MAT 473 LECTURES

JOHN QUIGG

CONTENTS

1. Linear maps 1
2. Differentiation 4
3. Lebesgue measure 15
4. Measurable functions 24
5. Integrating nonnegative functions 28
6. Integrable functions 33
7. Iterated integrals 41
8. Change of variables 45
9. Manifolds 52
Index 59

1. LINEAR MAPS

Definition 1.1. $L(\mathbb{R}^p, \mathbb{R}^q)$ denotes the set of linear functions from $\mathbb{R}^p$ to $\mathbb{R}^q$. If $p = q$ we write $L(\mathbb{R}^p)$. For each $A \in L(\mathbb{R}^p, \mathbb{R}^q)$ we define

$$\|A\| = \sup\{\|A(x)\| : \|x\| = 1\}.$$ 

Proposition 1.2.  
(i) For all $A \in L(\mathbb{R}^p, \mathbb{R}^q)$,

$$\|A\| = \min\{c \in \mathbb{R} : \|Ax\| \leq c\|x\| \text{ for all } x \in \mathbb{R}^p\}.$$ 

(ii) Every $A \in L(\mathbb{R}^p, \mathbb{R}^q)$ is uniformly continuous.

(iii) $\| \cdot \|$ is a norm on $L(\mathbb{R}^p, \mathbb{R}^q)$.

(iv) For all $A \in L(\mathbb{R}^p, \mathbb{R}^q)$ and $B \in L(\mathbb{R}^q, \mathbb{R}^r)$, $\|AB\| \leq \|A\|\|B\|$.

Proof. (i) First, if $x \neq 0$ then

$$\|Ax\| = \left\| A\left(\frac{x}{\|x\|}\right) \right\| = \left\| x\left(\frac{x}{\|x\|}\right) \right\| = \left\| x\left(\frac{x}{\|x\|}\right) \right\| = \|x\|\left\| A\left(\frac{x}{\|x\|}\right) \right\| \leq \|x\|\|A\|,$$

since $x/\|x\|$ is a unit vector (that is, has norm 1). This inequality also holds (trivially) for $x = 0$. 

Date: May 2, 2000.
Next, if \( c \) satisfies the inequality for all \( x \in \mathbb{R}^p \), then in particular if \( \|x\| = 1 \) we have
\[
\|Ax\| \leq c,
\]
so taking the supremum over \( x \) we get \( \|A\| \leq c \).

(ii) For all \( x, y \in \mathbb{R}^p \),
\[
\|Ax - Ay\| = \|A(x - y)\| \leq \|A\| \|x - y\|,
\]
so in fact \( A \) is Lipschitz, with (best) Lipschitz constant \( \|A\| \).

(iii) Let \( A, B \in L(\mathbb{R}^p, \mathbb{R}^q) \). Since \( \|Ax\| \geq 0 \) for all \( x \in \mathbb{R}^p \), we have \( \|A\| \geq 0 \).
If \( c \in \mathbb{R} \) and \( x \in \mathbb{R}^p \) then
\[
\|cAx\| = |c| \|Ax\|.
\]
Taking the supremum over \( \|x\| = 1 \), we get
\[
\|cA\| = |c| \|A\|.
\]
If \( x \in \mathbb{R}^p \) then
\[
\|(A + B)x\| = \|Ax + Bx\| \leq \|Ax\| + \|Bx\|
\leq \|A\| \|x\| + \|B\| \|x\| = (\|A\| + \|B\|) \|x\|,
\]
so
\[
\|A + B\| \leq \|A\| + \|B\|.
\]

(iv) For all \( x \in \mathbb{R}^p \),
\[
\|ABx\| \leq \|A\| \|Bx\| \leq \|A\| \|B\| \|x\|,
\]
so
\[
\|AB\| \leq \|A\| \|B\|.
\]

\( \square \)

**Example 1.3.** (i) The norm of the identity operator \( I \) on \( \mathbb{R}^p \) is 1.

(ii) Fix \( x \in \mathbb{R}^p \), and define \( A \in L(\mathbb{R}, \mathbb{R}^p) \) and \( B \in L(\mathbb{R}^p, \mathbb{R}) \) by \( At = tx \) and \( By = x \cdot y \). Then
\[
\|A\| = \|B\| = \|x\|.
\]

**Definition 1.4.** Every \( A \in L(\mathbb{R}^p, \mathbb{R}^q) \) is uniquely represented by a \( q \times p \) matrix \( [A] = [a_{ij}] \) relative to the standard bases of \( \mathbb{R}^p \) and \( \mathbb{R}^q \), so that \( A(x_1, \ldots, x_p)_i = \sum_{j=1}^p a_{ij} x_j \). The map \( A \mapsto [A] \) is an isomorphism of \( L(\mathbb{R}^p, \mathbb{R}^q) \) onto the vector space of \( q \times p \) matrices. If we identify this latter vector space with \( \mathbb{R}^{pq} \) in one of the obvious (but it doesn’t really matter which) ways, the Euclidean norm on \( \mathbb{R}^{pq} \) gives a norm on the matrices, hence another norm
\[
\|A\|_2 := \left( \sum_{ij} a_{ij}^2 \right)^{1/2}
\]
on \( L(\mathbb{R}^p, \mathbb{R}^q) \).
Proposition 1.5. For all $A \in L(\mathbb{R}^p, \mathbb{R}^q)$,
\[
\|A\| \leq \|A\|_2 \leq \sqrt{pq}\|A\|.
\]
Consequently, any of the natural isomorphisms of $L(\mathbb{R}^p, \mathbb{R}^q)$ onto $\mathbb{R}^q$ is a homeomorphism.

Proof. First, if $x \in \mathbb{R}^p$ then
\[
\|Ax\|^2 = \sum_i \left(\sum_j a_{ij}x_j\right)^2
\leq \sum_i \left(\sum_j a_{ij}^2 \sum_j x_j^2\right) \quad \text{(by the Cauchy-Schwarz Inequality)}
= \sum_{ij} a_{ij}^2 \sum_j x_j^2 = \|A\|^2_2 \|x\|^2,
\]
so $\|A\| \leq \|A\|_2$.

For the other inequality, let $\{e_j\}_1^p$ and $\{u_i\}_1^q$ be the standard bases for $\mathbb{R}^p$ and $\mathbb{R}^q$. Then for all $i, j$,
\[
|a_{ij}| = |(Ae_j) \cdot u_j| \leq \|Ae_j\| \|u_i\| \leq \|A\|,
\]
so
\[
\|A\|_2^2 = \sum_{ij} a_{ij}^2 \leq pq \|A\|^2.
\]

Corollary 1.6. (i) $\det : L(\mathbb{R}^p) \to \mathbb{R}$ is continuous.
(ii) Let $GL(p)$ denote the set of invertible linear operators on $\mathbb{R}^p$. Then $GL(p)$ is open in $L(\mathbb{R}^p)$, and the map $A \mapsto A^{-1}$ on $GL(p)$ is a homeomorphism.

Proof. (i) Identifying the $p \times p$ matrices with $\mathbb{R}^{p^2}$, det is a polynomial, hence continuous. Composing with the homeomorphism of $L(\mathbb{R}^p)$ onto $\mathbb{R}^{p^2}$, det is continuous on $L(\mathbb{R}^p)$.

(ii) First, $GL(p) = \det^{-1}(\mathbb{R} \setminus \{0\})$, the pre-image of an open set under a continuous function, hence is open. Next, by the adjoint formula for inverting matrices, each entry of the matrix $[A]^{-1}$ representing the linear map $A^{-1}$ is a rational function, hence continuous. Therefore, inversion is continuous on $GL(p)$.

Proposition 1.7. For all $A \in L(\mathbb{R}^p)$, if $\|A - I\| < 1$ then $A$ is invertible.

Proof. It suffices to show that $A$ is 1-1, and for this it suffices to show that $\ker A = \{0\}$. If $x \neq 0$ then
\[
\|Ax\| = \|x - (x - Ax)\| \geq \|x\| - \|x - Ax\| = \|x\| - \|(I - A)x\| \geq \|x\| - \|I - A\| \|x\| = (1 - \|I - A\|) \|x\| > 0,
\]
since \( \|x\| > 0 \).

2. Differentiation

**Definition 2.1.** Let \( E \subseteq \mathbb{R}^p, f : E \to \mathbb{R}^q \), and \( x \in E^o \). A derivative of \( f \) at \( x \) is a linear function \( A : \mathbb{R}^p \to \mathbb{R}^q \) such that

\[
\lim_{h \to 0} \frac{f(x + h) - f(x) - Ah}{\|h\|} = 0.
\]

Such an \( A \) is unique if it exists, and is denoted \( f'(x) \). \( f \) is differentiable at \( x \) if \( f'(x) \) exists. \( f \) is differentiable if it is differentiable at every point of \( E \).

**Remark 2.2.** If \( f \) is differentiable, then \( \text{dom } f \) is open.

**Example 2.3.** (i) If \( A \in L(\mathbb{R}^p, \mathbb{R}^q) \) then \( A'(x) = A \) for all \( x \in \mathbb{R}^p \).

(ii) If \( U \subseteq \mathbb{R}^p \) is open and \( f : U \to \mathbb{R}^q \) is constant then \( f'(x) = 0 \) for all \( x \in U \).

**Lemma 2.4.** \( f = (f_1, \ldots, f_q) \) is differentiable if and only if each \( f_i \) is.

**Proof.** If \( A = (A_1, \ldots, A_q) \in L(\mathbb{R}^p, \mathbb{R}^q) \) then

\[
\frac{f(x + h) - f(x) - Ah}{\|h\|} = \left( \frac{f_1(x + h) - f_1(x) - A_1h}{\|h\|}, \ldots, \frac{f_q(x + h) - f_q(x) - A_qh}{\|h\|} \right),
\]

and the result follows since limits of vectors can be taken coordinate-wise.

**Lemma 2.5.** \( f'(x) = A \) if and only if there exists a function \( r \) such that

\[ f(x + h) = f(x) + Ah + r(h)\|h\| \quad \text{and} \quad \lim_{h \to 0} r(h) = 0. \]

**Proof.** For each \( h \in \mathbb{R}^p \) such that \( x + h \in \text{dom } f \), define

\[ r(h) = \begin{cases} \frac{f(x + h) - f(x) - Ah}{\|h\|} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0. \end{cases} \]

The result follows from the definition of derivative.

**Proposition 2.6.** If \( f \) is differentiable at \( x \), then \( f \) is continuous at \( x \).

**Proof.** We have

\[ f(x + h) = f(x) + f'(x)h + r(h)\|h\|, \]

where \( r(h) \to 0 \) as \( h \to 0 \). Then certainly \( r(h)\|h\| \to 0 \). Also, by continuity of linear functions, \( f'(x)h \to 0 \) as \( h \to 0 \). Thus \( f(x + h) \to f(x) \) as \( h \to 0 \), so \( f \) is continuous at \( x \).

**Proposition 2.7.** If \( f \) and \( g \) are both differentiable at \( x \), then:

(i) \( (f + g)'(x) = f'(x) + g'(x) \);

(ii) \( (fg)'(x) = f'(x)g(x) + f(x)g'(x)h \) if \( f \) or \( g \) is real-valued;
(iii) \((cf)'(x) = cf'(x)\) if \(c \in \mathbb{R}\).

(iv) \((f \cdot g)'(x)h = (f'(x)h) \cdot g(x) + f(x) \cdot (g'(x)h)\).

Proof. (i) Write
\[
\begin{align*}
    f(x + h) &= f(x) + f'(x)h + r(h)\|h\|
    \\
g(x + h) &= g(x) + g'(x)h + s(h)\|h\|
\end{align*}
\]
with \(r(h), s(h) \to 0\) as \(h \to 0\). Then
\[
\begin{align*}
    (f + g)(x + h) &= f(x) + f'(x)h + r(h)\|h\| + g(x) + g'(x)h + s(h)\|h\|
    \\
    &= (f + g)(x) + (f'(x) + g'(x))h + (r(h) + s(h))\|h\|
\end{align*}
\]
and \(r(h) + s(h) \to 0\). By Lemma 2.5, the result follows.

(ii) Without loss of generality \(f\) is real-valued. Then
\[
\begin{align*}
    (fg)(x + h) &= (f(x) + f'(x)h + r(h)\|h\|)(g(x) + g'(x)h + s(h)\|h\|)
    \\
    &= f(x)g(x) + (f'(x)h)g(x) + f(x)(g'(x)h)
    \\
    &\quad + \frac{(f'(x)h)(g'(x)h)}{\|h\|}\|h\|
    \\
    &\quad + (f(x) + f'(x)h)s(h)\|h\| + r(h)\|h\|g(x + h).
\end{align*}
\]
We have
\[
\left\| \frac{(f'(x)h)(g'(x)h)}{\|h\|} \right\| \leq \|f'(x)\|\|g'(x)\|\|h\| \to 0.
\]
Since \(f(x) + f'(x)h \to f(x)\) and \(s(h) \to 0\),
\[
(f(x) + f'(x)h)s(h) \to 0.
\]
Since \(r(h) \to 0\) and \(g(x + h) \to g(x)\) by continuity,
\[
r(h)g(x + h) \to 0.
\]
The result now follows from Lemma 2.5.

(iii) Immediate from (ii), since the derivative of a constant function is 0.

(iv) Similar to (ii). \(\square\)

**Proposition 2.8 (Chain Rule).** If \(f\) is differentiable at \(x\) and \(g\) is differentiable at \(f(x)\), then \(g \circ f\) is differentiable at \(x\) and
\[
(g \circ f)'(x) = g'(f(x))f'(x).
\]

Proof. Put \(y = f(x)\), and write
\[
\begin{align*}
    f(x + h) &= f(x) + f'(x)h + r(h)\|h\|
    \\
g(y + k) &= g(y) + g'(y)k + s(k)\|k\|
\end{align*}
\]

where \( \lim_{h \to 0} r(h) = 0 \) and \( \lim_{k \to 0} s(k) = 0 \). Letting \( k = f(x + h) - f(x) \), we have
\[
g \circ f(x + h) = g \circ f(x) + g'(f(x))f'(x)h + g'(f(x))r(h) \frac{\|h\|}{\|h\|} + s(k) \frac{\|f'(x)h + r(h)\|}{\|h\|} \|h\|.
\]
Since \( r(h) \to 0 \), so does \( g'(f(x))r(h) \). Note that
\[
\frac{\|f'(x)h + r(h)\|}{\|h\|} \leq \|f'(x)\| + \|r(h)\| \xrightarrow{h \to 0} \|f'(x)\|.
\]
Also, \( s(k) \to 0 \) as \( h \to 0 \) since \( k \to 0 \) as \( h \to 0 \) by continuity. Hence
\[
s(k) \frac{\|f'(x)h + r(h)\|}{\|h\|} \xrightarrow{h \to 0} 0,
\]
and we are done by Lemma 2.5.

**Definition 2.9.** If \( U \subseteq \mathbb{R}^p \) is open, \( f : U \to \mathbb{R}^q \), and \( x \in U \), then for each \( i = 1, \ldots, q \) and \( j = 1, \ldots, p \) the partial derivative of \( f_i \) with respect to the \( j \)th variable is
\[
D_j f_i(x) = \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t},
\]
provided this limit exists, where \( \{e_j\}_1^p \) denotes the standard basis for \( \mathbb{R}^p \).

**Remark 2.10.** Note that if we define \( g(t) = f_i(x + te_j) \) then
\[
D_j f_i(x) = g'(0).
\]
In other words, \( D_j f_i(x) \) is the derivative at \( x_j \) of the one-variable function we get from \( f \) by holding all the other variables \( \{x_k : k \neq j\} \) constant.

**Proposition 2.11.** If \( f \) is differentiable at \( x \), then \( D_j f_i(x) \) exists for all \( i,j \), and \( f'(x) \) is represented by the matrix \([D_j f_i(x)]\).

**Proof.** Let \( A \) be the matrix representing \( f'(x) \). Then
\[
f(x + te_j) = f(x) + Ate_j + r(te_j) |t|,
\]
where \( r(h) \to 0 \) as \( h \to 0 \). Taking \( i \)th coordinates, we get
\[
f_i(x + te_j) = f_i(x) + a_{ij}t + r_i(te_j)|t|.
\]
Thus \( a_{ij} \) is the derivative of \( t \mapsto f_i(x + te_j) \) at \( 0 \), hence coincides with \( D_j f_i(x) \).

**Example 2.12.** (i) Define \( f : \mathbb{R}^2 \to \mathbb{R} \) by
\[
f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}
\]

Then \( f \) has partial derivatives everywhere, but is not continuous at \( (0,0) \).
MAT 473 LECTURES

(ii) Change the above $f$ so $f(x, y) = x^2y/(x^2 + y^2)$ when $(x, y) \neq (0, 0)$. Then $f$ has partial derivatives everywhere, and is continuous everywhere, but is not differentiable at $(0, 0)$.

**Example 2.13.** (i) If $g = (g_1, \ldots, g_p) : (a, b) \to \mathbb{R}^p$ is differentiable, then for each $t \in (a, b)$ the derivative

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

is represented by the column matrix with $i$th entry $g_i'(t)$. Identify $p \times 1$ matrices with elements of $\mathbb{R}^p$. Then $g'(t) = (g_1'(t), \ldots, g_p'(t))$, and this is called the **tangent vector** to the curve $g$ at $t$.

(ii) If $U \subseteq \mathbb{R}^p$ is open and $f: U \to \mathbb{R}$ is differentiable, then for each $x \in U$, $f'(x)$ is represented by the row matrix with $j$th entry $D_j f(x)$. Associate to this the **gradient vector** $\nabla f(x) = (D_1 f(x), \ldots, D_p f(x))$.

Then

$$f'(x) h = \nabla f(x) \cdot h = \sum_{1}^{p} D_j f(x) h_j.$$  

(iii) Now let $U \subseteq \mathbb{R}^p$ be open, and let $g: (a, b) \to U$ and $f: U \to \mathbb{R}$ both be differentiable. If $t \in (a, b)$ then

$$(f \circ g)'(t) = f'(g(t))g'(t) = \nabla f(g(t)) \cdot g'(t) = \sum_{1}^{p} D_j f(g(t)) g_j'(t).$$

Fix $x \in U$ and a unit vector $u \in \mathbb{R}^p$, and define $g: \mathbb{R} \to \mathbb{R}^p$ by $g(t) = x + tu$. The **directional derivative** is

$$D_u f(x) = \nabla f(x) \cdot u = (f \circ g)'(0) = \lim_{t \to 0} \frac{f(x + tu) - f(x)}{t},$$

measuring the rate of change of $f$ at $x$ in the direction $u$. At $x$, $f$ increases most rapidly in the direction of the gradient $\nabla f(x)$.

Also, $\nabla f(x)$ is orthogonal to the level hypersurface $S := \{y \in \mathbb{R}^p : f(y) = f(x)\}$ of $f$ through $x$, since if $g: (a, b) \to S$ is any differentiable curve in $S$ with $g(0) = x$ then $f \circ g$ is constant, hence

$$0 = (f \circ g)'(0) = \nabla f(x) \cdot g'(0),$$

and $g'(0)$ can be any vector tangent to the surface at $x$.

**Theorem 2.14 (Mean Value Theorem).** Let $U \subseteq \mathbb{R}^p$ be open and convex, and let $f: U \to \mathbb{R}$ be differentiable. Then for all $x, y \in U$, there exists $c \in (0, 1)$ such that

$$f(x) - f(y) = f'(cx + (1 - c)y)(x - y).$$

**Proof.** Fix $x, y \in U$, and define $g: [0, 1] \to U$ by

$$g(t) = tx + (1 - t)y.$$
Then $f \circ g$ is continuous on $[0, 1]$ and differentiable on $(0, 1)$. By the one-variable Mean Value Theorem, there exists $c \in (0, 1)$ such that

$$f(x) - f(y) = f(g(x)) - f(g(y)) = (f \circ g)'(c)(y - x) = f'(g(c))g'(c)(y - x).$$

\[\Box\]

**Corollary 2.15** (Mean Value Inequality). Let $U \subseteq \mathbb{R}^p$ be open and convex, and let $f \colon U \to \mathbb{R}^q$ be differentiable. Then for all $x, y \in U$,

$$\|f(x) - f(y)\| \leq \left( \sup_{U} \|f'(\cdot)\| \right) \|x - y\|.$$ 

**Proof.** Fix $x, y \in U$ and a unit vector $v \in \mathbb{R}^q$, and define $h \colon \mathbb{R}^q \to \mathbb{R}$ by

$$h(w) = v \cdot w.$$ 

By the Mean Value Theorem there exists $z$ on the line segment joining $x$ and $y$ such that

$$v \cdot (f(x) - f(y)) = h \circ f(x) - h \circ f(y) = (h \circ f)'(z)(x - y) = v \cdot f'(z)(x - y).$$

Hence

$$\| v \cdot (f(x) - f(y)) \| \leq \|v\| \|f'(z)\| \|x - y\| \leq \left( \sup_{U} \|f'(\cdot)\| \right) \|x - y\|.$$ 

Taking the sup over $v$, the desired inequality follows. \[\Box\]

**Definition 2.16.** Let $U \subseteq \mathbb{R}^p$ be open and $f \colon U \to \mathbb{R}^q$ be differentiable. $f$ is **continuously differentiable**, or $C^1$, if $f' \colon U \to L(\mathbb{R}^p, \mathbb{R}^q)$ is continuous.

**Proposition 2.17.** With the above notation, $f$ is $C^1$ if and only if every partial derivative $D_j f_i$ exists and is continuous on $U$.

**Proof.** First if $f$ is $C^1$, then the entries $D_j f_i$ of the matrix function representing $f'$ are continuous.

Conversely, assume the condition regarding the partials. We first show that $f$ is differentiable. Without loss of generality $f$ is real-valued. Let $x \in U$, and pick $r > 0$ such that $B_r(x) \subseteq U$, and then take any $h \in \mathbb{R}^p$ with $\|h\| < r$. Define points $x_0, \ldots, x_p \in B_r(x)$ by

$$x_j = \begin{cases} x & \text{if } j = 0 \\ x + \sum_{i=1}^{j} h_i e_i & \text{if } j = 1, \ldots, p. \end{cases}$$
Then
\[ f(x + h) - f(x) = \sum_{i=1}^{p} (f(x_j) - f(x_{j-1})) \]
\[ = \sum_{i=1}^{p} D_j f(x_{j-1} + t_j e_j) h_j, \]
for some \( t_j \) between 0 and \( h_j \), so
\[
\left| \frac{f(x + h) - f(x) - \nabla f(x) \cdot h}{\|h\|} \right| = \frac{1}{\|h\|} \left| \sum_{i=1}^{p} (D_j f(x_{j-1} + t_j e_j) - D_j f(x)) h_j \right|
\leq \sum_{i=1}^{p} |D_j f(x_{j-1} + t_j e_j) - D_j f(x)|
\xrightarrow{h \to 0} 0,
\]
since
\[
\|x_{j-1} + t_j e_j - x\| = \left\| \sum_{i=1}^{j-1} h_i e_i + t_j e_j \right\|
\leq \sum_{i=1}^{j-1} |h_i| + |t_j| \leq \sum_{i=1}^{p} |h_i| \leq p \|h\| \to 0.
\]
Thus \( f \) is differentiable at \( x \) and
\[
[f' (x)] = [D_1 f(x) \cdots D_p f(x)].
\]

In the general case where \( f \) is \( \mathbb{R}^m \)-valued, \( f'(x) \) is represented by the matrix \([D_j f_i(x)]\), and by hypothesis this is continuous in \( x \). Hence \( f \) is \( C^1 \). \( \blacksquare \)

**Remark 2.18.** The above result gives a sufficient, but not necessary, condition for \( f \) to be differentiable. For example, the function \( f : \mathbb{R} \to \mathbb{R} \) defined by
\[
f(x) = \begin{cases} 
x + 2x^2 \sin \frac{1}{x} & \text{if } x \neq 0 \\
0 & \text{if } x = 0
\end{cases}
\]
is differentiable on \( \mathbb{R} \), but \( f' \) is discontinuous at 0.

**Theorem 2.19** (Contraction Mapping Principle). *Let \( X \) be a nonempty complete metric space and \( f : X \to X \). If there exists \( \lambda < 1 \) such that \( d(f(x), f(y)) \leq \lambda d(x, y) \) for all \( x, y \in X \), then there exists a unique \( x \in X \) such that \( f(x) = x \).*

**Proof.** Pick \( x_0 \in X \), and inductively define \( x_n = f(x_{n-1}) \) for \( n \in \mathbb{N} \). Claim: \( (x_n) \) is Cauchy. If \( n \in \mathbb{N} \) then
\[
d(x_n, x_{n+1}) \leq \lambda d(x_{n-1}, x_n) \leq \lambda^2 d(x_{n-2}, x_{n-1}) \leq \cdots \leq \lambda^n d(x_0, x_1).
\]
Thus, if \( n, k \in \mathbb{N} \) then
\[
d(x_n, x_{n+k}) \leq \sum_{i=0}^{k-1} d(x_{n+i}, x_{n+i+1}) \leq \sum_{i=0}^{k-1} \lambda^{n+i} d(x_0, x_1)
\]
\[
= \lambda^n \left( \frac{1 - \lambda^k}{1 - \lambda} \right) d(x_0, x_1) \leq \frac{\lambda^n d(x_0, x_1)}{1 - \lambda}
\]
\[
n \to \infty, 0,
\]
proving the claim.

By completeness, there exists \( x \in X \) such that \( x_n \to x \). By continuity,
\[
f(x) = \lim f(x_n) = \lim x_{n+1} = x.
\]

For the uniqueness, if \( f(y) = y \) then
\[
d(x, y) = d(f(x), f(y)) \leq \lambda d(x, y),
\]
and since \( \lambda < 1 \) we must have \( d(x, y) = 0 \), so \( x = y \).

**Definition 2.20.** If \( U \subseteq \mathbb{R}^p \) is open, then a function \( f: U \to \mathbb{R}^p \) is a **diffeomorphism** if it is 1-1 and both \( f \) and \( f^{-1} \) are \( C^1 \).

**Remark 2.21.** A tacit part of the above definition is that the range \( f(U) \) is open. Also, by the Chain Rule, for all \( x \in U \) we have
\[
I = (\text{id}_U)'(x) = (f^{-1} \circ f)'(x) = (f^{-1})'(f(x)) f'(x).
\]

It follows from the theory of linear algebra that \( f'(x) \) is invertible and
\[
(f^{-1})'(f(x)) = f'(x)^{-1}.
\]

It is one of the miracles of calculus that continuity and invertibility of the derivative (roughly speaking) make else everything happen:

**Theorem 2.22** (Inverse Function Theorem). Let \( U \subseteq \mathbb{R}^p \) be open, \( f: U \to \mathbb{R}^p \) be \( C^1 \), and \( a \in U \). If \( f'(a) \) is invertible, then there exists an open set \( V \subseteq U \) such that \( a \in V \) and \( f \) is a diffeomorphism on \( V \).

**Proof.** Replacing \( f \) by \( f'(a)^{-1} \circ f \), without loss of generality \( f'(a) = I \). Then replacing \( f \) by \( x \mapsto f(x + a) - f(a) \), without loss of generality \( a = 0 \) and \( f(0) = 0 \).

By continuity of \( f' \), there exists \( r > 0 \) such that \( \overline{B_r(0)} \subseteq U \) and
\[
\|(f - I)'\| \leq \frac{1}{2} \quad \text{on} \quad \overline{B_r(0)}.
\]

Let \( y \in B_{r/2}(0) \), and define \( \phi: \overline{B_r(0)} \to \mathbb{R}^p \) by
\[
\phi(x) = y + x - f(x).
\]

If \( x \in \overline{B_r(0)} \) then
\[
\|\phi(x)\| \leq \|y\| + \|x - f(x)\| = \|y\| + \|(I - f)(x) - (I - f)(0)\|
\]
\[
\leq \frac{r}{2} + \frac{1}{2}\|x\| \leq \frac{r}{2} + \frac{r}{2} = r,
\]
MAT 473 LECTURES

where the inequality at * follows from continuity and the Mean Value
Inequality, so \( \phi(B_r(0)) \subseteq B_r(0) \).

Also, for all \( x, z \in \overline{B_r(0)} \),
\[
\| \phi(x) - \phi(z) \| = \| (x - z) - (f(x) - f(z)) \| \\
= \| (I - f)(x) - (I - f)(z) \| \leq \frac{1}{2} \| x - z \|.
\]

Since \( \overline{B_r(0)} \) is complete, by the Contraction Mapping Principle there exists
a unique \( z \in \overline{B_r(0)} \) such that \( \phi(z) = x \), and then \( f(x) = y \).

Moreover, the above computation also shows that if \( z, w \in \overline{B_r(0)} \) then
\[
\| f(z) - f(w) \| = \| (z - w) - (\phi(z) - \phi(w)) \| \\
\geq \| z - w \| - \| \phi(z) - \phi(w) \| \\
\geq \| z - w \| - \frac{1}{2} \| z - w \| = \frac{1}{2} \| z - w \|.
\]

In particular, if \( \| x \| = r \) then
\[
\| f(x) \| = \| f(x) - f(0) \| \geq \frac{1}{2} \| x \| = \frac{r}{2},
\]
so \( f(x) \notin B_{r/2}(0) \). Consequently, for all \( y \in B_{r/2}(0) \) there exists a unique
\( x \in \overline{B_r(0)} \) such that \( f(x) = y \). Put
\[
V = B_r(0) \cap f^{-1}(B_{r/2}(0)).
\]

Then \( f \) is 1-1 on \( V \) and \( f(V) = B_{r/2}(0) \) is open.

Note that if \( x \in V \) then \( \| f'(x) - I \| \leq 1/2 < 1 \), so \( f'(x) \) is invertible. Thus, reasoning similar to the above shows that for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that \( f(V \cap B_{\epsilon}(x)) \supseteq B_{\delta}(f(x)) \). Thus \( f^{-1}: f(V) \rightarrow V \) is continuous.

Now let \( y, y + k \in f(V) \), and put
\[
x = f^{-1}(y) \text{ and } h = f^{-1}(y + k) - f^{-1}(y).
\]

Then \( x, x + h \in V \), and \( f(x + h) = y + k \). Also put \( A = f'(x) \). Then \( A \) is
invertible, and we have
\[
\| f^{-1}(y + k) - f^{-1}(y) - A^{-1}k \| = \| h - A^{-1}k \| = \| A^{-1}(Ah - k) \| \\
\leq \| A^{-1} \| \| Ah - k \| \\
= \| A^{-1} \| \| f(x + h) - f(x) - Ah \|.
\]

also,
\[
\| k \| = \| f(x + h) - f(x) \| \geq \frac{1}{2} \| h \|,
\]
so
\[
\frac{\| h \|}{\| k \|} \leq 2.
\]
As $k \to 0$, we have $h \to 0$ by continuity of $f^{-1}$, hence
\[
\frac{\| f^{-1}(y + k) - f^{-1}(y) - A^{-1} k \|}{\| k \|} \leq \frac{\| A^{-1} \| \| f(x + h) - f(x) - Ah \|}{\| k \|} = \frac{\| A^{-1} \| \| f(x + h) - f(x) - Ah \|}{\| h \|} \leq 2 \| A^{-1} \| \frac{\| f(x + h) - f(x) - Ah \|}{\| h \|}
\]
\[
k \to 0 \implies (f^{-1})'(y) = A^{-1} = f'(f^{-1}(y))^{-1}.
\]
Thus $(f^{-1})'$ is continuous on $f(V)$ since $f^{-1}$, $f'$, and the inversion map $A \mapsto A^{-1}$ are continuous.

**Remark 2.23.** It follows from the Inverse Function Theorem that if $f$ is $C^1$ and $f'(x)$ is invertible for all $x$ then ran $f$ is open. However, $f$ need not be 1-1; the Inverse Function Theorem only tells us $f$ is locally 1-1.

**Notation and Terminology 2.24.** For $p, q \in \mathbb{N}$, it is often convenient to abuse notation by identifying $\mathbb{R}^p \times \mathbb{R}^q$ with $\mathbb{R}^{p+q}$, so that if $x = (x_1, \ldots, x_p)$ and $y = (y_1, \ldots, y_q)$ then
\[
(x, y) = (x_1, \ldots, x_p, y_1, \ldots, y_q).
\]
Similarly, if $f$ is a $C^1$ function defined on an open subset of $\mathbb{R}^{p+q}$, it is sometimes convenient to write
\[
f'(x, y) = \begin{bmatrix} D_1 f(x, y) & D_2 f(x, y) \end{bmatrix},
\]
where $D_1 f$ means the derivative of $f$ with respect to the first $p$ coordinates and $D_2 f$ the derivative with respect to the last $q$ coordinates.

**Remark 2.25.** If $f$ is an $\mathbb{R}^q$-valued function defined on a subset of $\mathbb{R}^{p+q}$, then for $c \in \mathbb{R}^q$ the equation $f(x, y) = c$ is really a system of $q$ simultaneous equations in $p + q$ unknowns. If $p = 0$ the Inverse Function Theorem gives a sufficient condition for the equation to have a unique solution. More generally, the following theorem, certainly one of the most important in multivariable calculus, gives a sufficient condition for us to solve the equation for $y$ as a function of $x$:

**Theorem 2.26 (Implicit Function Theorem).** Let $U \subseteq \mathbb{R}^{p+q}$ be open, $f: U \to \mathbb{R}^q$ be $C^1$, and $(a, b) \in U$. Suppose $f(a, b) = 0$ and $D_2 f(a, b)$ is invertible. Then there exists an open set $V \subseteq U$ such that $(a, b) \in V$ and $V \cap f^{-1}(\{0\})$ is the graph of a $C^1$ function $g$ defined on some open subset of $\mathbb{R}^p$. Moreover,
\[
g'(a) = -D_2 f(a, b)^{-1} D_1 f(a, b).
\]
MAT 473 LECTURES

Proof. More precisely, we must show that there exist open sets $V \subseteq U$ and $W \subseteq \mathbb{R}^p$ and a $C^1$ function $g: W \to \mathbb{R}^q$ such that if $(x, y) \in V$ then $f(x, y) = 0$ if and only if $x \in W$ and $y = g(x)$.

Define $\phi: U \to \mathbb{R}^{p+q}$ by

$$\phi(x, y) = (x, f(x, y)).$$

Then $\phi$ is $C^1$ and

$$\phi'(a, b) = \begin{bmatrix} I & 0 \\ D_1 f(a, b) & D_2 f(a, b) \end{bmatrix}$$

is invertible. By the Inverse Function Theorem, there exists an open set $V \subseteq U$ such that $(a, b) \in V$ and $\phi$ is a diffeomorphism on $V$.

Note that for all $(s, t) \in \phi(V)$, $\phi^{-1}(s, t)$ is the unique $(x, y) \in V$ such that $(s, t) = \phi(x, y) = (x, f(x, y))$.

Thus $s = x$, and

$$\phi^{-1}(x, t) = (x, \psi(x, t))$$

for a unique function $\psi: \phi(V) \to \mathbb{R}^q$, and moreover $\psi$ is $C^1$.

Now put

$$W = \{ x \in \mathbb{R}^p : (x, 0) \in \phi(V) \}.$$  

Then $W$ is open and $a \in W$. Define $g: W \to \mathbb{R}^q$ by

$$g(x) = \psi(x, 0).$$

Then $g$ is $C^1$.

Claim: $V \cap f^{-1}(\{0\})$ coincides with the graph of $g$. To see this, first suppose $(x, y) \in V$ and $f(x, y) = 0$. Then $\phi(x, y) = (x, 0)$. Hence $x \in W$ and $(x, y) = \phi^{-1}(x, 0)$, so that $y = \psi(x, 0) = g(x)$. Conversely, we can just reverse these steps to argue that if $x \in W$ and $y = g(x)$, then $(x, y) = (x, \psi(x, 0)) = \phi^{-1}(x, 0)$, hence $(x, y) \in V$ and $(x, 0) = \phi(x, y)$, thus $f(x, y) = 0$.

Finally, for the derivative formula, it is convenient to introduce one more auxiliary function, namely

$$\eta(x) = (x, g(x)) \quad \text{for } x \in W.$$  

If $h \in \mathbb{R}^p$ then

$$0 = (f \circ \eta)'(a)h = f'(\eta(a))\eta'(a)h$$

$$= f'(a, b)(h, g'(a)h) = D_1 f(a, b)h + D_2 f(a, b)g'(a)h,$$

so $g'(a)h = -D_2 f(a, b)^{-1}D_1 f(a, b)h$. Since $h \in \mathbb{R}^p$ was arbitrary, we must have

$$g'(a) = -D_2 f(a, b)^{-1}D_1 f(a, b).$$

$\square$
Definition 2.27. Let \( U \subseteq \mathbb{R}^p \) be open and \( f: U \to \mathbb{R} \). For each \((i_1, \ldots, i_n) \in \{1, \ldots, p\}^n\) define
\[
D_{i_1 \cdots i_n} f = D_{i_1} \cdots D_{i_n} f.
\]
The \( D_{i_1 \cdots i_n} f \) are the \( n \)th-order partial derivatives of \( f \). We say \( f \) is \( C^n \) if every \( D_{i_1 \cdots i_n} f \) is continuous.

Proposition 2.28. With the above notation, if \( f \) is \( C^n \), then:
(i) \( f \) is \( C^k \) for all \( k < n \);
(ii) (Clairaut’s Theorem) \( D_{j_1 \cdots j_n} f = D_{i_1 \cdots i_n} f \) for every rearrangement \((j_1, \ldots, j_n) \) of \((i_1, \ldots, i_n)\).

Proof. (i) Let \( n > 1 \). For all \( i_1, \ldots, i_n = 1, \ldots, p \),
\[
D_{i_1}D_{i_2 \cdots i_n} f = D_{i_1 \cdots i_n} f
\]
is continuous. Thus \( D_{i_2 \cdots i_n} f \) is differentiable, hence continuous. Therefore, \( f \) is \( C^{n-1} \). Continue inductively.

(ii) Since every rearrangement of \((i_1, \ldots, i_n)\) can be obtained by switching pairs of coordinates finitely many times, and since partial derivatives are computed by holding the other coordinates constant, without loss of generality \( p = n = 2 \). Fix \((a, b) \in U \). Choose open intervals \( I \) and \( J \) such that
\[
(a, b) \in I \times J \subseteq U,
\]
and define \( \phi: I \times J \to \mathbb{R} \) by
\[
\phi(x, y) = f(x, y) - f(a, y) - f(x, b) + f(a, b) = \psi(y) - \psi(b),
\]
where \( \psi: J \to \mathbb{R} \) is defined by
\[
\psi(y) = f(x, y) - f(a, y).
\]
By the one-variable Mean Value Theorem, for all \((x, y) \in I \times J \) there exists \( t \) between \( b \) and \( y \) such that
\[
\phi(x, y) = \psi'(t)(y - b) = (D_2 f(x, t) - D_2 f(a, t))(y - b),
\]
and then for the same reason there exists \( s \) between \( a \) and \( x \) such that
\[
\phi(x, y) = D_1 D_2 f(s, t)(x - a)(y - b).
\]
As \((x, y) \to (a, b)\), so does \((s, t)\), hence
\[
\frac{\phi(x, y)}{(x - a)(y - b)} = D_{12} f(s, t) \to D_{12} f(a, b)
\]
by continuity. By symmetry we also have
\[
\frac{\phi(x, y)}{(x - a)(y - b)} \to D_{21} f(a, b),
\]
so we are done. \( \square \)
Theorem 2.29 (Taylor’s Theorem). Let $U \subseteq \mathbb{R}^p$ be open and convex, and let $f: U \to \mathbb{R}$ be $C^n$. Then for all $a, h$ such that $a, a + h \in U$, there exists $c$ on the line segment joining $a$ and $a + h$ such that

$$f(a + h) = f(a) + \sum_{k=1}^{n-1} \frac{1}{k!} \sum_{i_1, \ldots, i_k} D_{i_1 \cdots i_k} f(a) h_{i_1} \cdots h_{i_k} + \frac{1}{n!} \sum_{i_1, \ldots, i_n} D_{i_1 \cdots i_n} f(c) h_{i_1} \cdots h_{i_n}.$$  

Proof. Define $g: \mathbb{R} \to \mathbb{R}^p$ by $g(t) = a + th$, and put $V = g^{-1}(U)$, an open set containing $[0, 1]$. Then $f \circ g: V \to \mathbb{R}$ is $C^n$. By the one-variable Taylor Theorem, there exists $s \in (0, 1)$ such that

$$f(a + h) - f(a) = f \circ g(1) - f \circ g(0) = \sum_{k=1}^{n-1} \frac{(f \circ g)^{(k)}(0)}{k!} + \frac{(f \circ g)^{(n)}(s)}{n!}.$$  

Thus, it suffices to show that if $1 \leq k \leq n$ and $t \in V$, then

$$(f \circ g)^{(k)}(t) = \sum_{i_1, \ldots, i_k} D_{i_1 \cdots i_k} f(g(t)) h_{i_1} \cdots h_{i_k}.$$  

The equality holds for $k = 1$ by the Chain Rule. Let $1 \leq k < n$, and assume the equality holds for $k$. Then

$$(f \circ g)^{(k+1)}(t) = \sum_{i_1, \ldots, i_k} ((D_{i_1 \cdots i_k} f) \circ g)'(t) h_{i_1} \cdots h_{i_k}$$  

$$= \sum_{i_1, \ldots, i_k} (D_{i_1 \cdots i_k} f)'(g(t)) g'(t) h_{i_1} \cdots h_{i_k}$$  

$$= \sum_{i_1, \ldots, i_k} \sum_{j} D_j D_{i_1 \cdots i_k} f(g(t)) g_j(t) h_{i_1} \cdots h_{i_k}$$  

$$= \sum_{i_1, \ldots, i_k} \sum_{j} D_{j i_1 \cdots i_k} f(g(t)) h_j h_{i_1} \cdots h_{i_k}$$  

$$= \sum_{i_1, \ldots, i_{k+1}} D_{i_1 \cdots i_{k+1}} f(g(t)) h_{i_1} \cdots h_{i_{k+1}}.$$  

3. Lebesgue Measure

Definition 3.1. (i) A ring of sets is a nonempty family of sets which is closed under differences and finite unions.

(ii) A $\sigma$-ring of sets is a nonempty family of sets which is closed under differences and countable unions.

(iii) An algebra of sets is a nonempty family of sets which is closed under complements and finite unions.
(iv) An $\sigma$-algebra of sets is a nonempty family of sets which is closed under complements and countable unions.

(v) A function $\mu$ from a ring $\mathcal{R}$ to $\mathbb{R}$ (the extended real numbers) is
\begin{itemize}
\item[(finitely additive)] if for all disjoint $A, B \in \mathcal{R}$, $\mu(A \cup B) = \mu(A) + \mu(B)$.
\end{itemize}

(vi) A function $\mu$ from a $\sigma$-ring $\mathcal{R}$ to $\mathbb{R}$ is countably additive if for every (pairwise) disjoint sequence $\{A_n\}_{n=1}^{\infty}$ in $\mathcal{R}$,
\[ \mu\left( \bigcup_{n} A_n \right) = \sum_{n} \mu(A_n). \]

(vii) A measure on a $\sigma$-algebra $\mathcal{A}$ is a countably additive function $\mu : \mathcal{A} \to \mathbb{R}$ such that:
\begin{itemize}
\item[(a)] $\mu(\emptyset) = 0$, and
\item[(b)] $\mu(A) \geq 0$ for all $A \in \mathcal{A}$.
\end{itemize}

**Remark 3.2.** (i) In (v)-(vi) above we must assume $\infty - \infty$ does not occur.

(ii) In (vi) the value of the series $\sum \mu(A_n)$ must be independent of the order of summation, since the union $\bigcup_{n} A_n$ is. In particular, if $\mu(\bigcup_{n} A_n) \in \mathbb{R}$ we must require the series $\sum \mu(A_n)$ to converge absolutely.

(iii) If $\mathcal{R}$ is a ring of sets then $\emptyset \in \mathcal{R}$, and in fact this is usually what is proved when verifying that a proposed ring is nonempty.

**Lemma 3.3.** (i) Every ring is closed under finite intersections, and every $\sigma$-ring is closed under countable intersections.

(ii) Every algebra is a ring, and every $\sigma$-algebra is a $\sigma$-ring.

**Proof.** (i) Let $\{A_n\}_1^k$ be in some ring $\mathcal{R}$, and put $B = \bigcup_1^k A_n$. Then
\[ \bigcap_1^k A_n = B \setminus \bigcup_1^k (B \setminus A_n) \in \mathcal{R}. \]

Similarly for $\sigma$-rings and countable intersections.

(ii) Since $A \setminus B = A \cap B^c$, every algebra is closed under differences. \qed

**Proposition 3.4.** Let $\mu$ be a measure on a $\sigma$-algebra $\mathcal{A}$, and let $\{A_n\}_1^{\infty} \subseteq \mathcal{A}$.

(i) (Continuity from Below) If $A_1 \subseteq A_2 \subseteq \cdots$, then
\[ \mu(A_n) \to \mu\left( \bigcup_{n} A_n \right). \]

(ii) (Continuity from Above) If $A_1 \supseteq A_2 \supseteq \cdots$ and $\mu(A_1) < \infty$, then
\[ \mu(A_n) \to \mu\left( \bigcap_{n} A_n \right). \]
Proof. (i) Put \( B_n = A_n \setminus \bigcup_{k<n} A_k \). Then \( B_1, B_2, \ldots \in A \) are disjoint, \( A_n = \bigcup_1^n B_k \), and \( \bigcup_1^\infty A_n = \bigcup_1^\infty B_n \). Thus
\[
\mu(A_n) = \mu\left( \bigcup_1^n B_k \right) = \sum_1^n \mu(B_k) \\
\rightarrow \sum_1^\infty \mu(B_n) = \mu\left( \bigcup_1^\infty B_n \right) = \mu\left( \bigcup_1^\infty A_n \right).
\]
(ii) Put \( B_n = A_1 \setminus A_n \). Then \( B_1 \subseteq B_2 \subseteq \cdots \subseteq A_1 \) and \( \bigcap_1^n B_n = A_1 \setminus \bigcap_1^n A_n \), so
\[
\mu(A_n) = \mu(A_1 \setminus B_n) = \mu(A_1) - \mu(B_n) \quad \text{since} \quad \mu(A_1) < \infty
\]
\[
\rightarrow \mu(A_1) - \mu\left( \bigcup_1^n B_n \right) = \mu\left( A_1 \setminus \bigcup_1^n B_n \right) = \mu\left( \bigcap_1^n A_n \right).
\]

\( \square \)

Definition 3.5. A box in \( \mathbb{R}^p \) is a product \( \prod_{j=1}^p I_j \), where each \( I_j \) is a bounded interval (possibly a single point or empty).

Proposition 3.6. For all boxes \( A, B \),

(i) \( A \cap B \) is a box;
(ii) \( A \setminus B \) is a finite disjoint union of boxes;
(iii) \( A \cup B \) is a finite disjoint union of boxes.

Proof. Let \( A = \prod_i^p C_i \) and \( B = \prod_i^p D_i \) for intervals \( C_i \) and \( D_i \).

(i) We have
\[
A \cap B = \prod_1^p (C_i \cap D_i),
\]
and every intersection of intervals is an interval.

(ii) Without loss of generality \( B \subseteq A \). For each \( i \), \( C_i \) contains the interval \( D_i \), hence is a union of disjoint intervals \( \{D_i^k : k = 1, 2, 3\} \), with \( D_i^1 = D_i \). We have
\[
A = \prod_{i=1}^p \bigcup_{k=1}^3 D_i^k = \bigcup_{k_1, \ldots, k_p=1}^3 \prod_{i=1}^p D_i^{k_i},
\]
and the products \( \prod_{i=1}^p D_i^{k_i} \) are disjoint boxes. Since \( B \) is one of these latter boxes, \( A \setminus B \) is a disjoint union of the remaining boxes.

(iii) We have
\[
A \cup B = (A \cap B) \cup (A \setminus B) \cup (B \setminus A).
\]
The sets \( A \cap B \), \( A \setminus B \), and \( B \setminus A \) are disjoint, and each is a finite disjoint union of boxes, hence \( A \cup B \) is also a finite disjoint union of boxes. \( \square \)
Definition 3.7. For each box $A = \Pi^k I_j$, define
\[ |A| = \Pi^k (b_j - a_j) \]
if $I_j$ has endpoints $a_j \leq b_j$.

Lemma 3.8. For all disjoint boxes $B_1, \ldots, B_k$, if $\bigcup_1^k B_n$ is a box then
\[ \left| \bigcup_1^k B_n \right| = \sum_1^k |B_n|. \]

Proof. This is remarkably fussy, and we only indicate the outline of an argument. Put $A = \bigcup_1^k B_n$, and replace $A, B_1, \ldots, B_k$ by their closures. The $B_n$ are no longer disjoint, but the values of $|.|$ remain the same, and $|B_n \cap B_j| = 0$ whenever $n \neq j$. Subdivide $A$ into a grid of closed boxes $C_l$ such that each $B_n$ is a union of $C_l$. If $C_l \subseteq B_n$ and $C_i \subseteq B_j$ for distinct $n$ and $j$, then $|C_i \cap C_l| = 0$. Also, if $B_n = \bigcup_{l \in L_n} C_l$, then
\[ |B_n| = \sum_{l \in L_n} |C_l|, \]
as can be seen by a tedious algebraic manipulation. For the same reason,
\[ |A| = \sum_{n=1}^k |C_l| = \sum_{n=1}^k \sum_{l \in L_n} |C_l| = \sum_{n=1}^k |B_n|. \]

Definition 3.9. $\mathcal{E}$ denotes the family of all finite disjoint unions of boxes.

Proposition 3.10. $\mathcal{E}$ is a ring.

Proof. $\mathcal{E}$ is nonempty since it contains every box. Next, if $A$ and $B$ are boxes then $A \cup B$ is a finite disjoint union of boxes. Hence $\mathcal{E}$ coincides with the family of all finite unions of boxes. In particular, $\mathcal{E}$ is closed under finite unions.

Finally, given boxes $A_1, \ldots, A_k, B_1, \ldots, B_l$,
\[ \left( \bigcup_1^k A_n \right) \setminus \left( \bigcup_1^l B_j \right) = \bigcup_{n \neq j} A_n \setminus B_j \]
is a finite union of boxes since each $\bigcap_{j \neq n} A_n \setminus B_j$ is. Strictly speaking, we should verify the last statement; the general fact claimed here is that $\mathcal{E}$ is closed under finite intersections. But this follows from rewriting a finite intersection of finite unions of boxes as a finite union of finite intersections of boxes. 

Definition 3.11. Given disjoint boxes $A_1, \ldots, A_k$, define
\[ \left| \bigcup_1^k A_n \right| = \sum_1^k |A_n|. \]
Proposition 3.12. (i) \(|\cdot|: \mathcal{E} \rightarrow \mathbb{R}\) is well-defined and finitely additive.
(ii) For all \(A, B \in \mathcal{E}\), if \(A \subseteq B\) then \(|A| \leq |B|\).
(iii) If \(\{A_n\}_1^k \subseteq \mathcal{E}\) then \(\bigcup_1^k A_n \leq \sum_1^k |A_n|\).

Proof. (i) Let \(\{A_n\}_1^k\) and \(\{B_j\}_1^l\) be families of disjoint boxes such that \(\bigcup_1^k A_n = \bigcup_1^l B_j\). By Proposition 3.8

\[
\sum_1^k |A_n| = \sum_n \sum_j |A_n \cap B_j| = \sum_j \sum_n |A_n \cap B_j| = \sum_j |B_j|.
\]

Next let \(A_1, \ldots, A_k \in \mathcal{E}\) be disjoint, and for each \(n\) let \(A_n\) be the union of disjoint boxes \(B_{1n}, \ldots, B_{ln}\). Then

\[
\left| \bigcup_1^k A_n \right| = \left| \bigcup_1^k \bigcup_1^l B_{jn} \right| = \sum_1^k \sum_1^l |B_{jn}| = \sum_1^k |A_n|.
\]

(ii) \(B = A \cup (B \setminus A)\), and the elements \(A\) and \(B \setminus A\) of \(\mathcal{E}\) are disjoint, so

\[
|B| = |A| + |B \setminus A| \geq |A|,
\]

since \(|B \setminus A| \geq 0\).

(iii) Put \(B_n = A_n \setminus \bigcup_{k<n} A_k\). Then \(B_1, \ldots, B_k \in \mathcal{E}\) are disjoint, and \(\bigcup_1^k A_n = \bigcup_1^k B_n\), so

\[
\left| \bigcup_1^k A_n \right| = \left| \bigcup_1^k B_n \right| = \sum_1^k |B_n| \leq \sum_1^k |A_n|,
\]

since \(B_n \subseteq A_n\) for all \(n\).

Definition 3.13. The outer measure of \(A \subseteq \mathbb{R}^p\) is

\[
m^*(A) := \inf \left\{ \sum_1^\infty |B_n| : A \subseteq \bigcup_1^\infty B_n \text{ and each } B_n \text{ is an open box} \right\}.
\]

Proposition 3.14. (i) \(m^*(\emptyset) = 0\).
(ii) \(A \subseteq B\) implies \(m^*(A) \leq m^*(B)\).
(iii) (Countable Subadditivity) \(m^*(\bigcup_1^\infty A_n) \leq \sum_1^\infty m^*(A_n)\).

Proof. (i) The empty set \(\emptyset\) is an open box, for example \(\emptyset = \prod_1^\infty (0, 0)\), so

\[
0 \leq m^*(\emptyset) \leq |\emptyset| = 0.
\]

(ii) Take open boxes \(C_1, C_2, \ldots\) such that \(B \subseteq \bigcup_1^n C_n\). Then \(A \subseteq \bigcup_1^n C_n\), so

\[
m^*(A) \leq \sum_1^n |C_n|.
\]

Taking the infimum on the right side, we get \(m^*(A) \leq m^*(B)\).

(iii) Given \(\epsilon > 0\), for each \(n \in \mathbb{N}\) choose open boxes \(B_{1n}, B_{2n}, \ldots\) such that \(A_n \subseteq \bigcup_j B_{jn}\) and

\[
\sum_j |B_{jn}| \leq m^*(A_n) + 2^{-n}\epsilon.
\]
Then $\bigcup_n A_n \subseteq \bigcup_j B^n_j$, so

$$m^*\left(\bigcup_n A_n\right) \leq \sum_n \sum_j |B^n_j| \leq \sum_n (m^*(A_n) + 2^{-n}\varepsilon) = \sum_n m^*(A_n) + \varepsilon.$$ 

Letting $\varepsilon \to 0$, we get $m^*(\bigcup_n A_n) \leq \sum_n m^*(A_n)$. \hfill \Box

**Definition 3.15.** $A \subseteq \mathbb{R}^p$ is measurable if for all $B \subseteq \mathbb{R}^p$,

$$m^*(B) = m^*(B \cap A) + m^*(B \cap A^c).$$

$\mathcal{M}$ denotes the family of measurable subsets of $\mathbb{R}^p$.

**Remark 3.16.** To show measurability of $A$ it suffices to show that if $m^*(B) < \infty$ then

$$m^*(B) \geq m^*(B \cap A) + m^*(B \cap A^c).$$

**Proposition 3.17.** If $m^*(A) = 0$, then $A \in \mathcal{M}$.

**Proof.** If $B \subseteq \mathbb{R}^p$ then

$$m^*(B \cap A) + m^*(B \setminus A) \leq m^*(B \setminus A) \leq m^*(B).$$

\hfill \Box

**Theorem 3.18** (Carathéodory’s Theorem). With the above notation, $\mathcal{M}$ is a $\sigma$-algebra and the restriction $m := m^*|\mathcal{M}$ is a measure.

**Proof.** First, if $B \subseteq \mathbb{R}^p$ then

$$m^*(B \cap \emptyset) + m^*(B \setminus \emptyset) = m^*(B),$$

so $\emptyset \in \mathcal{M}$.

Next, if $A \in \mathcal{M}$ then $A^c \in \mathcal{M}$ since $A^{cc} = A$.

Now, if $A, B \in \mathcal{M}$ and $C \subseteq \mathbb{R}^p$ then

$$m^*(C) = m^*(C \cap A) + m^*(C \cap A^c)$$

$$= m^*(C \cap A \cap B) + m^*(C \cap A \cap B^c)$$

$$+ m^*(C \cap A^c \cap B) + m^*(C \cap A^c \cap B^c)$$

$$\geq m^*(C \cap (A \cup B)) + m^*(C \setminus (A \cup B)),$$

so $A \cup B \in \mathcal{M}$. Thus $\mathcal{M}$ is an algebra.

Note: an induction argument shows that if $A_1, \ldots, A_k \in \mathcal{M}$ are disjoint and $B \subseteq \mathbb{R}^p$ then

$$m^*\left(B \cap \bigcup_{1}^{k} A_n\right) = \sum_{1}^{k} m^*(B \cap A_n).$$

To show that $\mathcal{M}$ is closed under countable unions, let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{M}$, and without loss of generality the $A_n$’s are disjoint (because $\mathcal{M}$ is an algebra).
Let \( B \subseteq \mathbb{R}^p \). For all \( k \in \mathbb{N} \),

\[
m^*(B) \geq m^* \left( B \cap \bigcup_{1}^{k} A_n \right) + m^* \left( B \setminus \bigcup_{1}^{k} A_n \right) \\
\geq \sum_{1}^{k} m^* \left( B \cap A_n \right) + m^* \left( B \setminus \bigcup_{1}^{k} A_n \right).
\]

Letting \( k \to \infty \),

\[
m^*(B) \geq \sum_{1}^{\infty} m^* \left( B \cap A_n \right) + m^* \left( B \setminus \bigcup_{1}^{\infty} A_n \right) \\
\geq m^* \left( \bigcup_{1}^{\infty} (B \cap A_n) \right) + m^* \left( B \setminus \bigcup_{1}^{\infty} A_n \right) \\
= m^* \left( B \cap \bigcup_{1}^{\infty} A_n \right) + m^* \left( B \setminus \bigcup_{1}^{\infty} A_n \right).
\]

Thus \( \bigcup_{n} A_n \in \mathcal{M} \), so \( \mathcal{M} \) is a \( \sigma \)-algebra.

The above argument with \( B = \bigcup_{n} A_n \) shows that

\[
m \left( \bigcup_{n} A_n \right) = \sum_{n} m(A_n),
\]

so \( m \) is countably additive. Finally, \( m \) is nonnegative since \( m^* \) is, and \( m(\emptyset) = m^*(\emptyset) = 0 \). \( \square \)

**Definition 3.19.** \( m \) is Lebesgue measure.

**Theorem 3.20.** \( \mathcal{E} \subseteq \mathcal{M} \), and \( m(A) = |A| \) for all \( A \in \mathcal{E} \).

**Proof.** First we show that \( m^* \leq |\cdot| \) on \( \mathcal{E} \). If \( A \) is a box, then

\[
|A| = \inf \{|B| : B \text{ is an open box, } A \subseteq B\},
\]

so \( m^*(A) \leq |A| \). If \( A_1, \ldots, A_k \) are disjoint boxes, then

\[
m^* \left( \bigcup_{1}^{k} A_n \right) \leq \sum_{1}^{k} m^*(A_n) \leq \sum_{1}^{k} |A_n| = |A|.
\]

Now let \( A \in \mathcal{E} \) and \( B \subseteq \mathbb{R}^p \). For any open boxes \( C_1, C_2, \ldots \) such that \( B \subseteq \bigcup_{n} C_n \),

\[
\sum |C_n| = \sum (|C_n \cap A| + |C_n \setminus A|) \\
\geq \sum (m^*(C_n \cap A) + m^*(C_n \setminus A)) \\
= \sum m^*(C_n \cap A) + \sum m^*(C_n \setminus A) \\
\geq m^* \left( \bigcup C_n \cap A \right) + m^* \left( \bigcup C_n \setminus A \right) \\
\geq m^* (B \cap A) + m^* (B \setminus A).
\]
Hence $m^*(B) \geq m^*(B \cap A) + m^*(B \setminus A)$. Thus $A \in \mathcal{M}$.

It remains to show that $|\cdot| \leq m$ on $\mathcal{E}$. First, if $A$ is a closed box and $B_1, B_2, \ldots$ are open boxes such that $A \subseteq \bigcup B_n$, then by compactness there exists $k \in \mathbb{N}$ such that $A \subseteq \bigcup_{1}^{k} B_n$, and then

$$|A| \leq \sum_{1}^{k} |B_n| \leq \sum_{1}^{\infty} |B_n|.$$ 

Thus $|A| \leq m(A)$. Now if $A$ is any box, then $\bar{A} = A \cup B$ where $B$ is a finite union of boxes $C_n$ with $|C_n| = m(C_n) = 0$, hence $|B| = m(B) = 0$. Thus

$$|A| \leq |\bar{A}| = m(\bar{A}) \leq m(A) + m(B) = m(A).$$

Finally, if $A_1, \ldots, A_k$ are disjoint boxes, then

$$\left| \bigcup_{1}^{k} A_n \right| = \sum_{1}^{k} |A_n| = \sum_{1}^{k} m(A_n) = m \left( \bigcup_{1}^{k} A_n \right).$$

\[ \square \]

**Definition 3.21.** $B$ denotes the intersection of all $\sigma$-algebras containing the family of open subsets of $\mathbb{R}^p$. A *Borel set* is a member of $B$.

**Definition 3.22.** A $G_\delta$ is a countable intersection of open sets, and an $F_\sigma$ is a countable union of closed sets.

**Theorem 3.23.** (i) $B$ is a $\sigma$-algebra contained in $\mathcal{M}$.

(ii) For all $A \in \mathcal{M}$ and $\epsilon > 0$ there exist an open $B \supseteq A$ and a closed $C \subseteq A$ such that

$$m(B \setminus A), m(A \setminus C) < \epsilon.$$ 

(iii) For all $A \in \mathcal{M}$ there exist a $G_\delta$ $B \supseteq A$ and an $F_\sigma$ $C \subseteq A$ such that

$$m(B \setminus A) = m(A \setminus C) = 0.$$ 

In particular, $A$ is a union of a Borel set and a set of measure 0.

(iv) Every set of measure 0 is contained in a Borel set of measure 0.

(v) The family of sets of measure 0 is a $\sigma$-ring containing every countable set.

**Proof.** (i) Let $U \subseteq \mathbb{R}^p$ be open. For all $x \in U$ there exists $r > 0$ such that $B_r(x) \subseteq U$. Pick $s \in \mathbb{Q}$ such that $s < r / \sqrt{p}$. Then the open cube centered at $x$ with side $2s$, that is, the set

$$\left\{ y \in \mathbb{R}^p : \max |x_i - y_i| < \frac{s}{2} \right\},$$

is contained in $U$. By density there exists $y \in \mathbb{Q}^p$ which is contained in the open cube centered at $x$ with side $s$, and then $x$ is in the open cube centered at $y$ with side $s$, which in turn is contained in $U$. Thus, $U$ is a union of open cubes with rational sides and centers. Since there are only countably many of these cubes, $U$ is a countable union of cubes, hence is measurable.
MAT 473 LECTURES

To see that $\mathcal{B}$ is a $\sigma$-algebra, first note that $\emptyset \in \mathcal{B}$ since $\emptyset \in \mathcal{A}$ for every $\sigma$-algebra $\mathcal{A}$ containing the open sets. Next, if $A \in \mathcal{B}$ then $A^c \in \mathcal{A}$ for every $\sigma$-algebra $\mathcal{A}$ containing the open sets, so $A^c \in \mathcal{B}$. Finally, if $\{A_n\}_{i=1}^{\infty} \subseteq \mathcal{B}$ and $\mathcal{A}$ is any $\sigma$-algebra containing the open sets, then $A_n \subseteq \mathcal{A}$ for all $n$, so $\bigcup_{n} A_n \in \mathcal{A}$. Hence $\bigcup_{n} A_n \in \mathcal{B}$.

(ii) For each $n \in \mathbb{N}$ put

$$A_n = A \cap B_n(0).$$

Then the $A_n$’s are bounded and $A = \bigcup A_n$. For each $n$ choose open boxes $B_1^n, B_2^n, \ldots$ such that $A_n \subseteq \bigcup_j B_j^n$ and

$$\sum_j m(B_j^n) < m(A_n) + 2^{-n} \epsilon.$$

Then $U_n := \bigcup_j B_j^n$ is open and

$$m(U_n) \leq \sum_j m(B_j^n) < m(A_n) + 2^{-n} \epsilon.$$

Since $m(A_n) < \infty$,

$$m(U_n \setminus A_n) < 2^{-n} \epsilon.$$

Then $U := \bigcap_n U_n$ is open, $A \subseteq U$, and

$$m(U \setminus A) \leq m \left( \bigcup_n (U_n \setminus A_n) \right) \leq \sum_n m(U_n \setminus A_n) < \sum_{i=1}^{\infty} 2^{-n} \epsilon = \epsilon.$$

For the other part, given $\epsilon > 0$, choose open $B \supseteq A^c$ such that $m(B \setminus A^c) < \epsilon$. Then $B^c$ is closed, $B^c \subseteq A$, and

$$m(A \setminus B^c) = m(B \setminus A^c) < \epsilon.$$

(iii) For all $n \in \mathbb{N}$ choose open $B_n \supseteq A$ such that $m(B_n \setminus A) < 1/n$. Put $B = \bigcap_n B_n$. Then $B$ is a $G_\delta$ containing $A$, and for all $n \in \mathbb{N}$

$$0 \leq m(B \setminus A) \leq m(B_n \setminus A) < \frac{1}{n}.$$

Letting $n \to \infty$, we get $m(B \setminus A) = 0$.

Now choose a $G_\delta$ set $B \supseteq A^c$ such that $m(B \setminus A^c) = 0$. Then $B^c$ is an $F_\sigma$, $B^c \subseteq A$, and

$$m(A \setminus B^c) = m(B \setminus A^c) = 0.$$

(iv) If $A \in \mathcal{M}$ with $m(A) = 0$, then there exists a $G_\delta$ set $B \supseteq A$ such that $m(B) = m(A) = 0$.

(v) Put $\mathcal{N} = \{A \in \mathcal{M} : m(A) = 0\}$. First of all, $\emptyset \in \mathcal{N}$. If $A, B \in \mathcal{N}$ then

$$0 \leq m(A \setminus B) \leq m(A) = 0,$$

so $A \setminus B \in \mathcal{N}$. If $\{A_n\}_{i=1}^{\infty} \subseteq \mathcal{N}$ then

$$m \left( \bigcup_n A_n \right) \leq \sum_n m(A_n) = 0,$$
so $\bigcup_n A_n \in \mathcal{N}$. Thus $\mathcal{N}$ is a $\sigma$-ring. 

If $x \in \mathbb{R}^p$ then $\{x\}$ is a box with $|\{x\}| = 0$, so $\{x\} \in \mathcal{N}$. Since $\mathcal{N}$ is a $\sigma$-ring, every countable set is in $\mathcal{N}$. 

4. MEASURABLE FUNCTIONS

**Definition 4.1.** If $A \in \mathcal{M}$, then $f: A \to \mathbb{R}$ is measurable if 

$$\{x : f(x) > a\} \in \mathcal{M} \quad \text{for all } a \in \mathbb{R}.$$ 

When we say a function $f$ is measurable without specifying its domain, by default we assume it is defined on all of $\mathbb{R}^p$. If $B$ is a measurable subset of $\text{dom } f$, we say $f$ is measurable on $B$ if $f|B$ is measurable.

**Remark 4.2.** Note that if $A \in \mathcal{M}$ and $f: A \to \mathbb{R}$, then $f$ is measurable if and only if the extension of $f$ to $\mathbb{R}^p$ obtained by putting $f = 0$ on $A^c$ is a measurable function. We often find it convenient to tacitly extend $f$ in this way. Thus we can without loss of generality develop the general theory of measurable functions in the context of functions from $\mathbb{R}^p$ to $\mathbb{R}$.

**Lemma 4.3.** If $A, B \in \mathcal{M}$ and $\mathbb{R}^p = A \cup B$, then $f: \mathbb{R}^p \to \mathbb{R}$ is measurable if and only if it is measurable on both $A$ and $B$. In particular, if we put $A = f^{-1}(\{\infty\})$, $B = f^{-1}(\{-\infty\})$, and $C = (A \cup B)^c$, then $f$ is measurable if and only if $A, B \in \mathcal{M}$ and $f$ is measurable on $C$.

**Lemma 4.4.** For all $f: \mathbb{R}^p \to \mathbb{R}$ the following are equivalent:

(i) $f$ is measurable;
(ii) $\{x : f(x) \leq a\} \in \mathcal{M}$ for all $a \in \mathbb{R}$;
(iii) $\{x : f(x) \geq a\} \in \mathcal{M}$ for all $a \in \mathbb{R}$;
(iv) $\{x : f(x) < a\} \in \mathcal{M}$ for all $a \in \mathbb{R}$.

**Proof.** This follows from the following set equalities:

$[-\infty, a] = (a, \infty)^c$

$[a, \infty] = \bigcup_{n=1}^{\infty} \left( a - \frac{1}{n}, \infty \right)$

$[-\infty, a] = [a, \infty)^c$. 

**Lemma 4.5.** For all $f: \mathbb{R}^p \to \mathbb{R}$ the following are equivalent:

(i) $f$ is measurable;
(ii) $f^{-1}(\{(a, b)\}) \in \mathcal{M}$ for all $a, b \in \mathbb{R}$ with $a < b$;
(iii) $f^{-1}(A) \in \mathcal{M}$ for all open $A \subseteq \mathbb{R}$;
(iv) $f^{-1}(A) \in \mathcal{M}$ for all $A \in B$.

**Proof.** (i) $\Rightarrow$ (ii). This follows from the equality $(a, b) = (a, \infty) \cap (-\infty, b)$.

(ii) $\Rightarrow$ (iii). Every open subset of $\mathbb{R}$ is a countable union of open intervals.

(iii) $\Rightarrow$ (iv) It suffices to show that the family 

$$A := \{A \subseteq \mathbb{R} : f^{-1}(A) \in \mathcal{M}\}$$
is a $\sigma$-algebra. First, $\emptyset \in A$ since $\emptyset$ is open. If $A \in A$ then
\[ f^{-1}(A^c) = f^{-1}(A)^c \in \mathcal{M}, \]
so $A^c \in A$. If $\{ A_n \}_{n=1}^\infty \subseteq A$ then
\[ f^{-1}\left( \bigcup_n A_n \right) = \bigcup_n f^{-1}(A) \in \mathcal{M}, \]
so $\bigcup_n A_n \in A$.

(iv) $\Rightarrow$ (i) This is immediate since $(a, \infty) \in B$ for all $a \in \mathbb{R}$. \hfill $\Box$

**Corollary 4.6.** Let $f: \mathbb{R}^p \to \mathbb{R}$.

(i) If $f$ is continuous then $f$ is measurable.

(ii) If $f$ is measurable and $g: \mathbb{R} \to \mathbb{R}$ is continuous then $g \circ f$ is measurable.

**Proof.** (i) If $A \subseteq \mathbb{R}$ is open then $f^{-1}(A)$ is open, hence measurable.

(ii) If $A \subseteq \mathbb{R}$ is open then $g^{-1}(A)$ is open, so
\[ (g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A)) \in \mathcal{M}. \]
\hfill $\Box$

**Definition 4.7** (Arithmetic in $\mathbb{R}$). We allow the following arithmetic operations among extended real numbers:

(i) for all $x > -\infty$, $x + \infty = \infty + x = \infty$;
(ii) for all $x < \infty$, $x - \infty = -\infty + x = -\infty$;
(iii) for all $x > 0$, $x(\pm \infty) = (\pm \infty)x = \pm \infty$;
(iv) for all $x < 0$, $x(\pm \infty) = (\pm \infty)x = \mp \infty$;
(v) $0(\pm \infty) = (\pm \infty)0 = 0$.

**Remark 4.8.** Note that (v) above is usually not allowed, but it is convenient in the Lebesgue integration theory.

**Definition 4.9** (positive and negative parts). For each $x \in \mathbb{R}$, define
\[ x^+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad \text{and} \quad x^- = \begin{cases} -x & \text{if } x \leq 0 \\ 0 & \text{if } x > 0 \end{cases}. \]

**Lemma 4.10.** With the above notation,

(i) $x^+, x^- \geq 0$;
(ii) $x = x^+ - x^-$;
(iii) $x^+x^- = 0$;
(iv) $|x| = x^+ + x^-$.\[ Moreover, the pair $(x^+, x^-)$ is uniquely determined by properties (i)–(iii).\]

**Remark 4.11.** If $f$ is an extended-real-valued function, define $f^+$ and $f^-$ in the usual pointwise way (for example, $f^+(x) := (f(x))^+$). Then the properties listed in the above lemma continue to hold at the level of functions.

**Corollary 4.12.** If $f: \mathbb{R}^p \to \mathbb{R}$ is measurable, then so are

(i) $\text{cf for all } c \in \mathbb{R}$,
(ii) $f^n$ for all $n \in \mathbb{N}$, and
(iii) $f^+$, $f^-$, and $|f|$.

Proof. By Lemma 4.3, without loss of generality $f$ is real-valued. Then the result follows from the preceding corollary. \hfill \Box

Proposition 4.13. If $f$ and $g$ are measurable, then so are $f + g$ and $fg$ (where we assume $\infty - \infty$ does not occur).

Proof. For the first, if $a \in \mathbb{R}$ then
\[
(f + g)^{-1}((a, \infty)) = \bigcup_{r \in \mathbb{Q}} (f^{-1}((r, \infty)) \cap g^{-1}((a - r, \infty))) \in \mathcal{M}.
\]

For $fg$, note that
\[
f(x)g(x) = \infty \iff [f(x) = \infty \text{ and } g(x) > 0]
\]
\[
or [f(x) > 0 \text{ and } g(x) = \infty]
\]
\[
or [f(x) = -\infty \text{ and } g(x) < 0]
\]
\[
or [f(x) < 0 \text{ and } g(x) = -\infty],
\]
so $(fg)^{-1}([\infty)) \in \mathcal{M}$, and similarly for $-\infty$. Hence it suffices to show that $fg$ is measurable on $C := (fg)^{-1}(\mathbb{R})$.

Furthermore,
\[
f(x)g(x) \in \mathbb{R} \iff [f(x), g(x) \in \mathbb{R}]
\]
\[
or [f(x) = \pm \infty \text{ and } g(x) = 0]
\]
\[
or [f(x) = 0 \text{ and } g(x) = \pm \infty],
\]
so it suffices to show that $fg$ is measurable on $D := f^{-1}(\mathbb{R}) \cap g^{-1}(\mathbb{R})$. But on $D$,
\[
fg = \frac{1}{4}((f + g)^2 - (f - g)^2),
\]
so the result follows from Corollary 4.12. \hfill \Box

Proposition 4.14. If $f_1, f_2, \ldots$ are measurable, then so are
\[
\sup f_n, \quad \inf f_n, \quad \lim\sup f_n, \quad \lim\inf f_n, \quad \text{and}
\]
\[
\lim f_n \quad \text{if this exists (in \mathbb{R})}.
\]

Proof. If $a \in \mathbb{R}$ then
\[
\sup f_n(x) > a \iff \text{there exists } n \text{ such that } f_n(x) > a
\]
\[
\iff x \in \bigcup_n f_n^{-1}((a, \infty)),
\]
so \{ $x : \sup f_n(x) > a$ \} \in \mathcal{M}. Similarly for inf, hence lim sup and lim inf.

For the last part, if $f_n \to f$ pointwise then $f = \lim sup f_n$ is measurable. \hfill \Box
Definition 4.15. If $P$ is a propositional function defined on $\mathbb{R}^p$, we say $P(x)$ a.e. $x$, or just a.e. if $x$ is understood, or even just $P$ a.e., to mean

$$m(\{x : P(x) \text{ is false}\}) = 0.$$  

Remark 4.16. Since every subset of a set with measure 0 is measurable and has measure 0, we have $P$ a.e. if and only if there exists $A$ with $m(A) = 0$ such that $P(x)$ for all $x \notin A$.

Proposition 4.17. If $f$ is measurable and $f = g$ a.e., then $g$ is measurable.

Proof. Suppose $m(A) = 0$ and $f(x) = g(x)$ for all $x \notin A$. Let $a \in \mathbb{R}$, and put $B = \{x : g(x) > a\}$. Then $B \cap A \subseteq A$ and $m(A) = 0$, so $B \cap A \in \mathcal{M}$.

Also,

$$B \setminus A = \{x \in A^c : g(x) > a\} = \{x \in A : f(x) > a\} = \{x : f(x) > a\} \setminus A \in \mathcal{M}.$$  

Thus $B = (B \cap A) \cup (B \setminus A) \in \mathcal{M}$. 

Definition 4.18. A simple function is a linear combination $\sum c_n \chi_{A_n}$, where $c_n \in \mathbb{R}$ and $A_n \in \mathcal{M}$.

Theorem 4.19. If $f$ is measurable, then there exists a sequence $(\phi_n)$ of simple functions such that $\phi_n \to f$ (in $\mathbb{R}$). Moreover, if $f \geq 0$ we can take $0 \leq \phi_1 \leq \phi_2 \leq \cdots < f$.

Proof. First assume $f \geq 0$. For all $n \in \mathbb{N}$ define $\phi_n : \mathbb{R}^p \to \mathbb{R}$ by

$$\phi_n = \sum_{k=0}^{2^n n-1} \chi_{f^{-1}([k2^{-n}, (k+1)2^{-n}))} + n \chi_{f^{-1}([n, \infty])}.$$  

Then each $\phi_n$ is simple, and

$$0 \leq \phi_1 \leq \phi_2 \leq \cdots \leq f.$$  

To see that $\phi_n \to f$ pointwise, fix $x \in \mathbb{R}^p$. If $f(x) = \infty$ then

$$\phi_n(x) = n \to f(x),$$  

while if $f(x) < \infty$ then for all $n > f(x)$ we have

$$|f(x) - \phi_n(x)| < 2^{-n} \to 0.$$  

Now remove the restriction $f \geq 0$. Apply the first part to $f^+$ and $f^-$, getting simple functions $\psi_n$ and $\xi_n$ such that $\psi_n \to f^+$ and $\xi_n \to f^-$. Then each $\psi_n - \xi_n$ is a simple function, and

$$\psi_n - \xi_n \to f^+ - f^- = f.$$  

\qed
5. Integrating Nonnegative Functions

Definition 5.1. Let $S^+$ denote the set of nonnegative simple functions. For each $\phi \in S^+$, write $\phi = \sum_1^k c_n \chi_{A_n}$ with $c_1, \ldots, c_k \geq 0$ and $A_1, \ldots, A_k$ disjoint, and define:

(i) $\int \phi = \sum_1^k c_n m(A_n)$;
(ii) $\int_A \phi = \int \phi\chi_A$ if $A \in \mathcal{M}$.

Remark 5.2. Note that if $\phi = \sum_1^k c_n \chi_{A_n}$ with $c_1, \ldots, c_k \geq 0$ and $A_1, \ldots, A_k$ disjoint, and if $B \in \mathcal{M}$, then

$$\int_B \phi = \sum_n c_n (A_n \cap B).$$

Proposition 5.3. For all $\phi, \psi \in S^+$,

(i) $\int \phi$ is well-defined;
(ii) $\phi \leq \psi$ implies $\int \phi \leq \int \psi$;
(iii) $\int c\phi = c\int \phi$ if $c \in [0, \infty]$;
(iv) $\int (\phi + \psi) = \int \phi + \int \psi$;
(v) $A \mapsto \int_A \phi$ is a measure on $\mathcal{M}$.

Proof. Throughout this proof, let $\phi = \sum_1^k c_n \chi_{A_n}$, with $c_1, \ldots, c_k \geq 0$ and $A_1, \ldots, A_k \in \mathcal{M}$ disjoint, and similarly for $\psi = \sum_1^l d_j \chi_{B_j}$. If necessary, add terms with coefficients equal to 0 so that without loss of generality we have $\bigcup_n A_n = \bigcup_j B_j$.

We prove (i) and (ii) in one whack: assume $\phi \leq \psi$. Then

$$\sum_n c_n m(A_n) = \sum_n c_n m \left( \bigcup_j (A_n \cap B_j) \right)$$

$$= \sum_n c_n \sum_j m(A_n \cap B_j)$$

$$= \sum_{n,j} c_n m(A_n \cap B_j)$$

$$\leq \sum_{n,j} d_j m(A_n \cap B_j) \quad \text{since } A_n \cap B_j \neq \emptyset \Rightarrow c_n \leq d_j$$

$$= \sum_j d_j m(B_j) \quad \text{by symmetry.}$$

In particular, if $\phi = \psi$ then by symmetry we get

$$\sum_n c_n m(A_n) = \sum_j d_j m(B_j),$$

proving (i), and then (ii) follows from the above computation.

(iii) We have

$$\int c\phi = \sum_n c c_n m(A_n) = c \sum_n c_n m(A_n) = c \int \phi.$$
(iv) We have
\[ \int (\phi + \psi) = \sum_{n,j} (c_n + d_j)m(A_n \cap B_j) \]
\[ = \sum_n c_n \sum_j m(A_n \cap B_j) + \sum_j d_j \sum_n m(A_n \cap B_j) \]
\[ = \sum_n c_n m(A_n) + \sum_j d_j m(B_j) \]
\[ = \int \phi + \int \psi, \]
where the equality at * follows from
\[ \sum_n c_n \chi_{A_n} + \sum_j d_j \chi_{B_j} = \sum_n c_n \chi_{\bigcup_j (A_n \cap B_j)} + \sum_j d_j \chi_{\bigcap_n (A_n \cap B_j)} \]
\[ = \sum_n c_n \sum_j \chi_{A_n \cap B_j} + \sum_j d_j \sum_n \chi_{A_n \cap B_j} \]
\[ = \sum_n (c_n + d_j) \chi_{A_n \cap B_j}. \]

(v) First of all, \( \int_A \phi \geq 0 \) by construction. Next,
\[ \int_{\emptyset} \phi = \int \phi \chi_{\emptyset} = \int 0 = 0. \]
Finally, if \( \{C_j\}_{1}^{\infty} \) is a disjoint sequence in \( \mathcal{M} \), then
\[ \int_{\bigcup_j C_j} \phi = \sum_n c_n m \left( A_n \cap \bigcup_j C_j \right) \]
\[ = \sum_n c_n m \left( \bigcup_j (A_n \cap C_j) \right) \]
\[ = \sum_n c_n \sum_j m(A_n \cap C_j) \]
\[ = \sum_j \sum_n c_n m(A_n \cap C_j) \]
\[ = \sum_j \int_{C_j} \phi. \]

\[ \square \]

**Definition 5.4.** Let \( L^+ = L^+ (\mathbb{R}^p) \) denote the set of nonnegative measurable functions. For each \( f \in L^+ \) define:

(i) \( \int f = \sup \{ \int \phi : \phi \in S^+, \phi \leq f \}; \)
(ii) \( \int_A f = \int f \chi_A \) if \( A \in \mathcal{M} \).

**Lemma 5.5.** If \( \phi \in S^+ \) then the two definitions of \( \int \phi \) are consistent.
Proof. Temporarily write $\int \phi$ for the first definition and $\int' \phi$ for the second. Since $\phi \in S^+$ and $\phi \leq \phi$, we have

$$\int \phi \leq \int' \phi.$$ 

On the other hand, if $\psi \in S^+$ and $\psi \leq \phi$, then $\int \psi \leq \int \phi$. Taking the supremum over $\psi$, we get

$$\int' \phi \leq \int \phi.$$ 

\[ \square \]

Proposition 5.6. For all $f, g \in L^+$,

(i) $f \leq g$ implies $\int f \leq \int g$;

(ii) $\int cf = c \int f$ if $c \in [0, \infty)$;

(iii) $\int f = 0$ if and only if $f = 0$ a.e.;

(iv) $\int f < \infty$ implies $f < \infty$ a.e.

Proof. (i) If $\phi \in S^+$ with $\phi \leq f$, then $\phi \leq g$, so $\int \phi \leq \int g$. Taking the supremum over $\phi$, we get

$$\int f \leq \int g.$$ 

(ii) If $c = 0$ then both sides are 0, so without loss of generality $c > 0$. If $\phi \in S^+$ and $\phi \leq f$, then $c\phi \in S^+$ and $c\phi \leq cf$, so

$$c \int \phi = \int c\phi \leq \int cf.$$ 

Thus

$$c \int f \leq \int cf.$$ 

By the same reasoning,

$$\frac{1}{c} \int cf \leq \int f.$$ 

Thus

$$c \int f \leq \int cf \leq c \int f,$$

so we must have equality throughout.

(iii) First assume $\int f = 0$. For each $n \in \mathbb{N}$ put $A_n = \{x : f(x) > 1/n\}$, and put $A = \{x : f(x) \neq 0\}$. Then $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcup_n A_n$, so

$$m(A_n) \to m(A).$$

For every $n \in \mathbb{N}$, we have $f \geq \frac{1}{n} \chi_{A_n}$, so

$$0 = \int f \geq \frac{1}{n} m(A_n) \geq 0,$$

hence $m(A_n) = 0$. Therefore $m(A) = 0$, so $f = 0$ a.e.
Conversely, assume $f = 0$ a.e., and let $\phi \in S^+$ with $\phi \leq f$. If $\phi = \sum c_n \chi_{A_n}$ with $c_1, \ldots, c_n \geq 0$ and $A_1, \ldots, A_n$ disjoint, then $c_n = 0$ whenever $m(A_n) \neq 0$ since $f \geq c_n$ on $A_n$. Hence

$$\int \phi = \sum_{m(A_n) \neq 0} c_n m(A_n) = \sum 0 = 0.$$  

Thus $\int f = 0$.

(iv) For all $n \in \mathbb{N}$ put $A_n = \{x : f(x) > n\}$. Then $f \geq n\chi_{A_n}$, so

$$\infty > \int f \geq nm(A_n),$$

hence

$$m(A_n) \leq \frac{1}{n} \int f \to 0.$$  

Also, $m(A_1) < \infty$. Since $A_1 \supseteq A_2 \supseteq \cdots$, we have

$$m\left(\bigcap_n A_n\right) = \lim m(A_n) = 0.$$  

Therefore $f < \infty$ a.e., because

$$f^{-1}(\{\infty\}) = \bigcap_n A_n.$$

\[\square\]

**Theorem 5.7 (Monotone Convergence Theorem).** If $(f_n)$ is a sequence in $L^+$ and $f_n \uparrow f$ (in $\mathbb{R}$), then

$$\int f_n \to \int f.$$  

**Proof.** Let $\phi \in S^+$ with $\phi \leq f$. Let $0 < a < 1$, and put $A_n = \{x : f_n(x) \geq a\phi(x)\}$. Then $f_n \geq a\phi\chi_{A_n}$, so

$$\int f_n \geq \int_{A_n} a\phi = a\int_{A_n} \phi.$$  

Since $A_1 \subseteq A_2 \subseteq \cdots$ and $\bigcup_n A_n = \mathbb{R}^p$, we have

$$\int_{A_n} \phi \to \int \phi.$$  

Since $f_1 \leq f_2 \leq \cdots$, the sequence $(\int f_n)$ is increasing, so $\lim \int f_n$ exists in $\mathbb{R}$, and

$$\lim \int f_n \geq a\int \phi.$$  

Letting $a \to 1$, we get

$$\lim \int f_n \geq \int \phi.$$
Taking the supremum over \( \phi \), we find
\[
\lim \int f_n \geq \int f.
\]
But \( f_n \leq f \), so \( \int f_n \leq \int f \) for all \( n \), hence
\[
\lim \int f_n \leq \int f.
\]
Therefore we must have \( \lim \int f_n = \int f \). \( \square \)

**Corollary 5.8.**

(i) If \( \sum f_n \) is a series in \( L^+ \), then \( \int \sum f_n = \sum \int f_n \).

(ii) For all \( f \in L^+ \), \( A \mapsto \int_A f \) is a measure on \( \mathcal{M} \).

(iii) For all \( f, g \in L^+ \), if \( f \approx g \) a.e. then \( \int f = \int g \).

(iv) In the Monotone Convergence Theorem, \( f_n \uparrow f \) a.e. is ok.

**Proof.** (i) We first do it for finite sums, and for this it suffices by induction to show that \( f, g \in L^+ \) implies \( \int (f + g) = \int f + \int g \). Choose sequences \( (\phi_n), (\psi_n) \) in \( S^+ \) such that \( \phi_n \uparrow f \) and \( \psi_n \uparrow g \). Then \( \phi_n + \psi_n \uparrow f + g \). So, by the Monotone Convergence Theorem
\[
\int (f + g) = \lim \int (\phi_n + \psi_n) = \lim \left( \int \phi_n + \int \psi_n \right)
= \lim \int \phi_n + \lim \int \psi_n
= \int f + \int g.
\]

Now let \( \sum f_n \) be a series in \( L^+ \), and put \( g_k = \sum_{1}^{k} f_n \) and \( g = \sum_{1}^{\infty} f_n \). Then \( 0 \leq g_k \uparrow g \), so by the Monotone Convergence Theorem
\[
\int \sum_{1}^{\infty} f_n = \int g = \lim \int g_k = \lim \int \sum_{1}^{k} f_n
= \lim \sum_{1}^{k} \int f_n = \sum_{1}^{\infty} \int f_n.
\]

(ii) First of all, \( \int_A f \geq 0 \) by construction. Also
\[
\int_{\emptyset} f = \int f \chi_{\emptyset} = \int 0 = 0.
\]

If \( A_1, A_2, \ldots \in \mathcal{M} \) are disjoint, then by the Monotone Convergence Theorem
\[
\int_{\bigcup_{n} A_n} f = \int f \chi_{\bigcup_{n} A_n} = \int f \sum \chi_{A_n}
= \sum \int f \chi_{A_n} = \sum_{A_n} f.
\]
(iii) Suppose $m(A) = 0$ and $f = g$ on $A^c$. Then both $f_{\chi_A}$ and $g_{\chi_A}$ are 0 a.e., so

$$\int f = \int_A f + \int_{A^c} f = \int_{A^c} g = \int_A g + \int_{A^c} g = \int g.$$ 

(iv) If $f = \lim f_n$ a.e., then by (iii) and the Monotone Convergence Theorem

$$\int f = \int \lim f_n = \lim \int f_n.$$ 

$\Box$

**Theorem 5.9** (Fatou’s Lemma). If $(f_n)$ is a sequence in $L^+$, then

$$\int \lim \inf f_n \leq \lim \inf \int f_n.$$ 

In particular, if $f_n \to f$ a.e., then $\int f \leq \lim \inf \int f_n$.

**Proof.** Put $g_k = \inf_{n \geq k} f_n$. Then $0 \leq g_k \uparrow \lim \inf f_n$ and $g_k \leq f_k$ for all $k$, so by the Monotone Convergence Theorem

$$\int \lim \inf f_n = \lim \int g_k \leq \lim \inf \int f_k.$$ 

$\Box$

### 6. Integrable Functions

**Definition 6.1.** Let $f$ be a measurable function.

(i) The **Lebesgue integral** of $f$ is

$$\int f = \int f(x) \, dx := \int f^+ - \int f^-,$$

if this makes sense (that is, if the right hand side is not $\infty - \infty$).

(ii) For $A \in \mathcal{M}$ we define

$$\int_A f = \int f_{\chi_A}.$$

(iii) $f$ is integrable if $\int f \in \mathbb{R}$, and $L^1 = L^1(\mathbb{R})$ denotes the set of integrable functions.

**Proposition 6.2.** (i) If $f$ is measurable, then $f \in L^1$ if and only if $\int |f| < \infty$, in which case $|\int f| \leq \int |f|$.

(ii) If $f \in L^1$ then $f(x) \in \mathbb{R}$ a.e.

(iii) $L^1$ is a vector space, and $\int : L^1 \to \mathbb{R}$ is linear.

(iv) If $f \in L^1$ and $A \in \mathcal{M}$ then

$$\int_A f = \int_A f^+ - \int_A f^-.$$

(v) If $f, g \in L^1$, then $\int_A f = \int_A g$ for all $A \in \mathcal{M}$ if and only if $f = g$ a.e.
Proof. (i) First assume $f \in L^1$. Then $\int f^+, \int f^- < \infty$. Hence

$$\int |f| = \int (f^+ + f^-) = \int f^+ + \int f^- < \infty.$$ 

Further,

$$\left| \int f \right| = \left| \int f^+ - \int f^- \right| \leq \int f^+ + \int f^- = \int |f|.$$ 

Conversely, assume $\int |f| < \infty$. Then

$$\int f^+, \int f^- \leq \int f^+ + \int f^- = \int (f^+ + f^-) = \int |f| < \infty,$$

so $f \in L^1$.

(ii) If $f \in L^1$ then $\int |f| < \infty$, hence $|f| < \infty$ a.e., therefore $f(x) \in \mathbb{R}$ a.e.

(iii) First let $f, g \in L^1$. How are we going to form the function $f + g$? The obstacle is that it is possible that for some $x$ the values $f(x)$ and $g(x)$ could both be infinite with opposite signs. The solution to this conundrum is to note that since both $f$ and $g$ are integrable they are finite-valued a.e., so $f + g$ is defined a.e., and this is good enough for us to consider the integrability of $f + g$. Actually, this would be the case in any event, since $f$ and $g$ themselves are officially only required to be defined a.e. Anyway, $f + g$ is measurable and

$$\int |f + g| \leq \int (|f| + |g|) = \int |f| + \int |g| < \infty,$$

so $f + g \in L^1$. Further, if $h = f + g$ then

$$h^+ - h^- = f^+ - f^- + g^+ - g^-,$$

so

$$h^+ + f^- + g^- = h^- + f^+ + g^+,$$

hence

$$\int h^+ + \int f^- + \int g^- = \int h^- + \int f^+ + \int g^+,$$

thus

$$\int (f + g) = \int (h^+ - h^-) = \int h^+ - \int h^-$$

$$= \int f^+ - \int f^- + \int g^+ - \int g^- = \int f + \int g.$$ 

Next, if $f \in L^1$ and $c \in \mathbb{R}$, then $cf$ is measurable and

$$\int |cf| = \int |c||f| = |c| \int |f| < \infty,$$
so $cf \in L^1$. Further, if $c \geq 0$ then

$$
\int cf = \int (cf^+ - cf^-) = \int ((cf)^+ - (cf)^-) = \int cf^+ - \int cf^-
$$

$= c \int f^+ - c \int f^- = c \int f,$

and similarly if $c < 0$.

(iv) Since $f \in L^1$, we have $f^+ \chi_A, f^- \chi_A \in L^1$, so by linearity

$$
\int_A f = \int f \chi_A = \int (f^+ \chi_A - f^- \chi_A)
$$

$$
= \int f^+ \chi_A - \int f^- \chi_A
$$

$$
= \int_A f^+ - \int_A f^-.
$$

(v) Replacing $f$ by $f - g$, without loss of generality $g = 0$. Put

$$
P = \{x : f^+(x) \neq 0\} = \{x : f(x) > 0\} \quad \text{and}
$$

$$
N = \{x : f^-(x) \neq 0\} = \{x : f(x) < 0\},
$$

so that $f^+ = f \chi_P$ and $f^- = -f \chi_N$.

Assume $\int_A f = 0$ for all $A \in \mathcal{M}$. Then in particular $\int f^+ = \int f = 0$, so $f^+ = 0$ a.e. by Proposition 5.6 (iii). Integrating over $N$ instead, we find $f^- = 0$ a.e. Thus $f = 0$ a.e.

Conversely, assume $f = 0$ a.e. Then $m(P \cup N) = 0$, so $m(P) = m(N) = 0$, hence $f^+$ and $f^-$ are 0 a.e. If $A \in \mathcal{M}$ then

$$
\int_A f = \int_A f^+ - \int_A f^- = 0 - 0 = 0.
$$

Definition 6.3. (i) We loosen the definition of the phrase “$f$ is a measurable function” to mean there exists $A \subseteq \mathbb{R}^p$ such that $m(A^c) = 0$ and $f : A \to \mathbb{R}$ is measurable.

(ii) If $f$ and $g$ are measurable functions, we say $f$ is equivalent to $g$, written $f \sim g$, if $f = g$ a.e.

(iii) We modify the official definition of $L^1$ so that its elements are equivalence classes of integrable functions (but we abuse this by continuing to speak of integrable functions as elements of $L^1$; this causes no confusion in practice).

Theorem 6.4 (Dominated Convergence Theorem). Let $f_n \to f$ a.e., with each $f_n$ measurable, and suppose there exists $g \in L^1$ such that for all $n \in \mathbb{N}$ we have $|f_n| \leq g$ a.e. Then $f \in L^1$ and

$$
\int f_n \to \int f.
$$
Proof. Since \( f_n \to f \) a.e., \( f \) is measurable. Since \( |f_n| \leq g \) a.e. for all \( n \) and a countable union of sets of measure 0 has measure 0, we have \( |f| \leq g \) a.e. Thus \( f \in L^1 \).

For the other part, \( g + f_n \geq 0 \) for all \( n \), so by Fatou’s Lemma
\[
\int g + \int f = \int (g + f) = \int \lim (g + f_n)
\leq \lim \inf \int (g + f_n) = \int g + \lim \inf \int f_n,
\]
hence
\[
\int f \leq \lim \inf \int f_n.
\]
Similarly, \( g - f_n \geq 0 \), so
\[
\int g - \int f \leq \lim \inf \int (g - f_n) = \int g - \lim \sup \int f_n,
\]
so
\[
\int f \geq \lim \sup \int f_n.
\]
Therefore we must have
\[
\lim \sup \int f_n = \lim \inf \int f_n = \int f.
\]

\[\Box\]

Corollary 6.5. If \( \sum f_n \) is a series in \( L^1 \) such that \( \sum \int |f_n| < \infty \), then the series \( \sum f_n \) converges a.e. to an integrable function, and
\[
\int \sum f_n = \sum \int f_n.
\]

Proof. By the Monotone Convergence Theorem
\[
\int \sum |f_n| = \sum \int |f_n| < \infty,
\]
so \( \sum |f_n| < \infty \) a.e., hence \( \sum f_n \) converges a.e. Then \( \sum f_n \) is measurable. For each \( k \in \mathbb{N} \),
\[
\left| \sum_{1}^{k} f_n \right| \leq \sum_{1}^{k} |f_n| \leq \sum_{1}^{\infty} |f_n| \in L^1,
\]
so by the Dominated Convergence Theorem
\[
\sum_{1}^{k} \int f_n = \int \sum_{1}^{k} f_n \to \int \sum_{1}^{\infty} f_n,
\]
hence
\[
\sum_{1}^{\infty} \int f_n = \int \sum_{1}^{\infty} f_n.
\]
Corollary 6.6. \( L^1 \) is a Banach space with norm \( \| f \|_1 := \int |f| \), and \( f : L^1 \to \mathbb{R} \) is continuous.

Proof. First of all, if \( f \in L^1 \) then \( \| f \|_1 = \int |f| \geq 0 \), and

\[
\| f \|_1 = 0 \iff \int |f| = 0 \iff |f| = 0 \text{ a.e.} \iff f = 0 \text{ a.e.}
\]

Next, if \( f \in L^1 \) and \( c \in \mathbb{R} \) then

\[
\| cf \|_1 = \int |cf| = |c| \int |f| = |c| \| f \|_1,
\]

and if also \( g \in L^1 \) then

\[
\| f + g \|_1 = \int |f + g| \leq \int (|f| + |g|) = \int |f| + \int |g| = \| f \|_1 + \| g \|_1.
\]

Thus \( \| \cdot \| \) is a norm on \( L^1 \).

Also,

\[
\left| \int f \right| \leq \int |f| = \| f \|_1,
\]

and \( f \mapsto \int f \) is linear, so \( f \mapsto \int f \) is continuous.

For the completeness, let \((f_n)\) be a Cauchy sequence in \( L^1 \). It suffices to show that \((f_n)\) has a convergent subsequence. Choose \( n_1 \in \mathbb{N} \) such that

\[
j, k \geq n_1 \Rightarrow \| f_j - f_k \|_1 < 2^{-1}.
\]

Then choose \( n_2 > n_1 \) such that

\[
j, k \geq n_2 \Rightarrow \| f_j - f_k \|_1 < 2^{-2}.
\]

Continue inductively, getting \( n_1 < n_2 < \cdots \) such that

\[
\| f_{n_{k+1}} - f_{n_k} \|_1 < 2^{-k} \quad \text{for all } k \in \mathbb{N}.
\]

Put

\[
g_k = \begin{cases} f_{n_1} & \text{if } k = 1, \\ f_{n_k} - f_{n_{k-1}} & \text{if } k > 1. \end{cases}
\]

Then \((g_k)\) is a sequence in \( L^1 \), and

\[
\sum \| g_k \|_1 = \| g_1 \| + \sum_{k=2}^{\infty} \| g_k \| < \| g_1 \| + \sum_{k=1}^{\infty} 2^{-k} < \infty.
\]

Hence \( \sum g_k \) converges a.e., and \( \sum g_k \in L^1 \). Moreover,

\[
\left\| \sum_{k=1}^{l} g_k - \sum_{k=1}^{\infty} g_k \right\|_1 = \left\| \sum_{k=1}^{\infty} g_k \right\|_1 = \int \left| \sum_{k=1}^{\infty} g_k \right| \\
\leq \int \sum_{k=1}^{\infty} |g_k| = \sum_{k=1}^{\infty} \int |g_k| \xrightarrow{l \to \infty} 0,
\]

\( \square \)
since $\sum_1^{\infty} |g_k| < \infty$. Thus the series $\sum_1^{\infty} g_k$ converges in the normed space $L^1$. But for all $l \in \mathbb{N}$,

$$\sum_1^{l} g_k = f_{nl},$$

so $(f_{nl})$ converges in $L^1$. \hfill \square

**Corollary 6.7** (of the above proof). If $f_n \to f$ in $L^1$ then some subsequence of $(f_n)$ converges to $f$ a.e.

**Definition 6.8.**

(i) If $f: \mathbb{R}^p \to \mathbb{R}$ is continuous, the support of $f$ is

$$\text{supp } f = \{x \in \mathbb{R}^p : f(x) \neq 0\}.$$

(ii) $C_c(\mathbb{R}^p)$ denotes the set of continuous functions on $\mathbb{R}^p$ with compact support.

**Corollary 6.9.** The following sets are dense in $L^1$:

(i) the set of integrable simple functions;  
(ii) $C_c(\mathbb{R}^p)$.

**Proof.** (i) Given $f \in L^1$, choose sequences $(\phi_n), (\psi_n)$ in $S^+$ such that $\phi_n \uparrow f^+$ and $\psi_n \uparrow f^-$. Then by the Monotone Convergence Theorem

$$\int \phi_n \to \int f^+ \quad \text{and} \quad \int \psi_n \to \int f^-,$$

so

$$\|f - (\phi_n - \psi_n)\|_1 = \|(f^+ - \phi_n) - (f^- - \psi_n)\|_1$$

$$\leq \|f^+ - \phi_n\|_1 + \|f^- - \psi_n\|_1$$

$$= \int |f^+ - \phi_n| + \int |f^- - \psi_n|$$

$$= \int f^+ - \int \phi_n + \int f^- - \int \phi_n$$

$$\to 0.$$

(ii) By (i) it suffices to show that if $A \in \mathcal{M}$ and $m(A) < \infty$ then for all $\epsilon > 0$ there exists $f \in C_c(\mathbb{R}^p)$ such that $\|\chi_A - f\|_1 < \epsilon$. We first show that there exists $B \in \mathcal{E}$ such that $\|\chi_A - \chi_B\|_1 < \epsilon$. Choose open boxes $B_1, B_2, \ldots$ such that $A \subseteq \bigcup_1^{\infty} B_n$ and $\sum m(B_n) - m(A) < \epsilon/2$. Then choose $k \in \mathbb{N}$ such that

$$\sum_{k+1}^{\infty} m(B_n) < \frac{\epsilon}{2}.$$
Let \( B = \bigcup_{1}^{k} B_{n} \). Then
\[
\left\| \chi_{A} - \chi_{B} \right\|_{1} = m(A \setminus B) + m(B \setminus A)
\leq m\left( \bigcup_{k+1}^{\infty} B_{n} \right) + m\left( \bigcup_{1}^{\infty} B_{n} \setminus A \right)
\leq \sum_{k+1}^{\infty} m(B_{n}) + m\left( \bigcup_{1}^{\infty} B_{n} \right) - m(A)
\leq \frac{\epsilon}{2} + \sum_{1}^{\infty} m(B_{n}) - m(A)
< \epsilon,
\]
as desired.

Now it suffices to show that if \( B \) is a box then \( \chi_{B} \) can be approximated in the \( L^{1} \) norm by \( f \in C_{c}(\mathbb{R}^{p}) \). Without loss of generality \( B \) is nonempty and open, say
\[
B = \prod_{1}^{p}(a_{i}, b_{i}) \quad \text{with } a_{i} < b_{i}.
\]
Temporarily fix \( c \in (0, 1) \), and for each \( i \) let \( (a_{i}', b_{i}') \) be the open interval concentric with \( (a_{i}, b_{i}) \) such that
\[
b_{i}' - a_{i}' = c(a_{i} - b_{i}),
\]
and then put \( B' = \prod_{1}^{p}(a_{i}', b_{i}') \). Then
\[
m(B') = c^{p}m(B).
\]
Now for each \( i \) let \( f_{i} \) be the piecewise-linear function on \( \mathbb{R} \) which is 0 outside of \( (a_{i}, b_{i}) \), 1 on \( [a_{i}', b_{i}'] \), and linear on both \( [a_{i}, a_{i}'] \) and \( [b_{i}', b_{i}] \). Then define \( f: \mathbb{R}^{p} \to \mathbb{R} \) by
\[
f(x_{1}, \ldots, x_{p}) = \prod_{1}^{p} f_{i}(x_{i}).
\]
Then \( f \) is continuous, \( 0 \leq f \leq 1 \), \( f \) is 0 outside \( B \), and \( f \) is 1 on \( B' \). In particular, \( f \) has compact support. Since \( \chi_{B'} \leq f \leq \chi_{B} \), we have
\[
\| \chi_{B} - f \|_{1} = \int (\chi_{B} - f) \leq \int (\chi_{B} - \chi_{B}') = m(B) - m(B') \xrightarrow{c^{p}} 0.
\]

\[\square\]

**Corollary 6.10** (Lebesgue’s Theorem on Riemann Integrability). A bounded real-valued function \( f \) on \([a,b]\) is Riemann integrable if and only if it is continuous a.e. Moreover, in this case \( f \) is also Lebesgue integrable, and the Riemann and Lebesgue integrals of \( f \) coincide.
Proof. Choose partitions $P_n = \{x_i^n\}_{i=1}^{k_n}$ of $[a, b]$ such that $P_1 \subseteq P_2 \subseteq \cdots$ and $\|P_n\| \to 0$. For each $n \in \mathbb{N}$ and $i = 1, \ldots, k_n$, define

$$M_i^n = \sup_{[x_{i-1}^n, x_i^n]} f \quad \text{and} \quad m_i^n = \inf_{[x_{i-1}^n, x_i^n]} f,$$

and then define

$$h_n = \sum_{i=1}^{k_n} M_i^n \chi_{(x_{i-1}^n, x_i^n]} + f \chi_{P_n} \quad \text{and} \quad g_n = \sum_{i=1}^{k_n} m_i^n \chi_{(x_{i-1}^n, x_i^n]} + f \chi_{P_n}.$$ 

Note that

$$\int h_n = U(f, P_n) \quad \text{and} \quad \int g_n = L(f, P_n).$$

Moreover,

$$g_1 \leq g_2 \leq \cdots \leq f \leq \cdots \leq h_2 \leq h_1.$$ 

Put $g = \lim g_n$ and $h = \lim h_n$. Then $g, h \in L^1$ and $g \leq f \leq h$. Also

$$\int g_n \to \int g \quad \text{and} \quad \int h_n \to \int h$$ 

by the Dominated Convergence Theorem. Put $P = \bigcup P_n$. Then $P$ is a countable set, so $m(P) = 0$. Suppose $x \in P^c$ and $f$ is continuous at $x$.

Given $\epsilon > 0$ choose $\delta > 0$ such that

$$|y - x| < \delta \Rightarrow |f(y) - f(x)| < \epsilon.$$ 

Choose $n \in \mathbb{N}$ such that $\|P_n\| < \delta$, and choose $i$ such that $x_{i-1}^n < x < x_i^n$. Then

$$|f(y) - f(z)| < 2\epsilon \quad \text{for all } y, z \in (x_{i-1}^n, x_i^n),$$

so

$$0 \leq h(x) - g(x) \leq h_n(x) - g_n(x) \leq 2\epsilon.$$ 

Letting $\epsilon \to 0$, we get $h(x) - g(x) = 0$.

Thus, assuming $f$ is continuous a.e., we have $g = h$ a.e., so $\int g = \int h$. Thus

$$U(f, P_n) - L(f, P_n) = \int h_n - \int g_n \to 0,$$

so $f$ is Riemann integrable.

Conversely, assume $f$ is Riemann integrable, and temporarily denote the Riemann integral of $f$ by $R \int_a^b f$. Given $\epsilon > 0$, choose $n \in \mathbb{N}$ such that $U(f, P_n) - L(f, P_n) < \epsilon$. Then

$$0 \leq \int (h - g) \leq \int h_n - \int g_n < \epsilon.$$ 

Letting $\epsilon \to 0$, we get $\int (h - g) = 0$, so $g = f = h$ a.e. Thus $f \in L^1$ and

$$\int g = \int f = \int h.$$
If \( n \in \mathbb{N} \) then
\[
\int g_n \leq R \int_a^b f \leq \int h_n.
\]
Letting \( n \to \infty \), we get
\[
\int g \leq R \int_a^b f \leq \int h,
\]
hence we must have
\[
R \int_a^b f = \int f.
\]
To see \( f \) is continuous a.e., it suffices to show that if \( x \in P_c \) and \( f \) is not continuous at \( x \) then \( g(x) \neq h(x) \), since \( g = h \) a.e. and \( m(P) = 0 \). Choose \( \epsilon > 0 \) such that for all \( \delta > 0 \) there exists \( y \) such that
\[
|y - x| < \delta \quad \text{and} \quad |f(y) - f(x)| \geq \epsilon.
\]
Then for all \( n \in \mathbb{N} \), if \( x_{i-1}^n < x < x_i^n \) there exists \( y \in (x_{i-1}^n, x_i^n) \) such that
\[
|f(y) - f(x)| \geq \epsilon.
\]
Thus
\[
h_n(x) - g_n(x) \geq \epsilon.
\]
Letting \( n \to \infty \), we get
\[
h(x) - g(x) \geq \epsilon.
\]
\[
\square
\]

7. Iterated Integrals

**Notation and Terminology 7.1.** Throughout this section we identify \( \mathbb{R}^{p+q} \)
with \( \mathbb{R}^p \times \mathbb{R}^q \). Also, for a function \( f \) on \( \mathbb{R}^{p+q} \) and \( y \in \mathbb{R}^q \) we write \( f(\cdot, y) \)
for the function \( x \mapsto f(x, y) \), and similarly for \( f(x, \cdot) \).

**Theorem 7.2** (Tonelli’s Theorem). If \( f \in L^+(\mathbb{R}^{p+q}) \), then:

(i) \( f(\cdot, y) \in L^+(\mathbb{R}^p) \) a.e. \( y \in \mathbb{R}^q \),

(ii) \( y \mapsto \int f(x, y) \, dx \) is measurable, and

(iii) \( \iint f(x, y) \, dx \, dy = \int f \).

Similarly for \( f(x, \cdot) \), \( x \mapsto \int f(x, y) \, dy \), and \( \iint f(x, y) \, dy \, dx \).

**Proof.** Put
\[
\mathcal{F} = \{ f \in L^+ : (i)-(iii) \text{ hold} \},
\]
\[
\mathcal{C} = \{ A \in \mathcal{M} : \chi_A \in \mathcal{F} \}.
\]
First note that if \( A \in \mathcal{M} \) then \( A \in \mathcal{C} \) if and only if

- \( A_y := \{ x : (x, y) \in A \} \in \mathcal{M} \) a.e. \( y \),
- \( y \mapsto m(A_y) \) is measurable, and
- \( \int m(A_y) \, dy = m(A) \).
since

\[ \chi_A(x, y) = \chi_{A_y}, \]

\[ \int \chi_A(x, y) \, dx = \int \chi_{A_y} = m(A_y), \]

\[ \iint \chi_A(x, y) \, dx \, dy = \int m(A_y) \, dy, \quad \text{and} \]

\[ \int \chi_A = m(A). \]

We first show that \( C \) contains all boxes. Given a box \( A \), choose boxes \( B \subseteq \mathbb{R}^p \) and \( C \subseteq \mathbb{R}^q \) such that \( A = B \times C \). If \( y \in \mathbb{R}^q \) then

\[ A_y = \begin{cases} B & \text{if } y \in C \\ \emptyset & \text{if } y \notin C, \end{cases} \]

so \( A_y \in \mathcal{M} \). Then \( m(A_y) = m(B)\chi_C(y) \), so \( y \mapsto m(A_y) \) is measurable, and

\[ \int m(A_y) \, dy = \int m(B)\chi_C = m(B)m(C) = m(A). \]

Next we show that \( C \) is closed under finite disjoint unions, which will imply that \( \mathcal{E} \subseteq C \). By induction it suffices to show that if \( A, B \in \mathcal{C} \) with \( A \cap B = \emptyset \) then \( A \cup B \in \mathcal{C} \). We have

\[ (A \cup B)_y = A_y \cup B_y \in \mathcal{M} \quad \text{a.e. } y. \]

Also, \( A_y \cap B_y = \emptyset \), so

\[ y \mapsto m((A \cup B)_y) = m(A_y) + m(B_y) \quad \text{is measurable.} \]

Then

\[ \int m((A \cup B)_y) \, dy = \int m(A_y) \, dy + \int m(B_y) \, dy \]

\[ = m(A) + m(B) = m(A \cup B). \]

Next we show that \( C \) is closed under countable increasing unions, which will imply that \( C \) contains all open sets. Let \( \{A^n\}_{n=1}^\infty \subseteq \mathcal{C} \) with \( A^1 \subseteq A^2 \subseteq \cdots \), and put \( A = \bigcup_n A^n \). Then \( A^n \colon (A^n)_y \in \mathcal{M} \) for all \( n \) and a.e. \( y \), so

\[ A_y = \bigcup_n A^n_y \in \mathcal{M} \quad \text{a.e. } y. \]

Since \( m(A^n_y) \uparrow m(A_y) \) a.e. \( y \),

\[ y \mapsto m(A_y) \quad \text{is measurable.} \]

Then by the Monotone Convergence Theorem

\[ \int m(A_y) \, dy = \lim \int m(A^n_y) \, dy = \lim m(A^n) = m(A). \]
Next we show that $A, B \in \mathcal{C}$ with $A \subseteq B$ and $A$ bounded implies $B \setminus A \in \mathcal{C}$. We have

$$(B \setminus A)_y = B_y \setminus A_y \in \mathcal{M} \quad \text{a.e. } y.$$  

Choose boxes $C \subseteq \mathbb{R}^p$ and $D \subseteq \mathbb{R}^q$ such that $A \subseteq C \times D$. Then

$$A_y \subseteq C \quad \text{for all } y \in \mathbb{R}^q,$$

so

$$m(A_y) \leq m(C) < \infty \quad \text{a.e. } y.$$  

Hence

$$m((B \setminus A)_y) = m(B_y) - m(A_y) \quad \text{a.e. } y,$$

so $y \mapsto m((B \setminus A)_y)$ is measurable. Also

$$\int m(A_y) \, dy = m(A) < \infty,$$

so

$$\int m((B \setminus A)_y) \, dy = \int m(B_y) \, dy - \int m(A_y) \, dy = m(B) - m(A)$$

$$= m(B \setminus A).$$

Next we show that $\mathcal{C}$ is closed under bounded countable decreasing intersections, which will imply that $\mathcal{C}$ contains all bounded $G_{\delta}$'s, since every bounded $G_{\delta}$ can be expressed as a countable decreasing intersection of bounded open sets. Let $\{A_n\}^\infty_{n=1} \subseteq \mathcal{C}$ with $A_1$ bounded and $A_1 \supseteq A_2 \supseteq \cdots$. For each $n \in \mathbb{N}$ put $B_n = A_1 \setminus A_n$. Then $\{B_n\}^\infty_{n=1} \subseteq \mathcal{C}$ and $B_1 \subseteq B_2 \subseteq \cdots$, so $\bigcup_n B_n \in \mathcal{C}$. Since $\bigcup_n B_n$ is bounded,

$$\bigcap_n A_n = A_1 \setminus \bigcup_n B_n \in \mathcal{C}.$$

Next we show that $\mathcal{C}$ contains all bounded sets of measure 0. Given a bounded set $A$ such that $m(A) = 0$, choose a bounded $G_{\delta}$ $B \supseteq A$ such that $m(B) = 0$. Then $B \in \mathcal{C}$, so

$$0 = m(B) = \int m(B_y) \, dy,$$

hence $m(B_y) = 0$ a.e. $y$. Since $A_y \subseteq B_y$ for all $y \in \mathbb{R}^q$, $m(A_y) = 0$ a.e. $y$. Thus

$$\int m(A_y) \, dy = 0 = m(A).$$

Next we show that $\mathcal{C}$ contains all bounded measurable sets. Given a bounded set $A \in \mathcal{M}$, choose a bounded $G_{\delta}$ $B \supseteq A$ such that $m(B \setminus A) = 0$. Then

$$A = B \setminus (B \setminus A) \in \mathcal{C}.$$
We now show that \( \mathcal{C} = \mathcal{M} \). Given \( A \in \mathcal{M} \), for each \( n \in \mathbb{N} \) put \( A_n = A \cap B_n(0) \). Then each \( A_n \in \mathcal{M} \) is bounded, so \( A_n \in \mathcal{C} \). Hence \( A = \bigcup_n A_n \in \mathcal{C} \) since \( A_1 \subseteq A_2 \subseteq \cdots \).

Next we show that \( S^+ \subseteq \mathcal{F} \), and since \( \mathcal{C} = \mathcal{M} \) it suffices to show that \( f_1, \ldots, f_k \in \mathcal{F} \) and \( c_1, \ldots, c_k \geq 0 \) imply \( f := \sum_1^k c_n f_n \in \mathcal{F} \). We have:

- \( f(\cdot, y) = \sum_1^k c_n f_n(\cdot, y) \in L^+ \) a.e. \( y \),
- \( y \mapsto \int f(x, y) \, dx = \sum_1^k c_n \int f_n(x, y) \, dx \) is measurable, and
- \( \iint f(x, y) \, dx \, dy = \sum_1^k c_n \iint f_n(x, y) \, dx \, dy = \sum_1^k c_n \int f_n = \int f \).

Finally, given \( f \in L^+ \), choose \( \{\phi_n\} \subseteq S^+ \) such that \( \phi_n \uparrow f \). Then

\[
\phi_n(\cdot, y) \uparrow f(\cdot, y) \in L^+ \quad \text{a.e. } y,
\]

so by the Monotone Convergence Theorem

\[
y \mapsto \int f(x, y) \, dx = \lim \int \phi_n(x, y) \, dx \quad \text{is measurable.}
\]

Again by the Monotone Convergence Theorem (twice),

\[
\iint f(x, y) \, dx \, dy = \lim \iint \phi_n(x, y) \, dx \, dy = \lim \phi_n = \int f.
\]

A similar argument shows the other part. 

\[\square\]

**Theorem 7.3** (Fubini’s Theorem). **Same as Tonelli, but \( L^1 \) instead of \( L^+ \).**

**Proof.** Let \( f \in L^1 \). Then \( f^+ \in L^1 \), so by Tonelli’s Theorem

\[
\int f^+ = \iint f^+(x, y) \, dx \, dy < \infty,
\]

so

\[
y \mapsto \int f^+(x, y) \, dx \quad \text{is integrable},
\]

hence \( \int f^+(x, y) \, dx < \infty \) a.e. \( y \). Thus \( f^+(\cdot, y) \in L^1 \) a.e. \( y \), and similarly for \( f^- \). Then

\[
f(\cdot, y) = f^+(\cdot, y) - f^-(\cdot, y) \in L^1 \quad \text{a.e. } y,
\]

and then

\[
y \mapsto \int f(x, y) \, dx = \int f^+(x, y) \, dx - \int f^-(x, y) \, dx \quad \text{is integrable.}
\]

Then, by Tonelli’s Theorem again,

\[
\iint f(x, y) \, dx \, dy = \iint f^+(x, y) \, dx \, dy - \iint f^-(x, y) \, dx \, dy
\]

\[
= \int f^+ - \int f^- = \int f.
\]

\[\square\]
MAT 473 LECTURES

8. CHANGE OF VARIABLES

Lemma 8.1. If $T: \mathbb{R}^p \to \mathbb{R}^p$ is a rearrangement of coordinates and $f \in L^+$, then $f \circ T \in L^+$ and $\int f = \int f \circ T$, and similarly for $L^1$ instead of $L^+$.

Proof. We first show that for all $A \in \mathcal{M}$,

$$T(A) \in \mathcal{M} \quad \text{and} \quad m(T(A)) = m(A).$$

Note that since $T^{-1}$ is continuous, $T(A) \in \mathcal{M}$ for all $A \in \mathcal{B}$. If $A$ is a box, then $T(A)$ is also a box, and $m(T(A)) = m(A)$.

Let $A \in \mathcal{B}$. For any open boxes $B_1, B_2, \ldots \text{ with } A \subseteq \bigcup_n B_n$,

$$m(T(A)) \leq \sum_n m(T(B_n)) = \sum m(B_n),$$

so $m(T(A)) \leq m(A)$. Since $T^{-1}$ has the same properties as $T$, we also have

$$m(A) = m(T^{-1} \circ T(A)) = m(T(A)).$$

Thus $m(T(A)) = m(A)$. In particular, if $A \in \mathcal{B}$ and $m(A) = 0$, then $m(T(A)) = 0$.

Next let $A \in \mathcal{M}$ with $m(A) = 0$. Choose $B \in \mathcal{B}$ with $A \subseteq B$ and $m(B) = 0$. Then

$$m^*(T(A)) \leq m(T(B)) = 0$$

hence $T(A) \in \mathcal{M}$ and $m(T(A)) = 0 = m(A)$.

Let $A \in \mathcal{M}$. Then $A = B \cup C$ with $B \in \mathcal{B}$, $m(C) = 0$, and $B \cap C = \emptyset$.

Thus $T(A) = T(B) \cup T(C) \in \mathcal{M}$, and $T(A) \cap T(B) = \emptyset$ since $T$ is 1-1, so

$$m(T(A)) = m(T(B)) + m(T(C)) = m(T(B)) = m(B) = m(A).$$

Now, if $A \in \mathcal{M}$ then $T^{-1}(A) \in \mathcal{M}$ and $\chi_A \circ T = \chi_{T^{-1}(A)} \in L^+$, so by the above we have

$$\int \chi_A = m(A) = m(T^{-1}(A)) = \int \chi_{T^{-1}(A)} = \int \chi_A \circ T.$$

Let $\phi \in S^+$, and write $\phi = \sum_{i=1}^k c_n \chi_{A_n}$ with $c_1, c_2, \ldots c_n \geq 0$ and $A_1, A_2, \ldots, A_n \in \mathcal{M}$. Then $\phi \circ T = \sum_{i=1}^k c_n \chi_{A_n} \circ T \in L^+$ and

$$\int \phi = \sum_{i=1}^k c_n \int \chi_{A_n} = \sum_{i=1}^k c_n \int \chi_{A_n} \circ T = \int \sum_{i=1}^k c_n \chi_{A_n} \circ T = \int \phi \circ T.$$

Given $f \in L^+$, choose $(\phi_n) \subseteq S^+$ such that $\phi_n \uparrow f$. Then $\phi_n \circ T \uparrow f \circ T$, so $f \circ T \in L^+$. Moreover, by the Monotone Convergence Theorem

$$\int f = \lim \int \phi_n = \lim \int \phi_n \circ T = \int f \circ T.$$

For the other part, let $f \in L^1$. Then $f \circ T = f^+ \circ T - f^- \circ T$. Since $f^+ \circ T, f^- \circ T \in L^+$ and $f^+, f^- \in L^1$, we have $f^+ \circ T, f^- \circ T \in L^1$. Hence
\[ f \circ T \in L^1 \text{ and} \]
\[ \int f = \int f^+ - \int f^- = \int f^+ \circ T - \int f^- \circ T \]
\[ = \int (f^+ \circ T - f^- \circ T) = \int f \circ T. \]

\[ \square \]

**Lemma 8.2.** (i) Same as Lemma 8.1, but with \( T(x) = x + a \) for some \( a \in \mathbb{R}^p. \)

(ii) Same as Lemma 8.1, but with
\[ T(x_1, \ldots, x_j, \ldots, x_p) = (x_1, \ldots, cx_j, \ldots, x_p) \]
for some nonzero \( c \in \mathbb{R} \) and \( 1 \leq j \leq p, \) and
\[ \int f = |c| \int f \circ T. \]

**Proof.** (i) Same strategy as Lemma 8.1.

(ii) Same strategy as Lemma 8.1, after noting that if \( A \) is a box then \( T(A) \)
is a box and
\[ |T(A)| = |c||A|. \]

\[ \square \]

**Theorem 8.3** (Fubini). If either \( f \in L^+ \) or \( f \in L^1, \) and if \( (i_1, \ldots, i_p) \) is a rearrangement of \( (1, \ldots, p), \) then
\[ \int f = \int \cdots \int f(x_1, \ldots, x_p) \, dx_{i_1} \cdots dx_{i_p}. \]

**Proof.** By Lemma 8.1, without loss of generality \( (i_1, \ldots, i_p) = (1, \ldots, p). \) By Tonelli’s or Fubini’s Theorem (whichever is appropriate),
\[ \int f = \int \int f(x_1, (x_2, \ldots, x_p)) \, dx_1 \, d(x_2, \ldots, x_p). \]
The result now follows by an induction argument. \[ \square \]

**Lemma 8.4.** Same as Lemma 8.1, but with
\[ T(x_1, \ldots, x_j, \ldots, x_p) = (x_1, \ldots, x_j + x_k, \ldots, x_p) \]
for some \( j \neq k. \)

**Proof.** We use roughly the same strategy as Lemma 8.1, but we need to argue a little differently that if \( A \in B \) then \( m(T(A)) = m(A), \) because this time \( T \) does not map boxes to boxes. Let \( f = \chi_{T(A)}. \) Then \( f, f \circ T \in L^+, \)
and by Theorem 8.3 and the one-variable version of Lemma 8.2 (i) we have

\[ m(A) = \int \chi_A = \int f \circ T \]

\[ = \int \cdots \int f(x_1, \ldots, x_j + x_k, \ldots, x_p) dx_j dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_p \]

\[ = \int \cdots \int f(x_1, \ldots, x_j, \ldots, x_p) dx_j dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_p \]

\[ = \int f = m(T(A)). \]

Now continue as in Lemma 8.1. \( \square \)

**Corollary 8.5.** If \( T \in GL(p) \) and \( f \in L^+ \), then \( f \circ T \in L^+ \) and

\[ \int f = |\det T| \int f \circ T. \]

Similarly for \( L^1 \) instead of \( L^+ \).

**Proof.** By the preceding lemmas, the conclusion holds for \( T \) of certain types.

By linear algebra every \( T \in GL(p) \) is a finite composition of functions of these types. If the conclusion holds for both \( T_1 \) and \( T_2 \), then

\[ f \in L^+ \Rightarrow f \circ T_1 \in L^+ \Rightarrow f \circ T_1 \circ T_2 \in L^+, \]

and

\[ \int f = |\det T_1| \int f \circ T_1 \]

\[ = |\det T_1||\det T_2| \int f \circ T_1 \circ T_2 \]

\[ = |\det(T_1T_2)| \int f \circ (T_1T_2), \]

and similarly for \( L^1 \). The result follows from an induction argument. \( \square \)

**Lemma 8.6.** Every open set in \( \mathbb{R}^p \) is a countable union of closed cubes with disjoint interiors.

**Proof.** Let \( A \subseteq \mathbb{R}^p \) be open. For each \( n \in \mathbb{N} \) let \( C_n \) denote the family of closed cubes of side \( 2^{-n} \) with vertices in \( (2^{-n} \mathbb{Z})^p \). Note that for each \( n \in \mathbb{N} \) we have \( \mathbb{R}^p = \bigcup_{C \in C_n} C \), and whenever \( C \in C_n \) and \( D \in C_k \) with \( n \leq k \) we have either \( C \subseteq D \) or \( C^o \cap D^o = \emptyset \). Define

\[ B_n = \begin{cases} \bigcup \{ C \in C_1 : C \subseteq A \} & \text{if } n = 1 \\ \bigcup \{ C \in C_n : C \subseteq A \setminus \bigcup_{k < n} B_k \} & \text{if } n > 1. \end{cases} \]

Then \( \bigcup_n B_n \) is a countable union of closed cubes contained in \( A \) with disjoint interiors. Let \( x \in A \), and choose an open box \( D \) such that \( x \in D \subseteq A \). Put

\[ r = \min \{|y_i - x_i| : y \notin D, i = 1, \ldots, p\}. \]
Then $r > 0$. Now choose $n \in \mathbb{N}$ such that $2^{-n} < r$, and then choose $C \in C_n$ such that $x \in C$. If $y \in C$ then

$$|y_i - x_i| \leq 2^{-n} < r$$

for all $i = 1, \ldots, p$, so $y \in D$. Hence $C \subseteq A$. By construction,

$$C \subseteq \bigcup_{1}^{n} B_k.$$

Thus $x \in \bigcup_{1}^{\infty} B_k$. \hfill \square

**Theorem 8.7** (Change of Variables Theorem). Let $U \subseteq \mathbb{R}^p$ be open, $g : U \to \mathbb{R}^p$ be a diffeomorphism, and $f \in L^+(g(U))$. Then $f \circ g \in L^+(U)$ and

$$\int_{g(U)} f = \int_{U} f \circ g |\det g'|.$$

Similarly for $L^1$ instead of $L^+.$

**Proof.** Let $A \in \mathcal{M}$ with $A \subseteq U$. Our first objective is to show that $g(A) \in \mathcal{M}$ and

$$m(g(A)) \leq \int_{A} |\det g'|.$$

We achieve this in a sequence of steps for different classes of $A$. First assume $A$ is a closed cube, and let $t$ be the center of $A$. Since $A$ is compact and $g$ is continuous, $g(A) \in \mathcal{M}$. If $x \in A$ and $i \in \{1, \ldots, p\}$ then by the Mean Value Theorem there exists $y$ on the line segment joining $x$ and $t$ such that

$$g_i(x) - g_i(t) = g_i'(y)(x - t) = \sum_{j} D_j g_i(y)(x_j - t_j).$$

For any $p \times p$ matrix $S = [s_{ij}]$ put

$$\|S\|_* := \max_{i} \sum_{j} |s_{ij}|,$$

and for $T \in GL(p)$ put

$$\|T\|_* := \|T\|_*.$$

Then for all $x \in A$ we have

$$|g_i(x) - g_i(t)| \leq \sum_{j} |D_j g_i(y)||x_j - t_j|$$

$$\leq \|g'(y)\|_* \frac{\text{side of } A}{2}$$

$$\leq \left(\sup_{A} \|g'\|_*\right) \frac{\text{side of } A}{2}.$$

So, $g(A)$ is contained in a cube of side

$$\left(\sup_{A} \|g'\|_*\right) \text{(side of } A),$$
hence

\[ m(g(A)) \leq \left( \sup_A \| g' \|_* \right)^p m(A). \]

Let \( T \in GL(p) \), and apply the above to \( T^{-1} \circ g \) instead of \( g \):

\[ m(g(A)) = m(T \circ T^{-1} \circ g(A)) \]
\[ = | \det T | m(T^{-1} \circ g(A)) \]
\[ \leq | \det T | \left( \sup_{y \in A} \| T^{-1} g'(y) \|_* \right)^p m(A). \]

Since \( g \) is \( C^1 \), the function

\[ (x, y) \mapsto \left( \| g'(x)^{-1} g'(y) \|_* \right)^p \]

is uniformly continuous on \( A^2 \), so given \( \epsilon > 0 \) we can choose \( \delta > 0 \) such that for all \( (x, y), (u, v) \in A^2 \) we have

\[ \| (x, y) - (u, v) \| < \delta \Rightarrow \left| \left( \| g'(x)^{-1} g'(y) \|_* \right)^p - \left( \| g'(u)^{-1} g'(v) \|_* \right)^p \right| < \epsilon, \]

so that in particular, for all \( x, y \in A \) we have

\[ \| x - y \| < \delta \Rightarrow \left( \| g'(x)^{-1} g'(y) \|_* \right)^p < 1 + \epsilon, \]

since

\[ \| x - y \| = \| (x, y) - (x, x) \| \quad \text{and} \quad \left( \| g'(x)^{-1} g'(x) \|_* \right)^p = 1. \]

Divide \( A \) into closed subcubes \( \{ B_n \}_n^k \) with disjoint interiors, sides smaller than \( \delta / \sqrt{p} \), and centers \( \{ t_n \}_1^k \). Then

\[ m(g(A)) = m\left( \bigcup_1^k g(B_n) \right) \leq \sum_1^k m(g(B_n)) \]
\[ \leq \sum_1^k | \det g'(t_n) | \left( \sup_{y \in B_n} \| g'(t_n)^{-1} g'(y) \|_* \right)^p m(B_n) \]
\[ \leq \sum_1^k | \det g'(t_n) | (1 + \epsilon) m(B_n) \]
\[ = (1 + \epsilon) \int_A \sum_1^k | \det g'(t_n) | \chi_{B_n}. \]

Since \( g' \) is continuous, letting \( \delta \to 0 \) gives

\[ m(g(A)) \leq (1 + \epsilon) \int_A | \det g' |, \]

since the \( B_n \) have disjoint interiors.
Then letting $\epsilon \to 0$, we get

$$m(g(A)) \leq \int_A |\det g'|.$$  

Next, assume $A$ is open, and choose closed cubes $\{B_n\}_1^\infty$ with disjoint interiors such that $A = \bigcup_n B_n$. Since $g(B_n) \in \mathcal{M}$ for all $n$, we have $g(A) = \bigcup_n g(B_n) \in \mathcal{M}$ and

$$m(g(A)) = m\left(\bigcup_n g(B_n)\right) \leq \sum m(g(B_n))$$

$$\leq \sum \int_{B_n} |\det g'| = \int_A |\det g'|,$$

again since the $B_n$ have disjoint interiors.

Now, for all $k \in \mathbb{N}$ put

$$W_k = \{x \in U : |\det g'(x)| < k\}.$$  

Then each $W_k$ is open, $W_1 \subseteq W_2 \subseteq \cdots$, and $U = \bigcup_k W_k$. Assume $A$ is a bounded Borel subset of some $W_k$. Choose bounded open sets $\{B_n\}_1^\infty$ such that:

- $W_k \supseteq B_1 \supseteq B_2 \supseteq \cdots$,
- $B := \bigcap_n B_n \supseteq A$, and
- $m(B \setminus A) = 0$.

Since $A \in \mathcal{B}$ and $g^{-1}$ is continuous, $g(A) \in \mathcal{M}$. Further, since

$$\int_{B_1} |\det g'| \leq km(B_1) < \infty,$$

by the Dominated Convergence Theorem (or continuity from above of the measure $C \mapsto \int_C |\det g'|$) we have

$$m(g(A)) \leq m(g(B)) \leq \lim m(g(B_n))$$

$$\leq \lim \int_{B_n} |\det g'| = \int_B |\det g'|$$

$$= \int_A |\det g'|.$$

Next, assume $A$ is bounded and Borel. For each $k \in \mathbb{N}$ put $A_k = A \cap W_k$. Then each $A_k$ is bounded and Borel, $A_1 \subseteq A_2 \subseteq \cdots$, and $A = \bigcup_k A_k$, so by the Monotone Convergence Theorem (or continuity from below of the measure $C \mapsto \int_C |\det g'|$) we have

$$m(g(A)) = \lim m(g(A_k)) \leq \lim \int_{A_k} |\det g'| = \int_A |\det g'|.$$

Next, assume $A \in \mathcal{B}$. For each $n \in \mathbb{N}$ put $A_n = A \cap B_n(0)$. Then each $A_n$ is bounded and Borel, $A_1 \subseteq A_2 \subseteq \cdots$, and $A = \bigcup_n A_n$, so again by the
Monotone Convergence Theorem

\[ m(g(A)) = \lim_{n \to \infty} m(g(A_n)) \leq \lim_{n \to \infty} \int_{A_n} |\det g'| = \int_A |\det g'|. \]

Next, assume \( m(A) = 0 \), and choose \( B \in \mathcal{B} \) such that \( A \subseteq B \subseteq U \) and \( m(B) = 0 \). Then

\[ m(g(B)) \leq \int_B |\det g'| = 0 \]

and \( g(A) \subseteq g(B) \), so

\[ m(g(A)) = 0 = \int_A |\det g'|. \]

Now, in general we can choose \( B \in \mathcal{B} \) such that \( B \subseteq A \) and \( m(A \setminus B) = 0 \). Then \( g(B) \in \mathcal{M} \) and

\[ 0 = m(g(A \setminus B)) = m(g(A) \setminus g(B)). \]

So, \( g(A) \in \mathcal{M} \) and

\[ m(g(A)) = m(g(B)) + m(g(A \setminus B)) \leq \int_B |\det g'| = \int_A |\det g'|. \]

We have thus accomplished our first objective.

Next, let \( \phi = \sum_{n=1}^{k} c_n \chi_{A_n} \) with each \( c_n \geq 0 \) and \( A_n \subseteq g(U) \) measurable. By the above reasoning applied to \( g^{-1} \) instead of \( g \), each set \( g^{-1}(A_n) \) is measurable and

\[ \phi \circ g = \sum_{n=1}^{k} c_n \chi_{A_n} \circ g = \sum_{n=1}^{k} c_n \chi_{g^{-1}(A_n)}, \]

and then

\[ \int_{g(U)} \phi = \sum_{n=1}^{k} c_n \int_{g(U)} \chi_{A_n} = \sum_{n=1}^{k} c_n m(A_n) \]

\[ \leq \sum_{n=1}^{k} c_n \int_{g^{-1}(A_n)} |\det g'| \]

\[ = \sum_{n=1}^{k} c_n \int_{U} \chi_{g^{-1}(A_n)} |\det g'| \]

\[ = \int_{U} \phi \circ g |\det g'|. \]

Now let \( f \in L^+(g(U)) \), and choose \( (\phi_n) \subseteq S^+ \) such that \( \phi_n \uparrow f \). Then \( \phi_n \circ g \uparrow f \circ g \), so \( f \circ g \in L^+(U) \). By the Monotone Convergence Theorem

\[ \int_{g(U)} f = \lim_{n \to \infty} \int_{g(U)} \phi_n \leq \lim_{n \to \infty} \int_{U} \phi_n \circ g |\det g'| = \int_{U} f \circ g |\det g'|. \]
Apply this with $g^{-1}$ and $f \circ g \det g'$ instead of $g$ and $f$:

\[
\int_U f \circ g \det g' = \int_{g^{-1}(g(U))} f \circ g \det g' \\
\leq \int_{g(U)} f \circ g \circ g^{-1} \det g' \circ g^{-1} \| \det(g^{-1})' \| \\
= \int_{g(U)} f.
\]

This proves the result for $f \in L^+$, and for $f \in L^1$ we apply a familiar argument to $f^+$ and $f^-$.

9. MANIFOLDS

**Definition 9.1.** Let $U \subseteq \mathbb{R}^p$ be open and $f: U \to \mathbb{R}^q$ be $C^1$. We say $f$ is:

(i) an immersion if $f'(x)$ is 1-1 for all $x \in U$;

(ii) a submersion if $f'(x)$ is onto for all $x \in U$;

**Definition 9.2.** Let $V \subseteq \mathbb{R}^k$ be open and $f: V \to \mathbb{R}^p$. We say $f$ is a parameterization of $f(V)$ if $f$ is an immersion which is also a homeomorphism of $V$ onto $f(V)$.

**Definition 9.3.** Let $M \subseteq \mathbb{R}^p$. We say $M$ is a manifold if for all $a \in M$ there exists an open set $U \subseteq \mathbb{R}^p$ such that $a \in U$ and $U \cap M$ has a parameterization.

**Definition 9.4.** Let $M, U \subseteq \mathbb{R}^p$, with $U$ open and $M \subseteq U$, and let $f: U \to \mathbb{R}^p$ be a diffeomorphism. We say $f$ is a straightening of $M$ if $f(M) = f(U) \cap (\mathbb{R}^k \times \{0\})$ (for some $k$).

**Definition 9.5.** Let $M \subseteq \mathbb{R}^p$. We say $M$ can be locally straightened if for all $a \in M$ there exists an open set $U \subseteq \mathbb{R}^p$ such that $a \in U$ and $U \cap M$ has a straightening.

**Remark 9.6.** In the preceding definition, without loss of generality we can take $U$ to be the domain of a straightening $f$ of $U \cap M$, so that

\[
f(U \cap M) = f(U) \cap (\mathbb{R}^k \times \{0\}).
\]

Also, in the spirit of the preceding definition we could say $M$ is a manifold if it is “locally parameterizable”.

**Definition 9.7.** Let $U \subseteq \mathbb{R}^p$ be open, $f: U \to \mathbb{R}^l$ be a submersion, and $b \in f(U)$. We say $f^{-1}(\{b\})$ is a level set of $f$.

**Remark 9.8.** In the above definition, without loss of generality we can assume whenever convenient that $b = 0$.

**Definition 9.9.** Let $M \subseteq \mathbb{R}^p$. We say $M$ is locally a level set if for all $a \in M$ there exists an open set $U \subseteq \mathbb{R}^p$ such that $a \in U$ and $U \cap M$ is a level set.
Remark 9.10. The above properties are invariant under diffeomorphisms. More precisely, if $M, U \subseteq \mathbb{R}^p$, with $U$ open, and $f: U \to \mathbb{R}^p$ is a diffeomorphism, then $U \cap M$ is parameterizable if and only if $f(U \cap M)$ is, and similarly for being straightenable or being a level set.

Theorem 9.11. For $M \subseteq \mathbb{R}^p$, the following are equivalent:

(i) $M$ is a manifold;
(ii) $M$ can be locally straightened;
(iii) $M$ is locally a level set.

Proof. (i) ⇒ (ii). First assume $M$ is a manifold, and let $a \in M$. Choose open sets $V \subseteq \mathbb{R}^k$ and $U \subseteq \mathbb{R}^p$, and a parameterization $f: V \to \mathbb{R}^p$ such that $a \in U$ and $f(V) = U \cap M$. Let $b = f^{-1}(a)$. Since $f'(b)$ is 1-1, there exists an invertible $k \times k$ submatrix of $f'(b)$, and after a rearrangement of coordinates in $\mathbb{R}^p$ (which, as we have remarked above, does not change the important properties of $M$), without loss of generality we can assume the first $k$ rows of $f'(b)$ form an invertible matrix. Define $g: V \times \mathbb{R}^l \to \mathbb{R}^p$ by

$$g(x, y) = f(x) + (0, y),$$

where we identify $\mathbb{R}^p$ with $\mathbb{R}^k \times \mathbb{R}^l$. Then $g$ is $C^1$ and

$$g'(b, 0) = \begin{bmatrix} f_1'(b) & 0 \\ f_2'(b) & I \end{bmatrix},$$

where here $f_1'$ means the derivative of the first $k$ components of $f$ and $f_2'$ the derivative of the last $l$ components. Thus $g'(b, 0)$ is invertible. Use the Inverse Function Theorem to find an open set $W \subseteq V \times \mathbb{R}^l$ such that $(b, 0) \in W$ and $g$ is a diffeomorphism on $W$.

If we had

$$g(W \cap (\mathbb{R}^k \times \{0\})) = g(W) \cap M,$$

then the inverse of $g|W$ would be a straightening of $g(W) \cap M$. Sadly, it is possible for $g(W) \cap M$ to be too big. We must find an open subset $W_0$ of $W$ such that

$$g(W_0 \cap (\mathbb{R}^k \times \{0\})) = g(W_0) \cap M$$

and $a \in g(W_0) \cap M$.

Put

$$V_0 = \{x \in V : (x, 0) \in W\},$$

so that $W \cap (\mathbb{R}^k \times \{0\}) = V_0 \times \{0\}$. The trick is to note that, since $f^{-1}$ is continuous and $V_0$ is an open subset of $V$, $f(V_0)$ is an open subset of the metric space $f(V)$, so there exists an open set $Z \subseteq \mathbb{R}^p$ such that

$$f(V_0) = f(V) \cap Z,$$

and without loss of generality we can assume $Z \subseteq g(W) \cap U$. Put

$$W_0 = g^{-1}(Z).$$
Note that \( a \in g(W_0) \cap M \) since \( b \in V_0 \). Then \( W_0 \) is open in \( W \), and

\[
g(W_0) \cap M = Z \cap M = Z \cap U \cap M = Z \cap f(V)
\]

\[
f(V_0) = Z \cap f(V_0) = g(W_0) \cap g(V_0 \times \{ 0 \})
\]

\[
g(W_0) \cap g(W \cap (\mathbb{R}^k \times \{ 0 \}))
\]

\[
g(W_0 \cap W \cap (\mathbb{R}^k \times \{ 0 \})) \quad \text{since } g \text{ is 1-1 on } W
\]

\[
g(W_0 \cap (\mathbb{R}^k \times \{ 0 \}))
\]

as desired.

(ii) \( \Rightarrow \) (iii). Assume \( U \subseteq \mathbb{R}^p \) is open and \( f: U \to \mathbb{R}^p \) is a straightening of \( U \cap M \). Then \( f \) is a diffeomorphism and

\[
f(U \cap M) = f(U) \cap (\mathbb{R}^k \times \{ 0 \}).
\]

As we mentioned before the statement of the theorem, we can immediately conclude that \( U \cap M \) is a level set, since \( f(U) \cap (\mathbb{R}^k \times \{ 0 \}) \) is (namely, the latter is a level set of the projection onto the last \( p - k \) coordinates).

(iii) \( \Rightarrow \) (i). Finally, assume \( U \subseteq \mathbb{R}^p \) is open, \( f: U \to \mathbb{R}^l \) is a submersion, \( a \in M \cap U \), and \( M \cap U = f^{-1}(\{ 0 \}) \). Identifying \( \mathbb{R}^p \) with \( \mathbb{R}^k \times \mathbb{R}^l \), write

\[
f'(a) = \begin{bmatrix} D_1 f(a) & D_2 f(a) \end{bmatrix},
\]

where we recall that in a situation such as this \( D_1 f \) means the derivative of \( f \) with respect to the first \( k \) coordinates and \( D_2 f \) the derivative with respect to the last \( l \) coordinates. After a suitable rearrangement of coordinates in \( \mathbb{R}^p \), without loss of generality \( D_2 f(a) \) is invertible. Use the Implicit Function Theorem to find an open set \( W \subseteq U \) such that \( a \in W \) and \( W \cap U \) is the graph of a \( C^1 \) function \( g: V \to \mathbb{R}^l \), for some open set \( V \subseteq \mathbb{R}^k \). Define \( h: V \to \mathbb{R}^p \) by

\[
h(x) = (x, g(x)).
\]

Then \( h \) is \( C^1 \) and

\[
h'(x) = \begin{bmatrix} I \\ g'(x) \end{bmatrix}
\]

is 1-1 for all \( x \in V \). Since \( h \) has a continuous inverse (namely, projection onto the first \( k \) coordinates), \( h \) is a homeomorphism onto \( h(V) \). Thus \( h \) is a parameterization of \( h(V) = M \cap W \). \( \square \)

Remark 9.12. It is occasionally useful to allow the following rather trivial special cases of the concept of a manifold: every open subset of \( \mathbb{R}^p \) can be regarded as a manifold with the identity map as a parameterization, and at the opposite extreme any subset of \( \mathbb{R}^p \) with no cluster points in \( \mathbb{R}^p \) can be regarded as a manifold by parameterizing each point \( x \) using the map \( 0 \mapsto x \) from \( \mathbb{R}^0 := \{ 0 \} \) to \( \mathbb{R}^p \). However, in practice these extreme cases are usually tacitly excluded.
Definition 9.13. Let \( M \) be a manifold in \( \mathbb{R}^p \), \( a \in M \), and \( f : V \to M \) a parameterization such that \( a \in f(V) \). Then \( f^{-1} : f(V) \to \mathbb{R}^k \) is called a coordinate chart of \( M \) around \( a \).

Corollary 9.14. Let \( M \) be a manifold in \( \mathbb{R}^p \), \( a \in M \), and \( \phi \) a coordinate chart of \( M \) around \( a \). Then there exists an open set \( U \subset \mathbb{R}^p \) and a \( C^1 \) function \( \tilde{\phi} \) defined on \( U \) such that \( a \in U \) and \( \tilde{\phi} \) agrees with \( \phi \) on \( U \cap M \).

Proof. After applying a straightening, without loss of generality (shrinking the domain of \( \phi \) a little if necessary), the parameterization \( \phi^{-1} \) is of the form \( \phi^{-1}(x) = (x, 0) \) for \( x \) in some open set \( V \subset \mathbb{R}^k \). Then we can take \( \tilde{\phi} \) to be a restriction of the function \( (x, y) \mapsto x : \mathbb{R}^p \to \mathbb{R}^k \).

Corollary 9.15. Let \( M \) be a manifold, \( a \in M \), and \( \phi \) and \( \psi \) two charts of \( M \) around \( a \). Then \( \phi \) and \( \psi \) take values in the same Euclidean space, and

\[
\psi \circ \phi^{-1} : \phi(\text{dom } \phi \cap \text{dom } \psi) \to \psi(\text{dom } \phi \cap \text{dom } \psi)
\]

is a diffeomorphism.

Proof. Suppose \( \phi \) and \( \psi \) take values in \( \mathbb{R}^k \) and \( \mathbb{R}^j \), respectively. For convenience, without loss of generality \( \phi \) and \( \psi \) have the same domain (containing \( a \)). Let \( V = \text{ran } \phi \) and \( W = \text{ran } \psi \). Then \( V \) and \( W \) are open subsets of \( \mathbb{R}^k \) and \( \mathbb{R}^j \), respectively. Shrinking the common domain even further if necessary, we can assume \( \psi \) extends to a \( C^1 \) function \( \psi \) on some open subset of \( \mathbb{R}^p \) containing \( a \). Since \( \phi^{-1} \) is \( C^1 \), so is the composition

\[
g := \psi \circ \phi^{-1} = \psi \circ \phi^{-1} : V \to \mathbb{R}^j.
\]

Symmetrically,

\[
h := \phi \circ \psi^{-1} : W \to \mathbb{R}^k
\]

is also \( C^1 \). Since \( h = g^{-1} \), the Chain Rule implies that

\[
h'(g(x))g'(x) = I_k \quad \text{and} \quad g'(x)h'(g(x)) = I_j \quad \text{for each } x \in V.
\]

By linear algebra, we must have \( k = j \).

Now we see that the function \( g = \psi \circ \phi^{-1} : W \to \mathbb{R}^k \) is \( C^1 \) and 1-1, and has invertible derivative at each element of \( W \), so it is a diffeomorphism.

Definition 9.16. Let \( M \) be a manifold in \( \mathbb{R}^p \) and \( k \) a nonnegative integer. If every point of \( M \) is contained in the domain of a chart with values in \( \mathbb{R}^k \), we call \( k \) the dimension of \( M \).

Remark 9.17. The preceding corollary tells us the dimension of a manifold is well defined if it exists. The union of a line and a disjoint plane in \( \mathbb{R}^3 \) is an example of a manifold which has charts with values in different-dimensional Euclidean spaces, and consequently has no dimension. However:

Corollary 9.18. Every connected manifold has a dimension.
Proof. Let $M$ be a connected manifold in $\mathbb{R}^p$, and suppose there exists a chart with values in $\mathbb{R}^k$. Let

$$A = \{x \in M : \text{some chart around } x \text{ has values in } \mathbb{R}^k \} \quad \text{and} \quad B = \{x \in M : \text{some chart around } x \text{ has values in } \mathbb{R}^j \text{ for some } j \neq k \}.$$  

Then $A$ and $B$ are open subsets of $M$, and the preceding corollary tells us they are disjoint. Since $A \neq \emptyset$ by assumption, and $A \cup B = M$ by construction, we must have $B = \emptyset$ since $M$ is connected.

Remark 9.19. Thus, we can recognize the dimension of a manifold $M$ in $\mathbb{R}^p$ from its charts, or equivalently from its parameterizations. Unsurprisingly, we can also determine the dimension from straightenings and also from level sets. To see this, let $U$ be open in $\mathbb{R}^p$. If $f : U \to \mathbb{R}^p$ is a straightening of $U \cap M$, with $f(U \cap M) = f(U) \cap (\mathbb{R}^k \times \{0\})$, then the dimension of $M$ is $k$. On the other hand, if $U \cap M$ is a level set of $g : U \to \mathbb{R}^l$, then the dimension of $M$ is $p - 1$.

Remark 9.20. If $M$ is a $k$-dimensional manifold, coordinate charts are used to do calculus on $M$ by transferring to open subsets of $\mathbb{R}^k$. This process is called “working in local coordinates”. For example:

Definition 9.21. Let $M$ be a manifold and $f : M \to \mathbb{R}^q$. We say $f$ is $C^1$ if for all $a \in M$ there exists a chart $\phi$ around $a$ such that $f \circ \phi^{-1}$ is $C^1$.

Lemma 9.22. The above condition on $f$ is independent of the choice of the chart $\phi$.

Proof. Let $\psi$ be another chart around $a$, and without loss of generality assume $\text{dom } \phi = \text{dom } \psi$. Then

$$f \circ \phi^{-1} = f \circ \psi^{-1} \circ \psi \circ \phi^{-1}$$

and $\psi \circ \phi^{-1}$ is a diffeomorphism, so $f \circ \phi^{-1}$ is $C^1$ if and only if $f \circ \psi^{-1}$ is. \qed

Definition 9.23. Let $M$ and $N$ be manifolds and $f : M \to N$. We say $f$ is $C^1$ if for all $b \in f(M)$ there exists a chart $\phi$ of $N$ around $b$ such that $\phi \circ f$ is $C^1$.

Lemma 9.24. The above condition of $f$ is independent of the choice of the chart $\phi$.

Proof. The argument is very similar to the preceding one. \qed

Definition 9.25. Let $M$ be a manifold and $f : V \to M$ a parameterization, and suppose $a = f(b)$. The tangent space of $M$ at $a$ is

$$T_a(M) := \{(a, x) : x \in \text{ran } f'(b) \}.$$  

Lemma 9.26. The above set $T_a(M)$ is independent of the choice of the parameterization $f$.  


**Proof.** Let \( g: W \to M \) be another parameterization, and suppose \( g(c) = a \) for some \( c \in W \). Then

\[
f'(b) = (g \circ g^{-1} \circ f)'(b) = g'(g^{-1} \circ f(b)) (g^{-1} \circ f)'(b) = g'(c)(g^{-1} \circ f)'(b).
\]

Thus ran \( f'(b) \) = ran \( g'(c) \) since the linear map \( (g^{-1} \circ f)'(b) \) is invertible. \( \square \)

**Theorem 9.27.** Let \( M \) be a manifold in \( \mathbb{R}^p \) and \( (a, x) \in M \times \mathbb{R}^p \). Then the following are equivalent:

(i) \((a, x) \in T_a(M)\);
(ii) there exists an open interval \( I \subseteq \mathbb{R} \), a \( C^1 \) curve \( \phi: I \to M \), and \( s \in I \) such that

\[ \phi(s) = a \quad \text{and} \quad \phi'(s) = x; \]

(iii) there exists a submersion \( g: U \to \mathbb{R}^d \) such that \( a \in U \), \( M \cap U = g^{-1}(0) \), and \( x \in \ker g'(a) \).

**Proof.** (i) \( \Rightarrow \) (ii). First assume \( f: V \to M \) is a parameterization with

\[ f(b) = a \quad \text{and} \quad f'(b)y = x. \]

Define \( \psi: \mathbb{R} \to \mathbb{R}^k \) by

\[ \psi(t) = a + ty. \]

Then \( \psi(0) = a \) and \( \psi'(0) = y \). Choose an open interval \( I \subseteq \mathbb{R} \) containing 0 such that \( \psi(I) \subseteq V \). Then \( f \circ \psi: I \to M \) is a \( C^1 \) curve with

\[ f \circ \psi(0) = f(\psi(0)) = f(b) = a \]

and

\[ (f \circ \psi)'(0) = f'(\psi(0))\psi'(0) = f'(b)y = x. \]

(ii) \( \Rightarrow \) (iii). Next assume \( \phi: I \to M \) is a \( C^1 \) curve with \( \phi(s) = a \) and \( \phi'(s) = x \). Choose any submersion \( g: U \to \mathbb{R}^d \) such that \( a \in U \) and \( M \cap U = g^{-1}(0) \). Without loss of generality \( \phi(I) \subseteq U \). Since \( g \circ \phi \equiv 0 \),

\[ g'(a)x = g'(\phi(s))\phi'(s) = (g \circ \phi)'(s) = 0. \]

Thus \( x \in \ker g'(a) \).

(iii) \( \Rightarrow \) (i). Finally, assume \( g: U \to \mathbb{R}^d \) is a submersion, \( a \in U \), \( M \cap U = g^{-1}(0) \), and \( x \in \ker g'(a) \). Choose any parameterization \( f: V \to M \), with \( V \) open in \( \mathbb{R}^k \), such that \( a \in f(V) \). Note that we have \( p = k + l \). Let \( a = f(b) \), and without loss of generality \( f(V) \subseteq U \). Since \( g \circ f \equiv 0 \),

\[ 0 = (g \circ f)'(b) = g'(f(b))f'(b) = g'(a)f'(b). \]

Thus ran \( f'(b) \subseteq \ker g'(a) \). On the other hand, since \( g \) is a submersion and \( f \) is an immersion,

\[ \dim \text{ran } f'(b) = k \quad \text{and} \quad \dim \text{ran } g'(a) = l = p - k, \]

so

\[ \dim \text{ran } f'(b) = p - \dim \text{ran } g'(a) = \dim \ker g'(a). \]
Therefore, \( \text{ran } f'(b) = \ker g'(a) \), so there exists \( y \in \mathbb{R}^k \) such that \( f'(b)y = x \). \qed
INDEX

immersion, 52
Implicit Function Theorem, 12
integrable, 33
integral
  Lebesgue, 33
  positive function, 29
  simple function, 28
Inverse Function Theorem, 10

$L(R^p, R^q)$, 1
$L^+$, 29
$L^1$, 33
Lebesgue integral, 33
Lebesgue measure, 21
Lebesgue’s Theorem on Riemann Integrability, 39
level hypersurface, 7
level set, 52
locally a level set, 52
locally straightened, 52

$\mathcal{M}$, 20

manifold, 52
manifold: dimension, 55
Mean Value Inequality, 8
Mean Value Theorem, 7
measurable
  function, 24
  set, 20
measure, 16
  Lebesgue, 21
  outer, 19
Monotone Convergence Theorem, 31

norm of linear map
  Euclidean norm, 2
  operator norm, 1

outer measure, 19

parameterization, 52
partial derivative, 6
positive and negative parts, 25

ring of sets, 15
$S^+$, 28

set
  Borel, 22
  measurable, 20
$s$-algebra, 16
$s$-ring, 15
simple function, 27

Fubini’s Lemma, 33
$E'$, 22
Fubini’s Theorem, 44, 46
function
  integrable, 33
  measurable, 24
  simple, 27

$G_\delta$, 22
gradient vector, 7
grid, 18

higher order partial derivatives, 14

derivative, 4
  directional, 7
  partial, 6
diffeomorphism, 10
differentiable, 4
  at a point, 4
  continuously, 8
dimension of a manifold, 55
directional derivative, 7
Dominated Convergence Theorem, 35

$\mathcal{E}$, 18
equivalent functions, 35

Fatou’s Lemma, 33
$E'$, 22
Fatou’s Theorem, 44, 46
function
  integrable, 33
  measurable, 24
  simple, 27

$G_\delta$, 22
gradient vector, 7
grid, 18

higher order partial derivatives, 14

a.e., 27
additive, 16
algebra of sets, 15
almost everywhere, 27
Arithmetic in $\mathbb{R}$, 25

Borel set, 22
box, 17

$C^1$, 8
Carathéodory’s Theorem, 20
$C$, 38
Chain Rule, 5
Change of Variables Theorem, 48
Clairaut’s Theorem, 14
completeness of $L^1$, 37
Continuity from Above, 16
Continuity from Below, 16
continuously differentiable, 8
Contraction Mapping Principle, 9
coordinate chart, 55
countable subadditivity, 19
countably additive, 16
curve, 7

59
straightening, 52
submersion, 52
support, 38

tangent space, 56
tangent vector, 7
Taylor’s Theorem, 15
Tonelli’s Theorem, 41