Sard's Theorem
Riyaz Ahuja

1 Preliminaries

We start off by stating a few formal definitions and notions of local Euclidian-ness that will be critical in properly stating the theorem. We then move on to state some simple measure-theoretic facts and a variant of Fubini's theorem.

1.1 Regular and Critical Points

Definition. *For a smooth map $f : \mathbb{R}^n \to \mathbb{R}^n$, we say f is a submersion at x if $Df(x)$ is surjective.*

Remark 1.1. *For some intuition behind this definition, we can write:*

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$$

And if $Df(x)$ is surjective, then this can be described as the partials of f local to p reaching every direction in the tangent space around p. Thus we can see that the map f is "submerging" into the entire space of Y at p.

Submersions are a useful concept as they encode a sense of "completeness" into the local behavior of f, allowing it to locally model every possible direction in the target space.

Definition. *For smooth $f : \mathbb{R}^m \to \mathbb{R}^n$, we say a point $y \in \mathbb{R}^n$ is a regular value if f is a submersion for all $x \in f^{-1}(y)$.*

Remark 1.2. *It's called regular because it's a well-behaved and "normal" point in the image of f, as it maps smoothly in all directions of its target space.*

However, as it is well-behaved, it isn't nearly as interesting to analyze as its non-submersion counterpart.

Definition. *For smooth $f : \mathbb{R}^m \to \mathbb{R}^m$, we say a point $y \in \mathbb{R}^n$ is a critical value if it is not a regular value.*

Remark 1.3. *We also say that x is a critical point if $f(x)$ is a critical value, and a regular point if $f(x)$ is a regular value.*

1.2 Measure and Fubini

Lemma 1.4. *(Baby Fubini) For a measurable $C \subseteq \mathbb{R}^n$, if we have that for all $t \in \mathbb{R}$, $C \cap (\{t\} \times \mathbb{R}^{n-1})$ has measure zero, then C has measure zero.*

Proof. ■ We'll talk about it in class soon as a special case of Fubini's Theorem (the switching iterated integral theorem) □

Remark 1.5. *Although we're leaving this lemma as a blackbox, the intuition behind why it works is pretty clear. Essentially, we're showing that if all "slices" of C in one dimension have measure zero, then C itself is measure zero.*

1.3 Reals are Lidelof

This is in the notes as Theorem 5.3.46, but recall anyways that

Lemma 1.6. *If $\{U_\alpha\}_{\alpha \in A}$ is a set of open sets in $\mathbb{R}^m$ with index set A, then there is a countable subcollection $\{U_{\alpha_k}\}_{k\in\mathbb{N}}$ such that:*

$$\bigcup_{\alpha\in A} U_\alpha = \bigcup_{k\in\mathbb{N}} U_{\alpha_k}$$

Proof. Theorem 5.3.46, examples 5.3.47 □

1.4 Image of Union is Union of Image

Lemma 1.7. *For a map $f : \mathbb{R}^m \to \mathbb{R}^n$, if we have a (possibly uncountable) collection $\{U_\alpha\}_{\alpha\in A}$ over an index set A, with $U_\alpha \subseteq \mathbb{R}^m$, then:*

$$f\left(\bigcup_{\alpha\in A} U_\alpha\right) = \bigcup_{\alpha\in A} f(U_\alpha)$$

Proof. We know that if $y \in f\left(\bigcup_{\alpha\in A} U_\alpha\right)$, then there exists $x \in \bigcup_{\alpha\in A} U_\alpha$ such that $f(x) = y$. Then, we know that $x \in U_\alpha$ for some $\alpha \in A$, and as such $y \in f(U_\alpha)$. Thus, we conclude that $y \in \bigcup_{\alpha\in A} f)U_\alpha$.

Conversely, if we have $y \in \bigcup_{\alpha\in A} f)U_\alpha$, there exists $\alpha \in A$ such that $y \in f(U_\alpha)$, meaning that there exists $x \in U_\alpha$ such that $f(x) = y$. Then, we know that as $x \in U_\alpha$, $x \in \bigcup_{\alpha\in A} U_\alpha$, and thus, $y \in f\left(\bigcup_{\alpha\in A} U_\alpha\right)$.

Thus:

$$f\left(\bigcup_{\alpha\in A} U_\alpha\right) = \bigcup_{\alpha\in A} f(U_\alpha)$$

□

2 Sard's Statement and Proof

Theorem 2.1. *(Sard) For $f : \mathbb{R}^m \to \mathbb{R}^n$ a smooth map, if $C \subseteq \mathbb{R}^m$ is the set of all critical points of f, then the set of critical values $f(C) \subseteq \mathbb{R}^n$ is Lebesgue measure zero.*

Proof. We proceed via induction on m. Clearly, if $m = 0$, the $f : \mathbb{R}^0 \to \mathbb{R}^n$, and as per convention, we write $\mathbb{R}^0$ to be a zero-tuple of real numbers, i.e. the empty tuple. In this degenerate case, the image $f(C) \subseteq f(\mathbb{R}^0) \in \mathbb{R}^n$ is a singleton, which necesarrily has measure zero in $\mathbb{R}^n$. Thus, the theorem is true for $m = 0$.

As such, we can suppose that the theorem is true for $0 \leq m - 1$, and proceed to the inductive step of showing how that implies that the theorem is true for m.

Now, define:

$$C_i = \{x \in \mathbb{R}^m \mid \partial_\alpha f(x) = 0, |\alpha| \leq i\}$$

Here, α is a multi-index, such that C_i denotes the set of all $x \in \mathbb{R}^m$ where all partials of f with order at most i evaluate to zero at x. Observe that clearly:

$$C_1 \supseteq C_2 \supseteq \cdots C_k \supseteq C_{k+1} \supseteq \cdots$$

And additionally, suppose we have some $z \in C_1$. Then we know that $\frac{\partial f_i}{\partial x_j}(z) = 0$, and as such:

$$Df(z) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(z) & \cdots & \frac{\partial f_1}{\partial x_m}(z) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(z) & \cdots & \frac{\partial f_n}{\partial x_m}(z) \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

Which is clearly not surjective. Thus, $Df(z)$ is not a submersion, and as such, z is a critical point. Thus, $z \in C$. As such, $C_1 \subseteq C$.

We can combine these facts to build a chain:

$$C \supseteq C_1 \supseteq C_2 \supseteq \cdots \supseteq C_k$$

Where we let $k > m/n - 1$ (We will see why soon!).

Recall that via lemma 1.4, we know that since:

$$C = (C \setminus C_1) \cup \left(\bigcup_{i=1}^{k-1} C_i \setminus C_{i+1} \right) \cup C_k$$

We have that:

$$f(C) = f(C \setminus C_1) \cup \left(\bigcup_{i=1}^{k-1} f(C_i \setminus C_{i+1}) \right) \cup f(C_k)$$

As such, we can divide the proof of $\lambda^*(f(C)) = 0$ into the following three steps.

1. $\lambda^*(f(C \setminus C_1)) = 0$
2. $\lambda^*(f(C_i \setminus C_{i+1})) = 0$ for all $1 \leq i < k$
3. $\lambda^*(f(C_k)) = 0$

Note that we're using the Lebesgue outer measure instead of the standard Lebesgue measure as although C is certainly measurable in $\mathbb{R}^m$, we don't necessarily know that $f(C)$ or $f(C_i)$ are measurable. However, once we have shown that they are all (outer) measure zero, we know by completeness of the Lebesgue measure that they are all fully Lebesgue measurable anyways.

We now proceed to prove each of these facts:

1. $f(C \setminus C_1)$:
   For all $x \in C \setminus C_1$, we aim to find an open V_x such that $f(V_x \cap C)$ has measure zero. This is sufficient as we know that $C \setminus C_1 \subseteq \bigcup_{x \in (C \setminus C_1)} V_x$, and as such by Lemma 1.6, we know that there exists a countable subcollection $\{V_{x_k}\}_{k \in \mathbb{N}}$ such that:

   $$C \setminus C_1 \subseteq \bigcup_{k \in \mathbb{N}} V_{x_k}$$

   And thus:

   $$C \setminus C_1 \subseteq \bigcup_{k \in \mathbb{N}} V_{x_k} \cap C$$

   $$f(C \setminus C_1) \subseteq f\left(\bigcup_{k \in \mathbb{N}} V_{x_k} \cap C \right)$$

$$
\begin{aligned}
f(C \setminus C_1) &\subseteq \bigcup_{k\in\mathbb{N}} f\left(V_{x_k} \cap C\right) \\
\lambda^*(f(C \setminus C_1)) &\leq \lambda^*\left(\bigcup_{k\in\mathbb{N}} f\left(V_{x_k} \cap C\right)\right) \\
\lambda^*(f(C \setminus C_1)) &\leq \sum_{k\in\mathbb{N}} \lambda^*(f\left(V_{x_k} \cap C\right)) \\
\lambda^*(f(C \setminus C_1)) &\leq \sum_{k\in\mathbb{N}} 0 \\
\lambda^*(f(C \setminus C_1)) &\leq 0 \\
\lambda^*(f(C \setminus C_1)) &= 0
\end{aligned}
$$

Now, we can fix $x \in C \setminus C_1$, and show that there exists an open V_x with $\lambda^*(f(V_x \cap C)) = 0$.

As $x \notin C_1$, we know that there must be some first-order partial $\frac{\partial f_i}{\partial x_j}(x) \neq 0$. We can always reindex coordinates and swap the labels on $\mathbb{R}^m, \mathbb{R}^n$, so WLOG suppose $i = j = 1$. Then we have that $\frac{\partial f_1}{\partial x_1}(x) \neq 0$. As we adopt the convention that $f = (f_1, \cdots, f_n)$, we define a new function $g : \mathbb{R}^m \to \mathbb{R}^m$ via:

$$
g(z) = (f_1(z), z_2, \cdots, z_m)
$$

Then, we have that:

$$
\begin{aligned}
Dg(z) &= \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(z) & \frac{\partial g_1}{\partial x_2}(z) & \cdots & \frac{\partial g_1}{\partial x_m}(z) \\ \frac{\partial g_2}{\partial x_1}(z) & \frac{\partial g_2}{\partial x_2}(z) & \cdots & \frac{\partial g_2}{\partial x_m}(z) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(z) & \frac{\partial g_m}{\partial x_2}(z) & \cdots & \frac{\partial g_m}{\partial x_m}(z) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(z) & \frac{\partial f_1}{\partial x_2}(z) & \cdots & \frac{\partial f_1}{\partial x_m}(z) \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}
\end{aligned}
$$

Now, if we consider $z = x$, we have that:

$$
Dg(z) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_m}(x) \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}
$$

And as $\frac{\partial f_1}{\partial x_1}(x) \neq 0$, we know that $Dg(z)$ has full (column) rank, and thus is invertible.

Thus, as $Dg(z) \in \mathcal{L}(\mathbb{R}^m; \mathbb{R}^m)$ is a bijection, and thus a linear homeomorphism. Moreover, as f is smooth, g is smooth, and thus we conclude that via inverse function theorem, that for an open neighborhood V_x about x, g is a local smooth diffeomorphism around x. This means that g maps V_x (smoothly) diffeomorphically to $V'_x = g(V_x)$.

Thus, we have an inverse g^{-1} well defined on V'_x, and as such we can define $h : V'_x \to \mathbb{R}^n$ via $h = f \circ g^{-1}$. As g is a bijection, the image of $g^{-1}(V'_x) = V_x$, and as such, we know that the critical values of h are exactly the critical values of f restricted to V_x. Thus, the critical values of h are $f(C \cap V_x)$. Thus, we aim to show that the set of critical values of h has measure zero.

Now, let's analyze some properties of $h = f \circ g^{-1}$ for $z \in V_x' = g(V_x)$. However, firstly notice that as:

$$g(g^{-1}(z)) = z$$

We know that:

$$(f_1(g^{-1}(z)), g_2^{-1}(z), \cdots, g_m^{-1}(z)) = z$$

So $f_1(g^{-1}(z)) = z_1$. Thus, for $z = (t, z_2, \cdots, z_m)$:

$$h(z) = f(g^{-1}(z)) = (f_1(g^{-1}(z)), f_2(g^{-1}(z)), \cdots, f_n(g^{-1}(z)) = (t, z_2', \cdots, z_m')$$

In other words, h leaves the first coordinate of the input unchanged. That means that we can induce a map by fixing this first coordinate t, as $h_t : V_x' \cap (\{t\} \times \mathbb{R}^{m-1}) \to \{t\} \times \mathbb{R}^{n-1}$, and define it via:

$$h_t(t, (z_2, \cdots, z_m)) = (t, (h_2(t, z_2, \cdots, z_m), \cdots, h_n(t, z_2, \cdots, z_m)))$$

Note that this is basically just h, but forcing the first coordinate to be t. As such, we essentially consider $h_t : \mathbb{R}^{m-1} \to \mathbb{R}^n$, as the other coordinate is always just going to be t in both the input and output vectors.

Now, let's consider $Dh(z)$, with $z = (t, z_2, \cdots, z_m) \in V_x'$. Then we know that:

$$Dh(z) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1}(z) & \cdots & \frac{\partial h_1}{\partial x_m}(z) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1}(z) & \cdots & \frac{\partial h_n}{\partial x_m}(z) \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1}(z) & \cdots & \frac{\partial h_n}{\partial x_m}(z) \end{pmatrix}$$

And now writing in block matrix form:

$$= \begin{pmatrix} 1 & \begin{pmatrix} 0 & \cdots & 0 \end{pmatrix} \\ * & \begin{pmatrix} \frac{\partial h_2}{\partial x_2}(z) & \cdots & \frac{\partial h_2}{\partial x_m}(z) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_2}(z) & \cdots & \frac{\partial h_n}{\partial x_m}(z) \end{pmatrix} \end{pmatrix}$$

Where $*$ is some other partials, but they really aren't important to consider. We just know that they exist, and they're there. However, notice that this is by contruction equal to:

$$Dh(z) = \begin{pmatrix} 1 & 0 \\ * & Dh_t(z) \end{pmatrix}$$

Now, clearly we know that $Dh(z)$ is not surjective if and only if $Dh_t(z)$ is not surjective. As such, we conclude that (t, z) is a critical point of h if and only if z is a critical point of h_t (as the t in h_t is fixed by the domain $(\{t\} \times \mathbb{R}^{m-1}) \cap V_x'$).

By induction, we know that all functions $\mathbb{R}^{m-1} \to \mathbb{R}^n$ have their set of critical values measure zero. Thus, if $h_t(C_t)$ is the set of critical points of h_t, we know that $h_t(C_t)$ has measure zero. Thus, as we know that the critical points of h_t coincide with the critical points of h (i.e. $\{t\} \times C_t$ are critical points of h) and $h(\{t\} \times C_t)$ must have measure zero. Moreover, as we know that the critical values of h is exactly $f(C \cap V_x)$, we now simply claim that these slices are identical, in that:

$$f(C \cap V_x) \cap (\{t\} \times \mathbb{R}^{n-1}) = h(\{t\} \times C_t)$$

Clearly, as $h(t, z_2, \cdots, z_m) \in \{t\} \times \mathbb{R}^{n-1}$, we know that $h(\{t\} \times C_t) \subseteq \{t\} \times \mathbb{R}^{n-1}$. And as $h(\{t\} \times C_t)$ are critical values of h, and $f(C \cap V_x)$ is **all** the critical values of h, clearly $h(\{t\} \times C_t) \subseteq f(C \cap V_x)$.

For the other direction, consider some value $f(c) \in f(C \cap V_x) \cap (\{t\} \times \mathbb{R}^{n-1})$. We know that $f_1(c) = t$ and that $f(c)$ is a critical value of h. As $h = f \circ g^{-1}$, we know that as g is bijective and $c \in V_x$, there exists $d \in V_x'$ such that $g^{-1}(d) = c$, and thus $h(d) = (f \circ g^{-1})(d) = f(c)$, and by definition of h, we know that since $f_1(c) = h_1(d) = t$, then $d_1 = t$, so $d \in \{t\} \cap C_t$, as d is a critical point of h thus h_t. Thus, we conclude that $f(c) = h(d) \in h(\{t\} \times C_t)$. Thus, we conclude that:

$$f(C \cap V_x) \cap (\{t\} \times \mathbb{R}^{n-1}) = h(\{t\} \times C_t)$$

Thus, as we know that for all $t \in \mathbb{R}$, $h(\{t\} \times C_t)$ has measure zero, we know that $f(C \cap V_x) \cap (\{t\} \times \mathbb{R}^{n-1})$ has measure zero for all $t \in \mathbb{R}$. Thus, we can directly apply Lemma 1.4 to get that this implies that $f(C \cap V_x)$ has measure zero. This, as explained earlier, is sufficient to conclude that $f(C \setminus C_1)$ has measure zero.

2. $f(C_i \setminus C_{i+1})$:
   This case is very similar to the previous case, albiet simpler.

   For all $x \in C_i \setminus C_{i+1}$, we aim to find an open V_x with $x \in V_x$ such that $f(V_x \cap C_i)$ has measure zero. This is sufficient as we know that $C_i \setminus C_{i+1} \subseteq \bigcup_{x \in (C_i \setminus C_{i+1})} V_x$, and as such by Lemma 1.6, we know that there exists a countable subcollection $\{V_{x_k}\}_{k \in \mathbb{N}}$ such that:

$$C \setminus C_1 \subseteq \bigcup_{k \in \mathbb{N}} V_{x_k}$$

   And thus:

$$C_i \setminus C_{i+1} \subseteq \bigcup_{k \in \mathbb{N}} V_{x_k} \cap C_i$$

$$f(C_i \setminus C_{i+1}) \subseteq f\left(\bigcup_{k \in \mathbb{N}} V_{x_k} \cap C_i\right)$$

$$f(C_i \setminus C_{i+1}) \subseteq \bigcup_{k \in \mathbb{N}} f\left(V_{x_k} \cap C_i\right)$$

$$\lambda^*(f(C_i \setminus C_{i+1})) \leq \lambda^*\left(\bigcup_{k \in \mathbb{N}} f\left(V_{x_k} \cap C_i\right)\right)$$

$$\lambda^*(f(C_i \setminus C_{i+1})) \leq \sum_{k \in \mathbb{N}} \lambda^*(f\left(V_{x_k} \cap C_i\right))$$

$$\lambda^*(f(C_i \setminus C_{i+1})) \leq \sum_{k \in \mathbb{N}} 0$$

$$\lambda^*(f(C_i \setminus C_{i+1})) \leq 0$$
$$\lambda^*(f(C_i \setminus C_{i+1})) = 0$$

Now, we can fix $x \in C_i \setminus C_{i+1}$, and proceed to show that there exists an open V_x with $\lambda^*(f(V_x \cap C_i)) = 0$.

As $x \in C_i$, all partials of f at x of order up to i vanish, however, there exists some $i+1$th order partial that does not vanish. Thus, we can find a ith order partial of x, denoted w, such that $\frac{\partial w_i}{\partial x_j} \neq 0$. And just as before, we can always switch and reindex coordinates to WLOG suppose $i = j = 1$. Thus $\frac{\partial w_1}{\partial x_1} \neq 0$.

as we adopt the convention that $w = (w_1, \cdots, w_n)$, we define a new function g via:

$$g(z) = (w_1(z), z_2, \cdots, z_m)$$

Then, we have that:

$$Dg(z) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(z) & \frac{\partial g_1}{\partial x_2}(z) & \cdots & \frac{\partial g_1}{\partial x_m}(z) \\ \frac{\partial g_2}{\partial x_1}(z) & \frac{\partial g_2}{\partial x_2}(z) & \cdots & \frac{\partial g_2}{\partial x_m}(z) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(z) & \frac{\partial g_m}{\partial x_2}(z) & \cdots & \frac{\partial g_m}{\partial x_m}(z) \end{pmatrix}$$
$$= \begin{pmatrix} \frac{\partial w_1}{\partial x_1}(z) & \frac{\partial w_1}{\partial x_2}(z) & \cdots & \frac{\partial w_1}{\partial x_m}(z) \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Now, if we consider $z = x$, we have that:

$$Dg(z) = \begin{pmatrix} \frac{\partial w_1}{\partial x_1}(x) & \frac{\partial w_1}{\partial x_2}(x) & \cdots & \frac{\partial w_1}{\partial x_m}(x) \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

And as $\frac{\partial w_1}{\partial x_1}(x) \neq 0$, we know that $Dg(z)$ has full (column) rank, and thus is invertible.

Thus, as $Dg(z)$ is a bijection, and thus a linear homeomorphism. Moreover, as f is smooth, w is smooth, and thus g is smooth. As such we conclude that via inverse function theorem, that for an open neighborhood V_x about x, g is a local smooth diffeomorphism around x. This means that g maps V_x (smoothly) diffeomorphically to $V_x' = g(V_x)$.

Now, observe that as w is a ith partial of f, and by hypothesis we know that ith partials vanish, we conclude that $w(x) = 0$. Thus, we know that by definition:

$$g(x) = (0, x_2, \cdots, x_m)$$

Thus, we conclude that g maps $C_i \cap V_x$ onto the hyperplane $\{0\} \times \mathbb{R}^{m-1}$. Thus, as g is locally bijective, we can define $h = f \circ g^{-1}$, and note that the critical points of h are in the hyperplane $\{0\} \times \mathbb{R}^{m-1}$. Clearly, h has the same critical values as $f \restriction_{V_x}$ as $h : V_x' \to \mathbb{R}^n$ and g is bijective from $V_x \to V_x'$. Thus, we know that $f(C_k \cap V_x)$ are exactly the critical values of type C_k of h.

Thus, we can define $h' : (\{0\} \cap \mathbb{R}^{m-1}) \cap V'_x \to \{0\} \cap \mathbb{R}^{n-1}$ via the restriction of h and the same reasoning from the previous case to show well-definedness due to invariance of the first coordinate. As such, we can essentially consider $h' : \mathbb{R}^{m-1} \to \mathbb{R}^n$, we conclude that due to the IH, the critical values of h', denoted as $h'(C')$, have measure zero in $\mathbb{R}^n$. Thus, the critical values of h' of type C_i, denoted $h'(C'_i)$, also have measure zero in $\mathbb{R}^n$.

However, note that as all partial derivatives of f of order up to i vanish, we know that $g(C_k \cap V_x) \in \{0\} \times \mathbb{R}^{m-1}$ must be critical points for $h' : \{0\} \times \mathbb{R}^{m-1} \to \mathbb{R}^n$. This is because:

$$h'(g(C_k \cap V_x)) = (f \circ g^{-1})(g(C_k \cap V_x)) = f(C_k \cap V_x)$$

And $C_i \subseteq C$ is by construction the set of critical points of f of type C_i (first i orders of partials are zero). Thus, as $f(C_i \cap V_x)$ is a subset of the critical values of h', we conclude that $f(C_i \cap V_x)$ has measure zero.

Thus, as per the earlier reasoning, this is sufficient to show that $f(C_i \setminus C_{i+1})$ has measure zero.

3. $f(C_k)$:
Recall that we required $k \geq m/n - 1$. With that in mind, fix some $\delta > 0$, and $Q \subseteq \mathbb{R}^m$ be a cube of side length $0 < s < \delta$. We claim that $f(C_k \cap Q)$ is measure zero. This implies that $f(C_k)$ is measure zero as we know that we can cover C_k with a countable union of such cubes $\{Q_i\}_{i \in \mathbb{N}}$, and thus we can write:

$$f(C_k) = f\left(C_k \cap \bigcup_{i \in \mathbb{N}} Q_i\right) = f\left(\bigcup_{i \in \mathbb{N}} C_k \cap Q_i\right) = \bigcup_{i \in \mathbb{N}} f\left(C_k \cap Q_i\right)$$

And thus:

$$\lambda(f(C_k)) = \lambda\left(\bigcup_{i \in \mathbb{N}} f\left(C_k \cap Q_i\right)\right) \leq \sum_{i \in \mathbb{N}} \lambda(f(C_k \cap Q_i)) = \sum_{i \in \mathbb{N}} 0 = 0$$

So:

$$\lambda(f(C_k)) = 0$$

Now, we continue to prove that $f(C_k \cap Q)$ is measure zero. Note that Q is a cube, it is clearly compact, and as all partials of up to order k are zero in C_k.

By Taylor's theorem (Theorem 9.2.27), we know that as f is smooth, thus C^{k+1}, we can write for some $x \in C_k \cap Q$ and $x + h \in Q$:

$$f(x+h) = \sum_{j=0}^{k} \frac{1}{j!} D^j(x)(h)^{\otimes j} + \left(\int_0^1 \frac{(1-t)^k}{k!} D^{k+1} f(th)\ dt\right)(h)^{\otimes k}$$

And moreover, we know that for all $\varepsilon > 0$ there exists $\delta > 0$ such that $x + h \in B(x, \delta)$ implies that:

$$\left\| f(x+h) - \sum_{j=0}^{k} \frac{1}{j!} D^j(x)(h)^{\otimes j} \right\| \leq \varepsilon |h|^{k+1}$$

And thus, there exists $\varepsilon > 0$, such that $h < \delta$, we know that we can let $\varepsilon = c$ for a constant c. We can do this because Q is compact, so we know that the derivatives are bounded. Thus, we know that we can bound this error term via extreme value theorem. (Proof by Subhasish). Thus as the first k orders of partials vanish at x:

$$\left\| \left(\int_0^1 \frac{(1-t)^k}{k!} D^{k+1} f(th)\ dt \right) (h)^{\otimes k} \right\| \leq c|h|^{k+1}$$

And by writing the integral as $R(x,h)$, we conclude that:

$$f(x+h) = f(x) + R(x,h)$$

With $|R(x,h)| < a|h|^{k+1}$ for some constant a.

Now, with this in mind, we can fix r and subdivide Q into r^m subcubes of size at most δ/r (Just define a grid partition). We denote this subdivision as $\{Q_i'\}_{i=1}^{r^m}$ and observe that by construction $\{Q_i'\}$ is almost disjoint, and moreover $\bigcup_{i=1}^{r^m} Q_i' = Q$. Thus, as $Q \cap C_k \neq \varnothing$, we know that there is some $1 \leq i \leq r^m$ such that $Q_i' \cap C_k \neq \varnothing$. Let $x \in Q_i' \cap C_k$, and observe that as Q_i' has side length at most δ/r, and thus diameter at most $\sqrt{m}\frac{\delta}{r}$. Then we know that we can write any $y \in Q_i'$ as $x + h$ for some $|h| < \sqrt{m}\frac{\delta}{r}$. Thus:

$$f(Q_i') \subseteq f(\{x+h \mid |h| < \sqrt{m}\delta/r\}) = \{f(x) + R(x,h) \mid |h| < \sqrt{m}\delta/r\} \subseteq Q_i''$$

With $Q_i'' \subseteq \mathbb{R}^n$ defined to be a cube of side length at most $2a(\sqrt{m}\delta/r)^{k+1}$ centered at $f(x)$. As we can do this for all r^m cubes, we conclude that:

$$f(C_k \cap Q) \subseteq \bigcup_{i=1}^{r^m} f(C_k \cap Q_i') \subseteq \bigcup_{i=1}^{r^m} f(Q_i') \subseteq \bigcup_{i=1}^{r^m} Q_i''$$

(We can say this as if $C_k \cap Q_i' = \varnothing$ for some i, then $f(C_k \cap Q_i')$ ends up measure zero anyways so it can just be ignored, so can just define $Q_i'' = \varnothing$ in that case).

As each $\lambda(f(Q_i')) \leq \left(\frac{2a(\sqrt{m}\delta)^{k+1}}{r^{k+1}}\right)^n$, we know that:

$$\lambda\left(f(C_k \cap Q)\right) \leq \sum_{i=1}^{r^m} \lambda\left(f(Q_i')\right) \leq r^m \left(\frac{2a(\sqrt{m}\delta)^{k+1}}{r^{k+1}}\right)^n = \left(r^{m/n}\frac{2a(\sqrt{m}\delta)^{k+1}}{r^{k+1}}\right)^n = \left(\frac{2a(\sqrt{m}\delta)^{k+1}}{r^{k+1-m/n}}\right)^n$$

And as $k > m/n - 1$, we know that $k + 1 - m/n > 0$. Thus, we conclude that as $r \to \infty$, $\lambda(f(C_k \cap Q)) \to 0$. As such, we conclude that $\lambda(f(C_k \cap Q)) = 0$.

As mentioned before, we know that this is sufficient to show $f(C_k)$ is measure zero.

As such, we conclude that:

$$\begin{aligned}
\lambda(f(C)) &= \lambda\left(f(C \setminus C_1) \cup \left(\bigcup_{i=1}^{k-1} f(C_i \setminus C_{i+1})\right) \cup f(C_k)\right) \\
&\leq \lambda\left(f(C \setminus C_1)\right) + \lambda\left(\bigcup_{i=1}^{k-1} f(C_i \setminus C_{i+1})\right) + \lambda\left(f(C_k)\right) \\
&\leq \lambda\left(f(C \setminus C_1)\right) + \left(\sum_{i=1}^{k-1} \lambda(f(C_i \setminus C_{i+1}))\right) + \lambda\left(f(C_k)\right) \\
&= 0 + 0 + 0 = 0
\end{aligned}$$

Thus:

$$\lambda(f(C)) = 0$$

□

3 Applications

The following proofs aren't fully rigorous and utilize a fair amount of blackboxes and handwaving as we don't yet have the tools to properly work with these concepts (i.e. manifolds). However, many of these results are fundamental in the study of differential geometry and topology, so it's interesting to see how far reaching the applications of Sard's go.

3.1 Manifolds and Critical Points

Definition. *A manifold is a topological space that locally resembles euclidean space $\mathbb{R}^n$ at all points. Specifically, we define M to be a smooth n-dimensional manifold over a topological space X if for all $x \in M$, there exists a neighborhood $x \in U$ such that U is smoothly homeomorphic to an open subset V of R^n.*

Remark 3.1. *This is a very informal definition, and the proper way to define such a manifold is by passing to charts and transistion maps in order to define smooth atlases, and then define a smooth manifold to be a pair $(X, \mathcal{A})$ for a topological space X and the maximal smooth atlas $\mathcal{A}$. We don't use this definition here because (a) I don't want to prove all the interesting things that this entails, (b) we will discuss manifolds properly in class soon, (c) this is enough for the handwavy extension we tackle next.*

Fact 3.2. *(Sard on Manifolds) For $f : M \to N$ a smooth map, with M, N smooth manifolds of dimension m, n respectively, if $C \subseteq M$ is the set of all critical points of f, then the set of critical values $f(C)$ has Lebesgue measure zero.*

Remark 3.3. *Note that we define the derivative on manifolds slightly differently, as $Df(x) \in \mathcal{L}(T_M(x), T_N(f(x)))$, where T_M and T_N represent the tangent spaces of $x, f(x)$ in M, N respectively.*

Similarly, if $E \subseteq M$ for M a smooth m-dimensional manifold, we say that E has measure zero if there exists a chart (φ, U, V) with $E \subseteq U$ such that $\varphi(E) \subseteq \mathbb{R}^m$ has measure zero.

Proof. We know that as measure zero is a local concept, we know that it suffices to prove this under the assumption that $M \subseteq \mathbb{R}^m$ and $N \subseteq \mathbb{R}^n$. Thus, this is equivalent to the Euclidean Sard's theorem. □

3.2 Morse Theory

Definition. *For a smooth m-dimensional manifold M, and a map $f : M \to \mathbb{R}$, we say that $x \in M$ is a nondegenerate critical point if it is a critical point, and moreover the Hessian $H(f)(x)$ has nonzero determinant. We define the Hessian here as:*

$$H(f)(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_m} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_m x_1} & \frac{\partial^2 f}{\partial x_m \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_m \partial x_m} \end{pmatrix}$$

Remark 3.4. *These non-degenerate critical points are interesting as the Hessian condition forces these points to be locally quadratic. That means that the critical point is always either a local max, local min, or a saddle point. (Proof: Morse Lemma)*

Definition. *For a smooth m-dimensional manifold M, and a map $f : M \to \mathbb{R}$, f is a Morse function if all its critical points are nondegenerate.*

Remark 3.5. *The usual Morse function that you will find in the wild is none other than the height function.*

Theorem 3.6. *Given $M \subseteq \mathbb{R}^m$ and $f : M \to \mathbb{R}$, then $f_a = f + a_1x_1 + \cdots a_mx_m$ is a Morse function for all but a measure zero set of $a \in \mathbb{R}^m$.*

Proof. Define $g : M \to \mathbb{R}$ via:

$$g = \nabla f = \frac{\partial f}{\partial x_1} + \cdots + \frac{\partial f}{\partial x_m}$$

Similarly, for arbitrary $a \in \mathbb{R}^m$, define $f_a : M \to \mathbb{R}$ via:

$$f_a = f + a_1x_1 + \cdots a_mx_m$$

And observe that:

$$\begin{aligned} \nabla f_a &= \sum_{i=1}^{m} \frac{\partial f}{\partial x_i} + a_ix_i \\ &= g + a \end{aligned}$$

And thus:

$$\begin{aligned} H(f) &= \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_m} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_m x_1} & \frac{\partial^2 f}{\partial x_m \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_m \partial x_m} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \\ \vdots \\ \frac{\partial g}{\partial x_m} \end{pmatrix} = \nabla g \end{aligned}$$

And:

$$\begin{aligned} H(f_a) &= \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_m} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_m x_1} & \frac{\partial^2 f}{\partial x_m \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_m \partial x_m} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\nabla f_a}{\partial x_1} \\ \frac{\nabla f_a}{\partial x_2} \\ \vdots \\ \frac{\nabla f_a}{\partial x_m} \end{pmatrix} = \begin{pmatrix} \frac{g}{\partial x_1} \\ \frac{g}{\partial x_2} \\ \vdots \\ \frac{g}{\partial x_m} \end{pmatrix} = \nabla g \end{aligned}$$

Thus:

$$H(f) = H(f_a) = \nabla g$$

With this in mind, we now aim to categorize for which $a \in \mathbb{R}^m$ is ∇g nonsingular. Consider $a \in \mathbb{R}^m$ such that $-a$ is a regular value of g. Then we know that for critical points x of f_a, we have that $Df_a(x)$ is not a surjection, however, as the output space is $\mathbb{R}$, this demands that all partials must be zero. Thus, we know that $(\nabla f_a)(x) = 0$, and thus $g(x) = -a$. Thus, $(\nabla g)(x) \neq 0$ as $-a$ is a regular value of g, and thus $Dg(x)$ is surjective, which demands that there exists a nonzero partial as the output space of g is $\mathbb{R}$. Thus, we conclude that $H(f_a)(x)$ is nonsingular. Thus, all critical points of f_a are nondegenerate, and thus f_a is a Morse function.

As this holds for all a with $-a$ a regular value of g, and by Sard's we know that the set of non-regular values (critical values) is measure zero, we can say that **most** $a \in \mathbb{R}^m$ allow f_a to be a Morse function. $\square$

4 References

[1] Tice Notes

[2] https://websites.umich.edu/ alexmw/Sard.pdf

[3] Topology from the Differentiable Viewpoint, John Milnor

[4] https://cdn-uploads.piazza.com/paste/hl5k6pq875t6qv/eed3e81a756dcb23ffea72c9bc7dfe0fdb2a8a671

[5] https://stanford.edu/ sfh/morse.pdf