Real and Complex Analysis

Paul D. Mitchener

February 4, 2010

Contents

1 Metric and Normed Spaces

1.1 Metric Spaces

Recall that a *metric space* (X, d) is a set X equipped with a *distance function* or *metric* $d: X \times X \to \mathbb{R}$ such that, for all $x, y, z \in X$:

- $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$. **(Positivitity)**
- $d(x, y) = d(y, x)$. **(Symmetry)**
- $d(x, z) \leq d(x, y) + d(y, z)$. **(The Triangle Inequality)**

Example 1.1 The real line, $\mathbb{R}$, is a metric space, with metric $d(a, b) = |a - b|$.

The proof of this fact is completely straightforward.

Example 1.2 The complex numbers, $\mathbb{C}$, form a metric space, with metric

$$d(w, z) = |w - z| = \sqrt{(w_1 - z_1)^2 + (w_2 - z_2)^2}$$

The proof of this fact is left as an exercise.

Example 1.3 Let (X, d) be a metric space. Let $Y \subseteq X$. Then the set Y is also a metric space, with metric defined by restricting the metric on X, that is to say

$$d_Y(x, y) = d(x, y)$$

for all points $x, y \in Y$.

When we encounter subsets of metric spaces, we will in general assume that we have this metric.

Definition 1.4 Let X and Y be metric spaces. We call a function $f: X \to Y$ *continuous at the point* $x \in X$ if for all $\varepsilon > 0$ there exists $\delta > 0$ such that $d(f(x), f(y)) < \varepsilon$ whenever $d(x, y) < \delta$.

We call the function $f: X \to Y$ *continuous* if it is continuous at every point $x \in X$.

Definition 1.5 Let X be a metric space. A sequence of points $(x_n)_{n\in\mathbb{N}}$ in the space X is said to *converge* to the point $x \in X$ if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ whenever $n \geq N$.

We write

$$\lim_{n\to\infty} x_n = x$$

when a sequence (x_n) converges to a point x. We say the point x is the *limit* of the sequence (x_n). We slso write $x_n \to x$ as $n \to \infty$.

One also refers to limits of function. Let X and Y be metric spaces, and let $f: X\backslash\{x_0\} \to Y$. Then we say the function f has *limit* a at the point x_0, and write

$$\lim_{x\to x_0} f(x) = a$$

if for all $\varepsilon > 0$ there exists $\delta > 0$ such that $d(f(x), a) < \varepsilon$ whenever $d(x, x_0) < \delta$ (and $x \neq x_0$).

By definition, a function $f: X \to Y$ is continuous at the point $x_0 \in X$ if and only if the value $f(x_0)$ is defined, and

$$\lim_{x\to x_0} f(x) = f(x_0)$$

Proposition 1.6 *Let $f: X \to Y$ be a function between metric spaces. The function f is continuous at the point $x \in X$ if and only if, for any sequence (x_n) with limit x, the sequence $(f(x_n))$ in the space Y has limit $f(x)$.*

Proof: Let f be continuous at the point x. Let (x_n) be a sequence with limit x.

Choose $\varepsilon > 0$. Then we have $\delta > 0$ such that $d(f(x), f(y)) < \varepsilon$ whenever $d(x, y) < \delta$.

Now we can find $N \in \mathbb{N}$ such that $d(x_n, x) < \delta$ whenever $n \geq N$. Hence $d(f(x_n), f(x)) < \varepsilon$ whenever $n \geq N$. We see that the sequence $(f(x_n))$ converges to $f(x)$.

To prove the converse, suppose that the function f is *not* continuous at the point x. Then we can find $\varepsilon > 0$ such that there is no $\delta > 0$ with the property that $d(x, y) < \delta$ implies $d(f(x), f(y)) < \varepsilon$.

To rephrase this statement, for all $\delta > 0$ there is a point $y \in X$ such that $d(x, y) < \delta$, but $d(f(x), f(y)) \geq \varepsilon$.

Take $\delta = \frac{1}{n}$. Then we have a point $x_n \in X$ such that $d(x, x_n) < \frac{1}{n}$, but $d(f(x), f(x_n)) \geq \varepsilon$. We see that the sequence (x_n) converges to x, but the sequence $(f(x_n))$ does not converge to $(f(x))$. □

Let (X, d) be a metric space. Given a point $x \in X$ and a real number $\delta > 0$, define

$$B(x, \delta) = \{y \in X \mid d(x, y) < \delta\}$$

We call the set $B(x, \delta)$ the *open ball* at the point x with radius δ.

Definition 1.7 We call a subset $U \subseteq X$ *open* if for any point $x \in Y$, there exists $\delta > 0$ such that $B(x, \delta) \subseteq U$.

We call a subset $A \subseteq X$ *closed* if the complement $X \backslash A$ is open.

Example 1.8 In any metric space X, an open ball $B(x, \delta)$ is itself open. The punctured open ball $B'(x, \delta) = B(x, \delta) \backslash \{x\}$ is also open.

Example 1.9 In any metric space X, given $x \in X$ and $\delta > 0$ the *closed ball*

$$\overline{B(x, \delta)} = \{y \in X \mid d(x, y) \leq \delta\}$$

is closed.

Example 1.10 Let X be a metric space X, let $x \in X$, and let $\delta > 0$. Then the set

$$S(x, \delta) = \{y \in X \mid d(x, y) = \delta\}$$

is closed.

Example 1.11 The interval $(0, 1)$ is an open subset of the real line $\mathbb{R}$. The interval $[0, 1]$ is a closed subset.

Example 1.12 In any metric space X, the space X itself and the empty set $\emptyset$ are both open.

1.2 Norms

Write $\mathbb{K}$ to denote either the real numbers, $\mathbb{R}$, or the complex numbers, $\mathbb{C}$. In either case, given $\alpha \in \mathbb{K}$, we can define the modulus $|\alpha|$.

Definition 1.13 Let V be a $\mathbb{K}$-vector space. A *norm* on V is a mapping $V \to \mathbb{R}$, written $v \mapsto \|v\|$, with the following properites.

- Let $v \in V$. Then $\|v\| \geq 0$, and $\|v\| = 0$ if and only if $v = 0$. **(Positivity)**
- Let $\alpha \in \mathbb{K}$ and $v \in V$. Then $\|\alpha v\| = |\alpha|\|v\|$. **(Homogeneity)**
- Let $u, v \in V$. Then $\|u + v\| \leq \|u\| + \|v\|$. **(The Triangle Inequality)**

A vector space equipped with a norm is called a *normed vector space.*

We leave the proof of the following as an exercise.

Proposition 1.14 *Let $(V, \| - \|)$ be a normed vector space. Then V is a metric space, with metric defined by the formula*

$$d(v, w) = \|v - w\|$$

□

In this course, we focus on studying metric spaces that are open subsets of normed vector spaces.

Proposition 1.15 *Let V be a normed vector space. Then the norm $\| - \| \colon V \to \mathbb{R}$ is a continuous map.*

Proof: Choose $v \in V$. Let $\varepsilon > 0$. Consider a point $w \in V$. Then by the triangle inequality

$$\|v\| = \|(v - w) + w\| \leq \|v - w\| + \|w\|$$

and so $\|v - w\| \geq \|v\| - \|w\|$. Since $\|w - v\| = \|v - w\|$, we also have $\|v - w\| \geq \|w\| - \|v\|$. Thus

$$\|v - w\| \geq |\|v\| - \|w\||$$

So, suppose that $d(v, w) < \varepsilon$. Then $\|v - w\| < \varepsilon$, and by the above, $|\|v\| - \|w\|| < \varepsilon$. So if we take '$\delta = \varepsilon$' in the definition of continuity, we see that the norm is a continuous map, as required. □

Example 1.16 The following are norms on the vector space $\mathbb{K}^n$.

- The Euclidean norm

$$\|v\|_2 = \sqrt{|v_1|^2 + \cdots |v_n|^2} \qquad v = (v_1, \ldots, v_n)$$

- The sum norm

$$\|v\|_1 = |v_1| + \cdots |v_n| \qquad v = (v_1, \ldots, v_n)$$

- The max norm

$$\|v\|_\infty = \max(|v_1|, \ldots, |v_n|) \qquad v = (v_1, \ldots, v_n)$$

Note that the metric induced by the Euclidean norm on $\mathbb{R}^n$ is the Euclidean metric. The metric induced by the Euclidean norm on $\mathbb{C}^n$ is the same thing as the Euclidean metric on $\mathbb{R}^{2n}$.

We usually assume that the space $\mathbb{K}^n$ is equipped with the Euclidean norm. In this case, we simply write the norm as $\| - \|$, reserving the notation $\| - \|$ for when we need to be very careful and specific.

The proof that the above are norms is an exercise. The following result is also an exercise.

Proposition 1.17 *For any point $v \in \mathbb{K}^n$, we have:*

$$\|v\|_\infty \leq \|v\|_2 \leq \|v\|_1 \leq n\|v\|_\infty$$

□

One reason that the Euclidean norm is important is that it comes from an inner product. To be precise, recall that on $\mathbb{K}^n$ we can define an inner product by the formula

$$\langle u, v\rangle = u_1\overline{v_1} + \cdots + u_n\overline{v_n}$$

Here $\overline{z}$ denotes the complex conjugate of a complex number z; if the number z happens to be real, then $\overline{z} = z$.

Observe that the Euclidean norm, $\| - \|$, can be defined by writing

$$\|v\| = \sqrt{\langle v, v\rangle}$$

The following result is called the *Cauchy-Schwarz inequality*

Lemma 1.18 *Let $u, v \in \mathbb{K}^n$. Then*

$$|\langle u, v\rangle| \leq \|u\|\|v\|$$

□

1.3 Bounded Linear Maps

Definition 1.19 Let $T\colon V \to W$ be a linear map between normed vector spaces. We call T *bounded* if there exists $M \geq 0$ such that $\|Tv\| \leq M\|v\|$ for all $v \in V$.

Note that, if required, we can choose $M > 0$. If $M = 0$ in the above definition, then $\|Tv\| = 0$ for all $v \in V$. Consequently, $Tv = 0$ for all $v \in V$, that is to say $T = 0$.

Proposition 1.20 *Let $T\colon V \to W$ be a linear map between normed vector spaces. Then the following are equivalent:*

(i) *T is continuous.*

(ii) *T is continuous at the point 0.*

(iii) *T is bounded.*

Proof:

$(i) \Rightarrow (ii)$: This is trivial.

$(ii) \Rightarrow (iii)$: Let us take $\varepsilon = 1$ in the definition of continuity at the point 0. Then we can find $\delta > 0$ such that $d(v, 0) < \delta$ implies $d(Tv, 0) < 1$, ie: $\|v\| < \delta$ implies $\|Tv\| < 1$.

Take $M = \frac{2}{\delta}$. Let $v \in V \backslash \{0\}$. Then

$$\left\| \frac{\delta v}{2\|v\|} \right\| = \frac{\delta}{2} < \delta$$

Hence $\|T(\delta v/2\|v\|)\| < 1$. By linearity

$$\frac{\delta}{2\|v\|}\|Tv\| < 1 \quad \text{and} \quad \|Tv\| < \frac{2}{\delta}\|v\| = M\|v\|$$

Certainly, if $v = 0$, then $\|Tv\| = 0 \leq M\|v\|$. So the map T is bounded.

$(iii) \Rightarrow (i)$: Let $T\colon V \to W$ be bounded. Then we can find $M > 0$ such that $\|Tv\| \leq M\|v\|$ for all $v \in V$.

Choose $\varepsilon > 0$. Write $\delta = \varepsilon/M$. Then, for $v, w \in V$ with $d(v, w) < \delta$, we have

$$\|v - w\| < \frac{\varepsilon}{M}$$

and

$$\|Tv - Tw\| = \|T(v - w)\| \leq M\|v - w\| < \frac{M\varepsilon}{M} = \varepsilon$$

We see that the map T is continuous.

□

Definition 1.21 Let V and W be normed vector spaces, with $V \neq \{0\}$, and let $T\colon V \to W$ be a bounded linear map. Then we define the *operator norm* of the operator T by the formula

$$\|T\| = \sup_{v \in V, v \neq 0} \frac{\|Tv\|}{\|v\|}$$

Certainly $\|T\| \geq 0$. By the above definition, we have the inequality

$$\|Tv\| \leq \|T\|\|v\|$$

for all $v \in V$, and the value $\|T\|$ is the *smallest* number where we have such an inequality.

Let us write $Hom(V, W)$ to denote the space of all bounded linear maps from V to W. We leave the proof of the following result as an exercise.

Proposition 1.22 *The set $Hom(V, W)$ is a normed vector space. The operations of addition and scalar multiplication defined by the formulae*

$$(S + T)v = Sv + Tv \qquad S, T \in Hom(V, W),\ v \in V$$

and

$$(\alpha T)v = \alpha T(v) \qquad T \in Hom(V, W),\ \alpha \in \mathbb{K},\ v \in V$$

The norm is defined by the above operator norm. □

Example 1.23 Let $w \in W$. Let $T_w \colon \mathbb{K} \to W$ be the linear map defined by the formula $T_w(\alpha) = \alpha w$. Then, for any scalar $\alpha \in \mathbb{K}$, we have

$$\|T_w(\alpha)\| = \|\alpha w\| = |\alpha|\|w\|$$

Hence the linear map T_w is bounded, and $\|T_w\| = \|w\|$.

Note that any linear map $T \colon \mathbb{K} \to W$ takes the form T_w for some $w \in W$. Hence we can identify the space $Hom(\mathbb{K}, W)$ with the normed vector space W.

1.4 Finite-Dimensional Spaces

Lemma 1.24 *Equip the space $\mathbb{K}^n$ with the Euclidean norm. Let V be a normed vector space, and let $T \colon \mathbb{K}^n \to V$ be a linear map. Then the map T is bounded.*

Proof: Let $\{e_1, \ldots, e_n\}$ be the standard basis for $\mathbb{K}^n$. Let $v \in \mathbb{K}^n$, and write

$$v = \alpha_1 e_1 + \cdots + \alpha_n e_n \qquad \alpha_i \in \mathbb{K}$$

Let $M = \max(\|Te_1\|, \ldots, \|Te_n\|)$, where $\| - \|$ is the norm on V. Then

$$\begin{aligned} \|Tv\| &= \|\alpha_1 Te_1 + \cdots + \alpha_n Te_n\| \\ &\leq |\alpha_1|\|Te_1\| + \cdots + |\alpha_n|\|Te_n\| \\ &\leq |\alpha_1| M + \cdots + |\alpha_n| M \\ &= M\|v\|_1 \leq nM\|v\|_\infty \leq nM\|v\|_2 \end{aligned}$$

by proposition 1.17.
So the map T is bounded. □

Now, recall the *Heine-Borel theorem.*

Theorem 1.25 *Let $\mathbb{R}^n$ be equipped with the Euclidean metric. Then a subset $K \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded.* □

Given a compact space K, a continuous map $f \colon K \to \mathbb{R}$ is bounded and attains its bounds. Hence we can find a point $a_0 \in K$ such that $f(a_0) \leq f(a)$ for all $a \in K$.

Lemma 1.26 *Let K be a compact subset of a normed vector space such that $0 \notin K$. Then we can find $m > 0$ such that $\|a\| \geq m$ for all $a \in K$.*

Proof: We have already seen that the norm is a continuous map. Hence, by the above, we can find a point $a_0 \in K$ such that $\|a_0\| leq \|a\|$ for all $a \in K$.

Now $0 \notin K$, so $a_0 \neq 0$. Hence, if we set $m = \|a_0\|$, we have $m > 0$, and $\|a\| \geq m$ for all $a \in K$. □

Theorem 1.27 *Let V be an n-dimensional normed vector space. Then V is linearly homeomorphic to the space $\mathbb{K}^n$ equipped with the Euclidean norm.*

Proof: Since the spaces $\mathbb{K}^n$ and V have the same dimension, there is an invertible linear map $T\colon \mathbb{K}^n \to V$. By lemma 1.24, the map T is continuous. We want to show that the inverse T^{-1} is also continuous.

Since the norm is a continuous map, and the map T is continuous, the composition

$$\mathbb{K} \xrightarrow{T} V \xrightarrow{\|-\|} \mathbb{R}$$

is continuous.

Let $S = \{x \in \mathbb{K}^n \mid \|x\| = 1\}$ be the unit sphere in the space $\mathbb{K}^n$. Then the subset S is closed and bounded, and therefore compact by the Heine-Borel theorem. It follows by lemma 1.26 that we can find $m > 0$ such that $\|Tx\| \geq m$ for all $x \in S$.

Write $M = \frac{1}{m}$. Observe that $\|T^{-1}(0)\| = 0 \leq M\|0\|$. Choose $v \in V \backslash \{0\}$. Then $T^{-1}v \neq 0$, and $\frac{T^{-1}v}{\|T^{-1}v\|} \in S$. Hence

$$T\left(\frac{T^{-1}v}{\|T^{-1}v\|}\right) \geq m$$

and $\|v\| \geq m\|T^{-1}v\|$.

We see that $\|T^{-1}v\| \leq M\|v\|$ for all $v \in V$. Thus the map T^{-1} is bounded, and we are done. □

The above result tells us that a finite-dimensional vector space has the same open sets for any choice of norm. In particular, as we shall see more fully in the next section, when looking at properties such as continuity, limits, or which subsets are compact, the choice of norm is irrelevant.

Corollary 1.28 *Let V and W be normed vector spaces, where V is finite-dimensional. Let $T\colon V \to W$ be a linear map. Then T is continuous.*

Proof: By the above theorem, we have a linear homeomorphism $\phi\colon \mathbb{K}^n \to V$. By lemma 1.24, the map $T \circ \phi\colon \mathbb{K}^n \to W$ is continuous.

The inverse of the homeomorphism ϕ is also continuous. Hence the map $T = T \circ \phi \circ \phi^{-1}\colon V \to W$ is also continuous, and we are done. □

2 Topological Spaces and Connectedness

2.1 Topologies and Continuity

Let $f: X \to Y$ be a function between metric spaces. We can write the definition of continuity in terms of open balls. To be precise, we know that the function f is continuous at the point $x \in X$ if and only if for all $\varepsilon > 0$ there exists $\delta > 0$ such that $d(f(x), f(y)) < \varepsilon$ whenever $d(x, y) < \delta$.

Consider the open balls

$$B(x, \delta) = \{y \in X \mid d(x, y) < \delta\} \qquad B(f(x), \varepsilon) = \{z \in Y \mid d(f(x), z) < \varepsilon\}$$

Thus the function f is continuous at the point $x \in X$ if and only if for all $\varepsilon > 0$ there exists $\delta > 0$ such that $f(y) \in B(f(x), \varepsilon)$ whenever $y \in B(x, \delta)$.

Rephrasing slightly, for all $\varepsilon > 0$ we can find $\delta > 0$ such that $B(x, \delta) \subseteq f^{-1}[B(f(x), \varepsilon)]$.

Proposition 2.1 *Let $f: X \to Y$ be a map between metric spaces. The map f is continuous if and only if for any open subset $U \subseteq Y$, the inverse image $f^{-1}[U] \subseteq X$ is open.*

Proof: Let $f: X \to Y$ be continuous. Let $U \subseteq Y$ be open. We want to show that the inverse image $f^{-1}[U]$ is open.

If $f^{-1}[U] = \emptyset$, then we are done. Otherwise, pick $x \in f^{-1}[U]$. Then $f(x) \in U$. Since the set U is open, we can find $\varepsilon > 0$ such that $B(f(x), \varepsilon) \subseteq U$. As we mentioned above, by continuity there exists $\delta > 0$ such that

$$B(x, \delta) \subseteq f^{-1}[B(f(x), \varepsilon)] \subseteq f^{-1}[U]$$

Hence the inverse image $f^{-1}[U]$ is an open set, by definition.

Conversely, suppose that the inverse image $f^{-1}[U]$ is open whenever $U \subseteq Y$ is an open subset. Let $x \in X$, and let $\varepsilon > 0$. Then the open ball $B(f(x), \varepsilon) \subseteq Y$ is certainly open.

Now the inverse image $f^{-1}[B(f(x), \varepsilon)]$ is open and contains the point x. There thus exists $\delta > 0$ such that $B(x, \delta) \subseteq f^{-1}[B(f(x), \varepsilon)]$.

It follows that the function f is continuous at our chosen point $x \in X$, and we are done. □

The proof of the following result is similar to the above; we leave it as an exercise.

Proposition 2.2 *Let X and Y be metric spaces, let $x_0 \in X$, and $f: X \backslash \{x_0\} \to Y$ be a function. Then*

$$\lim_{x \to x_0} f(x) = a$$

if and only if for any open set $U \subseteq Y$ containing a there is an open set $U' \subseteq X$ containing a, such that $f(x) \in U$ whenever $x \in U' \backslash \{x_0\}$. □

The idea of open sets and the proposition 2.1 leads to a generalisation of the notion of continuity.

First note the following result; we leave the proof as an exercise.

Proposition 2.3 *Let X be a metric space. Then the collection of all open subsets of X has the following properties.*

- *The space X and the empty set $\emptyset$ are open.*
- *Let $U_1, U_2 \subseteq X$ be open sets. Then the intersection $U_1 \cap U_2$ is open.*
- *Let $\{U_\lambda \mid \lambda \in \Lambda\}$ be an arbitrary collection of open sets. Then the union $\cup_{\lambda\in\Lambda} U_\lambda$ is open.*

$\square$

Definition 2.4 Let X be a set. Then a *topology* on X is a distinguished collection, $\mathcal{T}$, of subsets, such that:

- $X \in \mathcal{T}$ and $\emptyset \in \mathcal{T}$.
- Let $U_1, U_2 \in \mathcal{T}$. Then $U_1 \cap U_2 \in \mathcal{T}$.
- Let $\{U_\lambda \mid \lambda \in \Lambda\}$ be a collection of sets belonging to the collection $\mathcal{T}$. Then $\cup_{\lambda\in\Lambda} U_\lambda \in \mathcal{T}$.

A set equipped with a topology is called a *topological space*.

We saw in the above proposition that any metric space can be considered to be a topological space; the topology is the collection of open subsets.

In fact, in general, given a topological space, we term subsets that belong to the topology *open sets*. As before, we call a subset $A \subseteq X$ *closed* if the complement $X\backslash A$ is open.

Proposition 2.5 *Let X be a metric space. Let $x, y \in X$ be points, where $x \neq y$. Then there are open sets $U_1, U_2 \subseteq X$ such that $x \in U_1$, $y \in U_2$, and $U_1 \cap U_2 = \emptyset$.*

Proof: Let $\delta = d(x, y)$. Then we can take $U_1 = B(x, \frac{\delta}{2})$ and $U_2 = B(y, \frac{\delta}{2})$. $\square$

Example 2.6 Let X be any set. Then we can define a topology, $\mathcal{T}$, on the set X by writing $\mathcal{T} = \{X, \emptyset\}$.

Suppose we have points $x, y \in X$ with $x \neq y$. Then there are *no* open sets $U_1, U_2 \subseteq X$ such that $x \in U_1$, $y \in U_2$, and $U_1 \cap U_2 = \emptyset$. It follows that the above topology does not come from a metric.

Definition 2.7 Let X and Y be topological spaces. A function $f\colon X \to Y$ is called *continuous* if for any open set $U \subseteq Y$, the inverse image $f^{-1}[U] \subseteq X$ is open.

Definition 2.8 Let X and Y be topological spaces. A map $f: X \to Y$ is called a *homeomorphism* if it is continuous and bijective, and the inverse map is also continuous.

We call two spaces *homeomorphic* if there is a homeomorphism between them.

Example 2.9 Define a map $f: (-1, 1) \to \mathbb{R}$ by the formula

$$f(x) = \frac{x}{1 - |x|}$$

Then the map f is continuous, and we have a continuous inverse $f^{-1}: \mathbb{R} \to (-1, 1)$ defined by the formula

$$f^{-1}(y) = \frac{x}{1 + |x|}$$

Therefore the spaces $(-1, 1)$ and $\mathbb{R}$ are homeomorphic.

Homeomorphic metric spaces are essentially the same as far as open sets are concerned. Thus properties depending on open sets, such as continuity of functions and compactness are the same for homeomorphic spaces.

2.2 Connected Spaces

It is no extra work to formulate the definitions in this section for topological spaces. However, most of our examples will be metric spaces; in fact subsets of normed vector spaces.

Definition 2.10 Let X be a topological space. A *partition* of X is a pair of open subsets, $U_1 \neq \emptyset$ and $U_2 \neq \emptyset$, such that $U_1 \cap U_2 = \emptyset$, and $U_1 \cup U_2 = X$.

The space X is called *connected* if it admits no partition.

An equivalent formulation of the above definition, using the definition of a closed set as the complement of an open set, is to say that a space X is connected if and only if the only subsets which are *both* open and closed are the set X itself and the empty set $\emptyset$.

Example 2.11 Let us consider the space $\{0, 1\}$ as a metric space: the metric is defined by considering the points 0 and 1 to be distance 1 apart, ie: we consider the space $\{0, 1\}$ as a subspace of the real line $\mathbb{R}$.

Observe that

$$\{0\} = B(0, \frac{1}{2}) \qquad \{1\} = B(1, \frac{1}{2})$$

So the subsets $\{0\}$ and $\{1\}$ are open. Certainly neither of these subsets are empty, $\{0\} \cap \{1\} = \emptyset$, and $\{0, 1\} = \{0\} \cup \{1\}$. So the space $\{0, 1\}$ is not connected.

Proposition 2.12 *A metric space X is connected if and only if any continuous map $f: X \to \{0, 1\}$ is constant.*

Proof: Let $f: X \to \{0, 1\}$ be a non-constant continuous map. Let $U_1 = f^{-1}[\{0\}]$ and $U_2 = f^{-1}[\{1\}]$. The sets $\{0\}$ and $\{1\}$ are open by the above. Thus by continuity, the inverse images U_1 and U_2 are open.

Since the map f is not constant, it is surjective. It follows that $U_1 \neq \emptyset$ and $U_2 \neq \emptyset$. Obviously, $U_1 \cap U_2 = \emptyset$, and $U_1 \cup U_2 = f^{-1}[\{0, 1\}] = X$. So the space X is not connected.

Conversely, suppose that the space X is not connected. Then it admits a partition $X = U_1 \cup U_2$, where U_1 and U_2 are open and non-empty, and $U_1 \cap U_2 = \emptyset$. It follows that we have a well-defined non-constant continuous map $f: X \to \{0, 1\}$ given by writing

$$f(x) = \begin{cases} 0 & x \in U_1 \\ 1 & x \in U_2 \end{cases}$$

$\square$

Theorem 2.13 *The interval $[0, 1] \subseteq \mathbb{R}$ is connected.*

Proof: Let $f: [0, 1] \to \{0, 1\}$ be a continuous map. Suppose that f is not constant. Then we can find $a, b \in [0, 1]$ such that $f(a) = 0$ and $f(b) = 1$.

By the intermediate value theorem, for some value c between a and b, we have $f(c) = \frac{1}{2}$. But the value $\frac{1}{2}$ does not lie in the codomain of f. Hence this statement is a contradiction.

It follows that the function f must be constant. Therefore the interval $[0, 1]$ is connected. $\square$

Proposition 2.14 *Let X be a connected space, let Y be a topological space, and let $f: X \to Y$ be a surjective continuous map. Then the space Y is connected.*

Proof: Let $g: Y \to \{0, 1\}$ be a continuous map. Then the composite $f \circ g: X \to \{0, 1\}$ is continuous, and therefore constant, since X is conntected.

It follows that the map g is constant, since the map f is surjective. Hence the space Y is connected. $\square$

In particular, for any continuous map $f: X \to Y$ between topological spaces, if the space X is connected, then so is the image $f[X]$.

2.3 Path-Connected Spaces

As in the previous section, it is no extra work to formulate our main definitions and abstract results for topological spaces rather than metric spaces. However, later on we need to specialise.

Definition 2.15 Let X be a topological space, and let $x_0, x_1 \in X$. Then a *path* in X from x_0 to x_1 is a continuous map $\gamma\colon [0,1] \to X$ such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$.

The space X is called *path-connected* if there is a path between any two points in X.

Example 2.16 Let V be a normed vector space. Let $x_0 \in V$, $\delta > 0$. Then the ball $B(x_0, \delta)$ is path-connected.

To see this, let $v_0, v_1 \in B(x_0, \delta)$. Write

$$\gamma(t) = v_0 + t(v_1 - v_0)$$

It is straightforward to check that the map $\gamma\colon [0,1] \to V$ defined by the above formula is continuous. Let $t \in [0,1]$, and observe

$$\gamma(t) - x_0 = ((1-t)v_0 + tv_1) - x_0 = (1-t)(v_0 - x_0) + t(v_1 - x_0)$$

Hence, since $\|v_0 - x_0\| < \delta$ and $\|v_1 - x_0\| < \delta$, we have

$$\|\gamma(t) - x_0\| \leq (1-t)\|v_0 - x_0\| + t\|v_1 - x_0\| < (1-t)\delta + t\delta = \delta$$

Thus the map γ is a path in the ball $B(x_0, \delta)$ from the point v_0 to the point v_1.

Proposition 2.17 *Let X be a path-connected topological space. Then X is also connected.*

Proof: Let X be path-connected. Suppose we have a non-constant continuous map $f\colon X \to \{0,1\}$. Then we can find points $x_0, x_1 \in X$ such that $f(x_0) = 0$ and $f(x_1) = 1$.

Since X is path-connected, we have a path, γ, in X from x_0 to x_1. So we have a continuous non-constant function $f \circ \gamma\colon [0,1] \to \{0,1\}$. But by theorem 2.13, the interval $[0,1]$ is connected. Hence this statement is a contradiction.

It follows that we have no non-constant continuous map $f\colon X \to \{0,1\}$, and hence that the space X is connected, as required. $\square$

The proof of the following fact is left as an exercise.

Example 2.18 Let $G = \{(x, \sin(1/x)) \mid x > 0\}$. Let

$$X = (\{0\} \times [-1,1]) \cup G \subseteq \mathbb{R}^2$$

Then the space X is connected, but not path-connected.

The following proposition is also an exercise.

Proposition 2.19 *Let X be a topological space. Then we can define an equivalence relation, $\sim$, on X, by writing $x \sim y$ when there is a path in X from x to y.* $\square$

The equivalence classes of the above equivalence relation are called the *path-components* of the space X. If the space X is path-connected, then there is only one path-component. Otherwise, the space X is the disjoint union of its path-components.

Lemma 2.20 *Let R be an open subset of a normed vector space. Then the path-components of R are open sets.*

Proof: Let $C \subseteq R$ be a path-component. Let $x_0 \in C$. Then, since the set R is open, we can find $\delta > 0$ such that $B(x_0, \delta) \subseteq R$. But the open ball $B(x_0, \delta)$ is path-connected by example 2.16. Hence $B(x_0, \delta) \subseteq C$.

It follows that the path-component C is open. □

Theorem 2.21 *Let R be a connected open subset of a normed vector space. Then the space R is path-connected.*

Proof: Suppose that the space R is not path-connected. The path-components of the space R, $\{C_\lambda \mid \lambda \in \Lambda\}$ are equivalence classes, and thus are pairwise disjoint, and $\cup_{\lambda\in\Lambda C_\lambda} = R$.

Since the space R is not path-connected, there are at least two components. Choose a component, C, and let D be the union of the other components. Then $C \neq \emptyset$ and $D \neq \emptyset$. Further, $C \cap D = \emptyset$ and $C \cup D = R$.

By the above lemma, each path-component is an open set. Hence the set C is open, and the set D is a union of open sets, and therefore open. We see that the space R admits a partition $R = C \cup D$, and is therefore not connected. □

3 Differentiation

3.1 The Total Derivative

Recall that a function $f\colon \mathbb{R} \to \mathbb{R}$ is *differentiable* at a point $x \in \mathbb{R}$ if the *derivative*

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

exists.

We can rewrite this equation

$$f(x+h) = f(x) + f'(x)h + |h|r(h) \qquad \lim_{r\to 0} r(h) = 0$$

In this form, we can extend the above definition to normed vector spaces.

Definition 3.1 Let V and W be normed vector spaces, and let $R \subseteq V$ be an open subset. We say that a function $\varphi\colon R \to W$ is *differentiable* at a point $x \in R$

if there is a bounded linear map $(D\varphi)_x: V \to W$, and a function $r: B(0,\delta) \to W$ for some $\delta > 0$ such that

$$\varphi(x+v) = \varphi(x) + (D\varphi)_x v + \|v\| r(v)$$

for all $v \in B(0,\delta)$, where $\lim_{v\to 0} r(v) = 0$.

The above linear map $(D\varphi)_x: V \to W$ is called the *total differential* of the function φ at the point x.

We refer to a function $\varphi: R \to W$ as *differentiable* if it is differentiable at every point of the domain R.

Note that if V and W are finite-dimensional normed vector spaces, the above notion does not depend upon the choice of norm.

Lemma 3.2 *Let V and W be normed vector spaces, let $R \subseteq V$ be an open subset, and let $\varphi: R \to W$ be a function that is differentiable at the point $x \in R$.*

Then the total derivative $(D\varphi)_x$ is uniquely determined.

Proof: Suppose we have linear maps $L_1, L_2: V \to W$, and functions $r_1: B(0,\delta_1) \to W$, $r_2: B(0,\delta_2) \to W$ such that

$$\varphi(x+v) = \varphi(x) + L_1 v + \|v\| r_1(v) \qquad \lim_{v\to 0} r_1(v) = 0$$

and

$$\varphi(x+v) = \varphi(x) + L_2 v + \|v\| r_2(v) \qquad \lim_{v\to 0} r_2(v) = 0$$

Let $\delta = \min(\delta_1, \delta_2)$. Subtracting, we see that

$$0 = (L_1 - L_2)v + \|v\|(r_1(v) - r_2(v))$$

whenever $v \in B(0,\delta)$. We need to show that $L_1 - L_2 = 0$. For convenience, write $L = L_1 - L_2$, and $r = r_1 - r_2$. Then

$$0 = Lv + \|v\| r(v) \qquad \lim_{v\to 0} r(v) = 0$$

Let $u \in V$, $u \neq 0$. Write $v = tu$, where $t \in \mathbb{K}$. Then

$$0 = L(tu) + \|tu\| r(tu)$$

Rearranging:

$$\frac{tL(u)}{|t|\|u\|} = -r(tu)$$

Let $t > 0$. Then the above becomes

$$\frac{L(u)}{\|u\|} = -r(tu)$$

Now, let $t \to 0$. Then $-r(tu) \to 0$, but the expression $\frac{L(u)}{\|u\|}$ remains constant. Hence

$$\frac{L(u)}{\|u\|} = 0$$

that is to say $L(u) = 0$, and we are done. □

Example 3.3 Let $T: V \to W$ be a bounded linear map. Then, for any point $x \in V$ and $v \in V$, we have

$$T(x + v) = T(x) + Tv = T(x) + Tv + \|v\| \cdot 0$$

It follows that the map T is differentiable, with total derivative T at any point $x \in V$.

The following is similarly easy to check.

Example 3.4 Let $c: R \to W$ be a constant map, where $R \subseteq V$ is an open set, and V and W are normed vector spaces. Then the map c is differentiable, and the total derivative is 0.

Proposition 3.5 *Let V and W be normed vector spaces, and let $R \subseteq V$ be an open subset. Let $\varphi_1, \varphi_2: R \to W$ be functions that are differentiable at the point $x \in R$.*

Let $\alpha_1, \alpha_2 \in \mathbb{K}$. Then the function $\alpha_1\varphi_1 + \alpha_2\varphi_2$ is differentiable at the point x, with total derivative

$$D(\alpha_1\varphi_1 + \alpha_2\varphi_2)_x = \alpha_1(D\varphi_1)_x + \alpha_2(D\varphi_2)_x$$

Proof: We have $\delta_1, \delta_2 > 0$, and functions $r_1: B(0, \delta_1) \to W$, $r_2: B(0, \delta_2) \to W$ such that

$$\varphi_1(x + v) = \varphi_1(x) + (D\varphi_1)_x v + \|v\| r_1(v) \qquad \lim_{v \to 0} r_1(v) = 0$$

and

$$\varphi_2(x + v) = \varphi_2(x) + (D\varphi_2)_x v + \|v\| r_2(v) \qquad \lim_{v \to 0} r_2(v) = 0$$

Let $\delta = \min(\delta_1, \delta_2)$, and define $r: B(0, \delta) \to W$ by the formula $r(v) = \alpha_1 r_1(v) + \alpha_2 r_2(v)$. Then

$$\begin{aligned} \alpha_1\varphi_1(x + v) + \alpha_2\varphi_2(x + v) &= \alpha_1\varphi_1(x) + \alpha_2\varphi_2(x) \\ &\quad + (\alpha_1(D\varphi_1)_x + \alpha_2(D\varphi_2)_x)(v) + \|v\| r(v) \end{aligned}$$

where

$$\lim_{v \to 0} r(v) = 0$$

The result now follows. □

We leave the following as an exercise.

Proposition 3.6 *Let V and W be normed vector spaces, and let $R \subseteq V$ be an open subset. Let $\varphi: R \to W$ be a function that is differentiable at the point $x \in R$. Then the function φ is also continuous at the point x.* □

The following result is known as the *chain rule*. It illustrates the beauty and potential utility of the concept of the total differential.

Theorem 3.7 *Let U, V, and W be normed vector spaces. Let $R \subseteq U$ and $S \subseteq V$ be open sets, and let $\varphi\colon R \to S$ and $\psi\colon S \to W$ be functions.*

Suppose that the function φ is differentiable at the point $x \in R$, and the function ψ is differentiable at the point $\varphi(x) \in S$. Then the composite $\psi\circ\varphi\colon R \to W$ is differentiable at the point $x \in R$, with total derivative

$$D(\psi \circ \varphi)_x = (D\psi)_{\psi(x)} \circ (D\varphi)_x$$

Proof: Set $y = \varphi(x)$. Then

$$\varphi(x+v) = \varphi(x) + (D\varphi)_x v + \|v\| r(v) \qquad \lim_{v\to 0} r(v) = 0$$

and

$$\psi(y+w) = \psi(y) + (D\psi)_y w + \|w\| s(w) \qquad \lim_{w\to 0} s(w) = 0$$

Observe

$$\begin{aligned}
(\psi \circ \varphi)(x+v) &= \psi(\varphi(x+v)) \\
&= \psi(\varphi(x) + (D\varphi)_x v + \|v\| r(v)) \\
&= \psi(\varphi(x)) \\
&\quad + (D\psi)_y((D\varphi)_x v + \|v\| r(v)) \\
&\quad + \|(D\varphi)_x v + \|v\| r(v)\ \| \cdot s((D\varphi)_x v + \|v\| r(v))
\end{aligned}$$

So

$$(\psi \circ \varphi)(x+v) = (D\psi)_y((D\varphi)_x(v)) + \|v\| t(v)$$

where

$$t(v) = \frac{(D\psi)_y(\|v\| r(v))}{\|v\|} + \frac{\|(D\varphi)_x v + \|v\| r(v)\ \|}{\|v\|} s((D\varphi)_x v + \|v\| r(v))$$

We must show that $t(v) \to 0$ as $v \to 0$. Observe that

$$\lim_{v\to 0} \frac{(D\psi)_y(\|v\| r(v))}{\|v\|} = \lim_{v\to 0} (D\psi)_y r(v) = 0$$

since the function r has limit 0 at 0, and the total derivative $(D\varphi)_y$ is continuous and linear.

Let

$$q(v) = \frac{\|(D\varphi)_x v + \|v\| r(v)\ \|}{\|v\|} s((D\varphi)_x v + \|v\| r(v))$$

Then it remains to show that $q(v) \to 0$ as $v \to 0$. Since the derivative $(D\varphi)_x$ is a bounded linear map, there is a constant $M \geq 0$ such that

$$\|(D\varphi)_x v\| \leq M \|v\|$$

for all $v \in V$.

By the triangle inequality, we have

$$\|q(v)\| \leq \frac{\|(D\varphi)_x v\| + \|v\| \cdot \|r(v)\|}{\|v\|} \|s((D\varphi)_x v + \|v\| r(v))\|$$

and so

$$\|q(v)\| \leq (M + \|r(v)\|\ ) \|s((D\varphi)_x v + \|v\| r(v))\|$$

Let $v \to 0$. Then $\|v\| r(v) \to 0$, and $(D\varphi)_x v \to 0$ by continuity. Since $\lim_{w \to 0} s(w) = 0$, it follows that $s((D\varphi)_x v + \|v\| r(v)) \to 0$.

Hence $\|q(v)\| \to 0$. Therefore

$$\lim_{v \to 0} q(v) = 0$$

and we are finally done. □

3.2 Differentiation of Paths

Let V be a normed vector space. Recall that a *path* in V is simply a continuous function $\gamma\colon [a, b] \to V$.

Definition 3.8 We say that the path γ is *differentiable* at the point $t \in [a, b]$ if the derivative

$$\gamma'(t) = \lim_{h \to 0} \frac{\gamma(t + h) - \gamma(t)}{h}$$

exists.

The above makes sense, since $h \in \mathbb{R}$. We can rewrite the above equation

$$\gamma(t + h) = \gamma(t) + \gamma'(t)h + |h| r(h) \qquad \lim_{h \to 0} r(h) = 0$$

Thus, when $t \in (a, b)$, the above is related to the total derivative by the formula

$$(D\gamma)_t(h) = \gamma'(t)h$$

Let $V = \mathbb{K}^n$. Write $\gamma(t) = (\gamma_1(t), \dots, \gamma_n(t)$. Then it is straightforward to check that

$$\gamma'(t) = (\gamma_1'(t), \dots, \gamma_n'(t))$$

Example 3.9 Define a path $\gamma\colon [0, 2\pi] \to \mathbb{R}^2$ by the formula

$$\gamma(t) = (\cos t, \sin t)$$

Then γ is differentiable, and $\gamma'(t) = (-\sin t, \cos t)$.

In the above example, note that $\gamma(0) = \gamma(2\pi) = (1, 0)$. So $\gamma(2\pi) - \gamma(0) = 0$. However, $\gamma'(t) \neq 0$ for all t, so there is no value $c \in [0, 2\pi]$ such that

$$\gamma(2\pi) - \gamma(0) = (2\pi - 0)\gamma'(c)$$

The nearest analogy of the mean value theorem in higher dimensions is the following result, which we call the *mean value inequality*.

Theorem 3.10 *Let $\gamma\colon [a, b] \to \mathbb{R}^n$ be a differentiable function. Then there exists $c \in [a, b]$ such that*

$$\|\gamma(b) - \gamma(a)\| \leq (b - a)\|\gamma'(c)\|$$

Proof: Suppose that $\gamma(a) \neq \gamma(b)$; if $\gamma(a) = \gamma(b)$, then the result is obvious. Define a vector

$$z = \frac{\gamma(b) - \gamma(a)}{\|\gamma(b) - \gamma(a)\|}$$

Then $\|z\| = 1$. Define a map $k\colon [a, b] \to \mathbb{R}$ by the formula

$$k(t) = \langle z, \gamma(t)\rangle$$

Then the function k is differentiable on the interval $[a, b]$, with derivative

$$k'(t) = \langle z, \gamma'(t)\rangle$$

By the mean value theorem in ordinary real analysis, we can find $c \in [a, b]$ such that $k(b) - k(a) = (b - a)k'(c)$. In particular,

$$|k(b) - k(a)| \leq (b - a)|k'(c)|$$

Now

$$k(b) - k(a) = \langle \frac{\gamma(b) - \gamma(a)}{\|\gamma(b) - \gamma(a)\|}, \gamma(b) - \gamma(a)\rangle = \|\gamma(b) - \gamma(a)\|$$

because of the relation between the inner product and the Euclidean norm. Hence, by the Cauchy-Schwarz inequality

$$\|\gamma(b) - \gamma(a)\| \leq (b - a)\|z\|\|\gamma'(c)\| = (b - a)\|\gamma'(c)\|$$

and we are done. $\square$

3.3 Directional Derivatives

Let V and W be normed vector spaces, and let $R \subseteq V$ be an open subset.

Definition 3.11 Let $x \in R$, and $v \in V$. Then we say that a function $\varphi\colon R \to W$ is *differentiable* in the *direction* v if the *directional derivative*

$$D_v(\varphi)_x = \lim_{t\to 0} \frac{\varphi(x + tv) - \varphi(x)}{t}$$

exists.

Note that the directional derivative $D_v(\varphi)_x$ is a vector in the space W. We sometimes write $D_v(\varphi)_x = D_v(\varphi)(x)$. Observe that we can also write

$$D_v(\varphi) = \frac{d}{ds}\varphi(x+sv)|_{s=0}$$

Proposition 3.12 *Let $\varphi: R \to W$ be a mapping that is differentiable at the point $x \in R$. Then for all vectors $v \in V$, the directional derivative $D_v(\varphi)_x$ exists, and is related to the total derivative by the formula*

$$D_v(\varphi)_x = (D\varphi)_x(v)$$

Proof: Let $v \in V$. Let $t \in \mathbb{K}$. Then

$$\varphi(x+tv) = \varphi(x) + (D\varphi)_x(tv) + \|tv\| r(tv) \qquad \lim_{t\to 0} r(tv) = 0$$

Hence

$$\lim_{t\to 0} \frac{\varphi(x+tv) - \varphi(x)}{t} = (D\varphi)_x(v) + \lim_{t\to 0} \frac{|t|\|v\| r(tv)}{t}$$

Now

$$\lim_{t\to 0} \|\frac{|t|\|v\| r(tv)}{t}\| = \|v\| \lim_{t\to 0} \|r(tv)\| = 0$$

and we are done. $\square$

Corollary 3.13 *Let $\gamma: [a,b] \to R$ be a differentiable function, and let $\varphi: R \to W$ be a differentiable function. Then the path $\varphi \circ \gamma: [a,b] \to W$ is differentiable at each point, and $(\varphi\circ\gamma)'(t) = D_{\gamma'(t)}(\varphi)_{\gamma(t)}$.*

Proof: Recall that $\gamma'(t)h = (D\gamma)_t h$ for any path γ. Hence, by the chain rule and the above, for a real number h:

$$\begin{array}{rcl} (\varphi\circ\gamma)'(t)h & = & D(\varphi\circ\gamma)_t h \\ & = & (D\varphi)_{\gamma(t)}(\gamma'(t)h) \\ & = & h(D\varphi)_{\gamma(t)}(\gamma'(t)) \\ & = & hD_{\gamma'(t)}(\varphi)_{\gamma(t)} \end{array}$$

The result is now clear. $\square$

Theorem 3.14 *Let V and W be normed vector spaces, and let $R \subseteq V$ be a* connected *open subset. Then a function $\varphi: R \to W$ is constant if and only if it is differentiable at each point $x \in R$, with $(D\varphi)_x = 0$.*

Proof: If the function is constant, it is easy to check that it is differentiable everywhere, and the total derivative is always zero.

Conversely, suppose that the function is differentiable everywhere on the domain R, and the total derivative is always zero. Then by the above proposition, the directional derivatives are all zero.

Let $x \in R$. Then, since R is open, we can find $\delta > 0$ such that $B(x, \delta) \subseteq R$. Let $y \in B(x, \delta)$. Then we can define a path, $\gamma: [0, 1] \to B(x, \delta)$, from x to y by the formula

$$\gamma(t) = x + t(y - x)$$

Let $v = y - x$. Then $\gamma'(t) = v$, and by corollary 3.13,

$$(\varphi \circ \gamma)'(t) = D_{\gamma'(t)}(\varphi)_{\gamma(t)} = 0$$

since the directional derivatives are all zero.

Hence, by the mean value inequality, $\|\varphi(\gamma(0)) - \varphi(\gamma(1))\| = 0$, and so $\varphi(x) = \varphi(y)$.

The above argument shows that for any value, a, of the function $f(x)$, the set

$$C_a = \{x \in R \mid f(x) = a\}$$

is open.

Suppose that the function X is not constant. Then there are at least two values, a and b. Hence we have non-empty open sets C_a, and

$$D = \bigcup_{w \in B \backslash a} C_w$$

It is obvious that $R = C_a \cup D$, and $C_a \cap D = \emptyset$. Hence the domain R admits a partition, and is therefore not connected, which is a contradiction. □

Now, let V and W be finite-dimensional vector spaces, let $u \in V$, and let $R \subseteq V$ be an open subset, and let $\varphi: R \to W$ be a map such that the directional derivative $D_u(\varphi)_x$ exists for all $x \in V$. Then we have a map

$$D_u(\varphi): R \to W \qquad D_u(\varphi)(x) = D_u(\varphi)_x$$

We can ask whether this new map is differentiable in another direction $v \in V$. The following result is known as *Schwarz's theorem* or *Clairaut's theorem.*

Theorem 3.15 *Let $\varphi: R \to W$ be a map such that for $u, v \in V$ the second directional derivatives $D_v D_u(\varphi)$ and $D_v D_u(\varphi)$ exist and are continuous. Then $D_v D_u(\varphi) = D_u D_v(\varphi)$.*

Proof: Since the space W is finite-dimensional, we may look at components, and assume that $W = \mathbb{K}$. We can further assume, by splitting into real and imaginary parts, that $\mathbb{K} = \mathbb{R}$.

Fix $x \in R$. Define $f(s,t) = \varphi(x+su+tv) - \varphi(s,t)$, for $s,t \in \mathbb{R}$ sufficiently small that the expression makes sense. For fixed s and t, define a function $g\colon [0,s] \to \mathbb{R}$ by the formula

$$g(\sigma) = \varphi(x+\sigma u+tv) - \varphi(x+\sigma u)$$

By the mean value theorem for real functions, there exists $s_1 \in [0,s]$ such that $g(s) - g(0) = sg'(s_1)$, that is

$$\begin{aligned}&\varphi(x+st+tv) - \varphi(x+su) - \varphi(x+tv) + \varphi(x)\\&= sD_u(\varphi)(x+s_1u+tv) - D_u(\varphi)(x+s_1u)\end{aligned}$$

Now, define a map $h\colon [0,t] \to W$ by the formula

$$h(\tau) = (D_u\varphi)(x+s_1u+\tau v)$$

By the mean value theorem, we obtain $t_1 \in [0,t]$ such that $h(t) - h(0) = th'(t_1)$, that is

$$D_u(\varphi)(x+s_1t+tv) - D_u(\varphi)(x+s_1u) = tD_vD_u(\varphi)(x+s_1u+t_1v)$$

Combining the above two equations arising from the mean value theorem, we see that

$$\varphi(x+su+tv) - \varphi(x+su) - \varphi(x+tv) + \varphi(x) = stD_vD_u(\varphi)(x+s_1u+t_1v)$$

Similarly, applying the above trick in the opposite order, we have $s_2 \in [0,s]$ and $t_2 \in [0,t]$ such that

$$\varphi(x+su+tv) - \varphi(x+su) - \varphi(x+tv) + \varphi(x) = stD_uD_v(\varphi)(x+s_2u+t_2v)$$

Hence

$$D_vD_u(\varphi)(x+s_1u+t_1v) = D_uD_v(\varphi)(x+s_2u+t_2v)$$

whenever $s \neq 0$ and $t \neq 0$.

If we take the limit as $s \to 0$ and $t \to 0$, then certainly $s_1 \to 0$, $t_1 \to 0$, $s_2 \to 0$, and $t_2 \to 0$. The derivatives $D_vD_u(\varphi)$ and $D_uD_v(\varphi)$ are assumed in the statement of the theorem to be continuous at the point x.

Hence

$$D_vD_u(\varphi)_x = D_vD_u(\varphi)_x$$

and we are done. □

3.4 Partial Derivatives

Let $R \subseteq \mathbb{K}^m$ be an open subset, and let $f\colon R \to \mathbb{K}$ be a map. Let $(x_1, \ldots, x_m)$ be coordinates in the space $\mathbb{K}^m$. Then the *partial derivative*, $\frac{\partial f}{\partial x_j}$ is defined by differentiating the function f with respect to the variable x_j while treating the other variables as constants.

More generally, if we have a function $f\colon R \to \mathbb{K}^n$, we can write

$$f(x_1, \ldots, x_m) = (f_1(x_1, \ldots, x_m), \ldots, f_n(x_1, \ldots, x_m))$$

and form the partial derivatives $\frac{\partial f_i}{\partial x_j}$.

Proposition 3.16 *Let $\{e_1, \ldots, e_m\}$ and $\{e_1', \ldots, e_n'\}$ be the standard bases of the space $\mathbb{K}^m$ and $\mathbb{K}^n$ respectively. Then*

$$D_{e_j}(f)_x = \frac{\partial f_1}{\partial x_j}(x)e_1' + \cdots + \frac{\partial f_n}{\partial x_j}(x)e_n'$$

Proof: Let $x \in R$. Then the partial derivative $\frac{\partial f_i}{\partial x_j}(x)$ is defined as the limit

$$\frac{\partial f_i}{\partial x_j}(x) = \lim_{t \to 0} \frac{f_i(x_1, \ldots, x_j + t, \ldots, x_m) - f_i(x_1, \ldots, x_m)}{t}$$

We can write this limit

$$\frac{\partial f_i}{\partial x_j}(x) = \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t}$$

and the result follows. $\square$

In particular, the partial derivatives exist if and only if the directional derivatives in the directions of the standard basis vectors exist.

The following is immediate, by Schwarz's theorem.

Corollary 3.17 *Let $R \subseteq \mathbb{K}^m$ be an open subset, and let $(x_1, \ldots, x_m)$ be coordinates in the space $\mathbb{K}^m$. let $i, j \in \{0, \ldots, m\}$, and let $f\colon R \to \mathbb{K}$ be a map such that the second partial derivatives*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \qquad \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$$

exist and are continuous for all $x \in R$.

Then

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

$\square$

Partial derivatives are useful for computations, but depend upon a choice of basis, and are therefore less useful theoretically.

Example 3.18 Define a function $f\colon \mathbb{R}^2 \to \mathbb{R}$ by the formula

$$f(x,y) = \frac{xy^3}{x^2+y^6} \qquad (x,y) \neq (0,0)$$

and $f(0,0) = 0$.

Then the partial derivatives $\frac{\partial f}{\partial x}(0,0)$ and $\frac{\partial f}{\partial y}(0,0)$ exist. However, the function f is not continuous, and hence not differentiable, at the point $(0,0)$.

We leave the computations needed to prove the statements in the above example as an exercise.

Now, let $R \subseteq \mathbb{K}^m$ be an open subset, and let $f\colon R \to \mathbb{K}^n$ be differentiable at the point $x \in R$. Let $\{e_1, \ldots, e_m\}$ and $\{e'_1, \ldots, e'_n\}$ be the standard bases of the spaces $\mathbb{K}^m$ and $\mathbb{K}^n$ respectively.

Then by proposition 3.12 and the above proposition, we have

$$(Df)_x(e_j) = \frac{\partial f_i}{\partial x_j}(x)e'_1 + \cdots + \frac{\partial f_n}{\partial x_j}(x)e'_n$$

Hence the total derivative $(Df)_x$ has matrix $\left(\frac{\partial f_i}{\partial x_j}(x)\right)$ with respect to the standard bases.

Let $y = f(x)$, and suppose we have a map $g\colon f[R] \to \mathbb{K}^p$ that is differentiable at the point y. Abusing notation somewhat, let us write $(f_1, \ldots, f_n)$ to denote coordinates in the space $\mathbb{K}^n$. Then the total derivative $(Dg)_y$ has matrix $\left(\frac{\partial g_i}{\partial f_j}(y)\right)$, and by the chain rule the matrix of the derivative $D(g\circ f)_x$ is obtained by matrix multiplication of the above two matrices.

Theorem 3.19 *let $R \subseteq \mathbb{K}^m$ be an open subset, and let $\varphi\colon R \to \mathbb{K}^n$. Let $(x_1, \ldots, x_m)$ be coordinates in $\mathbb{K}^m$, write $\varphi(x_1, \ldots, x_m) = (\varphi_1(x_1, \ldots, x_m), \ldots \varphi_n(x_1, \ldots, x_m))$, and let $x \in R$.*

Suppose that the partial derivatives of the function φ all exist on R, and are continuous at the point x. Then the total derivative $(D\varphi)_x$ exists, and is given by the matrix $\left(\frac{\partial \varphi_i}{\partial x_j}\right)$ with respect to the standard bases of the vector spaces $\mathbb{K}^m$ and $\mathbb{K}^n$.

Proof: Let $\{e_1, \ldots, e_m\}$ and $\{e'_1, \ldots, e'_n\}$ be the standard bases of the spaces $\mathbb{K}^m$ and $\mathbb{K}^n$ respectively. Then the directional derivatives

$$D_j(\varphi)_y = D_{e_j}(\varphi)(y) = \frac{\partial \varphi_1}{\partial x_1}(y)e'_1 + \cdots + \frac{\partial \varphi_1}{\partial x_n}(y)e'_n$$

exist for all points $y \in R$, and are continuous at the point x.

Define a linear map $D\colon \mathbb{K}^m \to \mathbb{K}^n$ by the formula

$$D(v_1, \ldots, v_m) = \sum_{j=1}^{m} v_j D_j(\varphi)(x)$$

We need to show that the linear map D is the total derivative of the function φ at the point x.

Let $v = (v_1, \dots, v_m)$, and set

$$X_j = x + v_1 e_1 + \cdots v_j e_j$$

so that $X_m = x + v$ and $X_0 = x$. Then, for any sufficiently small vector v:

$$\begin{array}{rcl} \|\varphi(x+v) - \varphi(x) - D(v)\| & = & \|\varphi(x+v) - \varphi(x) - \sum_{j=1}^m v_j D_j(\varphi)(x)\| \\ & = & \| \sum_{j=1}^m (\varphi(X_j) - \varphi(X_{j-1}) - v_j D_j(\varphi)(x))\| \\ & \leq & \sum_{j=1}^m \|\varphi(X_j) - \varphi(X_{j-1}) - v_j D_j(\varphi)(x)\| \end{array}$$

We can define a path $\gamma_j\colon [0,1] \to \mathbb{K}^n$ by the formula

$$\gamma_j(t) = \varphi(X_{j-1} + t v_j e_j) - t v_j D_j(\varphi)(x)$$

Existence of all partial derivatives in the region R means that the map γ_j is differentiable. By the chain rule, the derivative is given by the formula

$$\gamma_j'(t) = v_j D_j(\varphi)(X_{j-1} + t v_j e_j) - v_j D_j(\varphi)(x)$$

By the mean value inequality, there is a point $c_j \in [0,1]$ such that

$$\|\gamma_j(1) - \gamma_j(0)\| \leq \|\gamma_j'(c_j)\|$$

ie:

$$\|\varphi(X_{j-1} + v_j e_j) - \varphi(X_{j-1}) - v_j D_j(\varphi)(x)\| \leq \|v_j D_j(\varphi)(X_{j-1} + c_j v_j e_j) - v_j D_j(\varphi)(x)\|$$

Takin sums, we see that

$$\begin{array}{rcl} \|\varphi(x+v) - \varphi(x) - D(v)\| & \leq & \sum_{j=1}^n |v_j| \|D_j(\varphi)(X_{j-1} + c_j v_j e_j) - v_j D_j(\varphi)(x)\| \\ & \leq & \|v\|_\infty \sum_{j=1}^m \|D_j(\varphi)(X_{j-1} + c_j v_j e_j) - D_j(\varphi)(x)\| \\ & \leq & \|v\| \sum_{j=1}^m \|D_j(\varphi)(X_{j-1} + c_j v_j e_j) - D_j(\varphi)(x)\| \end{array}$$

by proposition 1.17.

Now the partial derivatives, and hence the functions $D_j(\varphi)$ are continuous at the point x. Certainly $X_{j-1} \to x$ as $v \to 0$. It follows that

$$\lim_{v \to 0} \|D_j(\varphi)(X_{j-1} + c_j v_j e_j) - D_j(\varphi)(x)\| = 0$$

For sufficiently small v, let

$$r(v) = \frac{\varphi(x+v) - \varphi(x) - D(v)}{\|v\|}$$

Then

$$\varphi(x+v) = \varphi(x) + D(v) + \|v\| r(v)$$

By the above,

$$\|r(v)\| \leq \sum_{j=1}^{m} \|D_j(\varphi)(X_{j-1} + c_j v_j e_j) - D_j(\varphi)(x)\|$$

and

$$\lim_{v \to 0} r(v) = 0$$

Hence the map D is the total derivative of the function φ at the point x, as required. □

Example 3.20 Define a map $f: \mathbb{R}^2 \to \mathbb{R}^2$ by the formula

$$f(x, y) = (e^{x+y}, x^2 + y^2)$$

Write $f_1 = e^{x+y}$ and $f_2 = x^2 + y^2$. Then

$$\frac{\partial f_1}{\partial x} = e^{x+y} \qquad \frac{\partial f_1}{\partial y} = e^{x+y}$$

and

$$\frac{\partial f_2}{\partial x} = 2x \qquad \frac{\partial f_2}{\partial y} = 2y$$

These partial derivatives are all continuous. Hence the total differential is the linear transformation defined by the matrix

$$(Df)_{(x,y)} = \begin{pmatrix} e^{x+y} & e^{x+y} \\ 2x & 2y \end{pmatrix}$$

Let $v = (1, 1)$. Then the directional derivative $D_v(f)(x, y)$ is, as a column vector:

$$D_v(f)(x, y) = \begin{pmatrix} e^{x+y} & e^{x+y} \\ 2x & 2y \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2e^{x+y} \\ 2(x + y) \end{pmatrix}$$

or in other words $D_v(f)(x, y) = (2e^{x+y}, 2(x + y))$.

Example 3.21 Let $h: \mathbb{R}^2 \to \mathbb{R}$ be defined by the formula

$$h(x, y) = \sin(e^{x+y} + x^2 + y^2)$$

Define a map $g: \mathbb{R}^2 \to \mathbb{R}$ by the formula $g(x, y) = \sin(x + y)$.
Then

$$\frac{\partial g}{\partial x} = \cos(x + y) \qquad \frac{\partial g}{\partial y} = \cos(x + y)$$

These partial derivatives are both continuous. Hence the map g has total differential defined by the matrix

$$(Dg)_{(x,y)} = \begin{pmatrix} \cos(x + y) & \cos(x + y) \end{pmatrix}$$

Observe that $h = g \circ f$, where f is the function in the above example. Hence, by the chain rule

$$(Dh)_{(x,y)} = (Dg)_{(e^{x+y}, x^2+y^2)}(Df)_{(x,y)}$$

that is

$$(Dh)_{(x,y)} = \begin{pmatrix} \cos(e^{x+y} + x^2 + y^2) & \cos(e^{x+y} + x^2 + y^2) \end{pmatrix} \begin{pmatrix} e^{x+y} & e^{x+y} \\ 2x & 2y \end{pmatrix}$$

so

$$(Dh)_{(x,y)} = \big(\cos(e^{x+y} + x^2 + y^2)(e^{x+y} + 2x) \quad \cos(e^{x+y} + x^2 + y^2)(e^{x+y} + 2y)\big)$$

The proof of the following result is an exercise using the above theorem.

Proposition 3.22 *Let V and W be finite-dimensional vector spaces, and let $R \subseteq V$ be an open subset. Let $\varphi\colon V \to W$ be a function such that for all $v \in V$, the direction derivative $D_v(\varphi)(x)$ exists for all $x \in R$ and defines a continuous function $D_v(\varphi)\colon R \to W$.*

Then for all $x \in R$ the total derivative $(D\varphi)_x$ exists, and is given by the formula $(D\varphi)_x(v) = D_v(\varphi)(x)$. □

3.5 Real and Complex Differentiability

let V and W be *complex* normed vector spaces. The spaces V and W can also be considered as real normed vector spaces (with double the dimension). Let $R \subseteq V$ be an open subset. When we consider a map $\varphi\colon R \to W$, we have to distinguish between *real differentiability* and *complex differentiability*.

Recall that differentiability at a point $x \in R$ is the condition

$$\varphi(x+v) = \varphi(x) + Dv + \|v\| r(v) \qquad \lim_{v \to 0} r(v) = 0$$

where $D\colon V \to W$ is a linear map.

However, complex differentiability means that the map D is $\mathbb{C}$-linear, whereas real differentiability means that the map D is $\mathbb{R}$-linear. Thus, real differentiability is weaker than complex differentiability. Any complex-differentiable map is also real-differentiable. The converse of this statement is *not* true, as we shall see below.

Lemma 3.23 *Let V and W be complex normed vector spaces, let $R \subseteq V$ be open, and let $\varphi\colon R \to W$ be a function that is real-differentiable at a point $x \in \mathbb{R}$, with real total derivative $(D\varphi)_x\colon V \to W$.*

The function φ is complex-differentiable at the point $x \in R$ if and only if $(D\varphi)_x(iv) = i(D\varphi)_x(v)$ for all $v \in V$. If this condition is fulfilled, the complex differential is also $(D\varphi)_x$.

Proof: The condition

$$\varphi(x+v) = \varphi(x) + (D\varphi)_x v + \|v\| r(v) \qquad \lim_{v \to 0} r(v) = 0$$

completely determines the total derivative $(D\varphi)_x$ at the point x.

It follows that, if the real derivative is $(D\varphi)_x$, then we have a complex derivative, which is also $(D\varphi)_x$, if and only if the $\mathbb{R}$-linear map $(D\varphi)_x$ is also $\mathbb{C}$-linear.

But the last condition is equivalent to saying that $(D\varphi)_x(iv) = i(D\varphi)_x(v)$ for all $v \in v$. □

Theorem 3.24 *Let $R \subseteq \mathbb{C}^m$ be an open subset, and let $\varphi \colon R \to \mathbb{C}^n$ be a function that is real-differentiable at a point $x \in R$.*

Let $(z_1, \ldots, z_m)$ be coordinates in the space $\mathbb{C}^m$, and write $z_j = x_j + iy_j$, where $x_j, y_j \in \mathbb{R}$. Then the function φ is complex-differentiable at the point x if and only if

$$\frac{\partial \varphi}{\partial y_j}(x) = i \frac{\partial \varphi}{\partial x_j}(x)$$

for all j.

Proof: Write $D = (D\varphi)_x$. By the above, we need to show that $D(iv) = iD(v)$ for all $v \in \mathbb{C}^m$ if and only if

$$\frac{\partial \varphi}{\partial y_j}(x) = i \frac{\partial \varphi}{\partial x_j}(x)$$

for all j.

Let $\{e_1, \ldots, e_m\}$ be the standard basis of $\mathbb{C}^m$. We know that

$$\frac{\partial \varphi}{\partial y_j}(x) = (D_{ie_j}\varphi)(x) = D(ie_j)$$

and

$$\frac{\partial \varphi}{\partial x_j}(x) = (D_{e_j}\varphi)(x) = D(e_j)$$

By $\mathbb{R}$-linearity the condition $D(iv) = iD(v)$ holds for all $v \in \mathbb{C}^m$ if and only if $D(ie_j) = iD(e_j)$ for each basis vector e_j. The result now follows. □

We leave the proof of the following as an exercise.

Corollary 3.25 *Let $R \subseteq \mathbb{C}^m$ be an open subset, and let $\varphi \colon R \to \mathbb{C}$ be a function that is real-differentiable at a point $x \in R$.*

Write $\varphi = g + ih$, where g and h are real-valued functions. Let $(z_1, \ldots, z_m)$ be coordinates in the space $\mathbb{C}^m$, and write $z_j = x_j + iy_j$, where $x_j, y_j \in \mathbb{R}$. Then the function φ is complex-differentiable at the point x if and only if

$$\frac{\partial g}{\partial y_j}(x) = -\frac{\partial h}{\partial x_j}(x) \qquad \frac{\partial g}{\partial x_j}(x) = \frac{\partial h}{\partial y_j}(x)$$

for all j. $\square$

The above equations are called the *Cauchy-Riemann equations.*

Example 3.26 Define $f: \mathbb{C} \to \mathbb{C}$ by the formula

$$f(z) = \overline{z}$$

If have have $z = x + iy$, then $\overline{z} = x - iy$.

Writing $f = g + ih$, we have $g(x, y) = x$ and $h(x, y) = -y$. We have partial derivatives

$$\frac{\partial g}{\partial y} = 0 = \frac{\partial h}{\partial x}$$

and

$$\frac{\partial g}{\partial x} = 1 \qquad \frac{\partial h}{\partial y} = -1$$

The partial derivatives are continuous, so the function f is real-differentiable. However, the Cauchy-Riemann equations are not satisfied, so the function f is *not* complex-differentiable.

Example 3.27 Define $f: \mathbb{C}^2 \to \mathbb{C}$ by the formula

$$f(u+iv, x+iy) = u^2x^2+v^2y^2-u^2y^2-v^2x^2-4uvxy+2i(u^2xy-v^2xy+uvx^2-uvy^2)$$

Write $f = g + ih$, where g and h are real-valued functions. Then

$$g(u, v, x, y) = u^2x^2 + v^2y^2 - u^2y^2 - v^2x^2 - 4uvxy$$

and

$$h(u, v, x, y) = 2(u^2xy - v^2xy + uvx^2 - uvy^2)$$

We have partial derivatives

$$\frac{\partial g}{\partial v} = 2vy^2 - 2vx^2 - 4uxy \qquad \frac{\partial h}{\partial u} = -2vy^2 + 2vx^2 + 4uxy$$

$$\frac{\partial g}{\partial u} = 2ux^2 - 2uy^2 - 4vxy \qquad \frac{\partial h}{\partial v} = 2ux^2 - 2uy^2 - 4vxy$$

$$\frac{\partial g}{\partial y} = 2v^2y - 2u^2y - 4uvx \qquad \frac{\partial h}{\partial x} = -2v^2y + 2u^2y + 4uvx$$

$$\frac{\partial g}{\partial x} = 2u^2x - 2v^2x - 4uvy \qquad \frac{\partial h}{\partial y} = 2u^2x - 2v^2x - 4uvy$$

The above are all continuous, so the function f is real-differentiable. Further, the Cauchy-Riemann equations are satisfied so the function f is also complex-differentiable.

4 Differential Forms

4.1 The Notion of a Differential Form

Let V and W be normed vector spaces. Recall that $Hom(V, w)$ is the vector space of bounded linear maps $T: V \to W$. Assuming $V \neq \{0\}$, we can define a norm on the space $Hom(V, W)$ by the formula

$$\|T\| = \sup_{v \in V, v \neq 0} \frac{\|Tv\|}{\|v\|}$$

As already mentioned, we call this norm the operator norm. If we need to talk about continuity or differentiability in the space $Hom(V, W)$, we assume that we have this norm.

If the spaces V and W are finite-dimensional, with dimensions m and n respectively, then the space $Hom(V, W)$ is also finite-dimensional, with dimension mn. In this case, we do not need to worry about the particular choices of norms.

Definition 4.1 Let V and W be finite-dimensional vector spaces. Let $R \subseteq V$ be an open set. Then a *differential form* of *degree* 1 on the set R with *values* in the vector space W is a map

$$\omega: R \to Hom(V, W)$$

We write $\Omega^1(R; W)$ to denote the set of degree 1 differential forms on R with values in W.

Actually, we only consider degree one differential forms in this course. Hence, from now on, we refer to a map $\omega: R \to Hom(V, W)$ just as a *differential form*.

Example 4.2 Let $\varphi: R \to W$ be a differentiable map. Then for each point $x \in R$ we can form the derivative $(D\varphi)(x) \in Hom(V, W)$.

Thus the derivative $D\varphi$ is a differential form.

Observe that the set $\Omega^1(R; W)$ (with pointwise addition and scalar multiplication) is a vector space over the field $\mathbb{K}$. Moreover, given a function $f: R \to \mathbb{K}$ and a differential form ω, we can define a differential form $f\omega$ by writing

$$(f\omega)(x) = f(x)\omega(x) \qquad x \in R$$

Here $f(x) \in \mathbb{K}$ is a scalar, and $\omega(x): V \to W$ is a linear map, so the above makes sense.

Definition 4.3 Let $R \subseteq \mathbb{K}^m$ be an open subset. Then we define the differential form $dx_j \in \Omega^1(R; \mathbb{K})$ by writing

$$dx_j(x)(v_1, \ldots, v_m) = v_j$$

for all $x \in R$ and $(v_1, \ldots, v_m) \in \mathbb{K}^m$.

Given functions $g_j\colon R \to \mathbb{K}$, we have a differential form

$$g_1 dx_1 + \cdots + g_m dx_m \in \Omega^1(R, W)$$

Proposition 4.4 *Let $R \subseteq \mathbb{K}^m$ be an open set, and let $\omega \in \Omega^1(R; \mathbb{K})$. Then there are functions $g_j\colon R \to \mathbb{K}$ such that*

$$\omega = g_1 dx_1 + \cdots + g_m dx_m$$

Proof: Let $x \in R$. Then we have a linear map $\omega(x)\colon \mathbb{K}^m \to W$.

Let $\{e_1, \ldots, e_m\}$ be the standard basis of the space $\mathbb{K}^m$. Define the function $g_j\colon R \to \mathbb{K}$ by writing $g_j(x) = \omega(x)(e_j)$. Then, for a vector $v = (v_1, \ldots, v_m) \in K^m$, we have

$$\begin{array}{rcl} \omega(x)(v_1, \ldots, v_n) & = & v_1 \omega(x)(e_1) + \cdots + v_m \omega(x)(e_m) \\ & = & g_1(x) v_1 + \cdots + g_m(x) v_m \\ & = & g_1(x) dx_1(x)(v) + \cdots + g_m(x) dx_m(x)(v) \end{array}$$

and are done. □

Proposition 4.5 *Let $f\colon R \to \mathbb{K}$ be a differentiable function. Then*

$$Df = \sum_{j=1}^{m} \frac{\partial f}{\partial x_j} dx_j$$

Proof: Let $v = (v_1, \ldots, v_m) \in \mathbb{K}^m$, and let $x \in R$. Since the derivative $(Df)(x)$ has matrix $\left(\frac{\partial f}{\partial x_j}(x)\right)$ with respect to the standard basis of the space $\mathbb{K}^m$, we see that

$$(Df)(x)(v) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) v_j = \left(\sum_{j=1}^{m} \frac{\partial f}{\partial x_j} dx_j \right)(x)(v)$$

and we are done. □

4.2 Exactness

We have already seen that for a differentiable function, f, the total derivative Df is a differentiable form. The fundamantal question we are going to examine in this course is precisely which differential forms arise in this way.

Definition 4.6 We call a differential form $\omega\colon R \to Hom(V, W)$ *exact* if there is a differentiable function $\varphi\colon R \to W$ such that $D\varphi = \omega$.

Such a function φ is called a *primitive function* for the differential form ω.

For example, let $R \subseteq \mathbb{R}$ be an open interval, and let $\omega: R \to Hom(\mathbb{R}, \mathbb{R})$ be a continuous differential form. Then we can write $\omega = f\ dx$, where f is a continuous function.

Pick $x_0 \in R$, and define a function $F: R \to \mathbb{R}$ by the formula

$$F(x) = \int_{x_0}^{x} f(t)\ dt$$

Then by the fundamental theorem of calculus, $F'(x) = f(x)$. It follows that $Df = f\ dx = \omega$. Thus the form ω is exact, with primitive function F.

In higher dimensions, this problem is far more difficult to solve, and not every continuous form is exact. The exactness question depends upon, amongst other things, the shape of the region R.

Proposition 4.7 *Let V and W be finite-dimensional vector spaces, let $R \subseteq V$ be a* connected *open subset, and let $\omega \in \Omega^1(R; W)$ be an exact differential form.*

Let φ_1 and φ_2 be primitive functions of the form ω. Then the difference $\varphi_1 - \varphi_2$ is constant.

Proof: The difference $\varphi_1 - \varphi_2$ is differentiable, with derivative

$$D(\varphi_1 - \varphi_2) = D\varphi_1 - D\varphi_2 = \omega - \omega = 0$$

The result now follows by theorem 3.14. □

Corollary 4.8 *Let $\omega \in \Omega^1(R; W)$ be an exact differential form, where $R \subseteq V$ is a connected open susbset.*

Let $x \in R$ and $w \in W$. Then the form ω has a unique primitive function, $\varphi: R \to W$ such that $\varphi(x) = w$.

Proof: Let $\psi: R \to W$ be a primitive function for the form ω. Define $\varphi: R \to W$ by the formula

$$\phi(x) = psi(x) + w - \psi(x_0)$$

Then $D\phi = D\psi = \omega$, since the derivative of a constant function is constant, and $\phi(x_0) = w$.

Uniqueness now follows from the above proposition. □

4.3 Pullbacks

Let V and W be finite-dimensional vector spaces. Consider a differential form $\omega \in \Omega^1(R; W)$ where $R \subseteq V$ is an open subset.

Let V' be another finite-dimensional vector space, let $R' \subseteq V'$ be open, and let $f: R' \to R$ be a differentiable map. Then the differential Df gives us a linear map $(Df)_x: V' \to V$ for each point $x \in R'$.

Definition 4.9 We define the *pullback* of the form ω *along* f to be the differential form $f^*\omega \in \Omega^1(R';W)$ defined by writing

$$(f^*\omega)(x)(v') = \omega(f(x))(Df)_x(v')) \qquad x \in R',\ v' in V'$$

Thus, given $x \in R'$, the linear map $(f^*\omega)(x)\colon V' \to W$ is the composition

$$V' \stackrel{Df_x}{\to} V \stackrel{\omega(f(x))}{\to} W$$

Proposition 4.10 *Let $\omega \in \Omega^1(R;W)$ be an exact differential form, with primitive function φ. let $f\colon R' \to R$ be a differentiable map. Then the pullback $f^*\omega \in \Omega^1(R;W)$ is also exact, with primitive function $\varphi \circ f$.*

Proof: Let $x \in R'$ and $v \in V'$. By the chain rule

$$D(\varphi \circ f)_x(v') = (D\varphi)_{f(x)}(Df)_x(v')$$

By definition of pullback

$$f^*\omega(x)(v') = \omega(f(x))(Df)_x(v') = (D\varphi)_{f(x)}(Df)_x(v')$$

$\square$

Let $\omega \in \Omega^1(R;\mathbb{K})$, where $R \subseteq \mathbb{K}^n$ is an open subset. Write

$$\omega = g_1 dx_1 + \cdots + g_n dx_n$$

Let $f\colon R' \to R$ be a differentiable map, where $R' \subseteq \mathbb{K}^m$. Write

$$f(u_1, \ldots, u_m) = (f_1(u_1, \ldots, u_m), \ldots, f_n(u_1, \ldots, u_m))$$

The derivative $(Df)(u)$ is the matrix $\left(\frac{\partial f_i}{\partial u_j}(u)\right)$. Hence the pullback $f^*\omega(u)$ is obtained from the expression

$$\omega = g_1 dx_1 + \cdots + g_n dx_n$$

by writing

$$dx_i = \frac{\partial f_i}{\partial u_1}(u)du_1 + \cdots + \frac{\partial f_i}{\partial u_m}(u)du_m$$

and

$$g_i(x) = g_i(f(u))$$

Example 4.11 Let $\omega = y\ dx - x\ dy \in \Omega^1(\mathbb{R}^2, \mathbb{R})$. Define a map $f\colon (0,\infty) \times \mathbb{R} \to \mathbb{R}^2$ by the the formula

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

Write $x = r\cos\theta$ and $y = r\sin\theta$. Then

$$\frac{\partial x}{\partial r} = \cos\theta \qquad \frac{\partial x}{\partial \theta} = -r\sin\theta$$

and

$$\frac{\partial y}{\partial r} = \sin\theta \qquad \frac{\partial y}{\partial \theta} = r\cos\theta$$

We see that

$$dx = \cos\theta\ dr - r\sin\theta\ d\theta \qquad dy = \sin\theta\ dr + r\cos\theta\ d\theta$$

Hence

$$\begin{aligned} f^*\omega &= r\sin\theta\cos\theta\ dr - r^2\sin^2\theta\ d\theta - r\cos\theta\sin\theta\ dr - r^2\cos^2\theta\ d\theta \\ &= -r^2(\sin^2\theta + \cos^2\theta)\ d\theta \\ &= -r^2\ d\theta \end{aligned}$$

4.4 Symmetric Forms

Let V and W be real finite-dimensional vector spaces. Let $R \subseteq V$ be open, and let $\omega \in \Omega^1(R; W)$. Then the differential form ω is a map $\omega\colon R \to Hom(V, W)$. It makes sense to ask whether the map ω, as written above, is differentiable. If this is the case, then the derivative $(D\omega)_x$ at the point $x \in R$ is a map

$$(D\omega)_x\colon V \to Hom(V, W)$$

Write $(D\omega)_x(u, v) = (D\omega)_x(u)(v)$. Then we can view the differential $(D\omega)_x$ as a map $(D\omega)_x\colon V \times V \to W$.

The map $(D\Omega)_x$ is a *bilinear map*, ie: if we fix $u, v \in V$, the maps $(D\omega)_x(u, -)\colon V \to W$ and $(D\omega)_x(-, v)\colon V \to W$ are linear.

Definition 4.12 let $\omega \in \Omega^1(R; w)$ be a differentiable differential form. Then we call the form ω *symmetric* if, given $x \in R$, we have

$$(D\omega)_x(u, v) = (D\omega)_x(v, u)$$

for all $u, v \in V$.

Note that we are only considering the notion of symmetry for real differential forms. The plan in this section is to show that any continuously differentiable exact differential form is symmetric. To do this, we first need a computational lemma.

Lemma 4.13 *Let $\omega \in \Omega^1(R; W)$ be a differentiable differential form.*

1. *Let $v \in V$. Define a function $f_v\colon R \to W$ by the formula $f_v(x) = \omega(x)(v)$. Then, for a vector $u \in V$, we have directional derivative $D_u(f_v)_x = (D\omega)_x(u, v)$.*

2. *Suppose that the form ω is exact, with primitive function φ. Then for all $x \in r$ and $u, v \in V$, we have the formula $(D\omega)_x(u, v) = D_u D_v(\varphi)(x)$.*

Proof:

1. Define a linear map $Ev_v\colon Hom(V,W) \to W$ by the formula $Ev_v(T) = T(v)$. Then the map f_v is the composition

$$R \overset{\omega}{\to} Hom(V,W) \overset{Ev_v}{\to} W$$

Since the map Ev_v is linear, we have total derivative $(DEv_v)(T) = Ev_v(T)$. Hence, by the chain rule

$$D_u(f_v)_x = (Df_v)_x(u) = Ev_v(D\omega)_x(u) = (D\omega)_x(u)(v) = (D\omega)_x(u,v)$$

2. By the above, we know that

$$(D\omega)_x(u,v) = D_u(f_v)_x$$

where

$$f_v(x) = \omega(x)(v) = (D\varphi)_x(v) = D_v(\varphi)_x$$

Hence

$$(D\omega)_x(u,v) = D_u D_v(\varphi)_x$$

as claimed.

□

Theorem 4.14 *let $\omega \in \Omega^1(R;W)$ be a continuously differentiable differentiable form. Suppose that the form ω is exact. Then the form ω is symmetric.*

Proof: Since the form ω is exact, we have a primitive function φ, and by the above lemma

$$D\omega(u,v) = D_u D_v(\varphi)$$

Since the form ω has continuous derivative, Schwarz's theorem applies, and $D_u D_v(\varphi) = D_v D_u(\varphi)$. Thus the form ω is symmetric, and we are done. □

The following proposition provides an easy way to test when a form is symmetric.

Proposition 4.15 *Let $R \subseteq \mathbb{R}^m$ be an open subset, and let $\omega \in \Omega^1(R;\mathbb{R})$ be a differentiable differential form. Write*

$$\omega = g_1 dx_1 + \cdots + g_m dx_m$$

Then the form ω is symmetric if and only if

$$\frac{partial g_i}{\partial x_j} = \frac{\partial g_j}{\partial x_i}$$

for all i and j.

Proof: Since the form ω is differentiable, the functions g_i are all differentiable. Let $\{e_1, \ldots, e_m\}$ be the standard basis of the space $\mathbb{R}^m$. Define a function $f_i\colon R \to \mathbb{K}$ by the formula $f_i(x) = \omega(x)(e_i)$.

Then by lemma 4.13, we have

$$\frac{\partial f_i}{\partial x_j}(x) = (D\omega)_x(e_j, e_i)$$

Observe that $\omega(x)(e_i) = g_i$. Hence

$$(D\omega)_x(e_j, e_i) = \frac{\partial g_i}{\partial x_j}(x)$$

Thus, if the the form ω is symmetric, the condition

$$\frac{\partial g_i}{\partial x_j} = \frac{\partial g_j}{\partial x_i}$$

holds for all i and j.

Conversely, if

$$\frac{\partial g_i}{\partial x_j} = \frac{\partial g_j}{\partial x_i}$$

for all i and j, then, by the above, given $x \in R$

$$(D\omega)_x(e_i, e_j) = (D\omega)_x(e_j, e_i)$$

It is now easy to check that $(D\omega)_x(u, v) = (D\omega)_x(v, u)$ for all $u, v \in \mathbb{K}^m$, since the set $\{e_1, \ldots, e_m\}$ is a basis of the space $\mathbb{K}^m$, and the map $(D\omega)_x\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is bilinear. □

Example 4.16 Let

$$\omega = x\ dx + y\ dy \in \Omega^1(\mathbb{K}^2, \mathbb{K})$$

Then we have

$$\omega = g_1\ dx + g_2\ dy \qquad g_1(x, y) = x,\ g_2(x, y) = y$$

Now

$$\frac{\partial g_1}{\partial y} = 0 = \frac{\partial g_2}{\partial x}$$

So the differential form ω is symmetric.

Example 4.17 Let

$$\omega = xy\ dx + xy\ dy \in \Omega^1(\mathbb{K}^2, \mathbb{K})$$

Then we have

$$\omega = g_1\ dx + g_2\ dy \qquad g_1(x, y) = xy,\ g_2(x, y) = xy$$

Now

$$\frac{\partial g_1}{\partial y} = x \qquad \frac{\partial g_2}{\partial x} = y$$

So the differential form ω is not symmetric.

4.5 Complex Differential Forms and Symmetry

Let $R \subseteq \mathbb{C}$ be an open subset, and let $\omega \in \Omega^1(R;\mathbb{C})$ be a complex differential form. Then we can write

$$\omega = f\ dz$$

for some function $f\colon R \to \mathbb{C}$. Here $dz(w)\colon \mathbb{C} \to \mathbb{C}$ is the identity map for all $w \in \mathbb{C}$.

Using the basis $\{1, i\}$ we can of course regard the complex numbers, $\mathbb{C}$, as real vector space, and the form ω is a real differential form.

Theorem 4.18 *Let $\omega = f\ dz \in \Omega^1(R;\mathbb{C})$ be continuously differentiable as a real differential form. Then the form ω is symmetric (as a real differential form) if and only if the function $f\colon R \to \mathbb{C}$ is complex-differentiable.*

Proof: Write $z = x + iy$, where $x, y \in \mathbb{R}$, and $f(z) = f(x + iy) = g(x, y) + ih(x, y)$, where g and h are real-valued functions.

We know that $dz = dx + i\ dy$. Hence

$$f\ dz = (g + ih)(dx + i\ dy) = (g + ih)dx + (ig - h)dy$$

By proposition 4.15, the form $f\ dz$ is symmetric if and only if

$$\frac{\partial(g + ih)}{\partial y} = \frac{\partial(ig - h)}{\partial x}$$

that is

$$\frac{\partial g}{\partial y} = -\frac{\partial h}{\partial x} \qquad \frac{\partial g}{\partial x} = \frac{\partial h}{\partial y}$$

But the above are the Cauchy-Riemann equations, and the functions g and h are continuously differentiable. The result therefore follows by corollary 3.25. □

The above result is the key to relating results concerning differential forms to complex analysis.

5 Integration of Differential Forms

5.1 Integration of Vector-Valued Functions

Let V be a finite-dimensional vector space over the field $\mathbb{K}$. Let $f\colon [a, b] \to V$ be a continuous map, and let $\{e_1, \ldots, e_n\}$ be a basis of the space V. Write $f(t) = f_1(t)e_1 + \cdots + f_n(t)e_n$. Then the functions $f_i\colon [a, b] \to \mathbb{R}$ are continuous, and we can define the integral

$$\int_a^b f(t)\ dt = \left(\int_a^b f_1(t)\ dt\right) e_1 + \cdots + \left(\int_a^b f_n(t)\ dt\right) e_n$$

It is easy to check that this concept is independent of the choice of basis. Further, the fundamental results concerning integrals in this setting can be deduced from the corresponding results for ordinary, scalar-valued, integrals.

Some such results, which we will use in this course, are the following.

- Linearity

  Let $\alpha, \beta \in \mathbb{K}$, and let $f, g\colon [a,b] \to V$ be continuous functions. Then

$$\int_a^b (\alpha f(t) + \beta g(t))\ dt = \alpha \int_a^b f(t)\ dt + \int_a^b g(t)\ dt$$

- Additivity

  Let $f\colon [a,b] \to V$ be a continuous function. Let $c \in (a,b)$. Then

$$\int_a^b f(t)\ dt = \int_a^c f(t)\ dt + \int_c^b f(t)\ dt$$

- The Fundamental Theorem of Calculus (version 1)

  Let $I \subseteq \mathbb{R}$ be an open interval, and let $f\colon I \to V$ be a continuous function. Choose $x_0 \in I$, and define

$$F(x) = \int_{x_0}^x f(t)\ dt$$

  Then the function F is differentiable, with derivative $F'(x) = f(x)$.

- The Fundemental Theorem of Calculus (version 2)

  Let $f\colon [a,b] \to V$. Suppose we have a differentiable function $F\colon [a,b] \to V$ such that $F'(t) = f(t)$ for all $t \in [a,b]$. Then

$$\int_a^b f(t)\ dt = F(b) - F(a)$$

- The Mean Value Inequality

  Suppose that the space W is equipped with a norm. Let $f\colon [a,b] \to V$ be a continuous function. Then

$$\left\| \int_a^b f(t)\ dt \right\| \leq \int_a^b \|f(t)\|\ dt$$

- Differentiation Under the Integral Sign

  Let $f\colon [a,b] \times [c,d] \to V$ be a continuous map such that the partial derivative $\frac{\partial f}{\partial t}(x,t)$ exists and is continuous in both the variables x and t. Then

$$\frac{d}{dt} \int_a^b f(x,t)\ dx = \int_a^b \frac{\partial f}{\partial t}(x,t)\ dx$$

5.2 Path Integrals

Let V and W be finite-dimensional vector spaces, and let $R \subseteq V$ be an open subset. Let $\omega \in \Omega^1(R; W)$ be a continuous differential form, and let $\gamma\colon [a,b] \to R$ be a differentiable path.

The pullback $\gamma^*\omega \in \Omega^1([a,b]; W)$ is a continuous map $\gamma^*\omega\colon [a,b] \to Hom(\mathbb{R}, W)$. As we discussed in example 1.23, we can identify the space $Hom(\mathbb{R}, W)$ with the space W by associating the vector $w \in W$ to the map $T_w\colon \mathbb{R} \to W$ defined by the formula $T_w(\alpha) = \alpha w$. Furthermore, with the operator norm, $\|T_w\| = \|w\|$.

If we make this association, we have a continuous map $\gamma^*\omega\colon [a,b] \to W$ defined by the formula

$$(\gamma^*\omega)(t) = \omega(\gamma(t))(\gamma'(t))$$

Definition 5.1 Let $\omega \in \Omega^1(R; W)$ be a continuous differential form, and let $\gamma\colon [a,b] \to R$ be a differentiable map. Then we define the *path integral* of ω along the path γ by writing

$$\int_\gamma \omega = \int_a^b \gamma^*\omega = \int_a^b \omega(\gamma(t))\gamma'(t)\ dt \in W$$

As a special case, let $R \subseteq \mathbb{R}^n$ be an open subset, and let $\omega \in \Omega^1(R; \mathbb{R})$ be a continuous differential form.

Write

$$\omega = g_1 dx_1 + \cdots + g_n dx_n$$

where each map $g_j\colon R \to \mathbb{R}$ is continuous. Let $\gamma\colon [a,b] \to R$ be a differentiable path. Write

$$\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t)$$

Then by the above formula for the pull-back,

$$(\gamma^*\omega)(t) = g_1(\gamma(t))\gamma_1'(t) + \cdots + g_n(\gamma(t))\gamma_n'(t)$$

and

$$\int_\gamma \omega = \int_a^b g_1(\gamma(t))\gamma_1'(t)\ dt + \cdots + \int_a^b g_n(\gamma(t))\gamma_n'(t)\ dt$$

Example 5.2 Let $R \subseteq \mathbb{C}$ be an open subset, and let $f\colon R \to \mathbb{C}$ be a continuous function. Then, by definition

$$\int_\gamma f(z)\ dz = \int_a^b f(\gamma(t))\gamma'(t)\ dt$$

which is the usual definition of a complex path integral.

Thus, integration of differential forms along a path is a generalisation of the notion of integration of complex functions along a path.

Example 5.3 Define a path $\gamma\colon [0, \pi] \to \mathbb{R}^2$ by the formula $\gamma(t) = (\cos t, \sin t)$. Let $\omega = x\ dy - y\ dx$.

Then

$$\gamma'(t) = (-\sin t, \cos t)$$

and

$$(\gamma^*\omega)(t) = \cos t(\cos t) - (\sin t)(-\sin t) = \cos^2 t + \sin^2 t = 1$$

We see that

$$\int_\gamma \omega = \int_0^\pi dt = \pi$$

Definition 5.4 Let V be a finite-dimensional vector space, and let $R \subseteq V$. We call a continuous path $\gamma\colon [a, b] \to R$ *piecewise differentiable* if there are values

$$a = c_0 < c_1 < \cdots < c_k = b$$

such that the restricted paths $\gamma|_{[c_i, c_{i+1}]}\colon [c_i, c_{i+1}] \to R$ are differentiable for all i.

Given a piecewise differentiable path as above, and a continuous differential form $\omega \in \Omega^1(R; W)$, then we define the path integral

$$\int_\gamma \omega = \int_{c_0}^{c_1} \gamma^*\omega + \int_{c_1}^{c_2} \gamma^*\omega + \cdots + \int_{c_{k-1}}^{c_k} \gamma^*\omega$$

We now present a number of results to help us evaluate and analyse path integrals. In all of these results, we consider finite-dimensional vector spaces V and W, and an open subset $R \subseteq V$.

Proposition 5.5 *Let $\alpha_1, \alpha_2 \in \mathbb{K}$, and let $\omega_1, \omega_2 \in \Omega(R; W)$ be continuous differential forms. Let $\gamma\colon [a, b] \to R$ be a piecewise differentiable path. Then*

$$\int_\gamma (\alpha_1\omega_1 + \alpha_2\omega_2) = \alpha_1 \int_\gamma \omega_1 + \alpha_2 \int_\gamma \omega_2$$

Proof: The result follows immediately be definition of the path integral and linearity of the integral of a vector-valued function. $\square$

Definition 5.6 Let $\gamma\colon [a, b] \to R$ and $\beta\colon [b, c] \to R$ be a path such that $\gamma(b) = \beta(b)$. Then we define the *concatenation* of γ and β to be the path $\gamma\beta\colon [a, c] \to R$ defined by writing

$$\gamma\beta(t) = \begin{cases} \gamma(t) & t \leq b \\ \beta(t) & t \geq b \end{cases}$$

By definition the concatenation $\gamma\beta$ is a continuous function, and therefore a path. If the paths γ and β are piecewise-differentiable, then so is the concatenation $\gamma\beta$.

Note that, even when the paths $\gamma\colon [a,b] \to R$ and $\beta\colon [b,c] \to R$ are differentiable, the concatenation $\gamma\beta$ is probably not differentiable, but only piecewise-differentiable; in general, we have a 'kink' at the point b.

The following result is completely straightforward to check.

Proposition 5.7 *Let $\omega \in \Omega^1(R;W)$ be a continuous differential form, and let $\gamma\colon [a,b] \to R$ and $\beta\colon [b,c] \to R$ be piecewise differentiable paths such that $\gamma(b) = \beta(b)$. Then*

$$\int_{\gamma\beta} \omega = \int_\gamma \omega + \int_\beta \omega$$

□

Definition 5.8 Let $\gamma\colon [a,b] \to R$ be a continuous path. Then we define the *reverse path*, $\overline{\gamma}\colon [a,b] \to R$, by the formula

$$\overline{\gamma}(t) = \gamma(a+b-t)$$

The reverse of a piecewise-differentiable path is again piecewise-differentiable. The following result is again straightforward.

Proposition 5.9 *Let $\gamma\colon [a,b] \to R$ be a piecewise-differentiable path, and let $\omega \in \Omega^1(R;W)$ be continuous. Then*

$$\int_{\overline{\gamma}} \omega = -\int_\gamma \omega$$

□

The following result shows that we can reparametrise paths when evaluating integrals. We leave the proof of the following as an exercise.

Proposition 5.10 *Let $\gamma\colon [a,b] \to R$ be a piecewise-differentiable path, and let $\theta\colon [c,d] \to [a,b]$ be a differentiable mapping such that $\theta(c) = a$ and $\theta(d) = b$.*

Let $\omega \in \Omega^1(R;W)$ be continuous. Then

$$\int_\gamma \omega = \int_{\gamma\circ\theta} \omega$$

□

Proposition 5.11 *Let U be a finite-dimensional vector space, let $S \subseteq U$ be open, and let $\psi\colon S \to R$ be a continuously differentiable map.*

Let $\beta\colon [a,b] \to S$ and $\gamma\colon [a,b] \to R$ be piecewise-differentiable paths such that $\gamma = \psi \circ \beta$. Then

$$\int_\gamma \omega = \int_\beta \psi^*\omega$$

Proof: By proposition 5.7, it suffices to check the result when the paths γ and β are differentiable. In this case, by the chain rule and the definition of pull-back

$$\begin{array}{rcl} \int_\gamma \omega & = & \int_a^b \omega(\gamma(t))\gamma'(t)\ dt \\ & = & \int_a^b \omega(\psi(\beta(t))(D\psi)_{\beta(t)}\beta'(t)\ dt \\ & = & \int_a^b \psi^*(\omega)(\beta(t))\beta'(t)\ dt \\ & = & \int_\beta \psi^*\omega \end{array}$$

and we are done. □

Now, let V and W be finite-dimensional normed vector spaces. Given a differential form $\omega \in \Omega^1(R;W)$, for a point $x \in R$, we have a bounded linear map $\omega(x): V \to W$. If we equip the space $Hom(V,W)$ with the operator norm, we can define the norm $\|\omega(x)\|$.

Proposition 5.12 *Let $\gamma: [a,b] \to R$ be a piecewise-differentiable path. Let $\omega \in \Omega^1(R;W)$ be a continuous differential form. Then*

$$\left\| \int_\gamma \omega \right\| \leq \int_a^b \|\omega(\gamma(t))\| \cdot \|\gamma'(t)\|\ dt$$

Proof: By the mean value inequality for integrals, we have

$$\begin{array}{rcl} \| \int_\gamma \omega \| & = & \| \int_a^b \omega(\gamma(t), \gamma'(t))\ dt \| \\ & \leq & \int_a^b \|\omega(\gamma(t))\gamma'(t)\|\ dt \\ & \leq & \int_a^b \|\omega(\gamma(t))\| \cdot \|\gamma'(t)\|\ dt \end{array}$$

Here the last inequality follows by the definition of the operator norm. □

5.3 Path Integrals and Exact Forms

Let V and W be finite-dimensional vector spaces, and let $R \subseteq V$ be an open subset. Recall that a differential form $\omega \in \Omega^1(R;W)$ is termed *exact* if there is a differentiable function $\varphi: R \to W$ such that $D\varphi = \omega$. Such a function φ is called a *primitive function* for the form ω.

Definition 5.13 We call a path $\gamma: [a,b] \to R$ *closed* if $\gamma(a) = \gamma(b)$.

A closed path is also called a *loop* or a *contour*. Given a continuous differential form $\omega \in \Omega^1(R;W)$, we sometimes write

$$\int_\gamma \omega = \oint_\gamma \omega$$

when the path γ is a loop.

Theorem 5.14 *Suppose that the subset $R \subseteq V$ is connected. Let $\omega \in \Omega^1(R; W)$ be a continuous differential form. Then the following are equivalent.*

1. *The form ω is exact.*
2. *For every piecewise differentiable path $\gamma\colon [a,b] \to R$, the path integral $\int_\gamma \omega$ depends only on the start point $\gamma(a)$ and end point $\gamma(b)$.*
3. *For every closed piecewise differentiable path γ, we have $\oint_\gamma \omega = 0$.*

Further, if the form ω is exact, and we choose a point $x_0 \in R$, then we have a primitive function

$$\varphi(x) = \int_\gamma \omega$$

where we choose γ to be any *path in the region R with start point x_0 and end point x.*

Proof:

- (1) $\Rightarrow$ (2):

  Let ω be exact, with primitive function φ. Let $\gamma\colon [a,b] \to V$ be a differentiable path.

  By proposition 4.10, we have $\gamma^*\omega = D(\varphi \circ \gamma)$. Observe:

  $$\int_\gamma \omega = \int_a^b \gamma^*\omega = \int_a^b D(\varphi \circ \gamma) = \int_a^b (\varphi \circ \gamma)'(t)\ dt$$

  Hence, by the fundamental theorem of calculus

  $$\int_\gamma \omega = \varphi(\gamma(b)) - \varphi(\gamma(a))$$

  which only depends on the points $\gamma(b)$ and $\gamma(a)$ since any two primitive functions only differ by a constant.

  The corresponding result for piecewise-differentiable paths follows by the above by considering such a path as a concatenation of differentiable paths.

- (2) $\Rightarrow$ (3):

  Let $\gamma\colon [a,b] \to R$ be a closed piecewise-differentiable path. Then by (2), the integral $\int_\gamma \omega$ only depends on the points $\gamma(a)$ and $\gamma(b)$. But $\gamma(b) = \gamma(a)$, so the integral along the path γ is the same as the integral along a constant path, which is zero.

- (3) $\Rightarrow$ (2):

  Let γ_1 and γ_2 be two piecewise-differentiable paths where γ_1 and γ_2 begin at the same point, and γ_1 and γ_1 and γ_2 end at the same point. Then the end-point of the path γ_1 is the same as the start-point of the reversed path $\overline{\gamma_2}$, so we can form the concatenation $\gamma_1\overline{\gamma_2}$.

The path $\gamma_1\overline{\gamma_2}$ is a closed path, so by (3):

$$0 = \int_{\gamma_1\overline{\gamma_2}} \omega = \int_{\gamma_1} \omega + \int_{\overline{\gamma_2}} \omega = \int_{\gamma_1} \omega - \int_{\gamma_2} \omega$$

- (2) $\Rightarrow$ (1):

  Pick $x_0 \in R$. Let $x \in R$. Since the open set R is a connected open set of a finite-dimensional vector space, it is also path-connected, and we have a path $\gamma_x\colon [a,b] \to R$ such that $\gamma_x(a) = x_0$ and $\gamma_x(b) = x$. It is an exercise to show that we can choose such a path to be piecewise-differentiable.

  By (2), the function
  $$\varphi(x) = \int_\gamma \omega$$
  is well-defined, that is to say it only depends on the point x, and not the precise choice of path γ_x.

  Now, let $v \in V$. Since the set R is open, we can find $s > 0$ such that $\overline{B(0, s\|v\|)} \subseteq R$. Define a path
  $$\gamma_{v,s}\colon [0,s] \to R \qquad \gamma_{v,s}(t) = x + tv$$
  Then the path $\gamma_{v,s}$ is a linear path from the point x to the point $x + sv$. Using the above path γ, and the definition of the function φ, we see
  $$\varphi(x+sv) - \varphi(x) = \int_{\gamma\gamma_{v,s}} \omega - \int_\gamma \omega = \int_{\gamma_{v,s}} \omega = \int_0^s \omega(x+tv)(v)\ dt$$
  Hence we have directional derivative
  $$D_v(\varphi)(x) = \lim_{s\to 0} \frac{1}{s} \int_0^s \omega(x+tv)(v)\ dt$$
  By the definition of the derivative, and the fundamental theorem of calculus, we deduce that $D_v(\varphi)(x) = \omega(x)(v)$. Since the differential form ω is continuous, the total derivative $D\varphi$ exists, and is defined by the formula $(D\varphi)(x) = \omega(x)$. This completes the proof.

The last part of the theorem follows by the method used to prove the implication (2) $\Rightarrow$ (1). □

One further point from the above proof is worth singling out. Namely, if ω is an exact differential form, with primitive function φ, and $\gamma\colon [a,b] \to R$ is a piecewise-differentiable path, then

$$\int_\gamma \omega = \varphi(\gamma(b)) - \varphi(\gamma(a))$$

5.4 Path Integrals and Symmetric Forms

Let V and W be real finite-dimensional vector spaces, and let $R \subseteq V$ be open. Let $\omega\colon R\colon Hom(V,W)$ be a differentiable differential form, and let us write $(D\omega)_x(v)(w) = (D\omega)_x(v,w)$.

Recall that we call the differential form ω *symmetric* if $(D\omega)_x(u,v) = (D\omega)_x(v,u)$ for all points $x \in R$ and vectors u and v. We have already seen that any exact differential form is symmetric. Our aim in this section is to prove a partial converse of this result.

The converse depends on the shape of the region R under consideration.

Definition 5.15 Let V be a vector space. We call a subset $R \subseteq V$ *star-shaped* if there is a point $x_0 \in R$ such that for all $x \in R$ the line segment $\gamma\colon [0,1] \to V$ defined by the formula $\gamma(t) = x_0 + t(x - x_0)$ is contained entirely in the set R.

Example 5.16 The vector space V is itself star-shaped.

Example 5.17 Let V be a normed vector space, let $x_0 \in V$, and let $\delta > 0$. Then the open ball $B(x_0, \delta)$ is star-shaped: for any point $x \in V$ the line segment from x_0 to x is contained within the set $B(x_0, \delta)$.

Example 5.18 The punctured complex plane $\mathbb{C}^x = \mathbb{C}\backslash\{0\}$ is not star-shaped.

Observe that a star-shaped set is certainly path-connected.

Now, let $\omega\colon R \to Hom(V,W)$ be a differentiable differential form. Write $\tilde{\omega}(x,v) = \omega(x)(v)$. Then the function $\tilde{\omega}$ is the composition

$$R \times V \overset{(\omega,1)}{\to} Hom(V,W) \times V \overset{Ev}{\to} W$$

where we define $Ev\colon Hom(V,W) \times V \to W$ by the formula $Ev(T,v) = T(v)$.

Lemma 5.19 *The map Ev is differentiable, with directional derivatives*

$$(D_{(S,u)}Ev)(T,v) = S(v) + T(u)$$

Proof: Observe

$$\begin{array}{rcl} Ev(S+T, u+v) & = & (S+T)(u+v) \\ & = & S(u) + S(v) + T(u) + T(v) \\ & = & Ev(S,u) + S(v) + T(u) + T(v) \end{array}$$

Write

$$r(T,v) = \frac{T(v)}{\|(T,v)\|}$$

where the norm on the space $Hom(V,W) \times V$ is defined by the formula

$$\|(T,v)\| = \max(\|T\|, \|v\|)$$

Observe

$$\|r(T,v)\| = \frac{\|T(v)\|}{\max(\|T\|,\|v\|)} \leq \frac{\|T\|\cdot\|v\|}{\max(\|T\|,\|v\|)} \leq \frac{(\max(\|T\|,\|v\|)^2}{\max(\|T\|,\|v\|)} = \|(T,v)\|$$

It follows that

$$\lim_{(T,v)\to 0} r(T,v) = 0$$

and by definition of the total derivative,

$$(DEv)_{(S,u)}(T,v) = S(v) + T(u)$$

as required. □

Since the form ω is differentiable, for each point $x \in R$ we can form the derivative

$$(D\omega)_x\colon V \to Hom(V,W)$$

For convenience, recall that we write

$$(D\omega)_x(u,v) = (D\omega)_x(u)(v)$$

Lemma 5.20 *The map $\tilde{\omega}\colon R\times V \to W$ has total derivative given by the formula*

$$D\tilde{\omega}(x,u)(v_1,v_2) = (D\omega)_x(v_1,u) + \omega(x)(v_2)$$

Proof: The result is immediate from the above lemma, the definition of the function $\tilde{\omega}$, and the chain rule. □

Theorem 5.21 *Let V and W be finite-dimensional vector spaces, and let $R \subseteq V$ be an open star-shaped subset. Let $\omega \in \Omega^1(R;W)$ be a continuously differentiable differential form. Suppose that ω is symmetric. Then ω is exact.*

Further, choose a point $x_0 \in R$ such that for all $x \in R$, the path $\gamma_x\colon [0,1] \to V$ defined by the formula $\gamma_x(t) = x_0 + t(x - x_0)$ is contained in R. Then the differential form ω has a primitive function defined by the formula

$$\varphi(x) = \int_{\gamma_x} \omega = \int_0^1 \omega(x_0 + tx)(x - x_0)\ dt$$

Proof: For convenience, let us assume that $x_0 = 0$. Then

$$\varphi(x) = \int_0^1 \omega(tx)(x)\ dt = \int_0^1 \tilde{\omega}(tx,x)\ dt$$

Let $u \in V$. Let $f_t(x) = \tilde{\omega}(tx,x)$ so that

$$\varphi(x) = \int_0^1 f_t(x)\ dt$$

By definition of the directional derivative

$$\frac{\partial}{\partial s} f_t(x+su)(0) = D_u(f_t)(x)$$

and

$$D_u(\varphi)(x) = \frac{d}{ds}\int_0^1 f_t(x+su)\ dt$$

By lemma 5.20, we have

$$D_u(f_t)(x) = (D\omega)_{tx}(tu, x) + \omega(tx)(u)$$

Since the form ω is continuously differentiable, we see that the directional derivative $D_u(f_t)(x)$ is continuous in the variables x and t. Hence we can differentiate under the integral sign to obtain the formula

$$D_u(\varphi)(x) = \int_0^1 D_u(f_t)(x)\ dt$$

By symmetry and bilinearity, we see that

$$D_u f_t(x) = t(D\omega)_{tx}(x, u) + \omega(tx)(u)$$

Let $g(t) = t\omega(tx)(u)$. Then by the chain rule and the product rule

$$g'(t) = \omega(tx)(u) + t(D\omega)_{tx}(x, u) = D_u f(t)$$

Hence

$$D_u(\varphi)(x) = \int_0^1 g'(t)\ dt = g(1) - g(0) = \omega(x)(u)$$

by the fundamental theorem of calculus.

These directional derivatives all exist and are continuous, by continuity of the differential form ω. It follows that we have total derivative $D\varphi = \omega$, and we are done. □

Definition 5.22 Let V and W be normed vector spaces, and let $R \subseteq V$ be open. Then we call a differential form $\omega \in \Omega^1(R; V)$ *closed* or *locally exact* if for each point $x \in R$ there is an open set $U \ni x$ such that the restriction $\omega|_U$ is exact.

Theorem 5.23 *Let V and W be finite-dimensional vector spaces. Let $R \subseteq V$ be open, and let $\omega\colon R \to Hom(V, W)$ be a continuously differentiable differential form.*

Then the form ω is locally exact if and only if it is symmetric.

Proof: Let ω be locally exact. Let $x \in R$. Then there is an open set $U \ni x$ such that the restriction $\omega|_U$ is exact. By theorem 4.14, the restriction $\omega|_U$ is symmetric. In particular, for all vectors $u, v \in V$, we have $(D\omega)_x(u, v) = (D\omega)_x(v, u)$.

Since the above equation holds for all $x \in R$, the differential form ω is symmetric.

Conversely, suppose that ω is symmetric. Let $x \in R$. Then there is an open ball $B(x, \delta) \subseteq R$. But open balls are star-shaped. Thus the restriction $\omega|_{B(x,\delta)}$ is exact by the above theorem. Since open balls are open sets, the above tells us that the differential form ω is locally exact. $\square$

6 The Fundamental Group

6.1 Homotopies

In this section, we will assume that a path in a topological space X is a continuous maps $\gamma\colon [0, 1] \to X$; we confine ourselves to the interval $[0, 1]$ rather than more general intervals.

Definition 6.1 Let X be a topological space. Let $\gamma_0, \gamma_1\colon [0, 1] \to X$ be two paths with the same start point $p = \gamma_0(0) = \gamma(0)$, and end point $q = \gamma_0(1) = \gamma_1(1)$.

We call the paths γ_0 and γ_0 *homotopic* if there is a continuous map $H\colon [0, 1] \times [0, 1] \to X$ such that

$$H(0, t) = \gamma_0(t) \qquad H(1, t) = \gamma_1(t)$$

for all $t \in [0, 1]$, and

$$H(s, 0) = p \qquad H(s, 1) = q$$

for all $s \in [0, 1]$.

A map H with the above properties is called a *homotopy* from the path γ_0 to the path γ_1.

For each point $s \in [0, 1]$, the map $H(s, -)\colon [0, 1] \to X$ is itself a path. Thus the idea of a homotopy from the path γ_0 to the path γ_1 is that of a continuous deformation.

Lemma 6.2 *The notion of paths in the space X with start point $p \in X$ and end point $q \in X$ being homotopic is an equivalence relation.*

Proof:

- Let $\gamma\colon [0, 1] \to X$ be a path such that $\gamma(0) = p$ and $\gamma(1) = q$. Then we can define a homotopy $H\colon [0, 1] \times [0, 1] \to X$ from the path γ to itself by the formula $H(s, t) = \gamma(t)$.

  Thus the notion of paths being homotopic is reflexive.

- Let γ_0, γ_1 be homotopic paths in X from p to q. Then we have a homotopy $H: [0,1] \times [0,1] \to X$ from γ_0 to γ_1. The *reversed homotopy*, $\overline{H}: [0,1] \times [0,1] \to X$, defined by the formula $\overline{H}(s,t) = H(1-s,t)$ is then a homotopy from γ_1 to γ_0.

  Thus the notion of paths being homotopic is symmetric.

- Let $\gamma_1, \gamma_2, \gamma_3: [0,1] \to X$ be paths such that γ_1 and γ_2 are homotopic, and γ_2 and γ_3 are homotopic. Let H_1 be a homotopy from γ_1 to γ_2, and let H_2 be a homotopy from γ_2 to γ_3. Then we can define a homotopy, H, from γ_1 to γ_3 by the formula

$$H(s,t) = \begin{cases} H_1(2s,t) & s \leq \frac{1}{2} \\ H_2(2s-1,t) & s \geq \frac{1}{2} \end{cases}$$

  Since $H_1(1,t) = \gamma_2(t) = H_2(0,t)$ for all $t \in [0,1]$, the map H is continuous; we leave it as an exercise to check the precise details.

  We see that the notion of paths being homotopic is transitive.

□

We call the following the *reparametrisation lemma*.

Lemma 6.3 *Let $\gamma: [0,1] \to X$ be a path in a topological space. Let $\theta: [0,1] \to [0,1]$ be a continuous map such that $\theta(0) = 0$ and $\theta(1) = 1$. Then the maps γ and $\gamma \circ \theta$ are homotopic.*

Proof: Let $s, t \in [0,1]$. Then $\theta(t) \in [0,1]$, and $(1-s)t + s\theta(t) \in [0,1]$. Further, for all $s \in [0,1]$, when $t = 0$, we have $(1-s)t + s\theta(t) = 0$, and when $t = 1$, we have $(1-s)t + s\theta(t) = 1$.

We can therefore define a homotopy from the path γ to the path $\gamma \circ \theta$ by the formula

$$H(s,t) = \gamma((1-s)t + s\theta(t))$$

□

Since we are confining our paths to the interval $[0,1]$ we need a slightly different notion of concatenation of paths. The following idea works; the slight differences to the earlier definition do not matter.

Definition 6.4 Let $\gamma, \beta: [0,1] \to X$ be paths in the space X such that $\gamma(1) = \beta(0)$. Then we define the *concatenation* $\gamma\beta: [0,1] \to X$ to be the path defined by the formula

$$\gamma\beta(t) = \begin{cases} \gamma(2t) & t \leq \frac{1}{2} \\ \beta(2t-1) & t \geq \frac{1}{2} \end{cases}$$

Suppose we have paths $\gamma, \beta, \alpha\colon [0,1] \to X$ such that $\gamma(1) = \beta(0)$, and $\beta(1) = \alpha(0)$. Then the order in which we take the concatenations matter. To be precise, the paths $(\gamma\beta)\alpha$ and $\gamma(\beta\alpha)$ are not equal; these are defined by the formulae

$$(\gamma\beta)\alpha = \begin{cases} \gamma(4t) & t \leq \frac{1}{4} \\ \beta(4t-1) & \frac{1}{4} \leq t \leq \frac{1}{2} \\ \alpha(2t-1) & t \geq \frac{1}{2} \end{cases}$$

and

$$\gamma(\beta\alpha) = \begin{cases} \gamma(2t) & t \leq \frac{1}{2} \\ \beta(4t-2) & \frac{1}{2} \leq t \leq \frac{3}{4} \\ \alpha(4t-3) & t \geq \frac{3}{4} \end{cases}$$

respectively.

Thus the operation of concatenation is *not* associative. However, we do have the following.

Lemma 6.5 *The paths $(\gamma\beta)\alpha$ and $\gamma(\beta\alpha)$ are homotopic.*

Proof: Define a continuous map $\theta\colon [0,1] \to [0,1]$ by the formula

$$\theta(t) = \begin{cases} 2t & t \leq \frac{1}{4} \\ t + \frac{1}{4} & \frac{1}{4} \leq t \leq \frac{1}{2} \\ \frac{1}{2}t + \frac{1}{2} & t \geq \frac{1}{2} \end{cases}$$

Then $\theta(0) = 0$, $\theta(1) = 1$, and $(\gamma\beta)\alpha = \gamma(\beta\alpha) \circ \theta$. The result now follows from the reparametrisation lemma. $\square$

The proof of the following is straightforward to check.

Lemma 6.6 *Let $\gamma_1, \beta_1\colon [0,1] \to X$ be homotopic paths, both with end point p. Let $\gamma_2, \beta_2\colon [0,1] \to X$ be homotopic paths, both with start point p. Then the concatenations $\gamma_1\gamma_2$ and $\beta_1\beta_2$ are homotopic.* $\square$

Definition 6.7 Let $\gamma\colon [0,1] \to X$ be a path from the point p to the point q. Then we define the *reverse path*, $\overline{\gamma}\colon [0,1] \to X$, from the point q to the point p, by the formula $\overline{\gamma}(t) = \gamma(1-t)$.

The following is again straightforward.

Lemma 6.8 *Let γ_1 and γ_2 be homotopic paths. Then the reverse paths $\overline{\gamma_1}$ and $\overline{\gamma_2}$ are homotopic.* $\square$

Now, choose a point $x_0 \in X$. Let $c\colon [0,1] \to X$ be the *constant path*, defined by writing $c(t) = x_0$ for all $t \in [0,1]$.

Lemma 6.9 *Let $\gamma\colon [0,1] \to X$ be a path in X with starting point x_0. Then the concatenation $\gamma\overline{\gamma}$ is homotopic to the constant path c.*

Proof: The concatenation $\gamma\overline{\gamma}$ is defined by the formula

$$\gamma\overline{\gamma}(t) = \begin{cases} \gamma(2t) & t \leq \frac{1}{2} \\ \gamma(1-2t) & t \geq \frac{1}{2} \end{cases}$$

We can define a homotopy from the constant path c to the concatenation $\gamma\overline{\gamma}$ by the formula

$$H(s,t) = \begin{cases} \gamma(2st) & t \leq \frac{1}{2} \\ \gamma(2s(1-t)) & t \geq \frac{1}{2} \end{cases}$$

□

6.2 The Fundamental Group

Let X be a topological space, and let $x_0 \in X$. A path $\gamma\colon [0,1] \to X$ such that $\gamma(0) = \gamma(1) = x_0$ is called a *loop* in X that is *based* at the point x_0.

We saw in the previous section that the notion of loops in the space X based at the point x_0 being homotopic is an equivalence relation. We call the equivalence classes of this relation the *homotopy classes* of loops based at the point x_0.

Definition 6.10 The *fundamental group* of X at the point x_0 is the set of homotopy classes of loops in X based at x_0.

We use the notation $\pi_1(X, x_0)$ to denote the above fundamental group. Given a path, γ, we write $[\gamma]$ to denote the homotopy class. Thus, when γ is a loop based at the point x_0, we have $[\gamma] \in \pi_1(X, x_0)$.

The proof of the result is left as an exercise. The most efficient method of proof is to use the various lemmas proved in the previous section.

Theorem 6.11 *The fundamental group $\pi_1(X, x_0)$ is a group. The group operation is defined by the formula*

$$[\gamma] \cdot [\beta] = [\gamma\beta]$$

where γ and β are loops in X based at x_0.

Let $c\colon [0,1] \to X$ be the constant map onto the point x_0. Then the fundamental group has identity element $[c]$.

Inverses are defined by the formula $[\gamma]^{-1} = [\overline{\gamma}]$. □

Let $\varphi\colon X \to Y$ be a continuous map between topological spaces. Fix $x_0 \in X$, and let $y_0 = \varphi(x_0) \in Y$. Then we have an induced group homomorphism

$$\varphi_*\colon \pi_1(X, x_0) \to \pi_1(Y, y_0)$$

where $y_0 = \varphi(x_0)$ is defined by the formula

$$\varphi_*([\gamma]) = [\varphi \circ \gamma]$$

Given another continuous map $\psi: Y \to Z$, we can check that $(\psi \circ \varphi)_* = \psi_* \circ \varphi_*$. Further, if $1_X: X \to X$ is the identity map, then $(1_X)_* = 1_{\pi_1(X,x_0)}$.

It follows that homeomorphic topological spaces have isomorphic fundemental groups.

Changing topic slightly, the proof of the following result is left as an exercise.

Theorem 6.12 *Let X be a path-connected topological space. Then, up to isomorphism, the fundamental group $\pi_1(X, x_0)$ does not depend on the point x_0.* □

Thus, when X is a path-connected space, we will sometimes write simply $\pi_1(X)$ to denote the fundamental group, omitting explicit mention of a chosen basepoint.

Definition 6.13 Let X be a path-connected topological space. Then we call X *simply-connected* if the fundamental group $\pi_1(X)$ is trivial.

Definition 6.14 Let X be a topological space. We call X *contractible* if there is a point $x_0 \in X$ and a continuous map $H: X \times [0,1] \to X$ such that $H(x,0) = x$ and $H(x,1) = x_0$ for all $x \in X$.

Example 6.15 A star-shaped subset of a normed vector space is contractible.

To see this, let X be star-shaped. Choose $x_0 \in X$ such that for any point $x \in X$, the straight line joining x_0 to x lies entirely within the set X. Then we can define a continuous map

$$H: [0,1] \times X \to X$$

by the formula

$$H(x,t) = x + t(x_0 - x)$$

Then $H(x,0) = x$ and $H(x,1) = x_0$ for all $x \in X$.

We leave the proof of the following as an exercise.

Proposition 6.16 *Any contractible space is path-connected.* □

Proposition 6.17 *Any contractible space is simply-connected.*

Proof: Let X be contractible. Then we have a point $x_0 \in X$ and a continuous map $H: [0,1] \times X \to X$ such that $H(0,x) = x$ and $H(1,x) = x_0$ for all $x \in X$.

Let $\gamma: [0,1] \to X$ be a loop based at x_0. Then the map $H': [0,1] \times [0,1] \to X$ defined by writing $H'(x,t) = H(s,\gamma(t))$ is a homotopy from the loop γ to the constant loop c. It follows that $[\gamma] = [c]$ is the identity element of the fundamental group.

We see that the fundamental group contains only the identity element, and we are done. □

6.3 Coverings

Let I be an arbitrary set, and let Y be a topological space. Let $\{A_i \mid i \in I\}$ be a collection of open subsets of the space Y. We call the above collection of sets *pairwise disjoint* if $A_i \cap A_j = \emptyset$ whenever $i \neq j$.

We use the notation $\amalg_{i\in I} A_i$ to denote the union $\bigcup_{i\in I} A_i$ when the the sets A_i are pairwise disjoint.

Definition 6.18 Let X be a topological space. A surjective continuous map $p: E \to X$ is called a *covering map* of the space X if for each point $x \in X$ we have an open set $U \ni x$ such that:

- We can write
$$p^{-1}[U] = \amalg_{i\in I} V_i$$
where each set $V_i \subseteq E$ is open.
- For each $i \in I$, the restriction $p|_{V_i}: V_i \to U_i$ is a homeomorphism.

The space E in the above definition is called a *covering space* for the space X.

Example 6.19 Let X be a topological space. Let $\{X_i \mid i \in I\}$ be disjoint spaces equipped with homeomorphisms $p_i: X_i \to X$.

Then the disjoint union $E = \amalg_{i\in I} X_i$ is a covering space. The covering map $p: E \to X$ is defined by writing $p(x_i) = p_i(x_i)$ whenever $x_i \in X_i$.

A covering space of the type defined in the above example is called *trivial*

Example 6.20 Let $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$ be the unit circle. Define a map $p: \mathbb{R} \to S^1$ by the formula
$$p(t) = (\cos t, \sin t)$$

The map p is continuous an surjective. Choose $t_0 \in \mathbb{R}$, and $0 < \delta < \pi$. Then the pre-image
$$p^{-1}[\{(\cos t, \sin t) \mid t_0 - \delta < t < t_0 + \delta\}]$$
is the disjoint union
$$\amalg_{k\in\mathbb{Z}}(t_0 - \delta + 2k\pi, t_0 + \delta + 2k\pi)$$

The map p restricted to the interval $(t_0 - \delta + 2k\pi, t_0 + \delta + 2k\pi)$ is a homeomorphism onto the set
$$\{(\cos t, \sin t) \mid t_0 - \delta < t < t_0 + \delta\}$$

Thus the space $\mathbb{R}$ is a covering space of the circle S^1; the map $p: \mathbb{R} \to S^1$ is a covering map.

6.4 Liftings

Definition 6.21 Let X and Y be topological spaces. Let $p: E \to Y$ be a covering map. Then a *lifting* of a continuous map $f: X \to Y$ is a continuous map $\tilde{f}: X \to E$ such that $p \circ \tilde{f} = f$.

Lemma 6.22 *Let X be a connected topological space. Let $p: E \to Y$ be a covering map, and let $f: X \to Y$ be a continuous map. Let $x_0 \in X$, and let $\tilde{f}_1$ and $\tilde{f}_2$ be liftings of the map f such that $\tilde{f}_1(x_0) = \tilde{f}_2(x_0)$. Then $\tilde{f}_1 = \tilde{f}_2$.*

Proof: Let $S = \{x \in X \mid \tilde{f}_1(x) = \tilde{f}_2(x)\}$. Then $x_0 \in S$, so $S \neq \emptyset$.

Choose a point $x \in S$. Then by definition of a covering space, we have an open set $U \subseteq Y$ such that $f(x) \in U$ and

$$p^{-1}[U] = \amalg_{i \in I} V_i$$

where the sets $V_i \subseteq E$ are open, and the restrictions $p_i = p|_{V_i}: V_i \to U$ are homeomorphisms.

Since $x \in S$, $\tilde{f}_1(x) = \tilde{f}_2(x)$. Let $\tilde{f}_1(x) = \tilde{f}_2(x) \in V_i$. By continuity, the set

$$W_x = \tilde{f}_1^{-1}[V_i] \cap \tilde{f}_2^{-1}[V_i]$$

is open. Certainly $x \in W_x$, so $W_x \neq \emptyset$.

Let $y \in W_x$. Then $\tilde{f}_1(y) \in V_i$ and $\tilde{f}_2(y) \in V_i$. Since the maps $\tilde{f}_1$ and $\tilde{f}_2$ are liftings of the map f, $p_i \tilde{f}_1(y) = p_i \tilde{f}_2(y) = f(y)$. Since the map p_i is homeomorphism, we see that $\tilde{f}_1(y) = \tilde{f}_2(y)$, and $y \in S$.

We have shown that

$$S = \bigcup_{x \in S} W_x$$

where each set W_x is open. Hence the set S is open.

Now, let $T = \{x \in X \mid \tilde{f}_1(x) \neq \tilde{f}_2(x)\}$. Suppose that $T \neq \emptyset$. Then as above, we can show that the set T is open, and we have a partition $X = S \cup T$, which contradicts the fact that the set X is connected.

Hence $T = \emptyset$, which means that $S = X$, and we are done. □

Theorem 6.23 *Let $p: E \to X$ be a covering. Let $\gamma: [0,1] \to X$ be a path. Choose a point $y_0 \in E$ such that $p(y_0) = \gamma(0)$. Then there is a unique lifting $\tilde{\gamma}: [0,1] \to E$ such that $\tilde{\gamma}(0) = y_0$.*

Proof: We need to show existence of the path $\tilde{\gamma}$; once we have shown existence, uniqueness follows by the above lemma.

Let S be the set of all points $s \in [0,1]$ where the restricted path $\gamma|_{[0,s]}: [0,s] \to X$ has a lifting $\tilde{\gamma}: [0,s] \to X$ such that $\tilde{\gamma}(0) = y_0$.

Certainly $0 \in S$. We need to show that $1 \in S$.

Suppose $s_0 \in S$, and $s_0 < 1$. Then we have a path $\tilde{\gamma}: [0,s_0] \to X$ such that $\tilde{\gamma}(0) = y_0$ and $p \circ \tilde{\gamma}(t) = \gamma(t)$ whenever $t \in [0,s_0]$.

By definition of a covering space, we have an open set $U \subseteq X$ such that $\gamma(s_0) \in U$, and

$$p^{-1}[U] = \amalg_{i \in I} V_i$$

where $V_i \subseteq E$ is an open set, and the restriction $p_i = p|_{V_i}: V_i \to U$ is a homeomorphism.

Since the path γ is continuous, we can find $\delta > 0$ such that $\gamma(t) \in U$ whenever $|t - s_0| < 2\delta$, and $t \in [0,1]$. Suppose that $\gamma(s_0) \in U_i$. Then we can define a path $\tilde{\gamma}_1: [0, t_0 + \delta] \to E$ such that $\tilde{\gamma}_1(0) = y_0$ and $p \circ \tilde{\gamma}_1(t) = \gamma(t)$ whenever $t \in [0, t_0 + \delta]$ by the formula

$$\tilde{\gamma}_1(t) = \begin{cases} \tilde{\gamma}(t) & t \leq s_0 \\ p_i^{-1}(\gamma(t)) & s_0 \leq t \leq s_0 + \delta \end{cases}$$

We have shown that, given $s_0 \in S$, we have $\delta > 0$ such that $s_0 + \delta \in S$. But this implies that $1 \in S$; the precise details are left as an exercise. □

The proof of the following result is similar to the above, but the extra variable present makes it more fiddly. We therefore omit the details.

Theorem 6.24 *Let $p: E \to X$ be a covering. Let $H: [0,1] \times [0,1] \to X$ be a continuous map. Fix $y_0 \in E$ such that $p(y_0) = H(0,0)$. Then there is a unique lifting $\tilde{H}: [0,1] \times [0,1] \to W$ such that $\tilde{H}(0,0) = y_0$.*

Further, if the map $H(-,0)$ is constant, then the map $\tilde{H}(-,0)$ is also constant. □

Corollary 6.25 *Let $p: E \to X$ be a covering. Let $\gamma_1, \gamma_2: [0,1] \to X$ be two paths such that $\gamma_1(0) = \gamma_2(0)$ and $\gamma_1(1) = \gamma_2(1)$.*

Let $\tilde{\gamma_1}$ be a lifting of the path γ_1, and $\tilde{\gamma_2}$ be a lifting of the path γ_2 such that $\tilde{\gamma_1}(0) = \tilde{\gamma_2}(0)$. Suppose that the paths γ_1 and γ_2 are homotopic. Then $\tilde{\gamma_1}(1) = \tilde{\gamma_2}(1)$, and the liftings $\tilde{\gamma_1}$ and $\tilde{\gamma_2}$ are homotopic. □

6.5 Computations of the Fundamental Group

We have already seen that the fundamental group of a contractible space is trivial. In this section we use the ideas of coverings and liftings to compute some non-trivial fundamental groups. We begin with the circle

$$S^1 = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$$

Lemma 6.26 *Let $\gamma: [0,1] \to S^1$ be a closed path, and let $p: \mathbb{R} \to S^1$ be the covering map $p(t) = (\cos t, \sin t)$.*

Let $\tilde{\gamma}$ be a lifting of the path γ. Then the map γ is homotopic to the constant path if and only if $\tilde{\gamma}(0) = \tilde{\gamma}(1)$.

Proof: Suppose that the map γ is homotopic to the constant path, c. The constant path, $\tilde{c}$, in $\mathbb{R}$ is a lifting of the constant path c. By corollary 6.25, since the path $\tilde{c}$ is constant

$$\tilde{\gamma}(1) = \tilde{c}(1) = \tilde{c}(0) = \tilde{\gamma}(0)$$

Conversely, suppose that $\tilde{\gamma}(0) = \tilde{\gamma}(1)$. Then $\tilde{\gamma}$ is a loop in the space $\mathbb{R}$. But the space $\mathbb{R}$ is contractible, so the loop $\tilde{\gamma}$ is homotopic to a constant loop. It follows that the loop $\gamma = p \circ \tilde{\gamma}$ is homotopic to a constant loop. $\square$

To compute the fundamental group of the circle S^1, we will assume that our loops are based at the point $x_0 = (1, 0)$.

Given an integer $k \in \mathbb{Z}$, we can define a loop $\gamma_k : [0, 1] \to S^1$ by the formula

$$\gamma_k(t) = (\cos 2k\pi t, \sin 2k\pi t)$$

Theorem 6.27 *There is an isomorphism $\theta: \mathbb{Z} \to \pi_1(S^1, x_0)$ defined by the formula*

$$\theta(k) = [\gamma_k]$$

Proof: We leave it as an exercise to prove that the map θ is a group homomorphism.

Let $p: \mathbb{R} \to S^1$ be the above covering map. Then the map γ_k has lifting $\tilde{\gamma_k}: [0, 1] \to \mathbb{R}$ defined by the formula

$$\tilde{\gamma_k}(t) = 2k\pi t$$

Observe that $\tilde{\gamma_k}(1) = 2k\pi$. Hence, by the above lemma, if $k \neq 0$, then $[\gamma_k] \neq [c]$. It follows that the homomorphism θ is injective.

Let $\gamma: [0, 1] \to S^1$ be a loop based at $(1, 0)$. Let $\tilde{\gamma}: [0, 1] \to \mathbb{R}$ be a lifting such that $\tilde{\gamma}(0) = 0$. Then

$$\tilde{\gamma}(1) \in p^{-1}(1, 0) = \{2n\pi \mid n \in \mathbb{Z}\}$$

Let $\tilde{\gamma}(1) = 2k\pi$. The loop γ_k has lifting $\tilde{\gamma_k}$ such that $\tilde{\gamma_k}(1) = 2k\pi$. Thus, we can form the concatenation $\tilde{\gamma}\overline{\tilde{\gamma_k}}$, and

$$\tilde{\gamma}\overline{\tilde{\gamma_k}}(1) = 0 = \tilde{\gamma}\overline{\tilde{\gamma_k}}(0)$$

The concatenation $\tilde{\gamma}\overline{\tilde{\gamma_k}}$ is a lifting of the concatenation $\gamma\overline{\gamma_k}$. By the above lemma, the loop $\gamma\overline{\gamma_k}$ is therefore homotopic to the constant loop, c.

It follows that

$$[\gamma\overline{\gamma_k}] = [\gamma][\gamma_k]^{-1} = [c]$$

so $[\gamma] = [\gamma_k]$ and have shown that the homomorphism θ is also surjective. $\square$

We leave the proof of the following as an exercise.

Proposition 6.28 *Let X and Y be path-connected topological spaces. Let $x_0 \in X$ and $y_0 \in Y$. Then the groups $\pi_1(X, x_0) \times \pi_1(Y, y_0)$ and $\pi_1(X \times Y, (x_0, y_0))$ are isomorphic.* $\square$

Let $\mathbb{C}^x = \mathbb{C}\backslash\{0\}$ be the *punctured plane*. Let $x_0 = (1, 0)$. Let $\gamma_k: [0, 1] \to \mathbb{C}^x$ be as defined above.

Theorem 6.29 *There is an isomorphism $\theta: \mathbb{Z} \to \pi_1(\mathbb{C}^x, x_0)$ defined by the formula*

$$\theta(k) = [\gamma_k]$$

Proof: We have a homeomorphism $h: S^1 \times (0, \infty) \to \mathbb{C}^x$ defined by the formula

$$h((\cos t, \sin t), r) = (r \cos t, r \sin t)$$

The result now follows from the above proposition and theorem, and the fact that the space $(0, \infty)$ is contractible. $\square$

7 The Monodromy Theorem

7.1 The Integral Covering

Let V and W be finite-dimensional vector spaces. Let $R \subseteq V$ be an open subset. Recall that we call a differential form $\omega \in \Omega^1(R; W)$ *locally exact* or *closed* if for all $x \in R$ there is an open set $U \ni x$ such that the restriction $\omega|_U \in \Omega^1(U; W)$ is exact.

Suppose that we have a locally exact differential form $\omega \in \Omega^1(R; W)$. Consider the *set* $R_\omega = R \times W$ along with the projection $p: R_\omega \to R$ defined by the formula $p(x, w) = x$. We want to define a topology on the set R_ω, ie: a collection of sets that we define to be open and satisfy a few of axioms. This topology should reflect the local exactness of the differential form ω and turn the space R_ω into a covering space of the space R. Note that the topology we define on the space R_ω is *not* the same as the topology on the product $R \times W$.

Let $U \subseteq R$ be an open subset such that the restriction $\omega|_U$ is exact. Let $F: U \to W$ be a primitive function of the differential form $\omega|_U$. Then we define a set

$$\Gamma(U, F) = \{(x, F(x)) \mid x \in U\} \subseteq R_\omega$$

Proposition 7.1 *Call a subset of the set R_ω open if it is a union of sets of the above form. Then the set R_ω is a topological space.*

Proof: We need to check the topological space axioms.

- By definition of an open set in the space R_ω, any union of open sets is an open set.
- Note that the empty set, $\emptyset \subseteq R$ is open. There is a unique well-defined function $F: \emptyset \to W$; this function is a primitive function for the restriction $\omega|_\emptyset$. Observe that $\emptyset = \Gamma(\emptyset, F)$; hence the empty set $\emptyset \subseteq G_\omega$ is open.

- Let $x_0 \in R$, and $w \in W$. Since the form ω is locally exact, there is an open set $U \subseteq X$ where $x_0 \in U$ and the restriction $\omega|_U$ is exact.

  Consider a primitive function $F: U \to W$ such that $F(x_0) = w$. Then our chosen point $(x_0, w) \in R_\omega$ belongs to the open set $\Gamma(U, F)$. Thus the union of all such sets, which is by definition open, is the entire space R_ω.

- Let $U_1, U_2 \subseteq R$ be open sets such that the restrictions $\omega|_{U_1}$ and $\omega|_{U_2}$ are open. Let F_1 and F_2 be primitive functions for the restrictions $\omega|_{U_1}$ and $\omega|_{U_2}$ respectively.

  If the intersection $\Gamma(U_1, F_1) \cap \Gamma(U_2, F_2)$ is the empty set, then it is open. Otherwise, let $(x, w) \in \Gamma(U_1, F_1) \cap \Gamma(U_2, F_2)$.

  The set $U_1 \cap U_2 \subseteq R$ is open. We therefore have an open ball $B = B \subseteq U_1 \cap U_2$ such that $x \in B$, and $B \subseteq U_1 \cap U_2$. The primitive functions F_1 and F_2 restrict to primitive functions on the ball B.

  We know that $F_1(x) = F_2(x) = w$. By corollary 4.8, the primitive function of an exact form on a connected open set is determined uniquely determined by its value at a point. The open ball B is connected. Hence, $F_1|_B = F_2|_B$. We see that

  $$(x, w) \in \Gamma(B, F_1|_B) \subseteq \Gamma(U_1, F_1) \cap \Gamma(U_2, F_2)$$

  Thus our chosen point $(x, w) \in \Gamma(U_1, F_1) \cap \Gamma(U_2, F_2)$ belongs to an open subset. Hence the intersection $\Gamma(U_1, F_1) \cap \Gamma(U_2, F_2)$ is a union of open subsets.

  More generally, we see that any intersection of two open sets is open, and we are done.

$\square$

Proposition 7.2 *The topological space R_ω is a covering space of the space R, with covering map $p: R_\omega \to R$.*

Proof: It is straightforward to check that the map $p: R_\omega \to R$ is continuous. The map p is obviously surjective.

Let $x_0 \in R$. Since the form ω is locally exact, we have an open set U such that $x_0 \in U$ and the restriction $\omega|_U$ is exact. By taking a subset if necessary (for instance, an open ball containing the point x_0), we can assume that the set U is connected.

Define S to be the set of primitive functions of the differential form $\omega|_U$. Let $x \in U$ and $w \in W$. Then by corollary 4.8, we have a unique primitive function $F \in S$ such that $F(x) = w$. The same argument tells us that $\Gamma(U, F) \cap \Gamma(U, G) = \emptyset$ if $F \neq G$.

Hence

$$p^{-1}[U] = \amalg_{F \in S} \Gamma(U, F)$$

The restriction $p|_{\Gamma(U,F)}\colon \Gamma(U,F) \to U$ is defined by the formula $p((x, F(x)) = x$. It is straightforward to check that this map is a homeomorphism. $\square$

Definition 7.3 We call the space R_ω the *integral covering* for the locally exact differential form ω.

7.2 Homotopy and Path Integrals

Let R be an open subset of a finite-dimensional vector space. In the previous section, we only talked about homotopy for paths $\gamma\colon [0,1] \to R$

We can generalise the definition to talk about homotopy between paths $\gamma\colon [a,b] \to R$. The following result is called the *monodromy theorem.*

Theorem 7.4 *Let V and W be finite-dimensional vector spaces, and let $R \subseteq V$ be an open subset. Let $\omega \in \Omega^1(R; W)$ be a continuous locally exact differential form. Let $\gamma_1, \gamma_2\colon [a,b] \to R$ be homotopic piecewise differentiable paths.*

Then

$$\int_{\gamma_1} \omega = \int_{\gamma_2} \omega$$

Proof: Let $\gamma\colon [a,b] \to R$ be a piecewise differentiable path. Then be theorem 6.23 we have a lifting $\tilde{\gamma}\colon [a,b] \to R_\omega$, where R_ω is the integral cover defined in the previous section. We can write $\tilde{\gamma}(t) = (\gamma(t), \tilde{\gamma}_W(t))$.

We claim that

$$\int_\gamma \omega = \tilde{\gamma}_W(b) - \tilde{\gamma}_W(a)$$

Since the interval $[a,b]$ is compact, we can find points

$$a = c_0 < c_1 < \cdots < c_{k-1} < c_k = b$$

and open sets U_i such that the restriction $\omega|_{U_i}$ is exact, the restriction $\gamma|_{[c_i,c_{i+1}]}$ is differentiable, and $\gamma[c_i, c_{i+1}] \subseteq U_i$; we leave the proof of this fact as an exercise. Choose a primitive function, F_i, of the differential form $\omega|_{U_i}$ such that $F_i(\gamma(c_i)) = \tilde{\gamma}_W(c_i)$.

The map $s(t) = (\gamma(t), F(\gamma(t))$ is a lifting of the path γ restricted to the interval $[c_i, c_{i+1}]$, and $s(c_i) = \tilde{\gamma}(c_i)$. Hence, by theorem 6.23, $s = \tilde{\gamma}$ on the interval $[c_i, c_{i+1}]$. In particular, $\tilde{\gamma}_W = F_i \circ \gamma$ on the interval $[c_i, c_{i+1}]$.

It follows that

$$\int_{\gamma_i} \omega = F_i(\gamma(c_{i+1})) - F_i(\gamma(c_{i+1})) = \tilde{\gamma}_W(c_{i+1}) - \tilde{\gamma}_W(c_i)$$

where γ_i is the restriction of the path γ to the interval $[c_i, c_{i+1}]$. Adding the integrals of the restrictions together, we see

$$\int_\gamma \omega = \tilde{\gamma}_W(b) - \tilde{\gamma}_W(a)$$

as claimed.

Now, let $\gamma_1, \gamma_2 \colon [a, b] \to R$ be two homotopic paths. Then by corollary 6.25, any two liftings $\tilde{\gamma}_1, \tilde{\gamma}_2 \colon [a, b] \to R_\omega$ with the same start point have the same end point. It follows that

$$\int_{\gamma_1} \omega = \tilde{\gamma}_{1W}(b) - \tilde{\gamma}_{2W}(a) = \tilde{\gamma}_{2W}(b) - \tilde{\gamma}_{1W}(a) = \int_{\gamma_2} \omega$$

and we are done. □

Since the path integral of a differential form over a constant path is always zero, we have the following corollary.

Corollary 7.5 *Let $\omega \in \Omega^1(R; W)$ be a locally exact continuous differential form. Let $\gamma \colon [a, b] \to W$ be a closed path that is homotopic to the constant path. Then*

$$\oint_\gamma \omega = 0$$

The first part of the following result also follows from the above; the second part is immediate by theorem 5.14.

Theorem 7.6 *Let V and W be finite-dimensional vector spaces, and let $R \subseteq V$ be simply-connected. Let $\omega \in \Omega^1(R; W)$ be a locally exact continuous differential form. Then ω is exact.*

Further, if we pick a point $x_0 \in R$, we can define a primitive function for the differential form ω by writing

$$\varphi(x) = \int_\gamma \omega$$

where γ is any piecewise-differentiable path in the region R with start point x_0 and end point x. □

We can use the above to show that certain paths are *not* homotopic to constant loops. For instance, the function $f(z) = \frac{1}{z}$ is complex-differentiable in the space $\mathbb{C}^x$. Therefore the form $f(z) = \frac{1}{z}\, dz$ is symmetric as a real differential form by theorem 4.18, and hence locally exact by theorem 5.21.

Let $\gamma \colon [0, 2\pi] \to \mathbb{C}^x$ be the unit circle defined by the formula $\gamma(t) = \cos(it) + i\sin(it)$. Using the *Euler formula*, we can write $\gamma(t) = e^{it}$, and

$$\int_\gamma \frac{1}{z}\, dz = \int_0^{2\pi} e^{-it} \cdot ie^{it}\, dt = 2\pi i \neq 0$$

Thus the circle is not homotopic to the constant path. In particular, the space $\pi_1(\mathbb{C}^x)$ is not simply-connected; the fundamental group is not trivial.

Proposition 7.7 *We have an isomorphism $\phi: \pi_1(\mathbb{C}^x) \to \mathbb{Z}$ defined by the formula*

$$\phi([\gamma]) = \frac{1}{2\pi i} \oint_\gamma \frac{1}{z}\, dz$$

Proof: Let $k \in \mathbb{Z}$, and define a loop $\gamma_k: [0,1] \to \mathbb{C}^x$ by the formula $\gamma_k(t) = e^{2\pi it}$. By theorem 6.29,we have an isomorphism $\theta: \mathbb{Z} \to \pi_1(S^1, x_0)$ defined by the formula

$$\theta(k) = [\gamma_k]$$

Choose a loop γ in the space $\mathbb{C}^x$. Since the map θ is surjective, the map γ is homotopic to the loop γ_k for some k. Hence, by the monodromy theorem

$$\frac{1}{2\pi i} \oint_\gamma \frac{1}{z}\, dz = \frac{1}{2\pi i} \oint_{\gamma_k} \frac{1}{z}\, dz$$

and

$$\int_{\gamma_k} \frac{1}{z}\, dz = \int_0^1 e^{-2\pi kt} \cdot ike^{2\pi kt}\, dt = 2\pi ik$$

Therefore $\phi([\gamma]) = \phi([\gamma_k]) = k$. The map ϕ is thus a well-defined inverse to the isomorphism θ, and therefore itself an isomorphism. □

Definition 7.8 Let $a \in \mathbb{C}$. Let $\gamma: [0,1] \to \mathbb{C}\backslash\{a\}$ be a piecewise-differentiable loop. Then we define the *winding number* of γ around a:

$$w(\gamma, a) = \frac{1}{2\pi i} \oint_\gamma \frac{1}{z-a}\, dz$$

By the above, the winding number is an integer. Observe that $w(\gamma_k, 0) = k$ for all $k \in \mathbb{Z}$. The winding number is a rigorous way to define the number of times the loop γ winds around a in an anticlockwise direction. Further properties of the winding number can be found in the exercises.

8 Complex Analysis

8.1 Holomorphic Functions

Let $R \subseteq \mathbb{C}$ be an open set, and let $f: R \to \mathbb{C}$ be a complex function. We call the function f *holomorphic* if it is continuously differentiable as a complex function. If $R = \mathbb{C}$, we call the function f *entire*

In fact it is a theorem that any complex-differentiable function is holomorphic; continuity of the derivative follows automatically from complex-differentiability. Many of our results on differential forms can be applied immediately to holomorphic functions. Note in particular the following.

- Write
$$f(x+iy) = g(x,y) + ih(x,y)$$
where x and y are real numbers, and g and h are functions such that the partial derivatives all exist and are continuous.

  Then by theorem 3.19, the function f is real-differentiable.

- By corollary 3.25, the function f is holomorphic if and only if the *Cauchy-Riemann equations*
$$\frac{\partial g}{\partial y}(x) = -\frac{\partial h}{\partial x}(x) \qquad \frac{\partial h}{\partial y}(x) = \frac{\partial g}{\partial x}(x)$$
are satisfied.

- By theorem 4.18, the differential form $\omega = f\ dz$ is symmetric, as a real differential form, if and only if the function f is holomorphic.

- By theorem 5.23, the form $\omega = f\ dz$ is locally exact if and only if it is symmetric.

Putting the above together, we have the following.

Theorem 8.1 *Let $f\colon R \to \mathbb{C}$ be a holomorphic function. Then f is locally exact.*
□

The Cauchy-Riemann equations, noted above, provide a means to check whether a complex function is holomorphic.

The following result, which is usually called *Cauchy's theorem*, follows immediately from theorem 7.6.

Theorem 8.2 *Let $f\colon R \to \mathbb{C}$ be a holomorphic function, where the region R is simply-connected. Then f is exact.*

In particular, for any closed loop, γ, in the region R, we have
$$\oint_\gamma f(z)\ dz = 0$$
□

The following result is called *Goursat's lemma*; it is a generalisation of Cauchy's theorem to complex-differentiable functions where the derivative need not be continuous. We do not prove it here.

Theorem 8.3 *Let $f\colon R \to \mathbb{C}$ be a complex-differentiable function, where the region R is simply-connected. Then f is exact.*

In particular, for any closed loop, γ, in the region R, we have
$$\oint_\gamma f(z)\ dz = 0$$
□

In fact, Goursat's lemma is the key to showing that any complex-differentiable function is holomorphic, and hence has continuous derivative.

8.2 Residue Calculus and the Integral Formula

Let $a \in \mathbb{C}$, and $\delta > 0$. Consider the punctured disk $B'(a, \delta)$, and let $\gamma\colon [0,1] \to B'(a, \delta)$ be a loop that winds once anticlockwise around the point a.

It is geometrically clear that all such paths are homotopic. The following is therefore well-defined by the monodromy theorem.

Definition 8.4 Let ω be a continuous complex-valued locally exact differential form defined on the punctured disk $B'(a, \delta)$. Then we define the *residue* of the form ω at the point w by the formula

$$Res(\omega, a) = \frac{1}{2\pi i} \oint_\gamma \omega$$

More generally, given an open set $R \subseteq \mathbb{C}$, a point $a \in R$, a continuous complex-valued locally exact differential form ω defined on the set $R \backslash \{a\}$, we define the residue $Res(\omega, a)$ as above, by restricting to some ball $B(a, \delta) \subseteq R$.

Proposition 8.5 *Let $R \subseteq \mathbb{C}$ be open, and $a \in R$. Let ω be a continuous complex-valued locally exact form defined on the set $R \backslash \{a\}$. Then the following hold.*

1. *Let ω' be another continuous complex-valued locally exact form defined on the set $R \backslash \{a\}$, and let $\alpha, \alpha' \in \mathbb{C}$. Then*

$$Res(\alpha\omega + \alpha'\omega', a) = \alpha Res(\omega, a) + \alpha' Res(\omega', a)$$

2. *Let ω be exact on some punctured open ball $B'(a, \delta) \subseteq R$. Then $Res(\omega, a) = 0$.*

3. *Suppose that ω extends to a continuous locally exact differential form, $\tilde{\omega}$ on the set R. Then $Res(\omega, a) = 0$.*

4. *$Res(\frac{dz}{z-w}, w) = 1$*

5. *Let $k \in \mathbb{Z} \backslash \{-1\}$. Then $Res((z-w)^k\, dz, w) = 0$*

Proof:

1. The result follows by definition of the residue and linearity of path-integrals.

2. Let γ be a loop in the neighbourhood $B'(a, \delta)$ that winds once anticlockwise around a. Then

$$Res(\omega, a) = \oint_\gamma \omega = 0$$

   by exactness.

3. By definition, we have $Res(\omega, a) = Res(\tilde{\omega}, a)$. The differential form $\tilde{\omega}$ is continuous and locally exact on some disk $B(a, \delta) \subseteq R$.

   The ball $B(a, \delta)$ is contractible, and therefore simply-connected. Hence the differential form $\tilde{\omega}$ is exact on $B(a, \delta)$ by theorem 7.6. The result now follows by the previous part.

4. Define a loop $\gamma: [0, 2\pi] \to \mathbb{C}\backslash\{w\}$ by writing $\gamma(t) = w + e^{it}$. Then γ is a closed loop that winds around the point w once anticlockwise, so

$$Res(\frac{dz}{z-w}, w) = \frac{1}{2\pi i}\oint_\gamma \frac{dz}{z-w} = \frac{1}{2\pi i}\int_0^{2\pi} \frac{1}{e^{it}} ie^{it}\ dt = 1$$

5. Let γ be the above loop. Then

$$Res((z-w)^k\ dz, w) = \frac{1}{2\pi i}\oint_\gamma (z-w)^k\ dz = \frac{1}{2\pi i}\int_0^{2\pi} e^{kit} ie^{it}\ dt = \frac{k}{2\pi} e^{(k+1)it}\ dt$$

   Now $k \neq -1$, and we know that $e^{it} = \cos t + i \sin t$. So the function $e^{(k+1)it}$ is periodic, with period a factor of 2π. It follows that the above integral is zero.

□

The following result is called the *Cauchy integral formula.*

Theorem 8.6 *Let $R \subseteq \mathbb{C}$ be an open subset. Let $f: R \to \mathbb{C}$ be a holomorphic function. Then, for each point $w \in R$:*

$$f(w) = Res\left(\frac{f(z)}{z-w}\ dz, w\right)$$

Proof: Since the function f is holomorphic, the function $z \mapsto \frac{f(z)}{z-w}$ is holomorphic on the set $R\backslash\{w\}$. Hence the differential form $\frac{f(z)}{z-w}\ dz$ is continuous and locally exact.

Write

$$\frac{f(z)}{z-w}\ dz = \frac{f(z)-f(w)}{z-w}\ dz + \frac{f(w)}{z-w}\ dz$$

Then by the above proposition

$$Res(\frac{f(z)}{z-w}\ dz, w) = Res(\frac{f(z)-f(w)}{z-w}\ dz, w) + f(w)Res(\frac{dz}{z-w}, w) = Res(\frac{f(z)-f(w)}{z-w}\ dz, w) + f(w)$$

So we need to show that $Res(\frac{f(z)-f(w)}{z-w}\ dz, w) = 0$. Define a path $\gamma_\epsilon: [0, 2\pi] \to R$ by the formula $\gamma_\epsilon(t) = w + \varepsilon e^{it}$. Then, for ϵ sufficiently small, γ

is a path in the open set R that winds once anticlockwise around the point w. Hence

$$Res(\frac{f(z)-f(w)}{z-w}\,dz, w) = \frac{1}{2\pi i}\oint_{\gamma_\epsilon} \frac{f(z)-f(w)}{z-w}\,dz = \frac{1}{2\pi i}\int_0^{2\pi} \frac{f(w+\varepsilon e^{it}) = f(\varepsilon e^{it})}{\varepsilon e^{it}} i\varepsilon e^{it}\,dt$$

By the mean value inequality

$$|Res(\frac{f(z)-f(w)}{z-w}\,dz, w)| \le \sup_{t\in[0,2\pi]} |f(w+\varepsilon e^{it}) - f(w)|$$

Since the function f is holomorphic, it is continuous at the point w. Hence $\sup_{t\in[0,2\pi]} |f(w+\varepsilon e^{it}) - f(w)|$ as $\varepsilon \to 0$. But the residue $Res(\frac{f(z)-f(w)}{z-w}\,dz, w)$ is independent of the value ε. We are forced to conclude that $Res(\frac{f(z)-f(w)}{z-w}\,dz, w) = 0$, and we are done. □

Cauchy's theorem and Cauchy's integral formula together allow us to compute many path integrals.

For example, let $R \subseteq \mathbb{C}$ be an open subset, let $f\colon R \to \mathbb{C}$ be holomorphic on R, let $w \in R$, and let γ be a loop in R.

Suppose that the loop γ lies within a simply connected subset $S \subseteq R$, and $w \notin S$. This is the case if the region R contains everything "inside" the loop γ, and the point w is not inside γ. Anyway, by Cauchy's theorem, we have

$$\oint_\gamma \frac{f(z)}{z-w}\,dz = 0$$

On the other hand, let $R \subseteq \mathbb{C}$ be a simply-connected open subset, let $w \in \mathbb{C}$, and let γ be a loop in R that winds once anticlockwise around the point w. Since the region R is simply-connected, the loop γ is homotopic to another loop $\tilde{\gamma}$ that winds once around w anticlockwise, and lies within some punctured disk $B(w,\delta)$. By the monodromy theorem

$$\oint_\gamma \frac{f(z)}{z-w}\,dz = \oint_{\tilde{\gamma}} \frac{f(z)}{z-w}\,dz = 2\pi i Res(\frac{f(z)}{z-w}\,dz, w)$$

Hence, by Cauchy's integral formula

$$\oint_\gamma \frac{f(z)}{z-w}\,dz = 2\pi i f(w)$$

Corollary 8.7 *Let $R \subseteq \mathbb{C}$ be an open subset. Let $f\colon R \to \mathbb{C}$ be a holomorphic function. Let $w \in \mathbb{C}$ and $\delta > 0$, and suppose that $\overline{B(w,\delta)} \subseteq R$.*

Then

$$|f(w)| \le \sup\{|f(z)| \mid z \in S(w,\delta)\}$$

Proof: Applying the above to the path $\gamma\colon [0,2\pi] \to R$ defined by the formula $\gamma(t) = e^{it}$, we have

$$f(w) = \frac{1}{2\pi i}\int_0^{2\pi} \frac{f(w+e^{it})}{re^{it}} ire^{it}\ dt = \frac{1}{2\pi}\int_0^{2\pi} f(w+re^{it})\ dt$$

Let $M = \sup\{|f(z)| \mid z \in S(w,\delta)\}$. Then

$$|f(w)| \leq \frac{1}{2\pi}\int_0^{2\pi} M\ dt = M$$

and we are done. □

8.3 Power Series and Analytic Functions

Let $a \in \mathbb{C}$. Recall that a *power series* is an expression of the form

$$\sum_{n=0}^{\infty} a_n (z-a)^n$$

where $a_n \in \mathbb{C}$.

Recall the following results on power series.

Theorem 8.8

- *There is a number $r \geq 0$, called the* radius of convergence*, such that the series $\sum_{n=0}^{\infty} a_n(z-a)^n$ converges whenever $z \in B(w,r)$, and diverges whenever $|z-a| > r$.*

  Here we allow the possibility "$r = \infty$", which means that the series converges for all $z \in \mathbb{C}$.

- *The power series $\sum_{n=0}^{\infty} na_n(z-a)^{n-1}$ has the same radius of convergence as the above power series.*

- *Let $R > 0$. Then we can define a holomorphic function $f\colon B(a,R) \to \mathbb{C}$ by the formula $f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$. The derivative is given by the formula $f'(z) = \sum_{n=0}^{\infty} na_n(z-a)^{n-1}$, ie: we can differentiate "term-by-term".*

□

Example 8.9 The *geometric series* $\sum_{n=0}^{\infty} z^n$ has radius of convergence 1. In fact, if $|z| < 1$, then $\sum_{n=0}^{\infty} z^n = \frac{1}{1-z}$.

Example 8.10 The *exponential series* $\sum_{n=0}^{\infty} \frac{1}{n!} z^n$ has radius of convergence ∞. The *exponential function* is defined by the formula

$$e^z = \sum_{n=0}^{\infty} \frac{1}{n!} z^n$$

for all $z \in \mathbb{C}$.

Two fundamental properties of the exponential function are the formulae

$$e^{w+z} = e^w e^z \qquad \frac{d}{dz}e^z = e^z$$

Example 8.11 The cos and sin functions are defined by writing

$$\cos z = \sum_{n=0}^{\infty}(-1)^n \frac{z^{2n}}{(2n)!} \qquad \sin z = \sum_{n=0}^{\infty}(-1)^n \frac{z^{2n+1}}{(2n+1)!}$$

Both of these functions are defined for all $z \in \mathbb{C}$, since the relevant series have radius of convergence ∞.

Let $\theta \in \mathbb{R}$. Then, using the above series we immediately obtain the *Euler formula*

$$e^{i\theta} = \cos\theta + i\sin\theta$$

Definition 8.12 Let $R \subseteq \mathbb{C}$ be an open subset. A function $f\colon R \to \mathbb{C}$ is called *analytic* if given a point $a \in R$, we can find $r > 0$, and a power series that converges to the value $f(z)$ for all $z \in B(a,r)$.

Thus, given a point $a \in R$, we can find $r > 0$ and $a_n \in C$ such that

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$$

for all $z \in B(a,r)$.

Proposition 8.13 *Let $f\colon R \to \mathbb{C}$ be an analytic function. Then the function f is differentiable of arbitrary order. If on a disk $B(a,r)$, the function f is defined by the power series*

$$f(z) = \sum_{n=0}^{\infty} a_n(z-w)^n$$

then

$$a_k = \frac{f^{(k)}(a)}{k!}$$

Proof: Let

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$$

when $z \in B(a,r)$. Then by theorem 8.8, the function f is holomorphic on the disk $B(a,r)$, and

$$f'(z) = \sum_{n=0}^{\infty} na_n(z-a)^{n-1}$$

We see that the derivative f' is again analytic. Repeating the above, we see that derivatives of all orders exist.

Now, observe that the formula $f^{(k)}(w) = k!a_k$ certainly holds when $k = 0$. Fix $k \in \mathbb{N}$. Suppose, given a function g defined on $B(w, r)$ by a power series $\sum_{n=0}^{\infty} b_n(z-a)^n$, we have $g^{(k)}(a) = k!b_k$.

By theorem 8.8, we have

$$f'(z) = \sum_{n=0}^{\infty} n a_n (z-a)^{n-1} = \sum_{n=0}^{\infty} (n+1) a_{n+1} (z-a)^n$$

Applying the above to this series, we see that

$$f^{(k+1)}(a) = (f')^{(k)}(a) = k!(k+1)a_{k+1} = (k+1)!a_{k+1}$$

and the desired formula is true by induction. □

The following result is easy to check; we leave the proof as an exercise. Note that the first part of the proposition already follows from the fact that an analytic function is holomorphic.

Proposition 8.14 *Let $f: R \to \mathbb{C}$ be an analytic function. Then the differential form $f(z)\ dz$ is locally exact. If, on the disk $B(a, r)$, we have $f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n$, then we have a primitive function on the same disk defined by the formula*

$$F(z) = \sum_{n=0}^{\infty} \frac{a_n}{n+1}(z-a)^{n+1}$$

□

8.4 Taylor's Theorem

The concept of *uniform convergence* turns out to be very useful when dealing with power series.

Definition 8.15 Let A be any set, and let $\sum_{n=0}^{\infty} f_n$ be a series of functions $f_n: A \to \mathbb{C}$. We say the series $\sum_{n=0}^{\infty} f_n$ *converges uniformly* to a function $f: A \to \mathbb{C}$ on the set A if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $|\left(\sum_{n=0}^{m} f_n(x)\right) - f(x)| < \varepsilon$ for all $m \geq N$ and all $x \in A$.

Another way to phrase the above definition is to say that the series of functions $\sum_{n=0}^{\infty} f_n$ converges uniformly to the function f on the set A is to say

$$\lim_{N \to \infty} \sup\{\left|\sum_{n=0}^{N} f_n(x) - f(x)\right| = 0 \mid x \in A\}$$

One use of uniform convergence is in swapping limits and integrals.

Theorem 8.16 *Let $\sum_{n=0}^{\infty} f_n$ be a uniformly convergent series of continuous functions $f_n: [a,b] \to \mathbb{C}$. Then the function, f, defined as the limit of the series is continuous, and*

$$\int_a^b f = \sum_{n=0}^{\infty} \int_a^b f$$

In particular, the sum on the right converges. □

The following is straightforward, and left as an exercise.

Corollary 8.17 *Let $R \subseteq \mathbb{C}$ be an open subset, and let $\sum_{n=0}^{\infty} f_n$ be a series of continuous functions $f_n: R \to \mathbb{C}$ that converges to a function $f: R \to \mathbb{C}$.*

Let $\gamma: [a,b] \to R$ be a piecewise-differentiable path, and suppose that the above series converges uniformly on the image of the path γ. Then

$$\int_\gamma f(z)\ dz = \sum_{n=0}^{\infty} \int_\gamma f_n(z)\ dz$$

In particular, the sum on the right converges. □

Power series uniformly converge in the following sense.

Theorem 8.18 *Let $a \in \mathbb{C}$, and let $\sum_{n=0}^{\infty} a_n(z-a)^n$ be a power series with radius of convergence $r > 0$. Let $K \subseteq B(a,r)$ be a closed subset. Then the series converges uniformly on the set K.* □

The following result is called *Taylor's theorem.*

Theorem 8.19 *Let $R \subseteq \mathbb{C}$ be an open subset, and let $f: R \to \mathbb{C}$ be holomorphic. Then the function f is analytic.*

Further, given a point $a \in R$, in any disk $B(a,\delta) \subseteq R$, we have

$$f(z) = \sum_{n=0}^{\infty} a_n(z-a)^n \qquad a_n = Res\left(\frac{f(z)}{(z-a)^{n+1}}\ dz, a\right)$$

Proof: It suffices to prove the second part of the result; the first part than follows by definition of a function being analytic. For convenience, suppose $a = 0$.

Let $w \in B(a,\delta)$. Choose r between $|w|$ and δ, and define a loop $\gamma: [0,2\pi] \to B(0,\delta)$ by the formula $\gamma(t) = re^{it}$. Then by Cauchy's integral formula

$$f(w) = \frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z-w}\ dz$$

Let z be lie on the loop γ. Then $|z| > |w|$, so

$$\frac{f(z)}{z-w} = \frac{f(z)}{z(1-w/z)} = \sum_{n=0}^{\infty} \frac{f(z)}{z}\left(\frac{w}{z}\right)^n$$

In fact, we can check that the above series converges uniformly on the loop γ, so by corollary 8.17, we have

$$f(w) = \sum_{n=0}^{\infty} \frac{1}{2\pi i} \oint_{\gamma} \frac{f(z)}{z} \left(\frac{w}{z}\right)^n \ dz = \sum_{n=0}^{\infty} Res\left(\frac{f(z)}{z^{n+1}}\ dz\right) w^n$$

The result now follows. □

The following is called Cauchy's formula for derivatives. It is immediate by the above and proposition 8.13.

Corollary 8.20 *Let $R \subseteq \mathbb{C}$ be an open subset, and let $f\colon R \to \mathbb{C}$ be a holomorphic function. Let $a \in R$. Then the n-th order partial derivative of the function f exists, and is given by the formula*

$$f^{(n)}(a) = Res\left(\frac{n! f(z)}{(z-a)^{n+1}}\ dz, a\right)$$

Hence, if $B(z,\delta) \subseteq R$, and γ is a loop in the disk $B(z,\delta)$ that winds once anticlockwise around z, then we have the formula

$$f^{(n)}(a) = \frac{n!}{2\pi i} \oint_{\gamma} \frac{f(z)}{(z-a)^{n+1}}\ dz$$

We call the following the *Cauchy inequality.*

Corollary 8.21 *let $0 < r < R$, let $a \in \mathbb{C}$, and let $f\colon B(a,R) \to \mathbb{C}$ be a holomorphic function. Let $M = \sup\{|f(z)| \mid |z-a| = r\}$. Then*

$$|f^{(n)}(w)| \leq \frac{n! M r}{(r - |w-a|)^{n+1}}$$

for all $w \in B(w,r)$.

Proof: Let $\gamma\colon [0,2\pi] \to \mathbb{C}$ be the loop $\gamma(t) = a + re^{it}$. Since $w \in B(a,r)$, the loop γ winds once anticlockwise around the point w. Hence, by Cauchy's formula for derivatives

$$|f^{(n)}(w)| = \left|\frac{n!}{2\pi i} \oint_{\gamma} \frac{f(z)}{(z-w)^{n+1}}\ dz\right|$$

Hence

$$|f^{(n)}(w)| \leq \frac{n!}{2\pi} \int_0^{2\pi} \frac{|f(a+re^{it})|}{|a+re^{it}-w|^{n+1}} |ire^{it}|\ dt$$

Now, $|f(w+re^{it})| \leq M$, and $|a+re^{it}-w| \geq r - |w-a|$. Hence, since $|w-a| < r$, we have

$$|f^{(n)}(z)| \leq \frac{n!}{2\pi} \int_0^{2\pi} \frac{Mr}{(r-|w-a|)^{n+1}}\ dt = \frac{n! M r}{(r-|w-a|)^{n+1}}$$

and we are done. □

8.5 Liouville's Theorem

The following result is called *Liouville's theorem.*

Theorem 8.22 *Let $f: \mathbb{C} \to \mathbb{C}$ be a bounded holomorphic function. Then f is constant.*

Proof: Since the function f is bounded, we have a constant $K \geq 0$ such that $|f(z)| \leq K$ for all $z \in \mathbb{C}$. Choose $w \in \mathbb{C}$, and let $r > |w|$. then by the Cauchy inequality

$$|f'(w)| \leq \frac{Mr}{(r - |w|)^2}$$

let $r \to \infty$. Then $Mr/(r - |w|)^2 \to 0$. It follows that $f'(w) = 0$. Hence, since the domain $\mathbb{C}$ is certainly connected, the function f is constant. □

Definition 8.23 Let $R_1, R_2 \subseteq \mathbb{C}$ be open. We call R_1 and R_2 *biholomorphic* if there is a bijective holomorphic mapping $f: R_1 \to R_2$ with holomorphic inverse.

Certainly, if two open sets are biholomorphic, then they are homeomorphic, since a holomorphic function is also continuous.

Example 8.24 Let

$$R_1 = \{z \in \mathbb{C} \mid \Re(z) > 0 \text{ and } \Im(z) > 0\}$$

and

$$R_2 = \{z \in \mathbb{C} \mid \Im(z) > 0\}$$

Then we can write

$$R_1 = \{re^{i\theta} \mid r \geq 0,\ 0 < \theta < \frac{\pi}{2}\}$$

and

$$R_2 = \{re^{i\theta} \mid r \geq 0,\ 0 < \theta < \pi\}$$

Let $f(z) = z^2$. Then $f(re^{i\theta}) = r^2 e^{2i\theta}$. It follows that the map $f: R_1 \to R_2$ is a holomorphic bijection. The inverse $g(z) = \sqrt{z}$ is also holomorphic. Thus the sets R_1 and R_2 are biholomorphic.

Proposition 8.25 *The sets $B(0, 1) \subseteq \mathbb{C}$ and $\mathbb{C}$ are homeomorphic, but not biholomorphic.*

Proof: We can define a continuous map $f: B(0, 1) \to \mathbb{C}$ by the formula

$$f(z) = \frac{z}{1 - |z|}$$

This map is a homeomorphism, since we have a continuous inverse given by the formula

$$f^{-1}(w) = \frac{w}{1+|w|}$$

Now, let $g: \mathbb{C} \to B(0,1)$ be a holomorphic map. Then the map g is bounded. Hence, by Liouville's theorem, g is constant. Thus the map g is not bijective, and the sets $\mathbb{C}$ and $B(0,1)$ cannot be biholomorphic. $\square$

Definition 8.26 Let $f: \mathbb{C} \to \mathbb{C}$ be a function. Then we say that $f(z) \to a$ as $z \to \infty$, or

$$\lim_{z\to\infty} f(z) = a$$

if for all $\varepsilon > 0$ we can find $R \geq 0$ such that $|f(z) - a| < \varepsilon$ whenever $|z| \geq R$.

We have a version of the above definition in any normed vector space.

Definition 8.27 Let $f: \mathbb{C} \to \mathbb{C}$ be a function. Then we say that $f(z) \to \infty$ as $z \to \infty$ or

$$\lim_{z\to\infty} f(z) = \infty$$

if for all $s > 0$ we can find $R \geq 0$ such that $|f(z)| > s$ whenever $|z| \geq R$.

Proposition 8.28 *Let $f: \mathbb{C} \to \mathbb{C}$ be a continuous function such that $\lim_{z\to\infty} f(z)$ exists. Then f is bounded.*

Proof: Let

$$\lim_{z\to\infty} f(z) = a$$

Taking '$\varepsilon = 1$' in the definition of the limit, we can find $R \geq 0$ such that $|f(z) - a| < 1$ whenever $|z| \geq R$. Hence $|f(z)| < 1 + |a|$ whenever $|z| \geq R$.

On the other hand, the set $\overline{B(0,R)}$ is a closed bounded subset, and hence compact by the Heine-Borel theorem. Thus there is a constant $M \geq 0$ such that $|f(z)| \leq M$ whenever $z \in \overline{B(0,R)}$.

Take $K = \max(M, 1 + |a|)$. For any complex number $z \in \mathbb{C}$, either $z \in \overline{B(0,R)}$ or $|z| > R$. It follows that $|f(z)| \leq K$ for all $z \in \mathbb{C}$. Thus the function f is bounded, and we are done. $\square$

The following example and proposition are left as exercises.

Example 8.29 Let $p: \mathbb{C} \to \mathbb{C}$ be a non-constant polynomial (with complex coefficients). Then $p(z) \to \infty$ as $z \to \infty$.

Proposition 8.30 *Let $f: \mathbb{C} \to \mathbb{C}$ be a function. Then $\lim_{z\to\infty} f(z) = \infty$ if and only if $\lim_{z\to\infty} \frac{1}{f(z)} = 0$.* $\square$

Our last consequence of Liouville's theorem is called the *fundamantal theorem of algebra*.

Theorem 8.31 *Let $p: \mathbb{C} \to \mathbb{C}$ be a polynomial with complex coefficients. Suppose that p is not constant (ie: $n \geq 1$ and $a_n \neq 0$). Then there is at least one point $z \in \mathbb{C}$ such that $p(z) = 0$.*

Proof: Suppose that $p(z) \neq 0$ for all $z \in \mathbb{C}$. Then the function $z \mapsto \frac{1}{p(z)}$ is a holomorphic function defined on the entire complex plane.

Since p is non-constant, by the above example and proposition, $\frac{1}{p(z)} \to 0$ as $z \to \infty$.

Hence, by proposition 8.28, the function $\frac{1}{p}$ is bounded, and therefore constant by Liouville's theorem. Thus the polynomial p is constant, which is a contradiction.

It follows that there exists $z \in \mathbb{C}$ such that $p(z) = 0$. $\square$

8.6 The Identity Theorem

Let X be a metric space, and let $A \subseteq X$. Recall that a point $x_0 \in X$ is called an *accumulation point* of the subset Z if and only if for all $\varepsilon > 0$, we can find a point $x \in A \cap B(x_0, \varepsilon)$, where $x \neq x_0$.

We leave the following results on accumulation points and closed sets as exercises.

Proposition 8.32 *Let X be a metric space, and let $A \subseteq X$ be a subset. Then A is closed if and only if it contains all of its accumulation points.* $\square$

Let us write $\overline{A}$ to denote the union of a subset A with its accumulation points. By the above, if the subset A is already closed, then $\overline{A} = A$.

Proposition 8.33 *Let X be a metric space, and let A be a subset. Then the set $\overline{A}$ is the smallest closed subset of the space X which contains A.* $\square$

We call the subset $\overline{A}$ the *closure* of the set A. Our notation is consistent with the earlier notation used for closed balls, as the following example (again left as an exercise) shows.

Example 8.34 Let $x \in X$ and $\delta > 0$. Then the closure of the open ball $B(x, \delta)$ is the closed ball

$$\overline{B(x, \delta)} = \{y \in X \mid d(x, y) \leq \delta\}$$

Now, the following result in complex analysis is called the *identity theorem*.

Theorem 8.35 *Let $R \subseteq \mathbb{C}$ be an open connected subset, and let $f: R \to \mathbb{C}$ be a holomorphic function. Then the following are equivalent.*

1. $f = 0$.

2. *The set of zeroes*

$$Z(f) = \{z \in R \mid f(z) = 0\}$$

has an accumulation point lying in the domain R.

3. *There is a point $w \in R$ such that $f^{(n)}(w) = 0$ for all $n \in \mathbb{N}$.*

Proof:

- (1) $\Rightarrow$ (2).

  This is trivial.

- (2) $\Rightarrow$ (3).

  Let $w \in R$ be an accumulation point of the set of zeroes $Z(f)$. We want to show that $f^{(n)}(w) = 0$ for all n.

  Suppose this is not true. Then we have a *minimal* $k \in \mathbb{N}$ such that $f^{(k)}(w) \neq 0$. By Taylor's theorem, we have $r > 0$ such that

  $$f(z) = \sum_{n=k}^{\infty} a_n (z - w)^n = (z - w)^k \tilde{f}(z) \qquad \tilde{f}(z) = \sum_{n=k}^{\infty} a_n (z - w)^{n-k}$$

  for all $z \in B(w, r)$.

  Since $f^{(k)}(w) \neq 0$, we have $\tilde{f}(w) = a_k \neq 0$. The function $\tilde{f}\colon B(w, r) \to \mathbb{C}$ is holomorphic, and therefore continuous, so we can find $\varepsilon > 0$ such that $\tilde{f}(z) \neq 0$ whenever $z \in B(w, \varepsilon)$.

  Hence, $f(z) \neq 0$ for all $z \in B'(w, \varepsilon)$. Thus the point w is not an accumulation point of the set of zeroes $Z(f)$.

- (3) $\Rightarrow$ (1).

  Since the function f is holomorphic, it is also analytic by Taylor's theorem, and the n-th derivative $f^{(n)}\colon R \to \mathbb{C}$ is continuous. Thus the set

  $$W_n = \{z \in R \mid f^{(n)}(z) = 0\}$$

  is closed.

  It follows that the intersection $W = \cup_{n=0}^{\infty} W_n$ is also closed. If we assume (3), then $W \neq \emptyset$.

  Let $w \in W$. Then by Taylor's theorem, the function f is zero in open ball $B(w, \delta)$. Hence $B(w, \delta) \subseteq W$, and the set W is open.

  Thus $W \neq \emptyset$, and W is both open and closed. Since the domain R is connected, we therefore have $W = R$, and $f = 0$.

□

The identity theorem has a number of immediate consequences.

Corollary 8.36 *Let $R \subseteq \mathbb{C}$ be an open connected subset, and let $f, g: R \to \mathbb{C}$ be holomorphic functions. Then the following are equivalent.*

1. *$f = g$.*
2. *The set*
$$\{z \in R \mid f(z) = g(z)\}$$
has an accumulation point lying in the domain R.
3. *There is a point $w \in R$ such that $f^{(n)}(w) = g^{(n)}(w)$ for all $n \in \mathbb{N}$.*

Proof: Apply the identity theorem to the function $h = f - g$. $\square$

Corollary 8.37 *Let $R \subseteq \mathbb{C}$ be an open connected set. Then a holomorphic function $f: R \to \mathbb{C}$ is determined by its values on an arbitrarily small local ball lying in R.*

Proof: Let $w \in R$, and $\varepsilon > 0$. Let $g: R \to \mathbb{C}$ be a holomorphic function such that $g(z) = f(z)$ for all $z \in B(w, \varepsilon)$.

The point $w \in R$ is an accumulation point of the set

$$B(w, \varepsilon) \subseteq \{z \in R \mid f(z) = g(z)\}$$

Hence, by the above, $f = g$. $\square$

Let $f: \mathbb{R} \to \mathbb{R}$ be a function. A *holomorphic extension* of f is a holomorphic function $\overline{f}: \mathbb{C} \to \mathbb{C}$ such that $\overline{f}(x) = f(x)$ whenever $x \in \mathbb{R}$.

Holomorphic extensions do not always exist. For example, by Taylor's theorem, for a function to have a holomorphic extension it must be infinitely differentiable.

Corollary 8.38 *Let $f: \mathbb{R} \to \mathbb{R}$ be a function. Then f has at most one holomorphic extension.*

Proof: Let $\overline{f} = \overline{g}$ be holomorphic extensions of the function f. Let $S = \{z \in \mathbb{C} \mid \overline{f}(z) = \overline{g}(z)\}$.

Then $\mathbb{R} \subseteq S$, since the functions are both equal to the function f for real numbers. But the set of real numbers, $\mathbb{R}$, certainly has an accumulation point. Hence $\overline{f} = \overline{g}$. $\square$

The above corollary can be used to deduce certain properties of complex functions from the corresponding properties of real functions.

For instance, for all $\theta \in \mathbb{R}$, we have the Euler formula

$$e^{i\theta} = \cos\theta + i \sin\theta$$

Let $f(z) = e^{iz}$, and $g(z) = \cos z + i \sin z$. Then f and g are both holomorphic functions, and $f(x) = g(x)$ for all $x \in \mathbb{R}$. Hence $f(z) = g(z)$ for all $z \in \mathbb{C}$, that is
$$e^{iz} = \cos z + i \sin z$$
for all $z \in \mathbb{C}$.

8.7 Open Mappings

Definition 8.39 Let X and Y be topological spaces, and let $\varphi: X \to Y$ be a map. We call the map φ *open* if for any open subset $U \subseteq X$, the image $f[U] \subseteq Y$ is open.

The above is different to the definition of a continuous map, which is that the *inverse image* of an open set is open.

Observe that an open continuous bijection is a homeomorphism.

We want to show that any non-constant holomorphic map is open.

Lemma 8.40 *Let $R \subseteq \mathbb{C}$ be an open connected set, and let $f: R \to \mathbb{C}$ be a non-constant holomorphic function. Choose $w \in \mathbb{C}$. Then there exists $\delta > 0$ such that $f(z) \neq f(w)$ for all $z \in S(w, \delta)$.*

Proof: Suppose that for all $\delta > 0$ where $B(w, \delta) \subseteq R$, we can find $z \in S(w, \delta)$ such that $f(z) = f(w)$.

Let $c: R \to \mathbb{C}$ be the constant function, with $c(z) = f(w)$ for all $z \in R$. Then, by our assumption, the set
$$S = \{z \in \mathbb{C} \mid f(z) = c(z)\}$$
has an accumulation point, w.

It follows by the identity theorem that $f = c$, that is to say the function f is constant, which is a contradiction. □

Lemma 8.41 *Let $R \subseteq \mathbb{C}$ be an open set, and let $f: R \to \mathbb{C}$ be a holomorphic function. Let $w \in R$, and choose $\delta > 0$ such that $\overline{B(w, \delta)} \subseteq R$.*

Suppose that
$$|f(w)| < \inf\{|f(z)| \mid z \in S(w, \delta)\}$$
Then the function f has a zero in the ball $B(w, \delta)$.

Proof: Suppose that $f(z) \neq 0$ for all $z \in B(w, \delta)$. By the above inequality, $f(z) \neq 0$ for all $z \in \overline{B}(w, \delta)$.

Now the set of points $z \in R$ such that $f(z) \neq 0$ is the inverse image of the open set $\mathbb{C} \backslash \{0\}$. Hence, there is an open set $U \subseteq R$ such that $f(z) \neq 0$ for all $z \in U$, and $\overline{B(w, \delta)} \subseteq U$.

The function $g: U \to \mathbb{C}$ defined by the formula $g(z)/1/f(z)$ is holomorphic. By corollary 8.7 to Cauchy's integral formula, we have
$$|g(w)| \leq \sup\{|g(z)| \ z \in S(w, \delta)\}$$

Taking the reciprocal, we see

$$|f(w)| \geq \inf\{|f(z)|\ z \in S(w, \delta)\}$$

which contradicts our original inequality.

Thus the function f has at least one zero in the ball $B(w, \delta)$. □

Lemma 8.42 *Let $R \subseteq \mathbb{C}$ be an open set, and let $f \colon R \to \mathbb{C}$ be holomorphic. Let $w \in \mathbb{C}$, and choose $r > 0$ such that $\overline{B(w, r)} \subseteq R$.*

Let

$$\delta = \frac{1}{2} \inf\{|f(z) - f(w)| \mid z \in S(w, r)\}$$

Suppose $\delta > 0$. Then $B(f(w), \delta) \subseteq f[B(w, r)]$.

Proof: Let $u \in B(f(w), \delta)$. We want to show that $u \in f[B(w, r)]$. So we need to find a point $v \in B(w, r)$ such that $f(v) = u$.

Let $z \in S(w, r)$. By definition of the constant δ, we know that $|f(z) - f(w)| \leq 2\delta$. Since $u \in B(f(w), \delta)$, we know that $|f(w) - u| < \delta$. Hence by the triangle inequality

$$|f(z) - u| \geq |f(z) - f(w)| - |f(w) - u| > 2\delta - \delta = \delta$$

Therefore $|f(w) - u| < |f(z) - u|$ for all $z \in S(w, r)$. Since the sphere $S(w, r)$ is compact, we see that

$$|f(w) - u| < \inf\{|f(z) - u| \mid z \in S(w, \delta)\}$$

By the previous lemma, we can find $v \in B(w, r)$ such that $f(v) - u = 0$, that is to say $f(v) = u$, and we are done. □

Theorem 8.43 *Let $R \subseteq \mathbb{C}$ be an open connected subset, and let $f \colon R \to \mathbb{C}$ be a non-constant holomorphic mapping. Then f is open.*

Proof: Let $U \subseteq R$ be open. Let $w \in U$. Then by lemma 8.40, there exists $\delta > 0$ such that $\overline{B(w, r)} \subseteq U$, and $f(z) \neq f(w)$ for all $z \in S(w, \delta)$.

Since the sphere $S(w, \delta)$ is compact, if we define

$$\delta = \frac{1}{2} \inf\{|f(z) - f(w)| \mid z \in S(w, r)\}$$

then $\delta > 0$.

By the above lemma, $B(f(w), \delta) \subseteq f[B(w, r)] \subseteq f[U]$. Hence the image $f[U]$ is open, and we are done. □

Since it is impossible for an injective map defined an a non-empty open subset of $\mathbb{C}$ to be constant, we immediately obtain the following.

Corollary 8.44 *Let $R \subseteq \mathbb{C}$ be a connected open subset. Let $f \colon R \to S$ be a bijective holomorphic map. Then f is a homeomorphism.* □