PROOF OF CHANGE OF VARIABLES

We will give a proof of the following theorem.

**Theorem 0.1.** Let \( U \subset \mathbb{R}^n \) be an open set, \( \phi \in C^1(U, \mathbb{R}^n) \) and \( K \subset U \) be a compact set with content. Suppose that \( \phi \) is injective and \( \det J_\phi \neq 0 \) on \( K \setminus Z \) for some \( Z \subset K \) of content zero. Then \( \phi(K) \) has content and

\[
\int_{\phi(K)} f = \int_K f \circ \phi |\det J_\phi|
\]

for all continuous functions \( f : K \to \mathbb{R} \).

First, we show that it is enough to prove (1) for \( f \geq 0 \).

**Lemma 0.2.** Let \( U, \phi, K \) and \( Z \) be given in Theorem 0.1. Then (1) holds for all continuous functions \( f : K \to \mathbb{R} \) if and only if it holds for all nonnegative continuous functions \( f : K \to \mathbb{R} \).

**Proof.** Every function can be written as \( f = f_+ - f_- \) where

\[
f_+ = \frac{|f| + f}{2} \quad \text{and} \quad f_- = \frac{|f| - f}{2}.
\]

Obviously, \( f_+ \geq 0 \) and \( f_- \geq 0 \); both \( f_+ \) and \( f_- \) are continuous if \( f \) is. If (1) holds for all nonnegative continuous functions, it holds for \( f_+ \) and \( f_- \) and hence it holds for \( f \). \( \square \)

**Lemma 0.3.** Let \( K \subset \mathbb{R}^n \) be a compact set with content and \( f : K \to \mathbb{R} \) be a nonnegative continuous function. Then

\[
\int_K f = \mu(G_f)
\]

where \( G_f = \{(x, y) : x \in K, 0 \leq y \leq f(x)\} \subset \mathbb{R}^{n+1} \).

**Proof.** By Fubini’s theorem,

\[
\mu(G_f) = \int_{G_f} 1 = \int_K \int_0^{f(x)} dy dx = \int_K f.
\]

Now we can reduce (1) to the case that \( f \equiv 1 \).
Lemma 0.4. If $\phi(K)$ has content and

\[(3) \quad \mu(\phi(K)) = \int_K |\det J_\phi|\]

for all $U, \phi, K$ and $Z$ with the properties in Theorem 0.1, then (1) holds for all continuous functions $f : K \to \mathbb{R}$.

Proof. By Lemma 0.2, it is enough to prove (1) for $f \geq 0$. So let us assume that $f(x) \geq 0$ for all $x \in K$.

By Lemma 0.3,

\[\int_{\phi(K)} f = \mu(B)\]

where

\[B = \{(x, y) : x \in \phi(K), 0 \leq y \leq f(x)\} \subset \mathbb{R}^{n+1}.\]

We also let

\[A = \{(t, y) : t \in K, 0 \leq y \leq f(\phi(t))\}\]

and $h : U \times \mathbb{R} \to \mathbb{R}^{n+1}$ be the function given by

\[h(x, y) = (\phi(x), y).\]

It is easy to see that $h(A) = B$ and the Jacobian of $h$ is

\[J_h = \begin{bmatrix} J_\phi & 0 \\ 0 & 1 \end{bmatrix}.\]

Therefore, $h$ is injective and $J_h \neq 0$ on $(K \setminus Z) \times \mathbb{R}$. So by (3),

\[\mu(B) = \mu(\phi(A)) = \int_A |\det J_h|.\]

On the other hand, $\det J_h = \det J_\phi$ and

\[\int_A |\det J_h| = \int_A |\det J_\phi| = \int_K \int_0^{f(\phi(t))} |\det J_\phi|dydt = \int_K f \circ \phi |\det J_\phi|.\]

Hence (1) follows. \qed

Definition 0.5. Let $\rho : \mathbb{R}^n \to \mathbb{R}$ be the function given by

\[\rho(x_1, x_2, ..., x_n) = \max(|x_1|, |x_2|, ..., |x_n|).\]

Clearly, $\rho$ has the properties:

- $\rho(\lambda x) = |\lambda|\rho(x)$ for all $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^n$;
- Triangle inequality: $\rho(x + y) \leq \rho(x) + \rho(y)$. 

This is really another norm on the vector space $\mathbb{R}^n$. Actually, it is equivalent to the Euclidean norm $\| \cdot \|$ by
\[
\frac{1}{\sqrt{n}} \|x\| \leq \rho(x) \leq \|x\|
\]
for all $x \in \mathbb{R}^n$.

**Definition 0.6.** We call a compact interval
\[
I = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]
\]
a cube if $b_1 - a_1 = b_2 - a_2 = \ldots = b_n - a_n$. Then every cube is given by $I = \{\rho(x - p) \leq r\}$, where $p$ is the center of $I$. So we use the notation
\[
C_{p, r} = \{\rho(x - p) \leq r\}.
\]

**Lemma 0.7.** Let $U \subset \mathbb{R}^n$ be an open set, $p \in U$ be a point and $\phi: U \to \mathbb{R}^n$ be a map satisfying
\[
\lambda_1 \rho(x - p) \leq \rho(\phi(x) - \phi(p)) \leq \lambda_2 \rho(x - p)
\]
for all $x \in U$ and some nonnegative constants $\lambda_1$ and $\lambda_2$. Then
(4) $\phi(C_{p, r}) \subset C_{\phi(p), \lambda_2 r}$
for all $C_{p, r} \subset U$. If we further assume that $\phi$ is an open map, then
(5) $C_{\phi(p), \lambda_1 r} \subset \phi(C_{p, r})$
for all $C_{p, r} \subset U$.

**Proof.** WLOG, we may assume that $p = 0$ and $\phi(p) = 0$. For simplicity, we write $C_r = C_{0, r}$.

Clearly, since $\rho(\phi(x)) \leq \lambda_2 \rho(x)$, $\rho(\phi(x)) \leq \lambda_2 r$ for all $x \in C_{0, r}$. Therefore,
\[
\phi(C_r) \subset C_{\lambda_2 r}
\]
which is (4). It remains to show (5), i.e.,
\[
C_{\lambda_1 r} \subset \phi(C_r)
\]
when $\phi$ is open. Let $\lambda \geq 0$ be the largest number such that
\[
C_{\lambda r} \subset \phi(C_r).
\]
We claim that
\[
C_{\lambda r} \not\subset \phi(\text{int}(C_r))
\]
where $\text{int}(C_r) = \{\rho(x) < r\}$ is the interior of $I_r$. Otherwise, since $\phi$ is open, $\phi(\text{int}(C_r))$ is open and $C_{\lambda r}$ is a compact set contained in $\phi(\text{int}(C_r))$; it follows that $C_s \subset \phi(\text{int}(C_r))$ for some $s > \lambda r$. This contradicts the choice of $\lambda$. Therefore, we must have
\[
C_{\lambda r} \cap \phi(\partial C_r) \neq \emptyset
\]
where \( \partial C_r = \{ \rho(x) = r \} \) is the boundary of \( I_r \). That is, there is \( x \in \partial C_r \) such that \( \phi(x) \in C_{\lambda r} \). Namely, there is \( x \in U \) such that \( \rho(x) = r \) and \( \rho(\phi(x)) = \lambda r \). Since \( \rho(\phi(x)) \geq \lambda_1 \rho(x) \), we necessarily have \( \lambda \geq \lambda_1 \). Therefore, \( C_{\lambda r} \subset \phi(C_r) \). □

**Lemma 0.8.** Let \( U \subset \mathbb{R}^n \) be an open set and \( \phi \in C^1(U, \mathbb{R}^n) \). Let \( V \subset U \) be a bounded open set such that the closure \( \overline{V} \subset U \). Then there is a constant \( M_V \), depending on \( V \), such that

\[
\mu(\phi(K)) \leq M_V \mu(K)
\]

for all compact sets \( K \subset V \).

**Proof.** Let \( \phi = (\phi_1, \phi_2, ..., \phi_n) \). Since \( \overline{V} \) is compact and \( \partial \phi_i/\partial x_j \) is continuous on \( \overline{V} \) for \( i, j = 1, 2, ..., n \), there is a number \( M \) such that \( |\partial \phi_i/\partial x_j| \leq M \) on \( \overline{V} \) for all \( i, j = 1, 2, ..., n \).

For all \( e > 0 \), there exist compact intervals \( I_1, I_2, ..., I_m \) such that \( K \subset \bigcup I_v \) and

\[
\sum_{v=1}^{m} \mu(I_v) - e < \mu(K) \leq \sum_{v=1}^{m} \mu(I_v).
\]

By enlarging \( I_v \) and subdivision, we can always assume that \( I_v \) are cubes. For \( p, q \in I_v \),

\[
\rho(\phi(p) - \phi(q)) = \max_{1 \leq i \leq n} |\phi_i(p) - \phi_i(q)|
\]

and

\[
|\phi_i(p) - \phi_i(q)| = |(\nabla \phi_i(s), p - q)| \leq nM \rho(p - q)
\]

for some \( s \in pq \). Therefore,

\[
\rho(\phi(p) - \phi(q)) \leq nM \rho(p - q)
\]

for all \( p, q \in I_v \). Then by Lemma 0.7, \( \phi(I_v) \subset C_{\phi(p), nM_r} \) if \( I_v = C_{p,r} \).

Consequently,

\[
\mu(\phi(I_v)) \leq (nM)^n \mu(I_v)
\]

and hence

\[
\mu(\phi(K)) < (nM)^n (\mu(K) + e).
\]

Taking \( e \to 0 \) and \( M_V = (nM)^n \), we see that (6) holds. □

Now we can show that \( \phi(K) \) has content in Theorem 0.1.

Since \( K \) has content, \( \mu(\partial K) = 0 \). By Lemma 0.8, \( \mu(\phi(\partial K)) = 0 \).

We want to show that

\[
\mu(\partial \phi(K)) = 0.
\]

It is tempting to claim that \( \partial \phi(K) \subset \phi(\partial K) \).
However, this is not exactly true. But we do have
\begin{equation}
\partial\phi(K) \subset \phi(\partial K) \cup \phi(Z)
\end{equation}
which implies (7) since \(Z\) has content zero. It remains to justify (8).

Suppose that there is a point \(q \in \partial\phi(K)\) such that \(q \notin \phi(\partial K)\) and \(q \notin \phi(Z)\). Then \(\phi^{-1}(q) \subset \phi(\text{int}(K) \setminus Z)\). There is a point \(p \in \text{int}(K)\) and \(p \notin Z\) such that \(\phi(p)\). By the assumptions on \(Z\), \(\det J_\phi(p) \neq 0\). Then by Inverse Function Theorem (IFT), \(\phi\) is an open map locally at \(p\). Namely, there exists \(r > 0\) such that \(\phi(K) \supset B_r(q)\). This implies that \(q\) is an interior point of \(\phi(K)\), which is a contradiction.

Next, we will further reduce Theorem 0.1 to the case that \(K\) is a cube. Namely, we just have to prove (1) for \(f \equiv 1\) and \(K = C_{p,r}\).

Furthermore, we can even assume that \(Z = \emptyset\).

**Lemma 0.9.** With the same hypothesis in Theorem 0.1, (1) holds if and only if it holds for \(f \equiv 1\), \(K = C_{p,r}\), and \(Z = \emptyset\), i.e., (3) holds when \(K\) is a cube, \(\phi\) is injective and \(\det(J_\phi) \neq 0\) on \(K\).

**Proof.** Since \(K\) has content, for every \(e > 0\), there exists compact intervals \(I_1, I_2, \ldots, I_m\) such that \(\text{int}(I_u) \cap \text{int}(I_v) = \emptyset\) for \(u \neq v\), \(I = I_1 \cup I_2 \cup \ldots \cup I_m \subset K\) and
\[
\mu(K) - e < \mu(I) \leq \mu(K).
\]
By shrinking \(I_v\) and subdivision, we can assume that all \(I_v\) are cubes and \(I_v \cap Z = \emptyset\). Therefore,
\[
\int_{\phi(I_v)} 1 = \int_{I_v} |\det J_{\phi}|
\]
and hence
\[
\int_{\phi(I)} 1 = \int_I |\det J_{\phi}|.
\]
By Lemma 0.8
\[
\mu(\phi(K) \setminus \phi(I)) \leq \mu(\phi(K \setminus I)) \leq M\mu(K \setminus I) < Me
\]
for some constant \(M\) and hence
\[
\left| \int_{\phi(K)} 1 - \int_{\phi(I)} 1 \right| < Me.
\]
On the other hand, it is obvious that
\[
\left| \int_K |\det J_{\phi}| - \int_I |\det J_{\phi}| \right| \leq Ce
\]
where $C$ is the maximum of $|\det J_\phi|$ on $K$. As $\epsilon \to 0$, we see that

$$\int_{\phi(K)} 1 = \int_K |\det J_\phi|.$$

\[\square\]

Next, we will show that Theorem 0.1 holds for all non homogeneous linear maps $\phi$.

**Lemma 0.10.** Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be a non homogeneous linear map $\phi(x) = Ax + b$ for some $n \times n$ matrix $A$ and a vector $b$. Then (3) holds for every compact set $K$ with content.

**Proof.** Every non homogeneous linear map $\phi$ can be decomposed to $\phi = h \circ \varphi$ where $\varphi(x) = Ax$ is homogeneous and $h(x) = x + b$ is a translation. The identity (3) clearly holds for $\phi$ translations. Therefore, it is enough to treat the homogeneous case. That is, we may assume that $\phi(x) = Ax$.

If $J_\phi = A$ is singular, then $\phi(\mathbb{R}^n)$ is contained in some hyperplane $H = \{a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0\}$. Hence $\phi(K)$ has content zero and (3) holds.

Suppose that $\det(J_\phi) = \det(A) \neq 0$. Every nonsingular matrix can be written as a product of elementary matrices. That is, $A = A_1A_2A_3\ldots A_m$ where $A_k$ is an elementary matrix given by one of the following:

- $A_k = E_{ij}(\lambda)$ for $i \neq j$, where $E_{ij}(\lambda) = [e_{ab}]$ is the $n \times n$ matrix with
  $$e_{ab} = \begin{cases} 1 & \text{if } a = b \\ \lambda & \text{if } (a, b) = (i, j) \\ 0 & \text{otherwise} \end{cases}$$

- $A_k = F_{ij}$ for $i \neq j$, where $F_{ij} = [f_{ab}]$ is the $n \times n$ matrix with
  $$f_{ab} = \begin{cases} 1 & \text{if } a = b \neq i, j \\ 1 & \text{if } (a, b) = (i, j), (j, i) \\ 0 & \text{otherwise} \end{cases}$$

- $A_k = G_i(\lambda)$, where $G_i(\lambda) = [g_{ab}]$ is the $n \times n$ matrix with
  $$g_{ab} = \begin{cases} 1 & \text{if } a = b \neq i \\ \lambda & \text{if } (a, b) = (i, i) \\ 0 & \text{otherwise} \end{cases}$$
Correspondingly, we can decompose \( \phi \) into \( \phi = \phi_1 \circ \phi_2 \circ \ldots \circ \phi_n \) where \( \phi_k(x) = A_kx \) is given by an elementary matrix \( A_k \). So it is enough to show (3) for \( \phi = A_1x \) with \( A \) an elementary matrix. Also by Lemma 0.9, it is enough to verify it for \( K = C_{p,r} \) a cube. Let \( p = (a_1, a_2, \ldots, a_n) \). Then

\[
K = C_{p,r} = \{a_k - r \leq x_k \leq a_k + r \text{ for } 1 \leq k \leq n\}.
\]

When \( A = E_{ij}(\lambda) \), we may assume \( A = E_{12}(\lambda) \) WLOG. Then

\[
\phi(x_1, x_2, \ldots, x_n) = (x_1 + \lambda x_2, x_2, \ldots, x_n)
\]

and

\[
\phi^{-1}(x_1, x_2, \ldots, x_n) = (x_1 - \lambda x_2, x_2, \ldots, x_n).
\]

Hence

\[
\phi(K) = \{a_1 - r \leq x_1 - \lambda x_2 \leq a_1 + r, a_k - r \leq x_k \leq a_k + r \text{ for } 2 \leq k \leq n\}.
\]

Therefore,

\[
\int_{\phi(K)} 1 = (2r)^{n-2} \int_{a_2 - r}^{a_2 + r} \int_{a_1 + r + \lambda x_2}^{a_1 + r} dx_1 dx_2 = (2r)^n = \mu(K).
\]

And since \( \det(A) = 1 \), (3) holds.

When \( A = F_{ij} \), we may assume that \( A = F_{12} \) WLOG. Then

\[
\phi(x_1, x_2, x_3, \ldots, x_n) = (x_2, x_1, x_3, \ldots, x_n)
\]

and hence

\[
\phi(K) = [a_2 - r, a_2 + r] \times [a_1 - r, a_1 + r] \times [a_3 - r, a_3 + r] \times \ldots \times [a_n - r, a_n + r].
\]

Therefore, \( \mu(\phi(K)) = \mu(K) \). And since \( \det(A) = -1 \), (3) holds.

When \( A = G_{i}(\lambda) \), we may assume that \( A = G_1(\lambda) \) WLOG. Then

\[
\phi(x_1, x_2, \ldots, x_n) = (\lambda x_1, x_2, \ldots, x_n)
\]

and hence

\[
\phi(K) = [\lambda a_1 - r, \lambda a_1 + r] \times [a_2 - r, a_2 + r] \times \ldots \times [a_n - r, a_n + r].
\]

Therefore, \( \mu(\phi(K)) = |\lambda| \mu(K) \). And since \( \det(A) = \lambda \), (3) holds. \( \square \)

Every \( C^1 \) function is totally differentiable. So it is locally approximated by a linear function. We need the following lemma to compare \( \mu(\phi(K)) \) and \( \mu(T(K)) \) for a linear approximation \( T \) of \( \phi \).

Lemma 0.11. With the same hypothesis in Theorem 0.7, we further assume that \( K = C_{p,r} \) is a cube, \( \det(J_{\phi}) \neq 0 \) on \( K \) and \( T : \mathbb{R}^n \rightarrow \mathbb{R}^n \) is a non homogeneous linear map such that

\[
\rho(\phi(x) - T(x)) \leq c\rho(x - p)
\]
for a constant $e$ and all $x \in K$. Then

$$|\mu(\phi(K)) - \mu(T(K))| \leq |\det(L)|((1 + C_L e)^n - 1)\mu(K)$$

where $L = J_T$ and $C_L = n(\max|a_{ij}|)$ with $L^{-1} = [a_{ij}]_{n \times n}$.

Proof. For simplicity, we may assume that $p = 0$ and $\phi(0) = 0$ WLOG. By Lemma 0.10, (3) holds for $\phi = T^{-1}$. Therefore,

$$\mu(T^{-1} \circ \phi(K)) = \int_{\phi(K)} |\det(L)|^{-1} = |\det(L)|^{-1} \mu(\phi(K))$$

and

$$\mu(T^{-1} \circ T(K)) = \int_{T(K)} |\det(L)|^{-1} = |\det(L)|^{-1} \mu(T(K)).$$

Consequently,

$$|\mu(\phi(K)) - \mu(T(K))| = |\det(L)||\mu(T^{-1} \circ \phi(K)) - \mu(K)|. \quad (9)$$

We let $h = T^{-1} \circ \phi$. Obviously,

$$T^{-1}(x - y) = L^{-1}(x - y)$$

and hence

$$\rho(T^{-1}(x - y)) = \rho(L^{-1}(x - y)) \leq C_L \rho(x - y)$$

for all $x, y \in \mathbb{R}^n$. Therefore,

$$\rho(h(x) - x) = \rho(T^{-1}(\phi(x) - T(x))) \leq C_L \rho(\phi(x) - T(x)) \leq C_L e \rho(x)$$

for all $x \in K$. Then by the triangle inequality for $\rho$, we have

$$\max(1 - C_L e, 0) \rho(x) \leq \rho(h(x)) \leq (1 + C_L e) \rho(x)$$

for all $x \in K$. By Lemma 0.7,

$$C_{0, \lambda_1 r} \subset h(K) \subset C_{0, \lambda_2 r}$$

and hence

$$\lambda_1^n \mu(K) \leq \mu(h(K)) \leq \lambda_2^n \mu(K)$$

where $\lambda_1 = \max(1 - C_L e, 0)$ and $\lambda_2 = 1 + C_L e$. Therefore, we obtain

$$|\mu(h(K)) - \mu(K)| \leq (\lambda_2^n - 1)\mu(K) \quad (10)$$

by observing that $\lambda_2^n - 1 \geq 1 - \lambda_1^n$. Combining (9) and (10), we are done. \qed

Now we are ready to prove Theorem 0.1.
Proof of Theorem 0.1. We have reduced it to (3) with $K$ a cube, $\phi$ being injective and $\det(J_\phi) \neq 0$.

Since $K$ is compact and $J^{-1}_\phi = [h_{ij}(x)]_{n \times n}$ is continuous on $K$, there exists a number $C$ such that $n|h_{ij}(x)| \leq C$ for all $x \in K$ and all entries $h_{ij}(x)$ of $J^{-1}_\phi$.

We take a partition $P$ of $K$ which divides $K = \bigcup I_v$ with $I_v = C_{p_v,r}$ for some $r > 0$. By the uniform continuity of $\partial \phi_i/\partial x_j$, for every $e > 0$, there exists $r > 0$ such that

$$\left| \frac{\partial \phi_i}{\partial x_j}(p) - \frac{\partial \phi_i}{\partial x_j}(p_v) \right| < \frac{e}{n}$$

for all $p \in I_v$ and $1 \leq i, j \leq n$. For each $I_v$, we let $T_v$ be the non homogeneous linear map given by

$$T_v(x) = J_\phi(p_v)(x - p_v) + \phi(p_v).$$

Then by MVT, we see that

$$\rho(\phi(x) - T_v(x)) \leq e\rho(x - p_v).$$

It follows from Lemma 0.11 that

$$\left| \int_{\phi(I_v)} 1 - \int_{T_v(I_v)} 1 \right| \leq M((1 + Ce)^n - 1)\mu(I_v)$$

where $M$ is the maximum of $|\det(J_\phi)|$ on $K$. Therefore,

$$\left| \int_{\phi(K)} 1 - \sum_v \int_{T_v(I_v)} 1 \right| \leq M((1 + Ce)^n - 1)\mu(I_v)$$

where

$$\sum_v \int_{T_v(I_v)} 1 = \sum_v \int_{I_v} |\det J_\phi(p_v)| = \sum_v |\det J_\phi(p_v)|\mu(I_v)$$

is a Riemann sum of $\int_K |\det J_\phi|$. As $r, e \to 0$, we obtain (3). □