VARIATIONS ON SARD'S THEOREM

JARVIS KENNEDY

Contents

1. Sard's Theorem in Finite Dimensions

1.1. Hausdorff Measure and Dimension.

Let s and α be positive numbers, X a normed space, and $A \subset X$. Let $\delta(A)$ denote the *diameter* of A,

$$\delta(A) = \sup\{\|x - y\| : x, y \in A\}.$$

Let

$$\mu_{s,\alpha}(A) = \inf \left\{ \sum_{k\in\mathbb{N}} \delta(A_k)^s : \bigcup_{k\in\mathbb{N}} A_k \supset A, \delta(A_k) < \alpha \right\}.$$

Other than being sufficiently small and forming a countable cover of A, there is no restriction on what these sets might be. They can be open, closed, boxes, balls, etc.

Now, we define the *s-dimensional outer Hausdorff measure* of A to be

$$\mu_s(A) = \omega_s \lim_{\alpha\to 0} \mu_{s,\alpha}(A),$$

where $\omega_s = \pi^{s/2}/2^s\Gamma[(s+2)/2]$. When s is an integer, this is the volume of an s-dimensional unit sphere in s-space. The factor ω_s is a proportionality constant between the Hausdorff outer measure and the Lebesgue outer measure. That is, in s-dimensional space, $\mu_s(A) = \lambda_s(A)$, where $\lambda_s(A)$ is the outer Lebesgue measure of A.

We call a set A *s-null* if $\mu_s(A) = 0$, *s-finite* if $\mu_s(A)$ is finite, and *s-sigmafinite* if A is a countable union of s-finite sets.

If A is s-sigmafinite, then $\mu_\rho(A) = 0$ for every $\rho > s$. Equivalently, if $\mu_\rho(A) > 0$, then $\mu_s(A) = \infty$ for every $s < \rho$. It is enough to prove the result for s-finite sets, since the s-sigmafinite case will follow from countable decomposition into s-finite sets. For any $\alpha > 0$, since $\mu_{s,\alpha}(A)$ is an infimum, we can find a cover $\{A_k\}$ with diameter less than α such that $\sum_k \delta(A_k)^s < \mu_{s,\alpha}(A) + 1$. Now if $\rho > s$ then $\rho - s > 0$, so we have

$$\sum_k \delta(A_k)^\rho = \sum_k \delta(A_k)^{\rho-s}\delta(A_k)^s,$$

and since $\delta(A_k) < \alpha$, we have

$$\sum_k \delta(A_k)^\rho < \alpha^{\rho-s} \sum \delta(A_k)^s < \alpha^{\rho-s}(\mu_{s,\alpha}(A) + 1).$$

This gives the inequality

$$0 \le \mu_{\rho,\alpha}(A) \le \sum_k \delta(A_k)^\rho < \alpha^{\rho-s}(\mu_{s,\alpha}(A) + 1).$$

Since $\rho - s > 0$ and A is s-finite, we have $\alpha^{\rho-s}(\mu_{s,\alpha}(A) + 1) \to 0$ as $\alpha \to 0$, and so the by the squeeze theorem we have $\mu_\rho(A) = 0$.

1.1. *Remark.* As a special case of the above, any subset $A \subset \mathbb{R}^m$ is s-null for any $s > m$, since $\mathbb{R}^m$ is m-sigmafinite.

This gives us a well defined notion of the dimension of a set. There is exactly one value of s for which μ_s will give the "correct" measure of A, and this is the *Hausdorff dimension* of A.

$$\dim_\mu(A) = \inf\{\rho > 0 : \mu_\rho(A) = 0\} = \sup\{s > 0 : \mu_s(A) = \infty\},$$

with the convention that $\sup \emptyset = 0$.

1.2. **Change of Variables.**

Let

$$f : R \to \mathbb{R}^n \tag{1.2}$$

be a function from an open region $R \subset \mathbb{R}^m$ into $\mathbb{R}^n$. If f is differentiable at a point $x \in R$, then the rank of x is defined to be the rank of the Jacobian matrix of f at x. If this rank is less than maximal, we call x a critical point. Note that the usual definition of a critical point is that derivative is not surjective. With this usual definition, if $m < n$, then every point is critical, however we wish to differentiate between the points which have less than maximal rank and those which have maximal rank. Note that in the case of infinite dimensional Sard's Theorem, it will be necessary to use the usual definition of critical point.

The proof of Theorem 1.17 involves finding a neighbourhood N of a critical point x_0 of rank $r > 0$, and showing that $f(N)$ is a null set. Without loss of generality, we may assume that the Jacobian is in reduced form, so that the r^{th} principle minor is invertible, and the bottom $(m-r)$ rows are 0. Indeed, if M is the Jacobian matrix at x_0, then we can always find a linear isomorphism A such that MA is the reduced form of M. Then $f \circ A$ has MA as its Jacobian matrix at the point x', where $Ax' = x_0$. Then if we can find a neighbourhood N of x' such that $(f \circ A)(N)$ is null, it follows that $N' = A(N)$ is a neighborhood of x_0 such that $f(N')$ is null.

Now if x_0 as a critical point of rank $r > 0$, we define the change of variables

$$u = (u^1, \ldots, u^m) = (f^1(x), \ldots, f^r(x), x^{r+1}, \ldots, x^m). \tag{1.3}$$

The Jacobian of u is

$$\begin{pmatrix} D_1 f^1 & \ldots & D_m f^1 \\ D_1 f^2 & \ldots & D_m f^2 \\ \vdots & & \vdots \\ D_1 f^r & \ldots & D_m f^r \\ \mathbf{0} & & \mathbf{I} \end{pmatrix},$$

where $\mathbf{0}$ is the $(m-r) \times (m-r)$ zero matrix, and $\mathbf{I}$ is the $r \times r$ identity matrix.

At the point x_0 the determinant of this matrix is non-zero. By the inverse function theorem it is locally invertible with inverse of class C^q near $u_0 = u(x_0)$. Denoting its local inverse by ϕ, near u_0 we have

$$\phi(u) = (\phi^1(u), \ldots, \phi^m(u)) = (x^1, \ldots, x^m).$$

Then we can write $f(x)$ in terms of the new variables u. Define the function

$$F(u) = (u^1, \ldots, u^r, f^{r+1}(\phi(u)), \ldots, f^n(\phi(u))), \tag{1.4}$$

and near u_0, we have $f(x) = F(u)$. If we let J be the Jacobian matrix of u, M' the Jacobian matrix of F, and M the Jacobian matrix of f, we see that

$$M'J = M.$$

M has rank r if and only if M' has rank r, so the critical point x_0 of f corresponds to the critical point u_0 of F. Furthermore, we have $F(u_0) = f(x_0)$, so we may consider (1.4) near u_0 instead of (1.2) near x_0.

1.3. **Critical points of** rank $< s$.

1.5. **Theorem.** *[1] Suppose that* (1.2) *is of class* C^1 *on* R. *Then the critical points of* (1.2) *constitute an* m*-null set.*

Proof. □

In [1], Sard gives a direct proof of this theorem for $m \leq n$, and the case of $m > n$ follows from Remark 1.1. In [2], Sard generalized this result, dropping the condition of differentiability on R. The proof is simpler than that of Theorem 1.5, however it requires more tools.

1.6. **Lemma.** *[2] If* x *is a critical point of* (1.2) *of rank* $r < s$, *and if* $\varepsilon > 0$ *and* $\alpha > 0$ *are given, then there exists* $\eta > 0$ *such that*

$$\mu_{s,\alpha}(f(\Omega)) \leq \varepsilon(\delta(\Omega))^s,$$

whenever Ω *is a set containing* x *with diameter less than* η.

Before moving on, we require a few more definitions. By a cube, here we mean a cube which is parallel the coordinate axes and open from above. That is a set

$$K = \{y : a^j \leq y^j < a^j + \gamma, \quad j = 1, \ldots, n\}.$$

We denote the length of the edge of the cube by $eK = \gamma$.

Now define the *cubical measure* of A as

$$c_{s,\alpha}(A) = \inf \sum_{\nu} (eK_\nu)^s,$$

where $\{K_\nu\}$ is a countable cover of B by cubes all of which satisfy $eK < \alpha$, and we define the *binary measure* of A as

$$b_{s,\alpha}(A) = \inf \sum_{\nu} (K_\nu)^s,$$

where the inf is taken of all countable coverings of B by cubes of the form

$$K = \{y : k^j/2^h \leq y^j < (k^j + 1)/2^h, \quad j = 1, \ldots, n\},$$

where the k^j and h are integers and $eK = 1/2^h < \alpha$.

Then we have

$$n^{-s/2}\mu_{s,\alpha n^{1/2}}(A) \le c_{s,\alpha}(A) \le b_{s,\alpha}(A) \tag{1.7}$$

since the infima are taken over successively smaller classes of sets, and a cube of side length eK has diameter $\sqrt{n}(eK)$.

We also get the inequality

$$b_{s,4\alpha}(A) \le 2^{n+2s}\mu_{s,\alpha}(A) \tag{1.8}$$

from equation (9) in [2].

1.9. **Lemma.** *[2] Let U be a cube in $\mathbb{R}^n$. For any ascending sequence of sets,*

$$B_\nu \subset B_{\nu+1} \subset U \qquad\qquad \nu = 1, 2, \dots,$$

and for any positive integer p, we have

$$\lim_{\nu\to\infty} b_{s,2^{-p}}(B_\nu) = b_{s,2^{-p}}\left(\lim_{\nu\to\infty} B_\nu\right),$$

Now we have enough to prove the generalization of Theorem 1.5.

1.10. **Theorem.** *Suppose that A is an s-sigmafinite set of critical points of (1.2). If the points of A are all of* rank $< s$, *then $f(A)$ is s-null.*

Note we do not assume that it is differentiable or even continuous outside of A, nor that it satisfies any Lipschitz condition on A.

Proof. If we can prove the result for when $\mu_s(A) < \infty$, then the case when $\mu_s(A) = \infty$ will follow by the countable decomposition of A into s-finite sets.

Assume that $\mu_s(A) < \infty$ and let $k = \mu_s(A) + 1$. Let p be any given positive integer, and ε any given positive constant. Set $\alpha = 2^{-(p+2)}$. For each $\nu \in \mathbb{N}$, let A_ν be the set of points $x \in A$ for which

$$\mu_{s,\alpha}(f(\Omega)) \le (\varepsilon/k)(\delta(\Omega))^s,$$

whenever Ω is a set containing x and $\delta(\Omega) < 1/\nu$.

By Lemma 1.6, we have

$$A_\nu \subset A_{\nu+1} \subset A = \bigcup_\nu A_\nu = \lim_{\nu\to\infty} A_\nu.$$

Since $\mu_s(A_\nu) \le \mu_s(A) < k$, we can find a cover $\{A_{\nu,i} : i \in \mathbb{N}\}$ of A_ν such that

$$\delta(A_{\nu,i}) < 1/\nu, \qquad \sum_i \{\delta(A_{\nu,i})\}^s < k.$$

Then by the definition of A_ν, we have

$$\mu_{s,\alpha}(f(A_{\nu,i})) \le (\varepsilon/k)(\delta(A_{\nu,i}))^s,$$

and hence

$$\mu_{s,\alpha}(f(A_\nu)) \le \sum_i \mu_{s,\alpha}(f(A_{\nu,i})) \le (\varepsilon/k)\sum_i (\delta(A_{\nu,i}))^s < \varepsilon.$$

If U is any cube in $\mathbb{R}^n$, then $\mu_{s,\alpha}(U \cap f(A_\nu)) < \varepsilon$, and by (1.8), we have

$$b_{s,4\alpha}(U \cap f(A_\nu)) = b_{s,2^{-p}}(U \cap f(A_\nu)) \le 2^{n+2s}\mu_{s,\alpha}(U \cap f(A_\nu)) < 2^{n+2s}\varepsilon.$$

Since $U \cap f(A_\nu)$ is an ascending sequence of sets contained in the cube U, we can apply Lemma 1.9, and get that

$$b_{s,2^{-p}}(U \cap f(A)) = \lim_{\nu\to\infty} b_{s,2^{-p}}(U \cap f(A_\nu)) \le \lim_{\nu\to\infty} 2^{n+2s}\varepsilon = 2^{n+2s}\varepsilon.$$

Now, by (1.7), we have

$$\mu_{s,n^{1/2}2^{-p}}(U \cap f(A)) \le b_{s,2^{-p}}(U \cap f(A)) \le n^{s/2}2^{n+2s}\varepsilon,$$

and letting $p \to \infty$, we have $n^{1/2}2^{-p} \to 0$, which gives

$$\mu_s(U \cap f(A)) \le n^{s/2}2^{n+2s}\varepsilon.$$

Since $n^{s/2}2^{n+2s}$ is a constant and ε was arbitrary, we have $\mu_s(U \cap f(A)) = 0$. Moreover, since this holds for any cube U, we can cover $f(A)$ in countably many cubes, and conclude that $\mu_s(A) = 0$. This completes the proof. □

1.11. **Corollary.** *If B is any set of critical points of f, then $f(B)$ is m-null.*

Proof. Since $B \subset \mathbb{R}^m$ and $\mathbb{R}^m$ can be covered by countable many balls, B is m-sigmafinite. Moreover every critical point has rank strictly less than m, so Theorem 1.10 applies with $s = m$. □

Hence Theorem 1.5 follows from Corollary 1.11 without any further hypothesis on the differentiability of (1.2).

1.4. Critical Points of rank 0.

First, we state a theorem from Morse.

1.12. **Theorem.** *[3]*

Given a positive integer q and a set A in the space of the variables $x = (x^1, \dots, x^m)$, there exists a sequence $A_0, A_1, \dots$ of sets with the following properties:

(1) $A = \bigcup_k A_k$,

(2) A_0 *is countable,*

(3) A_k *is bounded for* $k \ge 1$,

(4) *if* $f : \mathbb{R}^m \to \mathbb{R}$ *is any* C^q *function such that its set of critical points contains* A*, then*

$$\lim_{x_1 \to x} \frac{f(x_1) - f(x)}{\|x_1 - x\|^q} = 0$$

where $x \in A_k$ *and* x_1 *approaches* x *in* A_k*.*

Now we prove a result about rank 0 critical points.

1.13. **Theorem.** *[2] Suppose that* A *is a* s*-sigmafinite set of critical points of* (1.2) *all of rank* 0*. If* f *is of class* C^q *where* $q \geq 1$*, then* $f(A)$ *is* (s/q)*-null.*

Proof. Again, it suffices to prove this for the case when A is s-finite.

Let $K = \mu_s(A) + 1$, and decompose A into the subsets A_k as in Theorem 1.12. Since A_0 is countable, it is (s/q)-null. Now if we can show that A_k is (s/q)-null for any $k \geq 1$, it will follow that A is (s/q)-null since

$$f(A) = \bigcup_k f(A_k).$$

Note that any point $x \in A$ is a critical point of each component function f^j $(j = 1, \ldots, n)$ of (1.2), since A consists only of rank 0 critical points. Now fix any integer $k \geq 1$ and let $\varepsilon > 0$. For convenience of notation let $B = A_k$. Applying property (4) of Theorem 1.12, for each $x \in B$ we can find some $\eta > 0$ such that

$$|f^j(x_1) - f^j(x)| \leq (\varepsilon^{q/s}/2nK^{q/s})|x_1 - x|^q,$$

whenever $|x_1 - x| < \eta$ and $x_1 \in B$. Then if Ω is any set containing x of diameter $\delta(\Omega) < \eta$, and $x_1, x_2 \in B \cap \Omega$, we have

$$\begin{aligned}
\|f(x_1) - f(x_2)\| &\leq \|f(x_1) - f(x)\| + \|f(x) - f(x_2)\| \\
&\leq \sum_j |f^j(x_1) - f^j(x)| + |f^j(x_2) - f^j(x)| \\
&\leq \sum_j (\varepsilon^{q/s}/2nK^{q/s})|x_1 - x|^q + (\varepsilon^{q/s}/2nK^{q/s})|x_2 - x|^q \\
&\leq 2n(\varepsilon^{q/s}/2nK^{q/s})\delta(\Omega)^q \\
&= (\varepsilon/K)^{q/s}\delta(\Omega)^q.
\end{aligned}$$

Hence, we have

$$\delta(f(B \cap \Omega)) \leq (\varepsilon/K)^{q/s}\delta(\Omega)^q. \tag{1.14}$$

For any set Ω containing x with $\delta(\Omega) < \eta$.

Now we define B_ν to be the set of points x in B for with (1.14) holds for any set Ω containing x of diameter less than $\eta = 1/\nu$. Then we have

$$B_\nu \subset B_{\nu+1} \subset B = \lim_{\nu \to \infty} B_\nu = \bigcup B_\nu.$$

Now since

$$B_\nu \subset B \subset A,$$

we have

$$\mu_s(B_\nu) \le \mu_s(B) \le \mu_s(A) < K.$$

Then by the definition of μ_s, for each $\alpha > 0$ and each $\nu = 1, 2, \ldots$, there exists sets $B_{\nu,i}$ $(i = 1, 2, \ldots)$ which cover B_ν and are such that

$$\delta(B_{\nu,i}) < 1/\nu,\ (K/\varepsilon)^{1/s}\alpha^{1/q};\ \sum_i \delta(B_{\nu,i})^s < K.$$

Then we have

$$\delta(f(B_{\nu,i})) \le (\varepsilon/K)^{q/s}\delta(B_{\nu,i})^q < (\varepsilon/K)^{q/s}\left[(K/\varepsilon)^{1/s}\alpha^{1/q}\right]^q = \alpha,$$

and

$$\delta(f(B_{\nu,i}))^{s/q} \le (\varepsilon/K)\delta(B_{\nu,i})^s$$

by (1.14) with $\Omega = B_{\nu,i} = B \cap B_{\nu,i} = B \cap \Omega$.

Now since $\mu_{s/q,\alpha}$ is an infimum, we have

$$\mu_{s/q,\alpha}(f(B_\nu)) \le \sum_i \delta(B_{\nu,i})^{s/q} \le \varepsilon \sum_i \delta(B_{\nu,i})^s/K < \varepsilon K/K = \varepsilon.$$

Since α was arbitrary and B_ν doesn't depend on α, we can take α to 0 and get

$$\mu_{s/q}(f(B_\nu)) < \varepsilon,$$

and since ε was arbitrary, this gives

$$\mu_{s/q}(f(B_\nu)) = 0.$$

Hence we have

$$\mu_{s/q}(f(B)) = \mu_{s/q}\left(\bigcup_\nu f(B_\nu)\right) \le \sum_\nu \mu_{s/q}(f(B_\nu)) = \sum_\nu 0 = 0.$$ [1]

□

Again, Theorem 1.13 is a generalization of an earlier theorem from Sard.

1.15. **Theorem.** *[1] Let A be the set of critical points of 1.2 of rank 0 and suppose that f is of class C^q, $q \ge 1$. Then $f(A)$ is s-null if $s \ge (m/q)$.*

Proof. A is an m-sigmafinite set of critical points of rank 0 since $A \subset \mathbb{R}^m$ and $\mathbb{R}^m$ is m-sigmafinite. Then Theorem 1.13 applies with $s = m$, so $f(A)$ is (m/q)-null and hence s-null for any $s \ge (m/q)$. □

1.16. **Corollary.** *If (1.2) is smooth, and A is the set of critical points of rank 0, then $f(A)$ has Hausdorff dimension 0.*

---

[1] In [2], Sard uses the regularity of the measure $\mu_{s/q}$ to get the equation $\mu_{s/q}(f(B)) = \lim_\nu \mu_{s/q}(f(B_\nu)) = 0$, however we believe that this would require further hypotheses on the sets A_k or B_ν. In order to use regularity you would need to approximate $f(B)$ by compact subsets $f(B_\nu)$ which would require the B_ν to be compact. This is avoided by using the sub-additivity of $\mu_{s/q}$.

Proof. If f is of class C^∞, then it is of class C^q for every q, and hence $f(A)$ is (m/q)-null for every q. Then for every $\varepsilon > 0$, we can find q large enough so that $f(A)$ is ε-null. Hence the Hausdorff dimension of $f(A)$ is 0. □

1.5. **Sard's Theorem.**

1.17. **Theorem.** *[1] Suppose that $m > n$ and let A be the set of critical points of rank $r < n$ of (1.2). Then $f(A)$ is n-null if $q \geq \frac{m-r}{n-r}$.*

Proof. If $r = 0$, then this reduces to Theorem 1.15 with $s = n$.

Suppose that $0 < r < n$. We will show that there is a neighbourhood around each point of A for which $f(N \cap A)$ is an n-null set. Covering A in countably many of these neighbourhoods will show that $f(A)$ is n-null.

Consider a point $x_0 \in A$ and introduce the change of variables (1.3). Let $\overline{N}$ be the closure of an open neighbourhood N of u_0 (the image of x_0 under the change of variables). We regard $u^1, \ldots, u^r$ as parameters for each permissible set of values of which

$$\hat{F} = (F^{r+1}, \ldots, F^n)$$

defines a map from the $(m-r)$-space $(u^{r+1}, \ldots, u^m)$ into the $(n-r)$-space $(y^{r+1}, \ldots, y^n)$. That is, we are fixing $(u^1, \ldots, u^r)$ and allowing the rest of the coordinates to vary as much as they can. Let M^* be the matrix of partials of $\hat{F}$. Then M^* consists of the bottom $(m - r)$ rows of the Jacobian matrix of F, say M'. Then if M' has rank r, M^* has rank 0. Thus, if $(u^1, \ldots, u^m)$ is a critical point of rank r, then $(u^{r+1}, \ldots, u^m)$ is a critical point of rank 0 of $\hat{F}$ with the parameters $(u^1, \ldots, u^r)$. For each set of values of the parameters, the critical points of rank 0 of $\hat{F}$ map into an $(n - r)$-null set by Theorem 1.10 and our hypothesis on q.

Let B be the set of critical points of rank r of F in $\overline{N}$. First we prove that B is a countable union of compact sets. Let $B_{\leq r}$ be the set of critical points of F of rank $\leq r$. We show this set is closed by showing its compliment, the set of critical points of rank $> r$, is open. Indeed, we can write it as the union over i of $(h_i \circ D_x F)^{-1}[(-\infty, 0) \cup (0, \infty)]$, where h_i is the determinant of the i'th $(r+1)$ minor, and $D_x F$ is the matrix of partial derivatives of F. Since both of these functions are continuous, and $(-\infty, 0) \cup (0, \infty)$ is open, we have written the set $B^c_{\leq r}$ as a union of open sets. Now the set of critical points of F of rank r is the set $B_{\leq r} \cap B^c_{\leq r-1}$, and intersecting this with $\overline{N}$ gives the set B. Since $B^c_{\leq r-1}$ is open, we can write it as a countable union of compact sets, say B_ν. Then we have $B = \bigcup_\nu \overline{N} \cap B_{\leq r} \cap B_\nu$, which is a countable union of compact sets since the intersection of a compact set with a closed set is compact.

For convenience, write $B = \bigcup C_\nu$, where the C_ν are the compact sets from above. Now the cross sections of $F(C_\nu)$ for each $(y_1, \ldots, y_r)$ are all $(n - r)$-null. Since C_ν is compact, $F(C_\nu)$ is compact, so it is closed and hence measurable. Then the measure of $F(C_\nu)$ is computed by integrating the characteristic function of $F(C_\nu)$. Applying Fubini's Theorem, we first integrate over the cross sections which gives 0 since they are $(n-r)$-null. Hence $F(C_\nu)$ is n-null, so $F(B) = \bigcup_\nu F(C_\nu)$ is n-null.[2] Since $f(N \cap A) \subset F(B)$, we conclude that $f(N \cap A)$ is n-null, and hence $f(A)$ is. □

Again, Sard generalized this theorem in [2], where he considered points of rank $r < s$ and replaced m with s in the condition on q. The idea of the proof is the same, but it involves finding a bound on a particular integral.

Everything we have done has been with a function between Euclidean space, but we could have done this with general smooth manifolds with the appropriate definition of s-null. If M is a smooth manifold and $A \subset M$, then A s-null if for every chart (U, ϕ), the set $\phi(U \cap A)$ is s-null. We don't actually have to check this condition for each chart however.

If (U_ν, ϕ_ν) is any countable collection of charts covering the set $A \subset M$ and satisfies the condition that $\phi_\nu(U_\nu \cap A)$ is s-null for each ν, then for any other chart (U, ϕ) we have

$$\phi(U \cap A) = \bigcup_\nu \phi(U \cap U_\nu \cap A) = \bigcup_\nu \phi \circ \phi_\nu^{-1} \circ \phi_\nu(U \cap U_\nu \cap A).$$

Each $\phi_\nu(U \cap U_\nu \cap A)$ is s-null. Denoting it as B_ν, we have

$$\phi(U \cap A) = \bigcup_\nu \phi \circ \phi_\nu^{-1}(B_\nu),$$

which is a countable union of the images of s-null sets under the maps $\phi \circ \phi_\nu$. The fact that these images are s-null follows from a Lemma due to Sard.

1.18. **Lemma.** *Suppose $f\colon R \to \mathbb{R}^n$ is differentiable on a set $B \subset R \subset \mathbb{R}^m$, then if B is s-null or s-sigmafinite, $f(B)$ is the same. If B is s-finite and the functional entries in the Jacobian of f are absolutely bounded on B, then $f(B)$ is the same.*

Proof. See Lemma 1 of [2]. □

Hence we see that if A is s-null with respect to any collection of charts in an atlas $\mathcal{A}$, then it is s-null with respect to the maximal atlas containing $\mathcal{A}$.

Now we state and prove what is generally known as Sard's Theorem (see Theorem 6.10 of [4]).

1.19. **Theorem.** *Sard's Theorem*

Let M and N be smooth manifolds with dimensions m and n respectively. Let $f\colon M \to N$ be a smooth function, and A the set of critical points of f. Then $f(A)$ is n-null.

[2] In [1], Sard claims that $F(B)$ is closed and hence measurable, then uses the same cross section argument to conclude $F(B)$ is n-null. The closedness of $F(B)$ seemed unjustified, so we reformulated the argument in terms of compact sets.

Proof. By covering M in countably many coordinate charts, we reduce to the case where $f: R \to \mathbb{R}^n$ as before.

If $m \leq n$, then by 1.5, $f(A)$ is m-null and hence n-null. If $m > n$, then decompose A into the sets A_r of critical points of rank r. By Theorem 1.17, each set $f(A_r)$ is n-null and hence $f(A) = \bigcup_r f(A_r)$ is n-null. □

Of course the condition on f could be weekend, we really only need f to be of class C^q with $q \geq m - n + 1$. However, when dealing with smooth manifolds the usual interest is with smooth functions.

Again, if $f: M \to N$ is a smooth function between manifolds, we call $y \in N$ a regular value if it is not the image of a critical point. A set $B \subset N$ is called dense if its closure is the whole space N. We now have the following Corollary to Sard's Theorem.

1.20. **Theorem.** *If $f: M \to N$ is smooth, and $B \subset N$ is the set of regular values of f, then B is dense in N.*

Proof. Let $O \subset N$ be any non-empty open subset. If $O \subset B^c$, then we can find some non-empty open set $\tilde{O} \subset O \subset B^c$ which is contained in a single coordinate chart (U, ϕ). Then $\phi(\tilde{O}) \subset \phi(U \cap B^c)$, and $\phi(\tilde{O})$ is open since ϕ is a homeomorphism. Since $\phi(\tilde{O})$ is non-empty and open it has positive measure which contradicts Theorem 1.19. Hence every non-empty open subset of N intersects B. Now if $y \in N$, then every open neighbourhood of y intersects B. This shows that y is in the closure of B and completes the proof. □

2. Sard's Theorem for Banach Spaces

Many of the definitions and theorems from finite dimensional calculus can be naturally extended to infinite dimensions. The first notion we need is that of being continuously differentiable, which was taken from [5].

Let U be an open subset of a Banach space E_1 and E_2 another Banach space. A function $f: U \to E_2$ is called *(Frechet) differentiable at* $x_0 \in U$ if there exists a continuous linear map $\Lambda: E_1 \to E_2$ such that

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - \Lambda h\|_{E_2}}{\|h\|_{E_1}} = 0.$$

f is called *differentiable* if it is differentiable at x for every $x \in U$. If the map Λ exists, then it is unique and we call it $Df(x)$.

f is called *continuously differentiable* or *of class* C^1 if the map

$$Df: U \to L(E_1, E_2), \quad x \mapsto Df(x)$$

is continuous, where $L(E_1, E_2)$ is the space of continuous linear operators from E_1 to E_2 in the norm topology. The space $L(E_1, E_2)$ is again a Banach space, so we now consider the map Df. If it is differentiable, then the map

$$D^2 f\colon U \to L(E_1, L(E_1, E_2)), \quad x \mapsto D(Df)(x)$$

exists. We denote $L(E_1, L(E_1, E_2))$ as $L^2(E_1, E_2)$.

Again, if this map is continuous then we say that f is of class C^2. We continue in this way. If the map f is of class C^k, then the map $D^k f\colon U \to L(E_1, L^{k-1}(E_1, E_2))$ exists and is continuous. We denote $L(E_1, L^{k-1}(E_1, E_2))$ as $L^k(E_1, E_2)$. If $D^k f$ is differentiable, then the map $D^{k+1} f\colon U \to L(E_1, L^k(E_1, E_2)$ exists, and we denote $L(E_1, L^k(E_1, E_2)$ as $L^{k+1}(E_1, E_2)$. If $D^{k+1} f$ is continuous, then we say that f is of class C^{k+1}. If the map f is of class C^k for every k, then we say f is of class C^∞.

We also wish to generalize the idea of "small" sets in infinite dimensional spaces. We use a topological definition rather than a measure. A subset $A \subset X$ is said to be *nowhere dense* if the interior of the closure of A is empty. A subset $B \subset X$ is said to be *meagre* if it can be written as a countable union of nowhere dense sets.

The closure condition in the definition of nowhere dense is important if we are trying to capture the notion of "small" sets. We would like the countable union of "small" sets to be "small", and if we drop the closure condition this fails. For example, the sets $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$ both have empty interior, however their union is $\mathbb{R}$.

In general, Sard's Theorem in infinite dimensions does not hold. Ivan Kupka gave a counter example in [6] by constructing a function $F\colon l_2 \to \mathbb{R}$ which is of class C^∞, yet has critical values $[0, 1]$. We must put an extra assumption on our function to generalize to the case of infinite dimensions.

A *Fredholm operator* is a continuous linear map $L\colon E_1 \to E_2$ between Banach spaces with the following properties:

(1) $\dim \operatorname{Ker} L < \infty$,
(2) $\operatorname{Range} L$ is closed,
(3) $\operatorname{Coker} L = E_2 / \operatorname{Range} L$ has finite dimension.

The *Index* of a Fredholm operator L is the integer $\dim \operatorname{Ker} L - \dim \operatorname{Coker} L$. The set of Fredholm operators between E_1 and E_2 is an open subset in the space of all bounded operators between E_1 and E_2 in the norm topology, and the index function is continuous on the set of Fredholm operators (Theorem (1.1) of [7]).

We consider functions f that are defined on an open connected subset U of a Banach space E_1. A C^q, $q \geq 1$, function $f\colon U \to E_2$ is called a *Fredholm map* if $Df(x)\colon E_1 \to E_2$ is a Fredholm operator for each $x \in U$. The *Index* of f is the Index of $Df(x)$. Since U is connected and both Df and the index function are continuous, it follows that the index is constant and the definition does not depend on the choice of x.

First, we prove some lemmas regarding the structure of a Banach space

2.1. **Lemma.** *Let E_1 be a Banach space, $C \subset E_1$ a finite dimensional subspace and $K \subset E_1$ a closed subspace. If the addition map $+\colon K \times C \to E_1$ is a linear isomorphism, then it is also a topological isomorphism.*

Proof. If $\|\cdot\|$ is the norm on E_1 then the norms we are considering on C and K are the induced norms, $\|\cdot\|_K$ and $\|\cdot\|_C$. On $K \times C$, we are considering the product norm $\|\cdot\|_K + \|\cdot\|_C$.

Suppose that such spaces C and K are given. If we can show that the norm on $K \times C$ and the norm on E_1 are equivalent, then the topologies they generate will be the same.

By the triangle inequality, we get that $\|k + c\| \leq \|k\| + \|c\| = \|k\|_K + \|c\|_C$.

Now we will show that there is a $\lambda > 1$ such that $\|k\|_K + \|c\|_C \leq \lambda \|k + c\|$. First consider $c \in S(C) = \{c \in C : \|c\|_C = 1\}$.

Since $\|c\|_C = 1$ and $K \cap C = \{0\}$ (since $+$ is an isomorphism), we have $\|k + c\| \neq 0$. Then consider

$$\frac{\|k\|_K + \|c\|_C}{\|k + c\|} = \frac{\|k\|_K + 1}{\|k + c\|}.$$

If $\|k\|_K > 2$, then $\|k + c\| \geq \|k\| - 1$ by the (reverse) triangle inequality and we have

$$\frac{\|k\|_K + 1}{\|k + c\|} \leq \frac{\|k\|_K + 1}{\|k\| - 1} = \frac{2}{\|k\|_K - 1} + 1 \leq 3.$$

Now suppose that $\|k\|_K \leq 2$. First note that the distance between $S(C)$ and K is positive. If it were not then there would exist a sequence $k_n - c_n$ with $k_n \in K$ and $c_n \in S(C)$ such that $\|k_n - c_n\| \to 0$. Then since $S(C)$ is compact, there would exist a subsequence $c_{n_l} \to c$ with $c \in S(C)$. Then we would have

$$\|k_{n_l} - c\| \leq \|k_{n_l} - c_{n_l}\| + \|c_{n_l} - c\| \to 0.$$

Hence we would have $c \in K$ since K is closed, and this is a contradiction since $S(C) \cap K = \emptyset$.

Hence we have,

$$\|k - (-c)\| \geq \mathrm{d}(K, c) \geq \mathrm{d}(K, S(C)) > 0,$$

where d is the distance function. This gives

$$\frac{\|k\|_K + 1}{\|k + c\|} \leq \frac{\|k\|_K + 1}{\mathrm{d}(K, c)} \leq \frac{\|k\|_K + 1}{\mathrm{d}(K, S(C))} \leq \frac{3}{\mathrm{d}(K, S(C))}.$$

Let $\lambda > \max\left\{3, \frac{3}{\mathrm{d}(K,S(C))}\right\}$.

Now if $\|c\|_C = 0$, then we have $\|k\|_K \leq \lambda \|k\|$ since $\lambda > 1$. If $\|c\| = t \neq 0$, then we let $c' = c/t$ and $k' = k/t$, and since $\|c'\|_C = 1$ we get

$$\|k\|_K + \|c\|_C = t\big[\, \|k'\|_K + \|c'\|_C \,\big] \leq t(\lambda \|k' + c'\|) = \lambda \|t(k' + c')\| = \lambda \|k + c\|.$$

Therefore we have

$$\|k\|_K + \|c\|_C \leq \lambda \|k + c\|$$

for every $k \in K$ and $c \in C$. This shows that the norms are equivalent, and hence the topologies they generate are the same.

□

2.2. **Lemma.** *Let E_1 be a Banach space and $A \subset E_1$ a finite dimensional subspace. Then there exists a Banach space $E \subset E_1$ such that the addition map $+\colon E \times A \to E_1$ is a topological linear isomorphism.*

Proof. Since A is finite dimensional, let $e_1, \ldots, e_k$ be a basis, and let $f_1, \ldots, f_k$ be the dual basis. Then each f_i is a bounded linear functional on A, and by the Hanh-Banach Theorem (Theorem 3.2 in [8]) can be extended to a bounded linear functional g_i on all of E_1. Since boundedness and continuity are equivalent in Banach spaces, each g_i is continuous.

Now set $E = \bigcap_{i=1}^k \operatorname{Ker} g_i$. Then for any $x \in E_1$, we have $y = x - \sum_{i=1}^k g_i(x) e_i \in E$ since $g_i(y) = g_i(x) - g_i(x) g_i(e_i) = 0$. Then we have

$$x = y + \sum_{i=1}^k g_i(x) e_i,$$

and moreover if $x \in E \cap A$, then $x = \sum_{i=1}^k \alpha_i e_i$, and $g_j(x) = \alpha_j = 0$ for every j, so $x = 0$. Hence we have a direct sum decomposition

$$E_1 = E \oplus A.$$

Moreover since each g_i is continuous and $\operatorname{Ker} g_i = g_i^{-1}(0)$, E is an intersection of closed sets and hence is closed (and so a Banach space). Then Lemma 2.1 applies, and $+$ is also a topological isomorphism. □

Recall that a map $f\colon X \to Y$ between topological spaces is called *proper* if the inverse image of any compact set is compact.

2.3. **Theorem.** *Let $U \subset E_1$ be an open connected subset of a Banach space, and $f\colon U \to E_2$ a Fredholm map. For every $x \in U$ there exists a neighbourhood N_1 of x and a neighbourhood N_2 of $f(x)$ such that $f(N_1) \subset N_2$ and $f\big|_{N_1}\colon N_1 \to N_2$ is proper.*

Proof. See Theorem 1.6 of [7]. □

Previously, by a critical point we meant a point x for which $Df(x)$ had less than maximal rank. We cared about differentiating the cases of maximal rank and less than maximal rank when the dimension of the target space was greater than the domain. Here, by a critical point we just mean a point for which $Df(x)$ is not surjective.

2.4. **Theorem.** *[**Inverse Function Theorem**] Let $U \subset E_1$ be an open subset of a Banach space and $f\colon E_1 \to E_2$ be a map of class C^q, $q \geq 1$. Let $x_0 \in U$. If $Df(x_0)\colon E_1 \to E_2$ is a topological linear isomorphism, then f is locally invertible around x_0 and its inverse is of class C^q.*

Proof. Theorem 1.2 in chapter XIV of [5]. □

Note that the hypothesis of $Df(x_0)$ being a topological isomorphism is not actually necessary. If $Df(x_0)$ is a linear isomorphism, then by the Open Mapping Theorem (Theorem 2.11 in [8]), it is also a topological isomorphism.

2.5. **Lemma.** *Let E_1, E_2 be banach spaces, $U \subset E_1$ open, and $x_0 \in U$. Let $f\colon U \to E_2$ be a C^r, $r \geq 1$, Fredholm map, and $A = Df(x_0)$. Then there exists a finite dimensional subspace $C \subset E_2$ such that $E_2 = \operatorname{Im}(A) \times C$.*

Writing $f(x) = (f_1(x), f_2(x))$ with respect to this decomposition, there exists a C^r diffeomorphism $\phi\colon V \times W \to U_0$, where $V \subset \operatorname{Im}(A)$, $W \subset \operatorname{Ker}(A)$, and $U_0 \subset U$ are open, and are such that

$$f\phi(v, w) = (v, f_2\phi(v, w)).$$

Proof. Pick a basis $[e_1], \ldots, [e_n]$ for the finite dimensional space $E_2/\operatorname{Im}(A)$. Then $e_1, \ldots, e_n$ span a finite dimensional space $C \subset E_2$. For any $x \in E_2$, we can write $[x] = \sum_i \alpha_i[e_i] \in E_2/\operatorname{Im}(A)$. Hence $x = \sum_i \alpha_i e_i + a$ for some $a \in \operatorname{Im}(A)$. Moreover if $x \in \operatorname{Im}(A) \cap C$, then $x = \sum_i \alpha_i e_i$, and $[x] = 0$, so $\sum_i \alpha_i[e_i] = 0$, and since $[e_i]$ is a basis, each $\alpha_i = 0$. Hence, we have a direct sum decomposition $E_2 = \operatorname{Im}(A) \oplus C$, and again Lemma 2.1 shows that $+\colon \operatorname{Im}(A) \times C \to E_2$ is a topological linear isomorphism.

Now split $E_1 = E \times \operatorname{Ker}(A)$ as in Lemma 2.2 and $E_2 = \operatorname{Im}(A) \times C$ as above. With respect to these decompositions we have $f(p, q) = (f_1(p, q), f_2(p, q))$, $x_0 = (p_0, q_0)$, and

$$A = \begin{pmatrix} D_1 f_1(p_0, q_0) & 0 \\ 0 & 0 \end{pmatrix}.$$

By our choice of splitting of E_1 and E_2 we have $\operatorname{Im}(D_1 f_1(p_0, q_0)) = \operatorname{Im}(A)$, and $\operatorname{Ker}(D_1 f_1(p_0, q_0)) = \{0\}$, and hence $D_1 f_1(p_0, q_0)\colon E \to \operatorname{Im}(A)$ is a linear isomorphism.

Now we define

$$F\colon U_1 \times U_2 \to \operatorname{Im}(A) \times \operatorname{Ker}(A), \quad F(p, q) = (f_1(p, q), q),$$

where $U_1 \subset E$, $U_2 \subset \operatorname{Ker}(A)$ are open neighbourhoods of p_0 and q_0 respectively, and $U_1 \times U_2 \subset U$. In block form we have

$$DF(p_0, q_0) = \begin{pmatrix} D_1 f_1(p_0, q_0) & 0 \\ 0 & 1 \end{pmatrix},$$

which is invertible since $D_1 f_1(p_0, q_0)$ is. Then by Theorem 2.4, there exists open sets $V \times W \subset \operatorname{Im}(A) \times \operatorname{Ker}(A)$ containing $F(x_0)$ and a C^r diffeomorphism

$$\phi\colon V \times W \to U_0$$

such that $F\phi(v, w) = (v, w)$. Then we have $f\phi(v, w) = (v, f_2\phi(v, w))$. □

2.6. **Theorem.** *Let E_1 and E_2 be Banach spaces, $U_1 \subset E_1$ an open and connected subset, and $U_2 \subset E_2$ open. Suppose that $f\colon U_1 \to U_2$ is a proper Fredholm map of class C^q where $q \geq \max\{\operatorname{Index} f, 0\}$. Then the set of critical values of f is meagre.*

Proof. Let B be the set of critical points of f. Consider B^c, the set of points x for which $Df(x)$ is surjective, and let $x_0 \in B^c$. Since the set of surjective linear operators is open in the set of all bounded linear operators with respect to the norm topology (see Theorem 3.4 in chapter XV of [5]), it follows by the continuity of Df that there exists a neighbourhood N of x_0 such that $Df(x)$ is surjective for every $x \in N$. Since x_0 was arbitrary, B^c is open so B is closed.

Let y be a regular value of f. Suppose that y is in the closure of $f(B)$. Then we can find a sequence of points $f(x_n) \to y$ with $x_n \in B$. The set $\{f(x_n) : n \in \mathbb{N}\} \cup \{y\}$ is compact, and since f is proper, the pre-image is compact. Hence the sequence $\langle x_n \rangle$ has a convergent subsequence $x_{n_k} \to x$, with $x \in B$. Since $f(x_n) \to y$, we also have $f(x_{n_k}) \to y$, so by the continuity of f we have $f(x) = y$. This contradicts the fact that y is a regular value and hence no regular values are in the closure of $f(B)$. Thus, if we can prove that for any $x \in B$ and any neighbourhood N of $f(x)$ there is a regular value $y \in N$, this will show that no point of the closure of $f(B)$ is interior and we will be finished.

Given a point $x_0 \in B$, it is enough to show that there exists a neighbourhood U_0 of x_0 such that the statement holds for $f\big|_{U_0}$. This is because we can cover B in countably many of these neighbourhoods, and a countable union of meagre sets is meagre.

Consider the open set and diffeomorphism given by Lemma 2.5. Recall the notation, $\phi\colon V \times W \to U_0$ and $f\phi(v,w) = (v, f_2\phi(v,w))$. let $f(x_0) = (n_1, n_2)$, and let $N_1 \times N_2$ be a neighbourhood of $f(x_0)$. In block matrix form we have,

$$Df\phi(v,w) = \begin{pmatrix} 1 & 0 \\ D_1 f_2\phi(v,w) & D_2 f_2\phi(v,w) \end{pmatrix}.$$

It follows that $Df\phi(v,w)$ is surjective if and only if $D_2 f_2\phi(v,w)$ is surjective.

Consider $f_2\phi(n_1, \cdot)\colon W \to C$. Then by our hypothesis on q and the finite dimensional Sard's Theorem, there is a regular value z of $f_2\phi(n_1,\cdot)$ in N_2. Then (n_1, z) is a regular value for $f\phi$ in $N_1 \times N_2$. Indeed if $f\phi(v,w) = (n_1, z)$ then $v = n_1$ and $f_2\phi(n_1, w) = z$, so $D_2 f_2\phi(n_1, w)$ is surjective and hence $Df\phi(n_1, w)$ is surjective also.

The same value is then a regular value for f. If $(p,q) \in U_0$ and $f(p,q) = (n_1, z)$, then there exists $(v,w) \in V \times W$ such that $\phi(v,w) = (p,q)$. Then $f\phi(v,w) = (n_1, z)$, so $Df\phi(v,w) = Df(\phi(v,w)) \circ D\phi(v,w)$ is surjective and hence $Df(\phi(v,w)) = Df(p,q)$ is surjective.

□

Now we drop the condition of f being proper and prove a version of Sard's Theorem for Banach spaces first given by Stephen Smale in 1965 [7].

2.7. **Theorem.** *Let E_1 and E_2 be Banach spaces, $U \subset E_1$ open and connected, and $f\colon U \to E_2$ a Fredholm map of class C^q where $q \geq \max\{\operatorname{Index} f, 0\}$. Then the set of critical values of f is meagre.*

Proof. By Theorem 2.3 and the fact that U has a countable base for its topology, we decompose U into countably many open subsets U_n for which there exists open sets $V_n \subset E_2$ and are such that $f\big|_{U_n}: U_n \to V_n$ is proper. The result then follows by Theorem 2.6 and the fact that a countable union of meagre sets is meagre. □

References

[1] Arthur Sard. "The Measure of the Critical Values of Differentiable Maps". In: *Bulletin of the American Mathematical Society, Vol. 48* (1942).
[2] Arthur Sard. "Images of Critical Sets". In: *Annals of Mathematics, Vol.68* (1958).
[3] Anthony P. Morse. "The Behavior of a Function on its Critical Set". In: *Annals of Mathematics, Vol. 40* (1939).
[4] John M. Lee. *Introduction to Smooth Manifolds, Second Edition*. Springer-Verlag New York, 2003.
[5] Serge Lang. *Real and Functional Analysis, Third Edition*. Springer-Verlag New York, 1993.
[6] Ivan Kupka. "Counterexample to the Morse-Sard Theorem in the Case of Infinite Dimensional Manifolds". In: *Proceedings of the American Mathematical Society, Vol. 16* (1965).
[7] Stephen Smale. "An Infinite Dimensional Version of Sard's Theorem". In: *American Journal of Mathematics, Vol. 87* (1965).
[8] Walter Rudin. *Functional Analysis, Second Edition*. McGraw-Hill, New York, 1991.