear map \( \lambda \) such that the function \( \varphi(x) = f(x) - f(a) - \lambda(x - a) \) satisfies \( |\varphi(x)|/|x - a| \rightarrow 0 \) as \( x \rightarrow a \). Therefore there is a \( \delta > 0 \) such that \( |\varphi(x)| \leq |x - a| \) for \( |x - a| < \delta \). Then \( |f(x) - f(a)| = |\lambda(x - a) + \varphi(x)| \leq (\|\lambda\| + 1)|x - a| \) for \( |x - a| < \delta \). Take \( k = \|\lambda\| + 1 \).

The conclusion of the Proposition is called Lipschitz continuity at \( a \); it implies that \( f \) is continuous at \( a \).

7. The chain rule

Theorem. If \( a \in U \subset \mathbb{R}^m \), \( b \in V \subset \mathbb{R}^n \), \( f : U \rightarrow V \), \( f(a) = b \), \( g : V \rightarrow \mathbb{R}^p \), \( f \) is differentiable at \( a \), and \( g \) is differentiable at \( b \); then \( g \circ f \) is differentiable at \( a \) and

\[
D(g \circ f)(a) = Dg(b) \circ Df(a).
\]

Proof. (See Spivak, p. 19.) Let \( \lambda = Df(a) \), \( \mu = Dg(b) \) and set

\[
\begin{align*}
\varphi(x) &= f(x) - f(a) - \lambda(x - a) \\
\psi(y) &= g(y) - g(b) - \mu(y - b) \\
\rho(x) &= g(f(x)) - g(b) - \mu(\lambda(x - a)).
\end{align*}
\]

We have

(i) \( |\varphi(x)|/|x - a| \rightarrow 0 \) as \( x \rightarrow a \),

(ii) \( |\psi(y)|/|y - b| \rightarrow 0 \) as \( y \rightarrow b \).

From the definitions,

\[
\begin{align*}
\rho(x) &= g(f(x)) - g(b) - \mu(f(x) - f(a) - \varphi(x)) \\
&= [g(f(x)) - g(b) - \mu(f(x) - f(a))] + \mu(\varphi(x)) \\
&= \psi(f(x)) + \mu(\varphi(x)).
\end{align*}
\]

First \( |\mu(\varphi(x))| \leq \|\mu\||\varphi(x)| \), so by (i) \( |\mu(\varphi(x))|/|x - a| \rightarrow 0 \) as \( x \rightarrow a \).

Second, by Proposition 6, there are \( k > 0, \delta > 0 \) such that

\( |x - a| < \delta \Rightarrow |f(x) - f(a)| \leq k|x - a| \).

By (ii), for any \( \varepsilon > 0 \) there is a \( \delta_1 > 0 \) such that

\( |f(x) - f(a)| < \delta_1 \Rightarrow |\psi(f(x))| < \varepsilon|f(x) - f(a)| \).

So for \( 0 \neq |x - a| < \min\{\delta, \delta_1/k\} \) we have \( |\psi(f(x))|/|x - a| < \varepsilon k \). Hence \( |\rho(x)|/|x - a| \rightarrow 0 \) as \( x \rightarrow a \) which gives the result.
8. Sample computations

(a) Let \( f(x) = x \cdot x = \beta \circ \Delta(x) \) where \( \Delta(x) = (x, x) \) is linear and \( \beta(x, y) = x \cdot y \). Then
\[
Df(a)(u) = D\beta(\Delta(a)) \circ D\Delta(a)(u) = D\beta(a)(u, u) = \beta(a, u) + \beta(u, a).
\]
Since \( \beta \) is symmetric, \( Df(a)(u) = 2a \cdot u \) and \( \text{grad } f(a) = 2a \).

If \( g(x) = |x - p| = \sqrt{f(x - p)} \),
\[
Dg(a)(u) = \frac{1}{2\sqrt{f(a - p)}} Df(a - p)(u) = \frac{a - p}{|a - p|} \cdot u \quad \text{for } a \neq p.
\]
So, for \( x \neq p \), \( \text{grad } g(x) = \frac{x - p}{|x - p|} \), the unit vector at \( x \) pointing away from \( p \).

(b) The derivative of a sum.

**Lemma.** Let \( f \) and \( g : U \rightarrow R^n \) be differentiable at \( a \in U \subset \mathbb{R}^m \).

Define \((f, g) : U \rightarrow \mathbb{R}^n \times \mathbb{R}^n \) by \((f, g)(x) = (f(x), g(x))\). Then
\[
D(f, g)(a) = (Df, Dg)(a).
\]

**Proof.** Let \( \lambda = Df(a) \), \( \varphi(x) = f(x) - f(a) - \lambda(x-a) \), \( \mu = Dg(a) \), and \( \psi(x) = g(x) - g(a) - \mu(x-a) \). Then \((\varphi, \psi)(x) = (f, g)(x) - (f, g)(a) - (\lambda, \mu)(x-a)\) and
\[
\frac{|(\varphi, \psi)(x)|}{|x-a|} = \sqrt{\frac{\varphi(x)^2}{|x-a|^2} + \frac{\psi(x)^2}{|x-a|^2}} \rightarrow 0 \text{ as } x \rightarrow a.
\]

Define the linear map \( s : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n \) by \( s(y_1, y_2) = y_1 + y_2 \). Now \((f + g)(x) = f(x) + g(x) = s \circ (f, g)(x)\). Hence the derivative of a sum is the sum of the derivatives:
\[
D(f + g) = Df + Dg.
\]

(c) The set \( M(n) \) of \( n \times n \)-matrices is an \( n^2 \)-dimensional vector space under addition and scalar multiplication and a ring under matrix multiplication. Let \( \beta(A, B) = AB \) and \( t(A) = A^t \) be the transpose. The maps \( t \) and \((I, t)\) are linear as maps of vector spaces where \( I \) is the identity linear map. On products \( t \) satisfies \( t(AB) = t(B)t(A) \). Define \( f : M(n) \rightarrow M(n) \) by \( f(A) = AA^t \), so \( f = \beta \circ (I, t) \).

Let \( O(n) \subset M(n) \) be the orthogonal group, \( O(n) = \{ A : f(A) = I \} \). Thus \( A \in O(n) \) means \( A \) is invertible and \( A^t = A^{-1} \).

**Exercise.** This is the computational part of a proof that \( O(n) \) is a manifold of dimension \( n(n-a)/2 \).

Show:
- \( f(A) \) is symmetric, \( f(A) = t(f(A)) \).
- \( Df(A)(M) = AM^t + MA^t \).
- If \( A \in O(n) \), then \( Df(A) \) maps \( M(n) \) onto the vector space of symmetric matrices.

[Hint: Given a symmetric \( S \), take \( M = \frac{1}{2} SA \).]
9. Differentiability of maps to $\mathbb{R}^n$

The results of §3 extend to maps to $\mathbb{R}^n$.

**Proposition.** If $f : U \rightarrow \mathbb{R}^n$ is differentiable at $a \in U$ then the partial derivatives of the components $D_jf_j$ exist at $a$ and are the entries in the matrix representing $Df(a)$. If all the partials are continuous at $a$ then $f$ is differentiable at $a$.

**Proof.** (See Spivak, p. 21, and for notation §§3, 5.) Define the linear projection map $\pi_j : \mathbb{R}^n \rightarrow \mathbb{R}$ by $\pi_j(y) = y \cdot \bar{e}_j$. The $j$th component of $f$ is $f_j = \pi_j \circ f$, $f(x) = \sum_j f_j(x)\bar{e}_j$ and

$$Df_j(a) = D\pi_j(f(a)) \circ Df(a) = \pi_j \circ Df(a).$$

The partial derivatives $\frac{\partial f}{\partial x_i}(a) = D_i f_j(a) = Df_j(a)(e_i) = Df(a)(e_i) \cdot \bar{e}_j$.

If $u = \sum_i u_i e_i$, then $Df(a)u = \sum_j \sum_i D_i f_j(a) u_i \bar{e}_j$.

Introducing the Jacobian matrix we write $Df(a)u$ as a matrix product:

$$Df(a)u = \begin{pmatrix} Df_1(a)u \\ \vdots \\ Df_n(a)u \end{pmatrix} = \begin{pmatrix} D_1f_1(a) & \ldots & D_mf_1(a) \\ \vdots & \ddots & \vdots \\ D_1f_n(a) & \ldots & D_mf_n(a) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}.$$  

If all the partials are continuous at $a$, by §3 each $D_i f(a)$ exists and by §8(b) $Df(a)$ exists.

When $m = 1$, $f(t)$ is a path in $\mathbb{R}^n$ and we define the velocity vector $f'(t) = Df(t)(e_1)$.

10. Mean value theorems

**Proposition.** If $U \subset \mathbb{R}^m$ is convex, $f : U \rightarrow \mathbb{R}$ is differentiable, and $a, x \in U$, then $f(x) - f(a) = Df(\zeta)(x - a)$ where $\zeta = a + t_0(x - a)$ for some $0 < t_0 < 1$.

**Proof.** Let $\varphi(t) = f(a + t(x - a))$. By the chain rule $\varphi'(t) = Df(a + t(x - a))(x - a)$.

By the one-variable mean value theorem

$$f(x) - f(a) = \varphi(1) - \varphi(0) = \varphi'(t_0) = Df(\zeta)(x - a)$$

where $\zeta = a + t_0(x - a)$ for some $0 < t_0 < 1$.

**Corollary.** If $\|Df(\zeta)\| \leq k$ for any $\zeta \in U$, then $|f(x) - f(a)| \leq k|x - a|$.

This follows from the Proposition and Corollary §5(i).

The Proposition is not true in general for maps to $\mathbb{R}^n$, $n > 1$. For example let $f : \mathbb{R} \rightarrow \mathbb{R}^3$ describe a helix about the vertical axis and take $x$ vertically above $a$. Then $x - a$ points straight up while $Df(t)(u)$ never does. The following Theorem extends the result of the Corollary to maps to $\mathbb{R}^n$. It says $f$ is Lipschitz continuous on $U$.

**Theorem.** If $U \subset \mathbb{R}^m$ is convex, $f : U \rightarrow \mathbb{R}^n$ is differentiable on $U$, $a, x \in U$, and

$$\left| \frac{\partial f_j}{\partial x_i}(a) \right| \leq \frac{k}{\sqrt{m}}$$

on $U$ for all $i, j$, then $|f(x) - f(a)| \leq k|x - a|$.

**Proof.** By the Proposition $f_j(x) - f_j(a) = Df_j(\zeta_j)(x - a)$. By §5 applied to the real-valued function $f_j$, $\|Df_j(\zeta_j)\| \leq \frac{k}{\sqrt{n}}$. By the Corollary, $|f_j(x) - f_j(a)| \leq \frac{k}{\sqrt{n}}|x - a|$. Then $|f(x) - f(a)| \leq k|x - a|$ as in §5.
10a. Alternate proof of the mean value theorem

In §10 we used the one-variable mean value theorem. The following proof gives both the Corollary and Theorem above without assuming the one-variable theorem and does not depend on bounds on the partial derivatives. See Loomis & Sternberg, p. 148, or Dieudonné, p. 153.

**Theorem.** Let \( f : [a, b] \to \mathbb{R}^n \) be continuous on \([a, b]\) and differentiable on \((a, b)\). Assume \( |f'(t)| \leq k \) for \( a < t < b \), where \((\text{see } \S \text{9})\) \( f'(t) = D_1 f(t)(e_1) \). Then

\[
|f(b) - f(a)| \leq k(b - a).
\]

**Proof.** Fix \( \varepsilon > 0 \). Let \( A = \{ x \in [a, b] : |f(x) - f(a)| \leq (k + \varepsilon)(x - a) + \varepsilon \} \).

(1) Since \( f \) is continuous at \( a \) there is a \( \delta > 0 \) such that

\[
|f(x) - f(a)| \leq \varepsilon \text{ for } a \leq x < a + \delta
\]

so \( x \in A \) for some \( x > a \).

(2) Set \( \ell = \sup A \). Either \( \ell \in A \) or for any \( \delta > 0 \) there is a \( t \) with \( \ell - \delta < t \leq \ell \) and \( t \in A \). But then by the continuity of \( f \) at \( \ell \), \( \ell \in A \).

(3) If \( \ell < b \) then \( f'(\ell) \) exists and \( |f'(\ell)| \leq k \). Hence there is a \( \delta > 0 \) such that

\[
\ell \leq t < \ell + \delta \Rightarrow |f(t) - f(\ell)| \leq (k + \varepsilon)(t - \ell).
\]

Then

\[
|f(t) - f(a)| \leq |f(t) - f(\ell)| + |f(\ell) - f(a)| \\
\leq (k + \varepsilon)(t - \ell) + (k + \varepsilon)(\ell - a) + \varepsilon \\
= (k + \varepsilon)(t - a) + \varepsilon.
\]

and hence \( t \in A \) for some \( t > \ell \), a contradiction. Therefore \( \ell = b \) and, as in (2), \( b \in A \).

Since \( \varepsilon > 0 \) is arbitrary, \( |f(b) - f(a)| \leq k(b - a) \).

**Corollary.** Let \( U \subseteq \mathbb{R}^n \) be convex, \( a, b \in U \), \( f : U \to \mathbb{R}^n \) be differentiable, and assume \( \|Df(x)\| \leq k \) for \( x \in U \). Then

\[
|f(b) - f(a)| \leq k|b - a|.
\]

**Proof.** Define \( c : \mathbb{R} \to \mathbb{R}^n \) by \( c(t) = tb + (1 - t)a \). Then \( c'(t) = b - a \) and \( f \circ c(1) - f \circ c(0) = f(b) - f(a) \). For \( 0 \leq t \leq 1 \), \( c(t) \in U \) and \( D(f \circ c)(t)(e_1) = Df(c(t))(b - a) \), so

\[
|(f \circ c)'(t)| \leq \|Df(c(t))\| \|b - a\| \leq k|b - a|.
\]

The result follows from the Theorem.
11. The inverse function theorem

**Definition.** A function $f : U \rightarrow \mathbb{R}^n$ is said to be of class $C^1$ if the partial derivatives exist and are continuous everywhere on $U$, $f$ is of class $C^k$ if the partial derivatives of orders $k$ and less are continuous, and $f$ is $C^\infty$ if it is $C^k$ for all positive integers $k$.

**Theorem.** Given $a \in U \subset \mathbb{R}^n$, $U$ open, and a $C^1$ function $f : U \rightarrow \mathbb{R}^n$ with $f(a) = b$ such that $Df(a)$ is invertible, there are neighborhoods $V$ of $a$, $V \subset U$, and $W$ of $b$ and a unique $C^1$ map $g : W \rightarrow V$ such that the restriction $f|V$ and $g$ are inverses. The derivative of $g$ is $Dg(y) = Df(g(y))^{-1}$. Further, if $f$ is $C^k$ ($1 \leq k \leq \infty$) then $g$ is also.

**Plan.** The map $g$ will need to satisfy $g(b) = a$. Let $g_0(y) = a$ be a first approximation to $g$. Since $Df(a)$ is invertible, the linear approximation to $f$, $y = f(x) \sim f(a) + Df(a)(x-a)$, can be solved for $x$. Let $g_1(y)$ be this solution: $g_1(y) = a + Df(a)^{-1}(y-b)$. We will define iteratively a sequence of functions $\{g_n\}$ converging to the local inverse of $f$.

**Proof.** (1) Define $F(x, y) = x + Df(a)^{-1}(y - f(x))$ on $U \times \mathbb{R}^n$. Let $D_1F(a, b)$ denote the derivative of the function $x \mapsto F(x, b)$ at $x = a$. Then
\[ F(a, b) = a + Df(a)^{-1}(b - f(a)) = a, \]
\[ D_1F(x, y) = I - Df(a)^{-1} \circ Df(x), \]
\[ D_1F(a, y) = I - Df(a)^{-1} \circ Df(a) = 0. \]

$D_1F(x, y)$ does not depend on $y$ and is the zero map for $x = a$. Hence for $x$ near $a$, $Df(x)$ is invertible and the entries in matrix $D_1F(x, y)$ are small. Choose $k > 0$ so that:

(i) $\overline{B_k(a)} \subset U$ and $Df(x)$ is invertible for $x \in \overline{B_k(a)}$, and
\[ \|D_1F(x, y)\| \leq \frac{1}{2} \text{ for } x \in \overline{B_k(a)}. \]

Then
\[ x, \xi \in \overline{B_k(a)} \Rightarrow |F(x, y) - F(\xi, y)| \leq \frac{1}{2}|x - \xi| \]

using the mean value theorem for the function $x \mapsto F(x, y)$. Since
\[ |F(a, y) - a| = |Df(a)^{-1}(y - b)| \leq \|Df(a)^{-1}\| |y - b|, \]

if we set $\delta = \frac{k}{2\|Df(a)^{-1}\|}$ we have:

(ii) $y \in \overline{B_\delta(b)} \Rightarrow F(a, y) \in B_{\delta/2}(a)$

and the same implication for the closed balls.

(2) Let $\mathcal{F}$ be the set of continuous functions $h : \overline{B_\delta(b)} \rightarrow \overline{B_k(a)}$ such that $h(b) = a$. For $h \in \mathcal{F}$ define $Th(y) = F(h(y), y)$. Then $Th(b) = F(a, b) = a$. For $y \in \overline{B_\delta(b)}$,
\[ |Th(y) - a| = |F(h(y), y) - a| \leq |F(h(y), y) - F(a, y)| + |F(a, y) - a| \leq \frac{1}{2}|h(y) - a| + \frac{k}{2} \leq k \quad \text{by (ii) and (iii)}. \]
Hence $Th(y) \in \overline{B_k(a)}$ so $Th \in \mathcal{F}$ and $T : \mathcal{F} \rightarrow \mathcal{F}$. The same argument, using the open version of (iii), shows $y \in B_\delta(b) \Rightarrow T_\gamma(y) \in B_k(a)$.

(3) $T$ has a fixed point.

Define a sequence of functions in $\mathcal{F}$ by $g_0(y) = a$ and $g_{n+1}(y) = Tg_n(y) = F(g_n(y), y)$. Note that $g_1$ is as defined in the plan. To shorten notation, temporarily fix $y$ and set $x_n = g_n(y)$. We have $x_0 = a$, $x_1 = F(a, y)$, and by (iii) $|x_1 - x_0| \leq k/2$.

\[
|x_{n+1} - x_n| = |F(x_n, y) - F(x_{n-1}, y)| \leq \frac{1}{2}|x_n - x_{n-1}| \leq \cdots \leq \frac{1}{2^n}|x_1 - x_0| \leq \frac{k}{2^{n+1}},
\]

\[
|x_m - x_n| \leq |x_m - x_{m-1}| + \cdots + |x_{n+1} - x_n| \leq \left( \frac{1}{2^m} + \cdots + \frac{1}{2^{n+1}} \right) k < \frac{k}{2^n},
\]

for $n < m$. Therefore $\{x_n\}$ is a Cauchy sequence.

Let $x = \lim x_n$. Since each $x_n \in B_k(a)$, $x \in \overline{B_k(a)}$. Define the map

$$g : \overline{B_\delta(b)} \rightarrow \overline{B_k(a)} \quad \text{by} \quad g(y) = x = \lim_{n \rightarrow \infty} g_n(y).$$

Since $|g(y) - g_n(y)| \leq \frac{k}{2^n}$, the sequence $\{g_n\}$ converges uniformly on $\overline{B_\delta(b)}$, so $g$ is continuous and $g \in \mathcal{F}$. Since $F$ is continuous, $Tg = g$:

$$g(y) = \lim g_n(y) = \lim F(g_n(y), y) = F(\lim g_n(y), y) = F(g(y), y) = Tg(y).$$

(4) $g$ is a unique local inverse of $f$.

Set $W = B_\delta(b)$ and $V = B_k(a) \cap f^{-1}(W) \subset U$. $V$ and $W$ are neighborhoods of $a$ and $b$ respectively. If $y \in W$, by (3) $Tg(y) = g(y)$ and by the definition of $Tg$, $g(y) = g(y) + Df(a)^{-1}(y - f(g(y)))$. Hence $f(g(y)) = y$. Then by (2), $g(y) \in V$, $g : W \rightarrow V$, and $f \circ g = 1_W$.

If $x, \xi \in V$ and $f(x) = f(\xi) = y \in W$, then $F(x, y) = x$, and $F(\xi, y) = \xi$. By (ii) $|x - \xi| \leq \frac{1}{2}|x - \xi|$, hence $x = \xi$. Therefore $f$ is one-to-one on $V$. If $x \in V$, let $y = f(x) \in W$ and let $\xi = g(f(x)) \in V$. Now $f(\xi) = f(g \circ f(x)) = f \circ g(f(x)) = f(x)$. Therefore $x = \xi$, $g(f(x)) = x$, and $g \circ f = 1_V$.

Let $h$ be another inverse of $f$ with $h(b) = a$. Let both $h$ and $g$ be defined on $W_1 \subset W$, and set $V_1 = B_k(a) \cap f^{-1}(W_1) \subset V$. For $y \in W_1$, let $x = g(y)$, and $\xi = h(y)$. Since $g$ and $h$ are right inverses of $f$, $f(x) = f(\xi)$. Since $f$ is 1-1, $x = \xi$ and hence $g = h$ on $W_1$.

(5) $g$ is Lipschitz continuous.

Let $g(y) = x$, $g(\eta) = \xi$ for $y, \eta \in B_\delta(b)$. Since $g = Tg$, $x = F(x, y)$ and $\xi = F(\xi, \eta)$. Then

$$|x - \xi| = |F(x, y) - F(\xi, \eta)|$$

$$\leq |F(x, y) - F(\xi, y)| + |F(\xi, y) - F(\xi, \eta)|$$

$$\leq \frac{1}{2}|x - \xi| + |Df(a)^{-1}(y - \eta)|$$

Therefore $\frac{1}{2}|x - \xi| \leq \|Df(a)^{-1}\| |y - \eta|$ and hence $|g(y) - g(\eta)| \leq 2\|Df(a)^{-1}\| |y - \eta|$. 

9
(6) $g$ is differentiable.
Since $f$ is $C^1$ and, by (i) $Df(\xi)$ is invertible for $\xi \in \overline{B_k(a)}$, we can choose $\kappa$ so that
\[
\|Df(\xi)^{-1}\| \leq \kappa \text{ for } \xi \in \overline{B_k(a)}.
\]
Let
\[
\varphi(x) = f(x) - f(\xi) - Df(\xi)(x - \xi).
\]
Then $|\varphi(x)|/|x - \xi| \to 0$ as $x \to \xi$, so for any $\epsilon > 0$, $|\varphi(x)| \leq \epsilon|x - \xi|$ for $x$ near $\xi$.
Let
\[
\psi(y) = g(y) - g(\eta) - Df(\xi)^{-1}(y - \eta) \\
= g(y) - g(\eta) - Df(\xi)^{-1}\{\varphi(x) + Df(\xi)(x - \xi)\} \\
= g(y) - g(\eta) - (x - \xi) - Df(\xi)^{-1}(\varphi(x)) \\
= -Df(\xi)^{-1}(\varphi(x)).
\]
Then
\[
|\psi(y)| \leq \kappa|\varphi(x)| \leq \kappa \epsilon|x - \xi| \text{ for } x \text{ near } \xi,
\]
\[
\leq 2\kappa^2 \epsilon|y - \eta| \text{ for } y \text{ near } \eta \text{ by (5)}.
\]
Hence $|\psi(y)|/|y - \eta| \to 0$ as $y \to \eta$. Therefore $g$ is differentiable at $\eta$ and $Dg(\eta) = Df(g(\eta))^{-1}$.

(7) If $f$ is $C^k$ so is $g$.
We can write $Dg$ as the composition $Dg = i \circ Df \circ g$ where $i(A) = A^{-1}$ is matrix inversion.
\[
B_\delta(b) \xrightarrow{g} U \xrightarrow{Df} G\ell(n) \xrightarrow{i} G\ell(n),
\]
where $g$ is continuous, $f$ is $C^k$ so that $Df$ is $C^{k-1}$, and $i$ is $C^\infty$ by Cramer’s rule. Since $g$ is continuous, the composition, $Dg$ is continuous, so $g$ is $C^1$. Now if $g$ is $C^j$ for any $j < k$, then similarly, $Dg$ is $C^j$, and $g$ is $C^{j+1}$. By induction $g$ is $C^k$, for $1 \leq k \leq \infty$.

This completes the proof of the inverse function theorem.

12. Applications of the inverse function theorem

**Implicit Function Theorem.** Let $(a, b) \in \mathbb{R}^k \times \mathbb{R}^n$. Let $f$ be a $C^1$ function from a neighborhood of $(a, b)$ to $\mathbb{R}^n$ with $f(a, b) = c$. Let $D_2f(a, b)$, the derivative of the function $y \mapsto f(a, y)$, be invertible.

Then there are neighborhoods $a \in U \subset \mathbb{R}^k$, $(a, b) \in V \subset \mathbb{R}^k \times \mathbb{R}^n$, and $c \in W \subset \mathbb{R}^n$ and a $C^1$ function $g : U \rightarrow \mathbb{R}^n$ such that $f(V) \subset W$ and

\[(x, y) \in V \text{ and } f(x, y) = c \iff x \in U \text{ and } y = g(x),
\]
\[Dg(x) = -D_2f(x, g(x))^{-1} \circ D_1f(x, g(x)).\]
Further there is a $C^1$ diffeomorphism $G : U \times W \to V$ such that, defining
\[ g_w(x) = \pi_2 \circ G(x,w), \quad \text{we have} \quad f(x,y) = w \iff y = g_w(x). \]
The function $\varphi_w : U \to V$ define by $\varphi_w(x) = G(x,w)$ parameterizes the level surface
\[ f^{-1}(w) = \{(x,y) \in V : f(x,y) = w\}. \]

**Proof.** Define $F$ on the domain of $f$ with values in $\mathbb{R}^k \times \mathbb{R}^n$ by $F(x,y) = (x,f(x,y))$. Then $F(a,b) = (a,c)$ and the Jacobian matrix of $DF(x,y)$ is
\[ \begin{pmatrix} I & 0 \\ L & M \end{pmatrix} \]
where
\[ L = D_1f = \frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_k)} \quad \text{and} \quad M = D_2f = \frac{\partial(f_1, \ldots, f_n)}{\partial(y_1, \ldots, y_n)}. \]

Since $M(a,b)$ is invertible, $DF(a,b)$ is invertible.

The inverse function theorem gives a map $G$ which we may assume is defined on a product neighborhood $U \times W \subset \mathbb{R}^k \times \mathbb{R}^n$ of $(a,c)$. Let $V = G(U \times W)$. Then $F|V$ and $G|U \times W$ are inverses. If $(x,y) \in V$ and $F(x,y) = (x,f(x,y)) = (x,w) \in U \times W$, then $G(x,w) = (x,y)$ and $f(x,y) = w$. Define $g_w(x) = \pi_2 \circ G(x,w) = y$. Then $f(x, g_w(x)) = f(x,y) = w$. For the case $f(x,y) = c$, take $g = g_c$.

Since $F$ has a $C^1$ inverse on $V$, it follows that $DF$ is invertible on $V$ and, from the form of its Jacobian matrix, that the matrix $M(x,y)$ of $D_2f(x,y)$ is also invertible. As a composition, $g_w(x)$ is differentiable. Differentiating $f(x,g_w(x)) = w$ with respect to $x$ using the chain rule we get
\[
D_1f(x,g_w(x)) + D_2f(x,g_w(x)) \circ Dg_w(x) = 0, \quad \text{hence} \quad Dg_w(x) = -D_2f(x,g_w(x))^{-1} \circ D_1f(x,g_w(x)).
\]

Notice that $V$ is not a product, the slice $\{y \in \mathbb{R}^n : (x,y) \in V\}$ depends on $x$.

**Proposition 1.** Let $p \in \mathbb{R}^m$ and let $f$ be a $C^1$ map on a neighborhood of $p$ to $\mathbb{R}^n$, $m \geq n$, with $Df(p)$ surjective. Then there is a neighborhood $p \in V \subset \mathbb{R}^m$ and a diffeomorphism $h : U \to V$, $U$ open in $\mathbb{R}^m$, such that $f \circ h(x_1, \ldots, x_m) = (x_{m-n+1}, \ldots, x_m)$ or $f \circ h = \pi_2$.

**Proof.** Let $m = k + n$. Since $Df(p)$ is surjective we can reorder the variables, i.e. the coordinates of $\mathbb{R}^m$, $x_1, \ldots, x_m$, so that the Jacobian matrix of derivatives with respect to the last $n$ variables is invertible. Then the implicit function theorem applies: the map $F(x) = (x_1, \ldots, x_k, f(x))$ restricted to a neighborhood $V$ of $a$ has an inverse $h : U \to V$. Then $F \circ h(z) = z$ and $f \circ h = \pi_2 \circ F \circ h = \pi_2$.
Proposition 2. Let \( a \in U \subset \mathbb{R}^m \) be open and \( f : U \rightarrow \mathbb{R}^n \) be a \( C^1 \) map, \( m \leq n \), with \( Df(a) \) injective. Then there are neighborhoods \( a \in U_1 \subset U \), \( V \subset \mathbb{R}^n \) with \( f(U_1) \subset V \), and \( b \in W \subset \mathbb{R}^n \) and a diffeomorphism \( h : V \rightarrow W \) such that \( h \circ f(x_1, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, 0) \).

Proof. The Jacobian matrix of \( Df(a) \) has an invertible \( m \times m \) submatrix \( A \). We may permute the coordinate functions, \( f_1, \ldots, f_n \), i.e. the coordinates in the range \( \mathbb{R}^n \), so that the first \( m \) rows of the Jacobian of \( f \) are an invertible matrix \( A \).

Define \( F : U \times \mathbb{R}^{n-m} \rightarrow \mathbb{R}^n \) by
\[
F(x_1, \ldots, x_n) = f(x_1, \ldots, x_m) + (0, \ldots, 0, x_{m+1}, \ldots, x_n)
\]
Then \( F(a, 0) = f(a) + 0 = b \) and
\[
DF(a, 0) = \begin{pmatrix} A & 0 \\ B & I \end{pmatrix}
\]
which is invertible. By the inverse function theorem there are neighborhoods \( (a, 0) \in V \subset U \times \mathbb{R}^{n-m} \) and \( b \in W \subset \mathbb{R}^n \) and a map \( h : W \rightarrow V \) inverse to \( F|V : V \rightarrow W \).

Set \( i(x_1, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, 0) \), so \( F \circ i = f \). Let \( U_1 = i^{-1}(V) \). On \( U_1 \)
\[
h \circ f = h \circ F \circ i = i.
\]

Think of \((h, W)\) as a new coordinate chart for \( \mathbb{R}^n \) with respect to which the map \( f \) has the simplest possible form: \( h \circ f = i \).

It follows that \( f|U_1 \) is a homeomorphism onto its image in the induced topology. That is \( O \) is open in \( U_1 \) if and only if \( f(O) \) is the intersection with \( f(U_1) \) of an open set in \( \mathbb{R}^n \).

13. Differential equations

For a continuous function \( g : J \rightarrow \mathbb{R}^n \) on an interval \( J \subset \mathbb{R} \) with \( t_0, t \in J \), we introduce the integral
\[
G(t) = \int_{t_0}^{t} g(s) \, ds
\]
defined componentwise by \( G_i(t) = \int_{t_0}^{t} g_i(s) \, ds \). \( G \) is \( C^1 \) and \( G'(t) = g(t) \) by the fundamental theorem of calculus. If \(|g(t)| \leq k\) on \( I \), then
\[
\left| \int_{t_0}^{t} g(s) \, ds \right| = |G(t) - G(t_0)| \leq k|t - t_0|
\]
by the mean value theorem, §10a. We will also need the following stronger result.

Lemma. \( \left| \int_{t_0}^{t} g(s) \, ds \right| \leq \int_{t_0}^{t} |g(s)| \, ds \).
Proof. Since \( g \) is continuous, \(|g|\) is integrable. Let \( P = \{t_0, \ldots, t_n\} \) be a partition of \([t_0, t]\) and let \( M_i = \sup\{|g(s)| : t_{i-1} \leq s \leq t_i\} \). Then, by (1),
\[
\left| \int_{t_0}^{t} g(s) \, ds \right| \leq \sum_{i=1}^{n} \left| \int_{t_{i-1}}^{t_i} g(s) \, ds \right| \leq \sum_{i=1}^{n} M_i(t_i - t_{i-1}) = U(|g|, P),
\]
hence the left hand term is a lower bound for the set of all upper sums for \(|g|\).

If \( f : U \rightarrow \mathbb{R}^n \) is continuous, \( 0 \in J \), and \( g : J \rightarrow U \) is differentiable, we say \( x = g(t) \) is a solution to the differential equation \( x' = f(x) \) with initial condition \( x_0 \) if
\[
(2') \quad g'(t) = f(g(t)) \quad \text{and} \quad g(0) = x_0.
\]

By the fundamental theorem of calculus, it is equivalent that \( g \) be continuous and satisfy the integral equation
\[
(2) \quad g(t) = x_0 + \int_0^t f(g(s)) \, ds.
\]

For a continuous function \( g : J \rightarrow U \) define a map \( T \) which takes \( g \) to a new function \( Tg \) defined by
\[
Tg(t) = x_0 + \int_0^t f(g(s)) \, ds.
\]
Then \( g \) is a solution to (2) if and only if \( Tg = g \). The plan is to use \( T \) to construct a sequence of functions which converges uniformly to a solution to (2). This method is known as Picard iteration.

Theorem. Let \( f : U \rightarrow \mathbb{R}^n \) be \( C^1 \) and let \( a \in U \). Then there is a \( \delta > 0 \) and a unique \( C^1 \) function from the interval \( J = (-\delta, \delta) \) to \( U \) satisfying (2).

Further there exists \( c > 0 \) and \( g : J \times \overline{B_c(a)} \rightarrow U \) such that the curve \( g_x(t) = g(t, x) \) is a solution to (2) with initial condition \( g_x(0) = x \).

Proof. Let \( a \in V \subset \overline{V} \subset U \) with \( V \) open and \( \overline{V} \) compact. Since \( f \) is \( C^1 \), \( f \) and \( Df \) are bounded on \( \overline{V} \). Let \( |f(x)| \leq k \) and \( \|Df(x)\| \leq L \) for \( x \in \overline{V} \). Choose \( \delta > 0 \) such that
\[
(i) \quad r = \delta L < 1 \quad \text{and} \quad B_{2\delta k}(a) \subset \overline{V}.
\]
Set \( c = \delta k \). Let \( \mathcal{F} \) be the set of continuous functions \( h : J \rightarrow \overline{B_{2c}(a)} \) such that
\[
(ii) \quad \text{if} \quad x = h(0), \quad \text{then} \quad x \in \overline{B_c(a)} \quad \text{and} \quad h(J) \subset \overline{B_c(x)}.
\]
For \( x \in \overline{B_c(a)} \), set \( \mathcal{F}_x = \{h \in \mathcal{F} : h(0) = x\} \). For \( h \in \mathcal{F} \) define
\[
(3) \quad Th(t) = h(0) + \int_0^t f(h(s)) \, ds.
\]
Then, by (1),
\[ |Th(t) - h(0)| = \left| \int_0^t f(h(s)) \, ds \right| \leq |t| \delta k = c, \]
therefore \( Th \in F_x \) and \( T : F_x \rightarrow F_x \).

For any continuous, bounded function \( g : J \rightarrow \mathbb{R}^n \) define the norm
\[ \|g\| = \sup \{|g(t)| : t \in J\}. \]

Warning: for linear functions \( \lambda \) will still use the norm defined in §5.

If \( g \) and \( h \) are functions in \( F_x \), then
\[ \|g - h\| = \|h - g\|, \]
\[ \|g - h\| = 0 \iff g = h, \]
\[ \|h_1 - h_3\| \leq \|h_1 - h_2\| + \|h_2 - h_3\|. \]

The third property is called the triangle inequality by analogy with the formula for distances between points in the plane. To prove it notice that for any \( t \in J \)
\[ |h_1(t) - h_3(t)| \leq |h_1(t) - h_2(t)| + |h_2(t) - h_3(t)| \leq \|h_1 - h_2\| + \|h_2 - h_3\|, \]
so the left hand side is bounded by the right hand side. These three properties make \( F \) a metric space with the distance between \( g \) and \( h \) given by \( \|g - h\| \).

For \( f, g \in F \) and \( t, s \in J \) we have
\[ |f(g(s)) - f(h(s))| \leq L|g(s) - h(s)| \leq L\|g - h\|, \quad \text{by §10a} \]
\[ \left| \int_0^t f(g(s)) - f(h(s)) \, dt \right| \leq L\|g - h\| |t| \leq L\delta \|g - h\|. \quad \text{by (1)} \]

If also \( g(0) = h(0) \), then
\[ |Tg(t) - Th(t)| \leq L\delta \|g - h\| \leq r\|g - h\|, \quad \text{by (i)} \]
\[ \|Tg - Th\| \leq r\|g - h\|. \quad \text{(5)} \]

Since \( r < 1 \), \( T \) is called a contraction map; \( T \) moves points (functions) closer together.

We will prove that the sequence \( g_n \in F_x \) defined inductively by
\[ g_0(t) = x \quad \text{and} \quad g_n = Tg_{n-1} \]
converges uniformly to a function \( g \) satisfying (3) with initial condition \( g(0) = x \). First \( g_0 \in F_x \), and hence \( g_n \in F_x \). Then (5) implies \( \|g_2 - g_1\| \leq r\|g_1 - g_0\| \) and inductively
\[ \|g_n - g_{n-1}\| \leq r^{n-1}\|g_1 - g_0\|. \]
Therefore, with \( m < n \),
\[
\|g_n - g_m\| \leq \|g_n - g_{n-1}\| + \cdots + \|g_{m+1} - g_m\|
\leq (r^{n-1} + \cdots + r^m)\|g_1 - g_0\|
\leq (r^m + r^{m+1} + \cdots)\|g_1 - g_0\|
\leq \frac{r^m}{1-r}\|g_1 - g_0\|.
\]

Since \( r^m \to 0 \) as \( m \to \infty \), this shows the sequence \( g_n \) is uniformly Cauchy and hence converges uniformly to a continuous function \( g \) which lies in \( \mathcal{F}_x \).

We need to show that \( g \) is a fixed point of \( T \). Since \( g_n \) converges uniformly to \( g \), it follows from (4) that \( f(g_n(t)) \) converges uniformly to \( f(g(t)) \). Then
\[
(Tg)(t) = x + \int_0^t \lim_{n \to \infty} f(g_n(s)) \, ds = \lim_{n \to \infty} \{ x + \int_0^t f(g_n(s)) \, ds \} = \lim_{n \to \infty} g_{n+1}(t) = g(t).
\]

Hence \( g \) is a solution to our differential equation on the interval \( J \).

If \( h \) were another solution to (2) on an interval \( J_1 \subset J \) with \( h(0) = g(0) \), then \( Th = h \) and \( Tg = g \) on \( J_1 \). Using the norm on \( J_1 \), \( \|g - h\| = \|Tg - Th\| \leq r\|g - h\| \) and \( r < 1 \) imply \( \|g - h\| = 0 \) and hence \( g = h \) on \( J_1 \).

Denote the constructed solution defined for \( t \in J \) and \( x \in \overline{B_c(a)} \) by \( g_x(t) \). By the fundamental theorem of calculus \( g_x(t) \) is differentiable in \( t \) and, since \( g_x'(t) = f(g_x(t)) \), \( g_x \) is \( C^1 \). Set \( g(t, x) = g_x(t) \).

14. Flows

The \( C^1 \) function \( f : U \to \mathbb{R}^n \) is pictured as a vector field on \( U \), that is, an assignment to each \( x \in U \) of a vector \( \vec{v}_x = f(x) \) “based” at the point \( x \). For any \( a \in U \), a solution \( g(t) : J \to U \) is pictured as a point moving along a path so that at time \( t \) the moving point is at \( g(t) \) and its velocity is \( g'(t) = f(g(t)) \). Each moving point that passes through a given point \( x \) has the same velocity, \( f(x) \), at the time it is at \( x \). This motion is called a steady flow. If \( f \) depended on \( t \) and \( x \), we would have a time-dependent flow.

In §13 we proved the existence, for any \( x \in U \) and for a short time depending on \( x \), of a unique flow \( C^1 \) in \( t \). In this section we will give some more global results on the flow for a given \( f \).

1. Let \( g \) satisfy \( g'(s) = f(g(s)) \) for \( s \in J \). Let \( s, s + t \in J \) and define \( h(t) = g(s + t) \). Since
\[
h'(t) = g'(s + t) \quad \text{by the chain rule}
\]
\[
= f(g(s + t))
\]
\[
= f(h(t)),
\]
h is a solution in an interval about 0 with \( h(0) = g(s) \).
(2) Let \( g_i(t) \) be a solution for \( t \in J_i, \ i = 1, 2 \), satisfying \( g_1(0) = g_2(0) \). Then \( g_1(t) = g_2(t) \) for all \( t \in J_1 \cap J_2 \).

Let \( J^* = \{ t \in J_1 \cap J_2 : g_1(t) = g_2(t) \} \). \( J^* \neq \emptyset \) since \( 0 \in J^* \). We will show that \( J^* \) is both open and closed in \( J_1 \cap J_2 \) and therefore \( J^* = J_1 \cap J_2 \). By the uniqueness result in §13, there is an open neighborhood \( J_0 \subset J^* \) containing 0. If \( s \in J^* \), by (1) there are solutions \( h_i(t) \) with \( h_i(0) = g_i(s) \). Since \( s \in J^* \), \( h_1(0) = h_2(0) \), and by uniqueness \( h_1(t) = h_2(t) \) in a neighborhood of 0. Hence \( g_1 = g_2 \) in a neighborhood of \( s \) and therefore \( J^* \) is open. If \( s \in J_1 \cap J_2 \) but \( s \notin J^* \) then, since \( \mathbb{R}^n \) is Hausdorff, there are disjoint neighborhoods \( U_i \) of \( g_i(s) \). Then \( s \in g_1^{-1}(U_1) \cap g_2^{-1}(U_2) \), an open set in \( J_1 \cap J_2 = J^* \). Therefore \( J^* \) is closed in \( J_1 \cap J_2 \). Since \( J_1 \cap J_2 \) is connected, \( J^* = J_1 \cap J_2 \).

(3) A maximal solution. Under the hypotheses of (2), define a \( C^1 \) map \( g : J_1 \cup J_2 \to U \) by

\[
g(t) = \begin{cases} 
  g_1(t) & \text{if } t \in J_1 \\
  g_2(t) & \text{if } t \in J_2.
\end{cases}
\]

This construction is just the union of the two functions \( g_1 \) and \( g_2 \) where a function is regarded as its graph in the product \( \mathbb{R} \times U \). For \( x \in U \), let \( S \) be the set of all graphs of solutions defined on intervals about 0 with initial point \( x \) and let \( g_x \) be the union of the elements of \( S \). This \( g_x \) is defined on the maximal interval \( J_x \) for a solution with initial point \( x \). Let \( \Omega = \{(t, x) \in \mathbb{R} \times U : t \in J_x \} \). Define \( g : \Omega \to U \) by \( g(t, x) = g_x(t) \).

(4) If \( g_a(t) \) is a solution defined on the maximal \( J_a \) with \( g_a(0) = a \), choose \( s \in J_a \) and let \( g_a(s) = b \in U \). For \( t \) such that \( s + t \in J_a \) define \( h(t) = g(s + t) \); \( h \) is defined on the interval \( \{ t : s + t \in J_a \} \). As in (1) \( h \) is a solution with \( h(0) = b \). Let \( g_b \) be the solution on the maximal interval \( J_b \) with \( g_b(0) = b \). By (2) \( g_b(t) = h(t) \) on the intersection of their intervals of definition. Then \( g_a(s + t) = g_b(t) \) where \( b = g_a(a) \). In terms of \( g : \Omega \to U \) we have \( g(t + s, a) = g(t, g(s, a)) \).

(5) The function \( \varphi_t(x) = g(t, x) \) is called the flow for time \( t \). For each \( x \in U \), \( \varphi_t(x) \) is defined for \( t \in J_x \). By §13 for each \( x \in U \) there is an interval \( (-\delta, \delta) \) and a neighborhood \( N_x \) of \( x \) such that \( y \in N_x \Rightarrow (-\delta, \delta) \subset J_y \). For all \( x \in U \), \( \varphi_0(x) = x \). The result of (4) restated in terms of \( \varphi \) and with \( x \) playing the role of \( a \) is:

\[
\varphi_{t+s}(x) = \varphi_t \circ \varphi_s(x) \text{ for all } x \text{ such that } s, t + s \in J_x,
\]

\( \varphi \) is said to be a local one-parameter group. When \( \varphi \) is defined it takes a neighborhood of 0 in the abelian group \( \mathbb{R} \) to the set of self maps of \( U \). We have not yet proved that \( \varphi_t \) is continuous. However, if \( s \in J_x \), then \( s + (-s) = 0 \in J_x \) and \( \varphi_{-s} \circ \varphi_s(x) = x \). Therefore \( \varphi_{-s} = \varphi_s^{-1} \), so \( \varphi_s \) is a bijection. The associative law is automatic for maps under composition, hence, except for the problem of where these maps are defined, they form an group and, for small \( t, t \mapsto \varphi_t \) is a homomorphism—hence a local group.

(6) We next show that \( \varphi_t(x) = g(t, x) \) is a continuous function of \( x \).
LEMMA 1. Given \( g : J \rightarrow U \) and \([0, t_1] \subset J \) there is an open set \( V \subset \nabla \subset U \) with \( \nabla \) compact and a \( c > 0 \) such that for any \( s \in [0, t_1] \), \( B_{2c}(g(s)) \subset V \).

PROOF. Since \( U \) is open, for each \( s \in [0, t_1] \) there is a \( c_s > 0 \) with \( B_{3c_s}(g(s)) \subset U \). The set of smaller, open balls, \( \{ B_{c_s}(g(s)) : s \in [0, t_1] \} \) covers \( g([0, t_1]) \). Since \( g([0, t_1]) \) is compact, a finite subset of these balls covers \( g([0, t_1]) \), say the balls corresponding to \( s \) in the finite set \( \{ s_1, \ldots, s_m \} \subset [0, t_1] \). Let

\[
c_i = c_{s_i}, \quad c = \min \{ c_i : 1 \leq i \leq m \}, \quad \text{and} \quad V = \bigcup_{i=1}^{m} B_{3c_i}(g(s_i))
\]

If \( |x - g(s)| \leq 2c \) there is an \( i \) with \( |g(s) - g(s_i)| < c_i \) hence \( |x - g(s)| < 3c_i \). Therefore \( x \in V \).

LEMMA 2. Let \( \nu : [0, t_1] \rightarrow \mathbb{R} \) be continuous, \( t_1 > 0 \), and \( \nu(t) \geq 0 \). If there is an \( L \geq 0 \) such that

\[
\nu(t) \leq \nu(0) + \int_{0}^{t} L \nu(s) \, ds \quad \text{for} \quad 0 \leq t \leq t_1.
\]

Then \( \nu(t) \leq \nu(0)e^{Lt} \) on \([0, t_1]\).

PROOF. First assume \( C = \nu(0) > 0 \). Set

\[
\mu(t) = C + \int_{0}^{t} L \nu(s) \, ds.
\]

Then \( \nu(t) \leq \mu(t) \), \( 0 < \mu(t) \), and \( \mu(0) = C \), hence:

\[
\frac{\mu'(t)}{\mu(t)} = \frac{L\nu(t)}{\mu(t)} \leq L,
\]

\[
\int_{0}^{t} \frac{\mu'(s)}{\mu(s)} \, ds \leq \int_{0}^{t} L \, ds = Lt,
\]

\[
\log \mu(t) \leq \log \mu(0) + Lt,
\]

\[
\mu(t) \leq Ce^{Lt}.
\]

The Lemma also holds for \( C = 0 \) because it holds for arbitrarily small \( C > 0 \).

PROPOSITION. Let \( a \in U \) and \( f : U \rightarrow \mathbb{R}^n \) be \( C^1 \). Let \( g_a : J_a \rightarrow U \) be a solution to §13(2') with initial value \( a \) on the maximal interval \( J_a \). Let \( [0, t_1] \subset J_a \). Then there exists \( \rho > 0 \) such that \( \varphi_t \) is defined and is Lipschitz continuous on \( \overline{B_{\rho}(a)} \) for \( t \in [0, t_1] \). Further, \( \Omega = \{ (t, x) \in \mathbb{R} \times U : t \in J_x \} \) is open and \( g : \Omega \rightarrow U \) is continuous.

PROOF. Let \( c > 0 \) and \( V \subset U \) with \( \overline{B_{2c}(g(s))} \subset V \) for \( s \in [0, t_1] \) be as constructed in Lemma 1. Let \( |f(x)| \leq k \) and \( \|Df(x)\| \leq L \) for \( x \in V \) as in Theorem §13. Choose \( \rho > 0 \) such that \( \rho e^{Lt_1} \leq 2c \). Let \( x \in \overline{B_{\rho}(a)} \) and \( g_x : J_x \rightarrow U \) be the maximal solution. Set \( \nu(t) = |g_a(t) - g_x(t)| \) on \([0, t_1] \cap J_x \). Then

\[
\nu(t) - \nu(0) = \int_{0}^{t} f(g_a(s)) - f(g_x(s)) \, ds \leq \int_{0}^{t} L \nu(s) \, dt.
\]
so, by Lemma 2, \( \nu(t) \leq \rho e^{Lt} \leq \rho e^{Lt_1} \leq 2c \), hence \( g_x(t) \in V \).

If \( t_1 \not\in J_x \), let \( t^* = \sup J_x \leq t_1 \). Then \( b = g_x(t^*) \in V \). By (2) the solution \( g_b(t) \) is defined in a neighborhood of 0 and can be used to extend \( g_x(t) \) to a neighborhood of \( t^* \). This contradicts \( t^* = \sup J_x \leq t_1 \) and hence \( t_1 \in J_x \). Hence for all \( x \in \overline{B}_\rho(a) \), \([0, t_1] \subset J_x \).

Now, given \( x, y \in B_\rho(a) \), we have \( g_x \) and \( g_y \) defined on \([0, t_1] \). Let \( \nu(t) = |g_x(t) - g_y(t)| \). Again \( |g_x(t) - g_y(t)| \leq |x - y|e^{Lt} \), so \( \varphi_t \) is Lipschitz on \( \overline{B}_\rho(a) \) for \( t \in [0, t_1] \).

Finally, for any \( (t, a) \in \Omega \), take \( t_1 > t \) with \( t_1 \in J_a \) and let \( s < t_1 \). Then

\[
|g_a(s) - g_x(t)| \leq |g_a(s) - g_x(s)| + |g_x(s) - g_x(t)| \leq e^{Lt_1}|a - x| + k|s - t|
\]

so \( g : \Omega \longrightarrow U \) is continuous at every point \( \Omega \).

The proof also shows for \( x \in B_\rho(a) \), \( J_a \subset J_x \) from which it follows that \( \Omega \) is open.