According to (**) and Radonich's Theorem,
\[ L^m(y - X) = 0. \]
Now if \( x \in X, Dg(f(x)) \) and \( Df(x) \) exist, and so
\[ Dg(f(x)) Df(x) = D(g \circ f)(x) \]
exists. Since \( (g \circ f)(x) - x = 0 \) on \( Y \), assertion (i) implies
\[ D(g \circ f) = I \quad L^m \text{ a.e. on } Y. \]

### 3.2 Linear maps and Jacobians

We next review some basic linear algebra. Our goal thereafter will be to define the Jacobian of a map \( f : \mathbb{R}^n \to \mathbb{R}^m \).

#### 3.2.1 Linear maps

**DEFINITIONS**

(i) A linear map \( O : \mathbb{R}^n \to \mathbb{R}^m \) is orthogonal if \( (Oz) \cdot (Oy) = z \cdot y \) for all \( x, y \in \mathbb{R}^n \).

(ii) A linear map \( S : \mathbb{R}^n \to \mathbb{R}^n \) is symmetric if \( x \cdot (Sy) = (Sx) \cdot y \) for all \( x, y \in \mathbb{R}^n \).

(iii) A linear map \( D : \mathbb{R}^n \to \mathbb{R}^m \) is diagonal if there exist \( d_1, \ldots, d_n \in \mathbb{R} \) such that
\[ Dx = (d_1x_1, \ldots, d_nx_n) \text{ for all } x \in \mathbb{R}^n. \]

(iv) Let \( A : \mathbb{R}^n \to \mathbb{R}^m \) be linear. The adjoint of \( A \) is the linear map \( A^* : \mathbb{R}^m \to \mathbb{R}^n \) defined by \( x \cdot (A^* y) = (Ax) \cdot y \) for all \( x \in \mathbb{R}^n, y \in \mathbb{R}^m \).

First we recall some routine facts from linear algebra.

**THEOREM 1**

(i) \( A^* = A \).

(ii) \( (A \circ R)^* = R^* \circ A^* \).

(iii) \( O^* = O^{-1} \) if \( O : \mathbb{R}^n \to \mathbb{R}^n \) is orthogonal.

(iv) \( S^* = S \) if \( S : \mathbb{R}^n \to \mathbb{R}^n \) is symmetric.

(v) If \( S : \mathbb{R}^n \to \mathbb{R}^m \) is symmetric, there exists an orthogonal map \( O : \mathbb{R}^n \to \mathbb{R}^n \) and a diagonal map \( D : \mathbb{R}^n \to \mathbb{R}^m \) such that
\[ S = O \circ D \circ O^{-1}. \]

### 3.3 Linear maps and Jacobians

(vi) If \( O : \mathbb{R}^n \to \mathbb{R}^m \) is orthogonal, then \( n < m \) and
\[ O^* \circ O = I \quad \text{on } \mathbb{R}^n, \]
\[ O \circ O^* = I \quad \text{on } O(\mathbb{R}^n) \]

**THEOREM 2 POLAR DECOMPOSITION**

Let \( L : \mathbb{R}^m \to \mathbb{R}^m \) be a linear mapping.

(i) If \( n \leq m \), there exists a symmetric map \( S : \mathbb{R}^n \to \mathbb{R}^n \) and an orthogonal map \( O : \mathbb{R}^n \to \mathbb{R}^n \) such that
\[ L = O \circ S. \]

(ii) If \( n > m \), there exists a symmetric map \( S : \mathbb{R}^n \to \mathbb{R}^n \) and an orthogonal map \( O : \mathbb{R}^n \to \mathbb{R}^n \) such that
\[ L = S \circ O^*. \]

**proof**

1. First suppose \( n \leq m \). Consider \( C = L^* \circ L : \mathbb{R}^n \to \mathbb{R}^n \). Now
\[ (Cx) \cdot y = (L^* \circ Ly) \cdot y = Lx \cdot Ly = x \cdot L^* \circ Ly = x \cdot Cy \]
and also
\[ (Cx) \cdot x = Lx \cdot Lx \geq 0. \]
Thus \( C \) is symmetric, nonnegative definite. Hence there exist \( \mu_1, \ldots, \mu_n \geq 0 \) and an orthogonal basis \( \{x_k\}_{k=1}^n \) of \( \mathbb{R}^n \) such that
\[ Cx_k = \mu_k x_k \quad (k = 1, \ldots, n). \]

Write \( \mu_k = \lambda_k^2, \lambda_k \geq 0 \quad (k = 1, \ldots, n) \).

2. **Claim**: There exists an orthonormal set \( \{x_k\}_{k=1}^n \) in \( \mathbb{R}^n \) such that
\[ Lx_k = \lambda_k x_k \quad (k = 1, \ldots, n). \]

**Proof of Claim**: If \( \lambda_k \neq 0 \), define
\[ z_k = \frac{1}{\lambda_k} Lx_k. \]
Then if \( \lambda_k \neq 0 \),
\[ z_k \cdot z_k = \frac{1}{\lambda_k} Lx_k \cdot Lx_k = \frac{1}{\lambda_k} \lambda_k (C x_k) \cdot z_k = \frac{1}{\lambda_k} \lambda_k \lambda_k = 1. \]
Thus the set \( \{ z_k \mid \lambda_k \neq 0 \} \) is orthonormal. If \( \lambda_k = 0 \), define \( z_k \) to be any unit vector such that \( \{ z_k \}_{k=1}^n \) is orthonormal.

3. Now define

\[ S : \mathbb{R}^n \to \mathbb{R}^n \text{ by } S_{vk} = \lambda_k z_k \quad (k = 1, \ldots, n) \]

and

\[ O : \mathbb{R}^m \to \mathbb{R}^m \text{ by } O_{vk} = z_k \quad (k = 1, \ldots, n). \]

Then \( O \circ S_{vk} = \lambda_k O_{vk} = \lambda_k z_k = L_{vk} \), and so

\[ L = O \circ S. \]

The mapping \( S \) is clearly symmetric, and \( O \) is orthogonal since

\[ O_{vk} \cdot O_{wl} = z_k \cdot z_l = \delta_{kl}. \]

4. Assertion (ii) follows from our applying (i) to \( L^* : \mathbb{R}^m \to \mathbb{R}^n \).

**Definition** Assume \( L : \mathbb{R}^n \to \mathbb{R}^m \) is linear.

1. If \( n \leq m \), we write \( L = O \circ S \) as above, and we define the Jacobian of \( L \) to be

\[ [L] = |\det S|. \]

2. If \( n > m \), we write \( L = S \circ O^* \) as above, and we define the Jacobian of \( L \) to be

\[ [L] = |\det S|. \]

**Remarks**

1. It follows from Theorem 3 below that the definition of \([L]\) is independent of the particular choices of \( O \) and \( S \).

2. Clearly,

\[ [L]^2 = [L^*]^2. \]

**Theorem 3**

1. If \( n \leq m \),

\[ [L]^2 = \det(L^* \circ L). \]

2. If \( n \geq m \),

\[ [L]^2 = \det(L \circ L^*). \]

### 3.2 Linear Maps and Jacobians

**Proof**

1. Assume \( n \leq m \) and write

\[ L = O \circ S, L^* = S^* \circ O^* = S \circ O^*, \]

then

\[ L^* \circ L = S \circ O^* \circ O \circ S = S^*, \]

since \( O \) is orthogonal, and thus \( O^* \circ O = I \). Hence

\[ \det(L^* \circ L) = (\det S)^2 = [L]^2. \]

2. The proof of (ii) is similar.

**Theorem 3** provides us with a useful Pappus formula for computing \([L]\), which we suggest with the Riesz-Fischer formula below.

**Definitions**

1. If \( n \leq m \), we define

\[ \Lambda(m, n) = \{ \lambda : \{1, \ldots, n\} \to \{1, \ldots, m\} \mid \lambda \text{ is increasing} \}. \]

2. For each \( \lambda \in \Lambda(m, n) \), we define \( P_\lambda : \mathbb{R}^m \to \mathbb{R}^n \) by

\[ P_\lambda (x_1, \ldots, x_m) = (x_{\lambda(1)}, \ldots, x_{\lambda(n)}). \]

**Remark** For each \( \lambda \in \Lambda(m, n) \), there exists an \( n \)-dimensional subspace

\[ S_\lambda = \text{span}(e_{\lambda(1)}, \ldots, e_{\lambda(n)}) \subset \mathbb{R}^m \]

such that \( P_\lambda \) is the projection of \( \mathbb{R}^m \) onto \( S_\lambda \).

**Theorem 4** Binet-Cauchy Formula

Assume \( n \leq m \) and \( L : \mathbb{R}^n \to \mathbb{R}^m \) is linear. Then

\[ \left[ [L]^2 \right] = \sum_{\lambda \in \Lambda(m, n)} (\det(P_\lambda \circ L))^2. \]

**Remark**

1. This is the formula \([L]^2\) we compute the sums of the squares of the determinants of each \((n \times n)\)-matrix of the \((m \times n)\)-matrix representing \( L \) (with respect to the standard bases of \( \mathbb{R}^n \) and \( \mathbb{R}^m \)).

2. In view of Lemma 1 in Section 3.3.1, this is a kind of higher dimensional version of the Pythagorean Theorem.
3.2 Linear Maps and Jacobians

3. For each \( \varphi \in \Phi \), we can uniquely write \( \varphi = \lambda \circ \theta \), where \( \theta \in \Sigma \) and \( \lambda \in \Lambda(m, n) \). Consequently,

\[
[L]^2 = \sum_{\sigma \in \Sigma} \text{sgn}(\sigma) \sum_{\lambda \in \Lambda(m, n)} \prod_{i=1}^{n} l_{\lambda(i), \lambda(\varphi(i))} \quad (\text{where we set } \rho = \sigma \circ \theta)
\]

\[
= \sum_{\lambda \in \Lambda(m, n)} \left( \sum_{\rho \in \Sigma} \text{sgn}(\rho) \prod_{i=1}^{n} l_{\lambda(i), \rho(i)} \right)^2
\]

\[
= \sum_{\lambda \in \Lambda(m, n)} (\det(F \circ L))^2.
\]

3.2.2 Jacobians

Now let \( f : \mathbb{R}^n \to \mathbb{R}^m \) be Lipschitz. By Rademacher's Theorem, \( f \) is differentiable \( C^1 \) a.e., and therefore \( Df(x) \) exists and can be regarded as a linear mapping from \( \mathbb{R}^n \) into \( \mathbb{R}^m \) a.e. \( x \in \mathbb{R}^n \).

**Notation**

If \( f : \mathbb{R}^n \to \mathbb{R}^m, f = (f_1, \ldots, f_m), \) we write the gradient matrix

\[
\frac{\partial f_1}{\partial x_1} \quad \ldots \quad \frac{\partial f_m}{\partial x_n}
\]

\[
Df = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \ldots & \frac{\partial f_m}{\partial x_n} \\
\frac{\partial f_1}{\partial x_1} & \ldots & \frac{\partial f_m}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_1}{\partial x_1} & \ldots & \frac{\partial f_m}{\partial x_n}
\end{bmatrix}
\]

**Definition**

The Jacobian of \( f \) is

\[
Jf(\sigma) = [Df(\sigma)] \quad (C^1 \text{ a.e. } x).
\]
3.3 The Area Formula
Throughout this section, we assume
\[ n \leq m. \]

3.3.1 Preliminaries

**LEMMA 1**

Suppose \( L : \mathbb{R}^n \to \mathbb{R}^m \) is linear, \( n \leq m \). Then
\[ \mathcal{H}^n (L(A)) = |L| \mathcal{L}^n (A) \]
for all \( A \subset \mathbb{R}^n \).

**PROOF**

1. Write \( L = O \circ S \) as in Section 3.1; \( |L| = |\det S| \).
2. If \( |L| = 0 \), then \( \dim S(\mathbb{R}^n) = n - 1 \) and so \( \dim L(\mathbb{R}^n) \leq n - 1 \). Consequently, \( \mathcal{H}^n (L(\mathbb{R}^n)) = 0 \).
3. If \( |L| > 0 \), then
\[
\frac{\mathcal{H}^n (L(B(x,r)))}{\mathcal{L}^n (B(x,r))} = \frac{\mathcal{L}^n (O' \circ L(B(x,r)))}{\mathcal{L}^n (L(B(x,r)))} = \frac{\mathcal{L}^n (O' \circ O \circ S(B(x,r)))}{\mathcal{L}^n (B(x,r))} = \frac{\mathcal{L}^n (S(B(x,r)))}{\mathcal{L}^n (B(0,1))},
\]
so
\[ \frac{\mathcal{H}^n (L(B(x,r)))}{\mathcal{L}^n (B(x,r))} = \frac{n(\theta)}{|\det S|} = \frac{n(\theta)}{|L|} \]
for \( B(x,r) \subset L^n (B(x,r)) \).
4. Define \( \mathcal{H}^n (A) = \mathcal{H}^n (L(A)) \) for all \( A \subset \mathbb{R}^n \). Then \( \mathcal{H}^n \) is a Radon measure, \( \nu \ll \mathcal{L}^n \), and
\[ D_{\mathcal{L}^n} \nu (r) = \frac{\mathcal{L}^n (L(B(x,r)))}{\mathcal{L}^n (B(x,r))} \]
for all \( x \in \mathbb{R}^n \). Theorem 2 in Section 1.6.2 implies
\[ \mathcal{H}^n (L(B)) = |L| \mathcal{H}^n (B). \]
Since \( \mathcal{H}^n \) and \( \mathcal{L}^n \) are Radon measures, the same formula holds for all sets \( A \subset \mathbb{R}^n \).

Henceforth we assume \( f : \mathbb{R}^n \to \mathbb{R}^m \) is Lipschitz.

**LEMMA 2**

Let \( A \subset \mathbb{R}^n \) be \( \mathcal{H}^n \)-measurable. Then
(i) \( f(A) \) is \( \mathcal{H}^n \)-measurable.
(ii) the mapping \( y \mapsto \mathcal{H}^n (A \cap f^{-1} \{ y \}) \) is \( \mathcal{H}^n \)-measurable on \( \mathbb{R}^m \), and
(iii) \( \int_{\mathbb{R}^m} \mathcal{H}^n (A \cap f^{-1} \{ y \}) \, d\mathcal{L}^n \leq (\text{Lip } f)^n \mathcal{H}^n (A) \).

**REMARK**
The mapping \( y \mapsto \mathcal{H}^n (A \cap f^{-1} \{ y \}) \) is called the *multiplicity function*.

**PROOF**

1. We may assume with no loss of generality that \( A \) is bounded.
2. By Theorem 5 in Section 1.1.1, there exist compact sets \( K_i \subset A \) such that \( \mathcal{L}^n (K_i) \leq \frac{1}{i} \) for \( i = 1, 2, \ldots \).

As \( \mathcal{L}^n (A) < \infty \) and \( A \) is \( \mathcal{L}^n \)-measurable, \( \mathcal{L}^n (A - K_i) < \infty \). Since \( f \) is continuous, \( f(K_i) \) is compact and thus \( \mathcal{H}^n \)-measurable. Hence \( f \left( \bigcup_{i=1}^{\infty} K_i \right) = \bigcup_{i=1}^{\infty} f(K_i) \) is \( \mathcal{H}^n \)-measurable. Furthermore,
\[
\mathcal{H}^n \left( f(A) - f \left( \bigcup_{i=1}^{\infty} K_i \right) \right) \leq \mathcal{H}^n \left( f \left( A - \bigcup_{i=1}^{\infty} K_i \right) \right) \leq \mathcal{H}^n \left( A - \bigcup_{i=1}^{\infty} K_i \right) = 0.
\]

Thus \( f(A) \) is \( \mathcal{H}^n \)-measurable: this proves (i).

3. Let
\[
B_k = \left\{ x \in \mathbb{R}^n \mid |Q| = (\alpha_1, b_1) \times \cdots \times (\alpha_n, b_n), \right\}
\]
and note
\[
\mathbb{R}^n = \bigcup_{Q \in B_k} Q.
\]

Now
\[
\mathcal{H}^n = \sum_{Q \in B_k} \mathcal{H}^n (f(A \cap f^{-1} \{ y \}) \subset \mathcal{H}^n \)-measurable by (i),
\]
and
\[
g_k (y) = \text{number of cubes } Q \in B_k \text{ such that } f^{-1} \{ y \} \cap (A \cap Q) \neq \emptyset.
\]

Thus
\[
g_k (y) \uparrow \mathcal{H}^n (A \cap f^{-1} \{ y \}) \quad \text{as } k \to \infty
\]
for each \( y \in \mathbb{R}^m \), and so \( y \mapsto \mathcal{H}^n (A \cap f^{-1} \{ y \}) \) is \( \mathcal{H}^n \)-measurable.
LEMMA 3
Let $t > 1$ and $B \equiv \{ x \mid Df(x) \text{ exists, } Jf(x) > 0 \}$. Then there is a countable collection $\{E_k\}_{k=1}^\infty$ of Borel subsets of $\mathbb{R}^n$ such that

(i) $B = \bigcup_{k=1}^\infty E_k$;
(ii) $f|E_k$ is one-to-one ($k = 1, 2, \ldots$); and
(iii) for each $k = 1, 2, \ldots$, there exists a symmetric automorphism $T_k : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\text{Lip} \left( f \mid E_k \right) \leq t^k \cdot \text{det} T_k \leq t^k \cdot \text{det} E_k.$$ 

PROOF

1. Fix $\epsilon > 0$ so that

$$\frac{1}{1 + \epsilon} \leq 1 < t - \epsilon.$$ 

Let $C$ be a countable dense subset of $B$ and let $S$ be a countable dense subset of symmetric automorphisms of $\mathbb{R}^n$.

2. Then, for each $c \in C$, $T \in S$, and $i = 1, 2, \ldots$, define $E(c, T, i)$ to be the set of all $x \in B \cap E(c, 1/i)$ satisfying

$$\left( \frac{1}{1 + \epsilon} \right)^i |Tv| \leq |Df(b)v| \leq (t - \epsilon)|Tv| \leq (t - \epsilon)\text{det} T_k$$

for all $v \in \mathbb{R}^n$ and

$$|f(a) - f(b) - Df(b) \cdot (a - b)| \leq \epsilon |T(a - b)|$$

for all $a \in B(b, 2/i)$. Note that $E(c, T, i)$ is a Borel set since $Df$ is Borel measurable. From (ii) and (iii) follows the estimate

$$\frac{1}{1 + \epsilon} |T(a - b)| \leq |f(a) - f(b)| \leq (t - \epsilon)|T(a - b)|$$

(***)

for $b \in E(c, T, i)$, $a \in B(b, 2/i)$.

3. Claim: If $b \in E(c, T, i)$, then

$$\left( \frac{1}{1 + \epsilon} \right)^i |\text{det} T| \leq |T(b)| \leq (t - \epsilon)^i |\text{det} T|$$

Proof of Claim: Write $Df(b) = L = O \circ S$, as above;

$$|Jf(b)| \equiv |\text{det} T| = |\text{det} S|.$$ 

By (‡)

$$\left( \frac{1}{1 + \epsilon} \right)^i |Tv| \leq |(O \circ S)v| = |Sv| \leq (t - \epsilon)|Tv|$$

for $v \in \mathbb{R}^n$, and so

$$\left( \frac{1}{1 + \epsilon} \right)^i |v| \leq |(S \circ T^{-1})v| \leq (t - \epsilon)|v| \quad (v \in \mathbb{R}^n).$$

Thus

$$[S \circ T^{-1}](B(0, 1)) \subseteq B(0, t - \epsilon);$$

whence

$$|\text{det}(S \circ T^{-1})|a(n) \leq \mathcal{L}^n(B(0, t - \epsilon)) = a(n)(t - \epsilon)^n,$$

and hence

$$|\text{det} S| < (t - \epsilon)^n|\text{det} T|$$

The proof of the other inequality is similar.

4. Relabel the countable collection $\{E(c, T, i) \mid c \in C, T \in S, i = 1, 2, \ldots\}$ as $\{E_k\}_{k=1}^\infty$. Select any $b \in E$, write $Df(b) = O \circ S$ as above, and choose $T \in S$ such that

$$\text{Lip} \left( T \circ S^{-1} \right) \leq \left( \frac{1}{1 + \epsilon} \right)^{-i} \leq (t - \epsilon) \leq t.$$ 

Now select $i \in \{1, 2, \ldots\}$ and $c \in C$ so that $|b - c| < 1/i$.

$$|f(c) - f(b) - Df(b) \cdot (c - b)| \leq \frac{t}{1 + \epsilon} |T(b - c)| \leq (t - \epsilon) |T(a - b)|$$

for all $a \in B(b, 2/i)$. Then $b \in E(c, T, i)$. As this conclusion holds for all $b \in B$, statement (i) is proved.
5. Next choose any set $E_k$, which is of the form $E(c; T, i)$ for some $c \in C$, $T \in S$, $i = 1, 2, \ldots$. Let $T_k = T$. According to (** *),

$$\frac{1}{t} |T_k(a - b)| = \left| f(a) - f(b) \right| < \left| T_k(a - b) \right|$$

for all $b \in E_k, a \in B(b, 2r)$. As $E_k \subset B(c, 2r) \subset B(b, 2r)$, we thus have

$$\frac{1}{t} |T_k(a - b)| \leq |f(a) - f(b)| \leq \left| T_k(a - b) \right| \quad (** *)$$

for all $a, b \in E_k$; hence $f \mid E_k$ is one-to-one.

6. Finally, notice (** ** *) implies

$$\text{Lip}(f \mid E_k) \leq t, \quad \text{Lip}(T_k \circ f \mid E_k) \leq t,$$

wheras the claim provides the estimate

$$t^{n-1} |\text{det} T_k| \leq |J f \mid E_k| \leq t^n |\text{det} T_k|.$$

Assertion (iii) is proved.

### 3.3 Proof of the Area Formula

**THEOREM 1 AREA FORMULA**

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz, $n \leq m$. Then for each $\mathcal{L}^m$-measurable subset $A \subset \mathbb{R}^n$,

$$\int_A f \, d\nu = \int_{\mathbb{R}^n} \nu(A \cap f^{-1}(y)) \, d\mathcal{L}^m(y)$$

**PROOF**

1. In view of Radon-Nikodym's Theorem, we may as well assume $D(f)$ and $J f$ exist for all $x \in A$. We may also suppose $\mathcal{L}^m(A) < \infty$.

2. Case 1. $A \subset \{ f > 0 \}$. Fix $t > 1$ and choose Borel sets $\{ E_k \}_{k=1}^\infty$ as in Lemma 2. We may assume the sets $\{ E_k \}_{k=1}^\infty$ are disjoint. Define $E_k$ as in the proof of Lemma 2.

3. Set

$$F_k = E_k \cap Q_i \cap A \quad (Q_i \in B_n, i = 1, 2, \ldots)$$

Then the sets $F_k$ are disjoint and $A = \bigcup_{k=1}^\infty F_k$.

3. **Claim #1:** Let

$$\lim_{k \to \infty} \sum_{i=1}^\infty \mathcal{L}^n(f(F_i)) = \int_{\mathbb{R}^n} \mathcal{L}^n(A \cap f^{-1}(y)) \, d\mathcal{L}^m(y).$$

**Proof of Claim #1:** Let

$$g_k = \sum_{i=1}^\infty \chi_{f(F_i)}$$

and

$$g_k = \mathcal{L}^n(f(F_i)) \leq t^n |\text{det} T_k|$$

and

$$\mathcal{L}^n(f(F_i)) \leq t^n |\text{det} T_k|$$

by Lemma 2. Thus

$$t^{-n} \mathcal{H}^n(f(F_i)) \leq t^{-n} \mathcal{L}^n(f(F_i))$$

and

$$\mathcal{L}^n(f(F_i)) \leq \mathcal{L}^n(f(F_i))$$

where we repeatedly used Lemmas 1 and 3. Now sum on $i$ and $j$:

$$t^{-n} \sum_{i=1}^\infty \mathcal{H}^n(f(F_i)) \leq \mathcal{L}^n(f(F_i)) \leq t^n |\text{det} T_k| \mathcal{L}^n(f(F_i)).$$

**FIGURE 3.3**

The Area Formula.

so that $g_k(y)$ is the number of the sets $\{ F_i \}$ such that $F_i \cap f^{-1}(y) \neq \emptyset$. Then $g_k(y) \mid \mathcal{L}^n(A \cap f^{-1}(y))$ as $k \to \infty$. Apply the Monotone Convergence Theorem.
Now let \( k \to \infty \) and recall Claim #1:
\[
\epsilon^{-\infty} \int_{A^{n}} \mathcal{H}^{n}(A \cap f^{-1}(y)) \, d\mathcal{H}^{n} \leq \int_{A} Jf \, dx
\]
\[
\leq \epsilon^{\infty} \int_{A^{n}} \mathcal{H}^{n}(A \cap f^{-1}(y)) \, d\mathcal{H}^{n}.
\]
Finally, send \( \epsilon \to 1^{+}\).

4. Case 2: \( A \in \{ Jf = 0 \} \). Fix \( \epsilon > 0 \). We factor \( f = p \circ g \), where
\[
g : \mathbb{R}^{n} \to \mathbb{R}^{n} \}, \quad g(x) = (g(x), x) \text{ for } x \in \mathbb{R}^{n},
\]
and
\[
p : \mathbb{R}^{n+1} \times \mathbb{R}^{n} \to \mathbb{R}^{n}, \quad p(y, z) = y \text{ for } y \in \mathbb{R}^{n}, z \in \mathbb{R}^{n}.
\]

5. Claim #2: There exists a constant \( C \) such that
\[
0 < Jg(x) \leq Ce^\epsilon
\]
for \( x \in A \).

Proof of Claim #2: Write \( g = (f^{1}, \ldots, f^{n}, x_{n}, \ldots, x_{m}) \); then
\[
Dg(x) = \left( \frac{Df^{1}(x)}{\epsilon^{1}}, \ldots, \frac{Df^{n}(x)}{\epsilon^{n}}, \frac{1}{\epsilon^{m+1}}, \ldots, \frac{1}{\epsilon^{n}} \right).
\]
Since \( Jf(x)^{2} \) equals the sum of the squares of the \((n \times n)\)-subdeterminants of \( Df(x) \) according to the Binet-Cauchy formula, we see
\[
Jg(x)^{2} = \text{sum of squares of } (n \times n) \text{-subdeterminants of } Dg(x) \geq \epsilon^{2n} > 0.
\]
Furthermore, since \( |Df| \leq \sup \{ f \} < \infty \), we may employ the Binet-Cauchy Formula to compute
\[
Jg(x)^{2} - Jf(x)^{2} + \left\{ \text{sum of squares of terms each involving at least one } \frac{1}{\epsilon} \right\} \leq \epsilon^{n} \epsilon^{2}
\]
for each \( x \in A \).

7. Since \( \mu : \mathbb{R}^{n+1} \times \mathbb{R}^{n} \to \mathbb{R} \) is a projection, we can compute, using Case 1 above,
\[
\mathcal{H}^{n}(f(A)) < \mathcal{H}^{n}(g(A))
\]
\[
\leq \int_{\mathbb{R}^{n+1}} \mathcal{H}^{n}(A \cap g^{-1}(y_{1})) \, d\mathcal{H}^{n}(y_{1})
\]
\[
= \int_{A} Jg(x) \, dx
\]
\[
\leq Ce^\epsilon \mathcal{H}^{n}(A)
\]

3.3. The Area Formula

Let \( \epsilon \to 0 \) to conclude \( \mathcal{H}^{n}(f(A)) = 0 \), and thus
\[
\int_{\mathbb{R}^{n}} \mathcal{H}^{n}(A \cap f^{-1}(y)) \, d\mathcal{H}^{n} = 0,
\]
since \( \text{sp} \, \mathcal{H}^{n}(A \cap f^{-1}(y)) \subset f(A) \). But then
\[
\int_{\mathbb{R}^{n}} \mathcal{H}^{n}(A \cap f^{-1}(y)) \, d\mathcal{H}^{n} = 0 = \int_{A} Jf \, dx
\]

8. In the general case, write \( A = A_{1} \cup A_{2} \) with \( A_{1} \subset \{ Jf > 0 \}, A_{2} \subset \{ Jf < 0 \} \), and apply Cases 1 and 2 above.

3.3.3 Change of variables formula

**Theorem 3**

Let \( f : \mathbb{R}^{m} \to \mathbb{R}^{n} \) be Lipschitz, \( n \leq m \). Then for each \( C^{0} \)-summable function \( g : \mathbb{R}^{n} \to \mathbb{R} \),
\[
\int_{\mathbb{R}^{n}} g(y) f(y) \, dx = \int_{\mathbb{R}^{n}} \left( \sum_{x \in \mathbb{R}^{m} \setminus \{0\}} \frac{g(x)}{x_{k}} \right) \, d\mathcal{H}^{n}(y).
\]

**Remark.** Using the Area Formula, we see \( f^{-1}(y) \) is at most countable for \( \mathcal{H}^{n} \) a.e. \( y \in \mathbb{R}^{m} \).

**Proof.**

1. Case 1. \( g \geq 0 \). According to Theorem 7 in Section 3.1.2 we can write
\[
g = \sum_{i=1}^{\infty} \chi_{A_{i}},
\]
for appropriate \( C^{0} \)-measurable sets \( \{ A_{i} \}^{\infty}_{i=1} \). Then the Monotone Convergence Theorem implies
\[
\int_{\mathbb{R}^{n}} g(y) f(y) \, dx = \sum_{i=1}^{\infty} \int_{A_{i}} \chi_{A_{i}}(y) f(y) \, dx
\]
\[
= \sum_{i=1}^{\infty} \int_{A_{i}} f(y) \, dx
\]
\[
= \sum_{i=1}^{\infty} \int_{\mathbb{R}^{n}} \mathcal{H}^{n}(A_{i} \cap f^{-1}(y)) \, d\mathcal{H}^{n}(y)
\]
3.3 The Area Formula

Then
\[ \mathcal{H}^n(C) = \text{"length" of } C = \int_a^b |f| \, dt. \]

B. Surface area of a graph \((n \geq 1, m = n + 1)\). Assume \(g : \mathbb{R}^n \to \mathbb{R}\) is Lipschitz and define \(f : \mathbb{R}^n \to \mathbb{R}^{n+1}\) by

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

Then
\[ Df = \begin{bmatrix}
1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 1 \\
\frac{\partial g}{\partial x_1} & \cdots & \frac{\partial g}{\partial x_n}
\end{bmatrix}_{(n+1) \times n} \]

so that

\[ (Jf)^2 = \text{sum of squares of } (n \times n)\text{-subdeterminants} \]

\[ = 1 + |Dg|^2. \]

For each open set \(U \subset \mathbb{R}^n\), define the graph of \(g\) over \(U\),

\[ G = G(g, U) = \{(x, g(x)) \mid x \in U\} \subset \mathbb{R}^{n+1}. \]

Then
\[ \mathcal{H}^n(G) = \text{"surface area" of } G = \int_U (1 + |Dg|^2)^{1} \, dx. \]

C. Surface area of a parametric hypersurface \((n \geq 1, m = n + 1)\). Suppose \(f : \mathbb{R}^n \to \mathbb{R}^{n+1}\) is Lipschitz and one-to-one. Write

\[ f = (f^1, \ldots, f^{n+1}). \]

Then
\[ Df = \begin{bmatrix}
\frac{\partial f^1}{\partial x_1} & \cdots & \frac{\partial f^1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f^{n+1}}{\partial x_1} & \cdots & \frac{\partial f^{n+1}}{\partial x_n}
\end{bmatrix}_{(n+1) \times n} \]

so that

\[ (Jf)^2 = \text{sum of squares of } (n \times n)\text{-subdeterminants} \]

\[ = \sum_{i=1}^{n+1} \left( \frac{\partial f^i}{\partial x_1}, \ldots, \frac{\partial f^i}{\partial x_n} \right)^2 + \left( \frac{\partial f^{n+1}}{\partial x_1}, \ldots, \frac{\partial f^{n+1}}{\partial x_n} \right)^2. \]
For each open set \( U \subseteq \mathbb{R}^n \), write

\[ S = f(U) \subseteq \mathbb{R}^{n+1}. \]

Then

\[ \mathcal{H}^{n+1}(S) = \text{"surface area" of } S = \int_U \left( \sum_{k=1}^{n+1} \left| \frac{\partial (f_1, \ldots, f_{k-1}, f_{k+1}, \ldots, f_{n+1})}{\partial (x_1, \ldots, x_n)} \right| \right)^{1 \over 2} \, dx. \]

D. Submanifolds. Let \( M \subseteq \mathbb{R}^n \) be a Lipschitz, \( n \)-dimensional embedded submanifold. Suppose that \( U \subseteq \mathbb{R}^n \) and \( f : U \to M \) is a chart for \( M \). Let \( A \subseteq f(U) \), a Borel, \( B \equiv f^{-1}(A) \). Define

\[ g_{ij} = \frac{\partial f_i}{\partial x_j}, \quad (1 \leq i, j \leq n), \]

\[ g = \det((g_{ij})). \]

Then

\[ (Df)^* \circ Df = (g_{ij})_{i,j=1}^n. \]

and so

\[ Jf = g^1. \]

Thus

\[ \mathcal{H}^n(A) = \text{"volume" of } A \text{ in } M = \int_B g^1 \, dx. \]
3.4 The Coarea Formula

Throughout this section we assume

\[ n > m. \]

3.4.1 Preliminaries

**LEMMA 1**

Suppose \( L : \mathbb{R}^n \to \mathbb{R}^m \) is linear, \( n \leq m \), and \( A \subset \mathbb{R}^n \) is \( L^\alpha \)-measurable. Then

(i) the mapping \( y \mapsto \mathcal{H}^{n-m}(A \cap L^{-1}(y)) \) is \( L^\alpha \)-measurable and

(ii) \( \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap L^{-1}(y)) \, dy = |L| \mathcal{H}^n(A) \).

**PROOF**

1. **Case 1.** \( \dim L(\mathbb{R}^n) < m \).

Then \( A \cap L^{-1}(y) \neq \emptyset \) and consequently \( \mathcal{H}^{n-m}(A \cap L^{-1}(y)) = 0 \) for \( L^\alpha \)-a.e. \( y \in \mathbb{R}^m \). Also, if we write \( L = S \circ O^* \) as in the Polar Decomposition Theorem (Section 3.2.1), we have \( L(\mathbb{R}^n) = S(\mathbb{R}^m) \). Thus \( \dim S(\mathbb{R}^m) < m \) and hence \( |L| = |\det S| > 0 \).

2. **Case 2.** \( L = P \) is the orthogonal projection of \( \mathbb{R}^n \) onto \( \mathbb{R}^m \).

Then for each \( y \in \mathbb{R}^m \), \( P^{-1}(y) \) is an \( (n-m) \)-dimensional affine subspace of \( \mathbb{R}^n \), a translate of \( P^{-1}(0) \). By Fubini’s Theorem,

\[ y \mapsto \mathcal{H}^{n-m}(A \cap P^{-1}(y)) \]

is \( L^\alpha \)-measurable and

\[ \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap P^{-1}(y)) \, dy = \mathcal{L}^m(A). \]

**LEMMA 2**

Let \( A \subset \mathbb{R}^n \) be \( L^\alpha \)-measurable, \( n \geq m \). Then

(i) \( f(A) \) is \( L^\alpha \)-measurable.

(ii) \( A \cap f^{-1}(y) \) is \( \mathcal{H}^{n-m} \)-measurable for \( L^\alpha \)-a.e. \( y \).

(iii) the mapping \( y \mapsto \mathcal{H}^{n-m}(A \cap f^{-1}(y)) \) is \( L^\alpha \)-measurable, and

(iv) \( \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap f^{-1}(y)) \, dy \leq (\alpha(n-m)\alpha(m)) / (\alpha(n)\alpha(m)) L^m f(A) \).

**PROOF**

1. **Statement (i) is proved exactly like the corresponding statement of Lemma 2 in Section 3.3.1.**

2. **For each \( j = 1, 2, \ldots \), there exist closed balls \( \{B_j\}_{j=1}^\infty \) such that**

\[ A \subset \bigcup_{j=1}^\infty B_j, \quad \text{and} \quad d_{B_j} \leq \frac{1}{j}, \]

3.4 The Coarea Formula

Proof of Claim: Let \( Q \) be any orthogonal map of \( \mathbb{R}^n \) onto \( \mathbb{R}^m \) such that

\[ Q^*(x_1, x_2, \ldots, x_m, 0, \ldots, 0) = Q(x_1, x_2, \ldots, x_m) \]

for all \( x \in \mathbb{R}^n \). Note

\[ P^* (x_1, \ldots, x_m) = (x_1, x_2, \ldots, x_m, 0, \ldots, 0) \in \mathbb{R}^n \]

for all \( x \in \mathbb{R}^n \). Thus \( Q = Q^* P^* \) and hence \( Q^* = P^* Q \).

Since \( L^{-1}(0) \) is an \( (n-m) \)-dimensional subspace of \( \mathbb{R}^n \) and \( L^{-1}(y) \) is a translate of \( L^{-1}(0) \) for each \( y \in \mathbb{R}^m \). Thus by Fubini’s Theorem, \( y \mapsto \mathcal{H}^{n-m}(A \cap L^{-1}(y)) \) is \( L^\alpha \)-measurable, and we may calculate

\[ \mathcal{L}^n(A) = \mathcal{L}^n(O(A)) \]

\[ = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(Q(A \cap P^{-1}(y))) \, dy \]

by (i)

\[ = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap Q^{-1} P^{-1}(y)) \, dy. \]

Now set \( z = L^m \) to compute using Theorem 2 in Section 3.3.2

\[ \mathcal{L}^n(A) = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap Q^{-1} \circ P^{-1}(z)) \, dz. \]

But \( f = S \circ O^* = S \circ P \circ Q \), and so

\[ \mathcal{L}^n(A) = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}(A \cap L^{-1}(z)) \, dz. \]

Henceforth we assume \( f : \mathbb{R}^n \to \mathbb{R}^m \) is Lipschitz.