The Morse-Sard Theorem Demystified

Andrew McLennan

Abstract. We provide a result that is stronger than the Morse-Sard theorem in its classic form. The proof is centered on an argument of Dubovitskiĭ [**2**] that clearly illuminates the driving force of the result. It does not require advanced background, and its demands on the reader are comparable to those of Milnor's [**6**] proof of the C^∞ case of the classic version. Applications of the Morse-Sard theorem in economic theory are explained, and Harsanyi's [**4**] application to noncooperative game theory is proved.

1. INTRODUCTION The Morse-Sard theorem is one of the most important theorems of real analysis. There is a popular proof due to Milnor [**6**] which is unfortunately restricted to C^∞ functions. The result has evolved quite a bit over time, and I recently wrote a paper [??] that brings almost all of the accumulated additional generality (except for results concerning Sobolev spaces) under one roof, with a relatively simple and unified proof.

The current paper aims at a broader audience. It has three main objectives:

- Explain the nature and importance of applications of the Morse-Sard theorem in theoretical economics to a mathematically literate audience with little prior economic background.
- Provide a "basic" version of the result and proof in [??] that is sufficient for almost all applications, and which can be understood without too much effort or high level background, and in particular without much more effort than is required by Milnor's argument, but which is more general and more principled in the sense of better illuminating the driving force of the result.
- Provide a suitable introduction and "appetizer" for [??].

At this point we develop just enough background to state the Morse-Sard theorem in its classic form. A set $S \subset \mathbb{R}^n$ *has Lebesgue measure zero* if, for any $\varepsilon > 0$, there is a countable collection of sets $S_1, S_2, \ldots \subset \mathbb{R}^n$ such that $S \subset \bigcup_i S_i$ and $\sum_i (\operatorname{diam} S_i)^n < \varepsilon$ where $\operatorname{diam} S_i$ is the supremum of the set of distances between pairs of points in S_i. A set of Lebesgue measure has an empty interior. In particular, it is easy to see that if $C \subset \mathbb{R}^n$ is a cube, then the sum above is not less than a constant (the ratio of the volume of a ball to the volume of the smallest cube containing it) times the usual notion of the volume of C. Since the theory of Lebesgue measure is fairly technical, it is fortunate that this is all we need to know about it!

Suppose that $U \subset \mathbb{R}^m$ is open and $f : U \to \mathbb{R}^n$ is a C^1 function. A point $x \in U$ is a *critical point* of f if the rank of $Df(x)$ is less than n, and otherwise x is a *regular point* of f. Let C_f be the set of critical points of f. The *critical values* of f are the elements of $f(C_f)$, and the *regular values* of f are the elements of $\mathbb{R}^n \setminus f(C_f)$. (Note that a regular value need not be a "value" of f in the usual sense.)

Theorem 1 (Morse [7], Sard [8]). *If $U \subset \mathbb{R}^m$ is open and $f : U \to \mathbb{R}^n$ is C^k where $k \geq \max\{m - n + 1, 1\}$, then $f(C_f)$ has Lebesgue measure zero.*

At this point it is appropriate to say a few words about smooth manifolds. If k is a positive integer, or if $k = \infty$, a set $M \subset \mathbb{R}^p$ is an *m-dimensional C^k manifold* if each $x \in M$ has a neighborhood $V \subset M$ such that there is an open $U \subset \mathbb{R}^m$ and a C^k diffeomorphism[1] $\varphi : U \to V$. Such a φ is a C^k *parameterization* for M, and φ^{-1} is a C^k *coordinate chart* for M. In addition to M, suppose that $N \subset \mathbb{R}^q$ is an n-dimensional C^k manifold. A function $f : M \to N$ is C^k if $\psi \circ f \circ \varphi$ is C^k whenever $\varphi : U \to V$ is a C^k parameterization for M, $\psi : W \to X$ is a C^k coordinate chart for N, and $f(V) \subset W$. (It is easy to verify that this definition is equivalent to the one given in the footnote.)

In this setting a point $p \in V$ is a *regular point* (*critical point*) of f if $\varphi^{-1}(p)$ is a regular (critical) point of $\psi \circ f \circ \varphi$. Let C_f be the set of critical points of f. A set $S \subset N$ *has Lesbesque measure zero* if $\psi(S \cap W)$ has Lebesgue measure zero whenever $\psi : W \to X$ is a C^k coordinate chart for N. That $f(C_f)$ has Lebesgue measure zero when $f : M \to N$ is C^k and $k \geq \max\{m - n + 1, 1\}$ is now an immediate consequence of the result above. Evidently the reason that Sard's theorem is usually expressed in terms of maps between Euclidean spaces, rather than the more general setting of maps between smooth manifolds, is not because the extension is either difficult or unimportant, but rather because it is immediate and obvious!

The remainder has two main parts, which are logically independent and can be read in either order. In the next section we explain the historical and conceptual importance of the applications of Sard's theorem in economic theory, and we state and prove one important example of such a result. Our version of the Morse-Sard theorem is stated in Section 3, and the remainder of the paper after that is devoted to the proof.

2. ECONOMIC APPLICATIONS The writings of the great classical economists, such as Smith, Ricardo, and Mills, are entirely verbal. In various cases the modern reader can sense a mathematical model underlying the many words, but with the very important exception of Cournot, it was only during the late 19th century that formal economic modelling got underway. We will briefly describe how Léon Walras began the study of general equilibrium theory, in which supply and demand is equalized simultaneously in multiple markets.

We consider one of the simplest possible frameworks. There are ℓ commodities, each of which is perfectly divisible. There are m consumers. Each consumer $i = 1, \ldots, m$ has an *initial endowment* $\omega_i \in \mathbb{R}^\ell_{++}$ of positive amounts of all commodities. (In economics $\mathbb{R}^\ell_+$ and $\mathbb{R}^\ell_{++}$ are the standard notations for the nonnegative and positive orthants of $\mathbb{R}^\ell$.) The *aggregate endowment* is $\omega = \sum_i \omega_i$. There is no production, so the only things that happen are that all of the agents take their endowments to the market and trade with each other, then each goes home and consumes what she has at the end of the day.

Suppose $p \in \mathbb{R}^\ell_+$ is a vector of prices[2] of the various commodities. We imagine that consumer i sells her endowment, obtaining the income $\langle p, \omega_i \rangle$, after which she purchases one of her favorite bundles in her *budget set*

$$B_i(p) = \{\, x \in \mathbb{R}^\ell_+ : \langle p, x \rangle \leq \langle p, \omega_i \rangle \,\}.$$

Note that $B_i(\alpha p) = B_i(p)$ for all $\alpha > 0$, which economists describe by saying that only "relative prices" (the ratios of prices of pairs of commodities) matter. Let

$$\Delta = \{\, p \in \mathbb{R}^\ell_+ : \sum_h p_h = 1 \,\} \quad \text{and} \quad \Delta^\circ = \{\, p \in \mathbb{R}^\ell_{++} : \sum_h p_h = 1 \,\}$$

[1]For an arbitrary set $S \subset \mathbb{R}^a$, a function $\psi : S \to \mathbb{R}^b$ is C^k, by definition, if there is an open $\tilde{S} \subset \mathbb{R}^a$ that contains S and a C^k function $\tilde{\psi} : \tilde{S} \to \mathbb{R}^b$ such that $\psi = \tilde{\psi}|_S$. Such a function is a C^k *diffeomorphism* if it is injective and its inverse is also C^k.

[2]It is more correct conceptually, but uncommon in the literature, to treat p as an element of the dual of $\mathbb{R}^\ell$.

be the standard unit simplex in $\mathbb{R}^\ell$ and its interior (relative to the hyperplane that is its affine hull). We assume that for each $p \in \Delta^\circ$ consumer i has a unique favorite element of $B_i(p)$, which we denote by $f_i(p)$, and that $f_i(p) \in \mathbb{R}^\ell_{++}$ and $\langle p, f_i(p)\rangle = \langle p, \omega_i\rangle$. In addition we assume that $f_i : \Delta^\circ \to \mathbb{R}^\ell$ is continuous, and if $\{p^k\}$ is a sequence in Δ° converging to a point in $\Delta \setminus \Delta^\circ$, then $|f_i(p^k)| \to \infty$. Hopefully the reader can easily appreciate the intuititive significance of these assumptions. It is possible to derive them from assumptions on an underlying preference ordering of $\mathbb{R}^\ell_+$, but we will not do so.

A *Walrasian equilibrium price vector* is a $p \in \Delta^\circ$ such that $\sum_i f_i(p) = \omega$: supply is equal to demand in each of the ℓ markets. Initially this looks like a system of ℓ equations, but Walras recognized that one of them is redundant: if supply and demand are equal in $\ell - 1$ markets, then in each of them total expenditure is equal to the value of that commodity's initial endowment, but each consumer spends all her income, so total expenditure across all markets is equal to $\langle p, \omega\rangle$, and consequently the expenditure in the final market is also equal to the value of that commodity's initial endowment, so supply is equal to demand in that market as well. Thus we have a system of $\ell - 1$ equations in $\ell - 1$ independent variables. This principle is known as *Walras' law*.

What is the scientific import of this model? Classical mechanics is deterministic and quantitative: for a given initial condition, the laws of motion give precise numerical predictions concerning the state of the system at any future time. General equilibrium theory cannot hope for anything so luxurious, because (among other reasons) we have very little information concerning the parameters of our model. An economic model is already quite successful if its qualitative features are in accord with our general sense of how the world works. While equality of supply and demand in any particular market is always at best approximate, large and persistent discrepancies seem to be rare, so we have the overall sense that general equilibrium theory "works" for arbitrary values of the parameters (endowments, preferences, and technology if the model includes production). Our general experience is that, for the most part, the equilibrium we observe changes only slightly as the given parameters vary from one day, month, year, or decade, to the next, and it seems intuitive reasonable to hope that this can be explained by some corresponding mathematical property of the theory.

The standards of rigor and precision of nineteenth century economics were much lower than we are accustomed to, so it is unlikely that Walras' writings contained exact formulations. Nevertheless he is credited with conjecturing that, for any particular values of the parameters of a model similar to the one above, there is at least one equilibrium price vector. Simple examples show that there can be more than one equilibrium, and in fact it not hard to construct examples with infinitely many equilibrium price vectors. Playing around with such examples suggests that finite multiplicity can be robust, in the sense that it can be the case that when the parameters are perturbed, there continue to be multiple equilibria, but infinite multiplicity is not. That is, when there are infinitely many equilibrium price vectors, after arbitrarily small perturbations of the parameters there are only finitely many equilibria. Walras conjectured that, in some sense, it is "typically" the case that there are finitely many equilibria that are well behaved, in the sense that for each of these there is a continuous function from nearby parameters to nearby equilibria.

That Walras was unable to prove his conjectures now seems completely unsurprising, since the requisite tools were major results of twentieth century mathematics. In the early 1950's Arrow-Debreu and McKenzie showed that the existence of an equilibrium price vector can be understood as a consequence of Brouwer's fixed point theorem. (Later work established a converse: if you know that there is always an equilibrium price vector, then you can easily prove Brouwer's fixed point theorem.) Almost twenty years later Debreu [**1**] used the Morse-Sard theorem to give a sense in which it is "typically" the case that there are finitely many equilibrium price vectors, each of which varies continuously under small perturbations of the parameters.

The other foundational model of mathematical economics is the theory of noncooperative games, for which there are similar results. One of the notational conventions of game theory is that there are n players. For each $i = 1, \dots, n$ let A_i be a nonempty finite set of *actions*, and let $A = A_1 \times \cdots \times A_n$ be the set of *action profiles*. When the game is played the players simultaneously (in the sense that each decides before learning the choices of others) choose actions, which results in an action profile. For each i there is a function $u_i : A \to \mathbb{R}$ that gives the payoff $u_i(a)$ that player i receives when the profile a occurs. A *pure Nash equilibrium* is an action profile a^* such that no player can improve her payoff by deviating: for all i and $a_i \in A_i$, $u_i(a^*_{-i}, a_i) \leq u_i(a^*)$. (Here a^*_{-i} denotes the projection of a^* onto $\prod_{j \neq i} A_j$.)

	C_r	D_r
C_c	$(2,2)$	$(0,3)$
D_c	$(3,0)$	$(1,1)$

(a) Prisoner's Dilemma

	H_r	T_r
H_c	$(1,-1)$	$(-1,1)$
T_c	$(-1,1)$	$(1,-1)$

(b) Matching Pennies

	B_r	S_r
B_c	$(3,2)$	$(0,0)$
S_c	$(0,0)$	$(2,3)$

(c) Battle of the Sexes

Figure 1

Figure 1 shows the three most famous examples of games in which two players each have two pure strategies. In such examples the first player is called the *row player* because she chooses which row of the "bimatrix" will occur, and the second player is the *column player*. The first component of the entry corresponding to an action profile is the payoff to the row player, and the second entry is the column player's payoff.

In the Prisoner's Dilemma each player chooses between "cooperate" and "defect." For each, defecting is a *strictly dominant strategy* in the sense that it gives a strictly greater payoff no matter what the other player does, so (D_c, D_r) is the only pure Nash equilibrium. This outcome is *strongly Pareto dominated* by (C_c, C_r): at the latter outcome both players receive strictly higher payoffs. This is in contrast with the *first welfare theorem of economics* which asserts that the allocation resulting from a Walrasian equilibrium is *Pareto optimal*; there is no reallocation of ω that makes no one worse off and some agent strictly better off. (The intuition of the proof is very simple: in an improving allocation everyone must be receiving a bundle of goods that costs at least as much as their equilibrium bundle, and the agents who are strictly better off are receiving bundles that cost strictly more, so the total value of the resources in the improving allocation must be greater than the value of ω, which is impossible.)

In Matching Pennies each player chooses either heads or tails, the row player wins 1 and column player loses 1 if their choices are the same, and it is the other way around when their choices are different. It is easy to see that Matching Pennies has no pure Nash equilibrium. Especially in view of examples such as penalty kicks in soccer and pitch selection in baseball, randomized strategies are intuitively appealing. For each i let Σ_i be the set of functions $\sigma_i : A_i \to [0,1]$ such that $\sum_{a_i \in A_i} \sigma_i(a_i) = 1$, and let $\Sigma = \Sigma_1 \times \cdots \times \Sigma_n$. (Insofar as σ_i is "really" a probability measure, the standard notational conventions of game theory are admittedly a bit idiosyncratic.) Elements of Σ_i are *mixed strategies* for player i, and elements of Σ are *mixed strategy profiles*.

Insofar as the players move simultaneously, we assume that a mixed strategy profile induces the product measure on A. For each i there is an induced payoff function $u_i : \Sigma \to \mathbb{R}$ that is computed by taking expectations:

$$u_i(\sigma) = \sum_{a \in A} \Big(\prod_j \sigma_j(a_j) \Big) u_i(a).$$

A *Nash equilibrium* is a mixed strategy profile σ^* such that no player can improve her payoff by deviating: for all i and $\sigma_i \in \Sigma_i$, $u_i(\sigma^*_{-i}, \sigma_i) \le u_i(\sigma^*)$. Applying the distributive law to the formula above gives

$$u_i(\sigma^*_{-i}, \sigma_i) = \sum_{a_i \in A_i} \sigma_i(a_i) u_i(\sigma^*_{-i}, a_i),$$

so σ^* is a Nash equilibrium if and only if each σ^*_i assigns all probability to i's *pure best responses* to σ^*, which are those a_i that maximize $u_i(\sigma^*_{-i}, a_i)$.

A two player game is *strictly competitive* or *zero-sum* if the sum of the two player's payoffs is the same at all pure strategy profiles. (Since adding a constant to a player's payoffs does not affect her incentives, any strictly competitive game is equivalent to one in which the sum of the players payoffs is actually always zero.) In 1928 von Neumann formulated and proved the minimax theorem, which implies that such games have Nash equilibria. In the early 1940's Kakutani pointed out that a much simpler proof of the minimax theorem could be obtained by applying a slight (but very important!) generalization of Brouwer's fixed point theorem, and a few years later Nash developed the notion that became known as Nash equilibrium, and used Kakutani's argument to prove that every game has a least one Nash equilibrium.

The story of the Battle of the Sexes is that on Valentine's day each member of a couple surprised the other by sending a ticket to a concert, and now they must simultaneously decide which concerts to attend. (Over time the "simultaneity" in this story becomes ever more difficult to reconcile with contemporary technology!) The row player prefers Bach and the column player prefers Stravinsky, but not being together is a disaster for both. Evidently (B_c, B_r) and (S_c, S_r) are pure Nash equilibria, and there is an additional equilibrium $(\frac{3}{5}B_c + \frac{2}{5}S_c, \frac{2}{5}B_r + \frac{3}{5}S_r)$. It is easy to see that for nearby payoffs there are also three equilibria that vary continuously.

A game in which some u_i is a constant function obviously has infinitely many Nash equilibria, and there are many other examples. But for any such example you might study, you will find that the infinite multiplicity is not robust: any neighborhood of the given payoff contains payoffs for which there are finitely many equilibria. In 1973 Harsanyi (a colleague of Debreu at UC Berkeley, and a fellow Nobel laureate) used the Morse-Sard theorem to prove this. We are now going to present his proof.

Let G be the set of $u = (u_1, \dots, u_n)$ in which each u_i is a real valued function on A. Elements of G are called *games*. As a mathematical object, G can be identified with $\mathbb{R}^{d^G}$ where $d^G = |A| \cdot |I|$. Let

$$E = \{\, (u, \sigma^*) \in G \times \Sigma : \sigma^* \text{ is a Nash equilibrium for } u \,\},$$

and let $\pi : E \to G$ be the natural projection. The main idea is to apply the Morse-Sard theorem to restrictions of π to well behaved subsets of E.

The *support* of a mixed strategy σ_i is the set

$$\operatorname{supp}(\sigma_i) = \{\, a_i \in A_i : \sigma_i(a_i) > 0 \,\}$$

of actions that are assigned positive probability. For a nonempty $A_i^* \subset A_i$ let

$$\Sigma_i^{A_i^*} = \{\, \sigma_i \in \Sigma_i : \mathrm{supp}(\sigma_i) = A_i^* \,\}.$$

For nonempty $A_1^* \subset A_1, \ldots, A_n^* \subset A_n$ let $\Sigma^{A^*} = \Sigma_1^{A_1^*} \times \cdots \times \Sigma_n^{A_n^*}$, and let

$$E^{A^*} = \{\, (u, \sigma^*) \in E : \sigma^* \in \Sigma^{A^*} \,\}.$$

Suppose that for each i, $A_i^* \subset A_i^\circ \subset A_i$. Let

$$E^{A^*,A^\circ} = \{\, (u, \sigma^*) \in E^{A^*} : \text{for each } i,\ A_i^\circ \text{ is the set of pure best responses to } \sigma^* \,\}.$$

Evidently the various sets E^{A^*,A° constitute a partition of E.

For each i we fix $a_i^* \in A_i^*$. We now have

$$E^{A^*,A^\circ} = \{\, (u, \sigma^*) \in\ G \times \Sigma^{A^*} : \text{for each } i,\ u_i(\sigma_{-i}^*, a_i) = u_i(\sigma_{-i}^*, a_i^*) \text{ for all } a_i \in A_i^\circ \setminus \{a_i^*\} \text{ and } u_i(\sigma_{-i}^*, a_i) < u_i(\sigma_{-i}^*, a_i^*) \text{ for all } a_i \in A_i \setminus A_i^\circ \,\}.$$

Of course $G \times \Sigma^{A^*}$ is a $(d^G + \sum_i |A_i^*| - n)$-dimensional C^∞ manifold. Since the payoffs associated with the different actions can be varied freely, it is easy to see that every point of $G \times \Sigma^{A^*}$ is a regular point of the function $f : G \times \Sigma^{A^*} \to \prod_i \mathbb{R}^{A_i^\circ \setminus \{a_i^*\}}$ whose components are the functions $u_i(\sigma_{-i}^*, a_i) - u_i(\sigma_{-i}^*, a_i^*)$, so the regular value theorem[3] implies that $f^{-1}(0)$ is a d^{A^*,A°-dimensional C^∞ manifold where $d^{A^*,A^\circ} = d^G - \sum_i |A_i^\circ \setminus A_i^*|$. The inequalities $u_i(\sigma_{-i}^*, a_i) < u_i(\sigma_{-i}^*, a_i^*)$ for $a_i \in A_i \setminus A_i^\circ$ define an open subset of $f^{-1}(0)$, so E^{A^*,A° is a d^{A^*,A°-dimensional C^∞ manifold.

If, for some i, A_i° is a proper superset of A_i^*, then $d^{A^*,A^\circ} < d^G$, and the Morse-Sard theorem (generalized to manifolds, as described in Section 1) implies that $\pi(E^{A^*,A^\circ})$ has Lesbesgue measure zero. A Nash equilibrium σ^* of u is *strict* if unused actions are not pure best responses to σ^*: $u_i(\sigma_{-i}^*, a_i) < u_i(\sigma^*)$ for all i and $a_i \in A_i \setminus \mathrm{supp}(\sigma_i^*)$. Taking the union over the finite number of possibilities, we have shown that the set of games that have non-strict Nash equilibria has Lebesgue measure zero.

Now suppose that $A_i^\circ = A_i^*$ for all i, so that $E^{A^*,A^*} = E^{A^*,A^\circ}$ is d^G-dimensional. In this case the Morse-Sard theorem asserts that the set of critical values of $\pi|_{E^{A^*,A^*}}$ has Lesbesgue measure zero. Since a finite union of sets of Lebesgue measure zero has Lebesgue measure zero, we know that for all u outside a set of Lebesgue measure zero, all Nash equilibria are strict and u is not a critical value of any $\pi|_{E^{A^*,A^*}}$. Fixing such a u, let σ^* be a Nash equilibrium of u. Then (u, σ^*) is an element of some E^{A^*,A^*}, and (u, σ^*) is a regular point of $\pi|_{E^{A^*,A^*}}$. The inverse function theorem (applied to C^∞ coordinate charts for neighborhoods of (u, σ^*) and u) implies that there is C^∞ diffeomorphism between an open neighborhood $V \subset E^{A^*,A^*}$ and an open neighborhood $U \subset G$ of u. In particular, $\pi|_V^{-1}$ induces a C^∞ function that maps u to σ^* and each game in U to a Nash equilibrium of that game.

[3]The *regular value theorem* asserts that if $M \subset \mathbb{R}^p$ is an m-dimensional C^k manifold, $N \subset \mathbb{R}^q$ is an n-dimensional C^k function, $g : M \to N$ is C^k, and q is a regular value of g, then $g^{-1}(q)$ is an $(m-n)$-dimensional C^k manifold. It can be thought of as a "coordinate free" version of the implicit function theorem, and its proof is a matter of constructing coordinate charts within which the implicit function theorem can be applied.

We have shown that each preimage of u in E^{A^*,A^*} has a neighborhood that contains no other preimage. If there was an infinite sequence of distinct preimages, it would have a subsequence that converged to a point in the closure of E^{A^*,A^*}, and the limit point $(u, \tilde{\sigma}^*)$ could not be in E^{A^*,A^*} itself, so it would necessarily be an element of $E^{\tilde{A},A^*}$ for some sets $\tilde{A}_1 \subset A_1^*, \ldots, \tilde{A}_n \subset A_n^*$ with $\tilde{A}_i$ a proper subset of A_i^* for at least one i, which means that $\tilde{\sigma}^*$ is a Nash equilibrium of u that is not strict, but we have assumed that u has no such Nash equilibria. Thus $(\pi|_{E^{A^*,A^*}})^{-1}(u)$ must be finite, and since there are finitely many possibilities for A^*, we conclude that the set of Nash equilibria of u is finite.

Summarizing:

Theorem 2. *For every $u \in G$ outside a set of Lebesgue measure zero there are finitely many equilibria, and for each equilibrium σ^* there is a neighborhood $U \subset G$ of u and a C^∞ map $f : U \to \Sigma$ such that $f(u) = \sigma^*$ and, for every $u' \in U$, $f(u')$ is a Nash equilibrium of u.*

3. STATEMENT OF THE THEOREM

In this section we provide the statement of the version of the Morse-Sard theorem we are going to prove, preceded by the requisite background.

Let Y be a *normed space* over $\mathbb{R}$, so Y is a (possibly infinite dimensional) real vector space endowed with a function on $|\cdot| : Y \to [0, \infty)$ (the *norm*) satisfying $|y| = 0$ if and only if $y = 0$, $|\alpha y| = |\alpha| \cdot |y|$ for all $\alpha \in \mathbb{R}$ and $y \in Y$, and $|y + y'| \leq |y| + |y'|$ for all $y, y' \in Y$. Defining the distance between y and y' to be $|y - y'|$ makes Y into a metric space.

Let E be m-dimensional Euclidean space. We need to discuss differentiation of functions from E to Y. It is possible to define partial derivatives, then to say that f is C^k if its partial derivatives up to order k are all defined and continuous, and then to prove Taylor's theorem, but for our purposes it is simpler to take the conclusion of Taylor's theorem as the definition of a C^k function.

A function $\mu \colon E^i \to Y$ is *multilinear* if, for all $j = 1, \ldots, i$ and all $v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_i \in E$,

$$\mu(v_1, \ldots, v_{j-1}, \cdot, v_{j+1}, \ldots, v_i) : E \to Y$$

is linear. Such a function is *symmetric* if $\mu(v_{\sigma(1)}, \ldots, v_{\sigma(i)}) = \mu(v_1, \ldots, v_i)$ for all $v_1, \ldots, v_i \in E$ and all permutations (i.e., bijections) $\sigma : \{1, \ldots, i\} \to \{1, \ldots, i\}$. We endow the space of symmetric multilinear functions $\mu \colon E^i \to Y$ with the norm

$$|\mu| = \max_{|v_1|, \ldots, |v_i| \leq 1} |\mu(v_1, \ldots v_i)|.$$

Fix an open $U \subset E$, a positive integer k, and a function $f : U \to Y$. By definition, f is C^k if, for each $i = 0, \ldots, k$, there are symmetric multilinear functions $D^i f(x) \colon E^i \to Y$ which vary continuously with x, such that for each $x \in U$ and $\varepsilon > 0$ there is a $\delta > 0$ such that

$$|f(x + h) - \sum_{i=0}^{k} \tfrac{1}{i!} D^i f(x)(h, \ldots, h)| \leq \varepsilon |h|^k$$

for all $h \in E$ such that $x + h \in U$ and $|h| < \delta$. Of course $D^0 f$ is simply f itself, and we always write Df in place of $D^1 f$.

Fix an $\alpha \in [0,1]$. The function f is $C^{k,\alpha+}$ if it is C^k and $D^k f$ satisfies the Hölder continuity condition that for each compact $K \subset U$ there is a nondecreasing continuous $\epsilon_K \colon \mathbb{R} \to \mathbb{R}$ with $\epsilon_K(0) = 0$ such that

$$|D^k f(x) - D^k f(y)| \le \epsilon_K(|x-y|)|x-y|^\alpha$$

for all $x, y \in K$. (A variant of this concept is defined in Section 8.) Note that, insofar as a continuous function is uniformly continuous on any compact subdomain, f is $C^{k,0+}$ if and only if it is C^k. For a general $S \subset E$ a function $f \colon S \to Y$ is, by definition, $C^{k,\alpha}$ ($C^{k,\alpha+}$) if it has a $C^{k,\alpha}$ ($C^{k,\alpha+}$) extension $\tilde{f} \colon U \to Y$ to an open $U \subset E$ containing S.

Compositions of $C^{k,\alpha+}$ functions are $C^{k,\alpha+}$, and the inverse and implicit function theorems hold for this class of functions, in the sense that the output of each of these theorems is a $C^{k,\alpha+}$ function if the given function is $C^{k,\alpha+}$. (See Appendix B of [??] for details.) In what follows we treat these theorems as known, as one usually would in connection with C^k functions.

We now introduce the theory of Hausdorff measure, which is fortunately much simpler (at the level we need) than Lebesgue measure, and in addition allows noninteger dimensions. For a metric space (X, d), a subset $S \subset X$, and numbers $\tau \ge 0$ and $\beta > 0$, let

$$\mathcal{H}^\tau_\beta(S) = \inf \sum_{i=1}^{\infty} (\operatorname{diam} T_i)^\tau$$

where $\operatorname{diam} T_i = \sup_{x,y \in T_i} d(x,y)$ and the infimum is over all countable collections $\{T_i\}$ of subsets of S such that $\operatorname{diam} T_i < \beta$ for all i and $S \subset \bigcup_i T_i$. Clearly $\mathcal{H}^\tau_\beta(S)$ is a nonincreasing function of β, and the τ-dimensional *Hausdorff measure* of S is

$$\mathcal{H}^\tau(S) = \lim_{\beta \to 0} \mathcal{H}^\tau_\beta(S) \in [0, \infty].$$

We say that S is τ-*null* if $\mathcal{H}^\tau(S) = 0$, τ-*finite* if $\mathcal{H}^\tau(S) < \infty$, and τ-*sigmafinite* if it is a countable union of τ-finite sets. Of course a countable union of τ-null sets is τ-null, and a countable union of τ-sigmafinite sets is τ-sigmafinite.

If $\tau' > \tau$, then $\mathcal{H}^{\tau'}_\beta(S) \le \beta^{\tau'-\tau}\mathcal{H}^\tau_\beta(S)$ for all β, so $\mathcal{H}^\tau(S) < \infty$ implies $\mathcal{H}^{\tau'}(S) = 0$ and $\mathcal{H}^{\tau'}(S) > 0$ implies $\mathcal{H}^\tau(S) = \infty$. The *Hausdorff dimension* of S is the supremum of the set of τ such that $\mathcal{H}^\tau(S) > 0$. This is an extremely interesting concept, which generates a lot of beautiful mathematics. For example, Shishikura [**10**] has shown that the boundary of the Mandelbroit set has Hausdorff dimension 2.

Around the same time he proved the Morse-Sard theorem, Sard [**9**] showed that for subsets of $\mathbb{R}^n$, $\mathcal{H}^n$ is a constant multiple of Lebesgue outer measure. (Some authors divide by this constant in their definition of Hausdorff measure in order to get exact agreement.) In particular, a subset of $\mathbb{R}^n$ that has Hausdorff measure zero necessarily has an empty interior. (Of course this can be proved directly, without referring to Lebesgue measure.)

In addition to the fixed objects U, k, f, and α, throughout we work with a fixed integer r such that $0 \le r < m$. When f is C^1 the r-*critical set* of f is

$$C_{f,r} = \{\, c \in U : \operatorname{rank} Df(c) \le r \,\}.$$

For $s \ge r$ let $d(s) = r + \frac{s-r}{k+\alpha}$. Our main goal is to prove:

Theorem 3. *If f is $C^{k,\alpha+}$, $s > r$, and $S \subset C_{f,r}$ is s-sigmafinite, then $f(S)$ is $d(s)$-null.*

Suppose that $Y = \mathbb{R}^n$, f is C^k, and $k \geq \max\{1, m-n+1\}$. Since it is a subset of E, $C_{f,n-1}$ is m-sigmafinite, and f is $C^{k,0+}$, so if $m > n-1$, then Theorem 3 implies that $\mathcal{H}^{n-1+(m-n+1)/k}(f(C_{f,n-1})) = 0$, and since $k \geq m-n+1$ it follows that $\mathcal{H}^n(f(C_{f,n-1})) = 0$. If $m < n$, then it is easy to show that $\mathcal{H}^n(f(U)) = 0$. (This is Proposition 1 below.) In either case $f(C_{f,n-1})$ has Lebesgue measure zero. Thus Theorem 3 may be regarded as a sharpening of the classic result.

4. STRAIGHTENING This and the following three sections contain the proof of Theorem 3. We begin by showing that a simplifying change of coordinates is possible. Throughout the remainder we work with a fixed $(m-r)$-dimensional linear subspace $Z \subset E$. The next argument applies a consequence of the Hahn-Banach theorem of functional analysis: if X is a finite dimensional linear subspace of Y and $\ell : X \to E$ is a linear function, then there is a continuous linear $L : Y \to E$ such that $L|_X = \ell$.

Lemma 1. *If f is $C^{k,\alpha+}$, $x_0 \in U$, and $\ker Df(x_0) = Z$, then there is an open neighborhood $V \subset U$ of x_0, an open $\tilde{V} \subset E$, and a $C^{k,\alpha+}$ diffeomorphism $g\colon V \to \tilde{V}$ such that $\tilde{f} = f \circ g^{-1}$ is $C^{k,\alpha+}$ and, for each $\tilde{x} \in \tilde{V}$, $\operatorname{rank} D\tilde{f}(\tilde{x}) \leq r$ if and only if $\ker D\tilde{f}(\tilde{x}) = Z$.*

Proof. Let W be a linear subspace of E such that $W \cap Z = \{0\}$ and $W + Z = E$, and let $p_W\colon w + z \mapsto w$ and $p_Z\colon w + z \mapsto z$ be the linear projections. Then $Df(x_0)|_W$ is injective, so there is a continuous linear $L\colon Y \to W$ that extends $(Df(x_0)|_W)^{-1}$. Let $g\colon U \to E$ be the function $g(x) = L(f(x)) + p_Z(x)$. Since $Dg(x_0)|_W$ and $Dg(x_0)|_Z$ are the respective identities, $Dg(x_0)$ is the identity, and the inverse function theorem gives a neighborhood $V \subset U$ of x_0 such that $g|_V$ is invertible with $C^{k,\alpha+}$ inverse. Let $\tilde{V} = g(V)$ and $\tilde{f} = f \circ (g|_V)^{-1}$.

Fix $x \in V$, and let $\tilde{x} = g(x)$, noting that $\tilde{f}(\tilde{x}) = f(x)$. We have

$$p_W(\tilde{x}) = p_W(L(f(x)) + p_Z(x)) = p_W(L(f(x))) = L(f(x)) = L(\tilde{f}(\tilde{x})),$$

so $L \circ \tilde{f} = p_W|_{\tilde{V}}$. Therefore $L \circ D\tilde{f}(\tilde{x}) = D(L \circ \tilde{f})(\tilde{x}) = p_W$, and consequently $\operatorname{rank} D\tilde{f}(\tilde{x}) \geq \operatorname{rank} L \circ D\tilde{f}(\tilde{x}) = r$, $D\tilde{f}(\tilde{x})W \cap \ker L = \{0\}$, and $D\tilde{f}(\tilde{x})Z \subset \ker L$. If $\operatorname{rank} D\tilde{f}(\tilde{x}) = r$, then $D\tilde{f}(\tilde{x})E = D\tilde{f}(\tilde{x})W$, so $D\tilde{f}(\tilde{x})Z \subset D\tilde{f}(\tilde{x})W \cap \ker L = \{0\}$. ■

Suppose X is a metric space, $S \subset X$, and each $x \in S$ has a neighborhood U such that $S \cap U$ is τ-null. If X is separable with countable dense subset D, then each $x \in S$ is contained in a τ-null ball of rational radius centered at a point in D. Since the set of such balls is countable, S is τ-null. The same line of reasoning applies to sets that are τ-sigmafinite, so to prove Theorem 3 it suffices to show that each $x_0 \in S$ has an open neighborhood V and a compact neighborhood $K \subset V$ such that $f(S \cap K)$ is $d(s)$-null or $d(s)$-sigmafinite, as required by the relevant assertion. Suppose f, V, g, $\tilde{V}$, and $\tilde{f}$ are as in Lemma 1, $K \subset V$ is a compact neighborhood of x_0, $S \subset V$, $\tilde{K} = g(K)$, and $\tilde{S} = g(S)$. Since $g|_K$ and its inverse are C^1, they are locally Lipschitz, hence Lipschitz[4], so $S \cap K$ is s-null (sigmafinite) if and only if $\tilde{S} \cap \tilde{K}$ is s-null (sigmafinite).

[4]Suppose X and Y are metric spaces, X is compact, and $f\colon X \to Y$ is locally Lipschitz. If $d(f(x_n), f(x'_n))/d(x_n, x'_n) \to \infty$, then $d(x_n, x'_n) \to 0$ because $\operatorname{diam} f(X) < \infty$, so there is a common limit point x, and f is Lipschitz in some neighborhood of x, contrary to assumption. Thus f is Lipschitz.

Since $d(s)$ is an increasing function of r it suffices to prove Theorem 3 with the condition $S \subset C_{f,r}$ replaced by $S \subset C_{f,r} \setminus C_{f,r-1}$. Since $\tilde{S} \subset \{ \tilde{c} : \ker D\tilde{f}(\tilde{c}) = Z \}$, the upshot of this discussion is that it suffices to prove Theorem 3 with the hypothesis $S \subset C_{f,r}$ replaced by $S \subset \{ c \in U : \ker Df(c) = Z \}$.

Throughout the remainder a $C^{k,\alpha+}$ *manifold* (with one coordinate chart) is a set $N \subset U$ such that, for some integer μ_N, there is a μ_N-dimensional linear subspace $W_N \subset E$, an open $U_N \subset W_N$, and a $C^{k,\alpha+}$ diffeomorphism $\psi_N \colon U_N \to N$. In this circumstance the *tangent space* T_xN of N at $x \in N$ is the image of $D\psi_N(\psi_N^{-1}(x))$. The following definition provides additional objects and restrictions in relation to Z. A tuple $(N, \mu_N, W_N, U_N, \psi_N, Z_N, A_N, \rho_N)$ is a *full featured* $C^{k,\alpha+}$ *manifold* if:

(a) W_N is a μ_N-dimensional linear subspace of E, $U_N \subset W_N$ is open, and $\psi_N \colon U_N \to N \subset U$ is a $C^{k,\alpha+}$ diffeomorphism. (Of course N is a $C^{k,\alpha+}$ manifold.)

(b) $T_xN + Z = E$ for each $x \in N$. (This implies that $\mu \geq r$.)

(c) Z_N is a $(\mu_N - r)$-dimensional linear subspace of W_N such that:
   (i) For all $\xi \in U_N$ and $\nu \in W_N$, $\nu \in Z_N$ if and only if $D\psi_N(\xi)\nu \in Z$.
   (ii) For all $\xi, \xi' \in U_N$, $\xi - \xi' \in Z_N$ if and only if $\psi_N(\xi) - \psi_N(\xi') \in Z$.

(d) $A_N \subset U$ is an open neighborhood of N, and $\rho_N \colon A_N \to N$ is a $C^{k,\alpha+}$ retraction such that $y - \rho_N(y) \in Z$ for all $y \in A_N$.

When speaking of such an object, N may refer either to the entire tuple or to the manifold itself, with the correct interpretation to be inferred from context.

5. MEASUREMENTS This section gives two results that measure the image of f under certain circumstances. The second may be regarded as the driving force of the result. The first is simple, and illustrates the methods that will be applied in the more substantial result that follows.

Proposition 1. *If N is a full featured C^1 manifold, $\mu_N = r$, $f|_N$ is C^1, and $\tau > r$, then $f(N)$ is τ-null.*

Proof. If $r = 0$, then $W_N = \{0\}$, so N and $f(N)$ are singletons. Otherwise U_N can be covered by a countable collection of cubes, and it suffices to show that $f(\psi_N(C))$ is τ-null when $C \subset U_N$ is a compact cube, say of sidelength ℓ. Since C is compact and $f \circ \psi_N$ is C^1, $f \circ \psi_N|_C$ is Lipschitz; let L be a Lipschitz constant. For a positive integer P, C can be covered by P^r subcubes C_i of sidelength ℓ/P. We have $\operatorname{diam} f(\psi_N(C_i)) \leq L \cdot \operatorname{diam} C_i = L\ell\sqrt{r}/P$. Of course $L\ell\sqrt{r}/P \to 0$ as $P \to \infty$, and

$$\mathcal{H}^\tau_{L\ell\sqrt{r}/P}(f(\psi_N(C))) \leq P^r(L\ell\sqrt{r}/P)^\tau = (L\ell\sqrt{r})^\tau P^{r-\tau}. \qquad \blacksquare$$

We introduce a function from [**3**, 3.4.2]. If N is a full featured C^1 manifold, $K \subset N$ is compact, $0 \leq \kappa < \infty$, and $\delta > 0$, let

$$\eta(N, K, f, \kappa, \delta) = \sup_{c \in K,\, x \in N,\, x-c \in Z,\, 0<|x-c| \leq \delta} \frac{|f(x) - f(c)|}{|x - c|^\kappa}.$$

For this definition, if there are no c and x with the required properties we convene that the defined number is zero. Since $\eta(N, K, f, \kappa, \delta)$ is a nondecreasing functions of δ, its limit as $\delta \to 0$ is well defined. We denote this limit by $\eta(N, K, f, \kappa)$.

The next result will be applied with $\kappa = k + \alpha$ and $\omega = \frac{s-r}{k+\alpha}$. The proof below is derived from the proof of [**2**, Lemma 8].

Proposition 2. *Suppose that N is a full featured C^1 manifold, $\mu_N > r$, $K \subset N$ is compact, and $f|_N$ is C^1. If $\kappa \geq 1$ and $\omega \geq 0$ are real numbers such that $r + \omega > 0$ and $\eta(N, K, f, \kappa) = 0$, and $S \subset K$ is $(r + \omega\kappa)$-finite, then $f(S)$ is $(r + \omega)$-null.*

Proof. Let $f' = f \circ \psi_N$, $K' = \psi_N^{-1}(K)$, and $S' = \psi_N^{-1}(S)$. The restriction of ψ_N to a compact neighborhood of K' is Lipschitz, because ψ_N is C^1, so $\eta(U_N, K', f', \kappa) < \infty$. Thus it suffices to prove the assertion with N, K, f, and S replaced by U_N, K', f', and S', so we may assume that N is an open subset of W_N, in which case (c) reduces to $Z_N = W_N \cap Z$.

Let $\tilde{K} \subset U_N$ be a compact neighborhood of K. Since f is C^1, $f|_{\tilde{K}}$ is Lipschitz; let L be a Lipschitz constant. Fix $\varepsilon > 0$ such that $\varepsilon\sqrt{\mu_N - r} < L\sqrt{r}$. Fix $\delta > 0$ small enough that $\tilde{K}$ contains the ball of radius δ around K and $\eta(N, K, f, \kappa, \delta) < T + \varepsilon$. Fix $\theta > 0$ small enough that $\mathcal{H}_\theta^{r+\omega\kappa}(S) < \mathcal{H}^{r+\omega\kappa}(S) + \varepsilon$, $\sqrt{\mu_N}\theta < \delta$, and $\varepsilon(2\sqrt{\mu_N - r})^\kappa \theta^{\kappa-1} < 2L\sqrt{r}$.

Let $\{S_i\}$ be a countable cover of S by sets S_i of diameter $d_i < \theta$ such that $\sum_i d_i^{r+\omega\kappa} < \mathcal{H}^{r+\omega\kappa}(S) + \varepsilon$. For each i let S_i be contained in a cube $C_i = A_i \times B_i$ of sidelength $2d_i$ centered at a point of S_i, where B_i is a cube in Z_N and A_i is a cube in its orthogonal complement. Note that $C_i \subset \tilde{K}$ because $\sqrt{\mu_N} d_i < \delta$. Let $D_i^A = 2L\sqrt{r} d_i$ and $D_i^B = 2\sqrt{\mu_N - r} d_i$. Let $G = 2^{r+\kappa\omega} r^{r/2} (\mu_N - r)^{\kappa\omega/2} L^r$, so that

$$\sum_i (D_i^A)^r (D_i^B)^{\kappa\omega} = G \sum_i d_i^{r+\kappa\omega} < G(\mathcal{H}^{r+\kappa\omega}(S) + \varepsilon).$$

Since

$$D_i^A / \varepsilon (D_i^B)^\kappa \geq 2L\sqrt{r} / \varepsilon (2\sqrt{\mu_N - r})^\kappa \theta^{\kappa-1} > 1$$

there is an integer P_i such that

$$\frac{D_i^A}{\varepsilon (D_i^B)^\kappa} \leq P_i < \frac{2D_i^A}{\varepsilon (D_i^B)^\kappa}.$$

We cover A_i with P_i^r subcubes A_{ij} of sidelength $2d_i/P_i$. We claim that

$$\operatorname{diam} f((A_{ij} \times B_i) \cap S) \leq 2\varepsilon (D_i^B)^\kappa.$$

Suppose that $c, c' \in (A_{ij} \times B_i) \cap S$ where $c = (a, b)$ and $c' = (a', b')$, and let $x = (a, b')$. Then $|f(c') - f(x)| \leq L|a' - a| \leq D_i^A / P_i$. We have $|x - c| = |b' - b| \leq D_i^B$, and $b' - b \in Z$ and $|b - b'| \leq 2\sqrt{\mu_N - r} d_i < \delta$, so

$$|f(x) - f(c)| \leq \varepsilon |x - c|^\kappa \leq (T + \varepsilon)(D_i^B)^\kappa,$$

and thus, as claimed,

$$|f(c') - f(c)| \leq D_i^A / P_i + \varepsilon (D_i^B)^\kappa \leq 2\varepsilon (D_i^B)^\kappa.$$

Let $\beta = 2\varepsilon(2\sqrt{\mu_N - r}\theta)^\kappa$, so that $\operatorname{diam} f((A_{ij} \times B_i) \cap S) < \beta$ for all i and j. Since $P_i < 2D_i^A / (T + \varepsilon)(D_i^B)^\kappa$ we now have

$$\mathcal{H}_\beta^{r+\omega}(f(S)) \leq \sum_{i,j} (\operatorname{diam} f((A_{ij} \times B_i) \cap S))^{r+\omega} \leq \sum_i \left(\frac{2D_i^A}{\varepsilon (D_i^B)^\kappa} \right)^r (2\varepsilon (D_i^B)^\kappa)^{r+\omega}$$

$$= 2^{2r+\omega} \varepsilon^\omega \sum_i (D_i^A)^r (D_i^B)^{\kappa\omega} < 2^{2r+\omega} G \varepsilon^\omega (\mathcal{H}^{r+\kappa\omega}(S) + \varepsilon).$$

Since β and ε may be arbitrarily small, the claim follows from this. ∎

6. ORDERS OF MAGNITUDE The remainder of the proof of Theorem 3 is a matter of finding a countable cover of $C_{f,r}$ by compact sets to which Propositions 1 and 2 can be applied. The hypothesis on $\eta(N, K, f, \kappa)$ resembles the conclusion of Taylor's theorem, and this section's argument obtains this via an induction in which the desired conclusion at one degree of differentiability is integrated, using Lemmas 2 and 3 below, to obtain the desired conclusion at the next degree.

We introduce a second function from [**3**, 3.4.2]. If N is a full featured C^1 manifold, $K \subset N$ is compact, $0 \le \kappa < \infty$, and $\delta > 0$, let

$$\zeta(N, K, f, \kappa, \delta) = \sup_{\substack{c \in K,\ x \in N,\ x-c \in Z,\ 0<|x-c|\le\delta,\\ z \in T_x N \cap Z,\ |z|\le 1}} \frac{|Df(x)z|}{|x-c|^\kappa}.$$

If there are no c, x, and z with the required properties we convene that the defined number is zero. As before, $\zeta(N, K, f, \kappa, \delta)$ is a nondecreasing functions of δ, and we denote its limit as $\delta \to 0$ by $\zeta(N, K, f, \kappa)$.

Lemma 2. *If N is a full featured C^1 manifold, $K \subset N$ is compact, $\kappa \ge 1$, and $\zeta(N, K, f, \kappa - 1) = 0$, then $\eta(N, K, f, \kappa) = 0$.*

Proof. Let $\varepsilon > 0$ be small enough that $\tilde{K}_\varepsilon = \{\, x \in E : \mathrm{dist}(x, K) \le \varepsilon \,\}$ is contained in A_N. Since ρ_N is C^1 and $\tilde{K}_\varepsilon$ is compact, $\rho_N|_{\tilde{K}_\varepsilon}$ is Lipschitz; let L be a Lipschitz constant. We will show that $\eta(N, K, f, \kappa) \le L^\kappa \zeta(N, K, f, \kappa - 1)$.

Suppose that $0 < \delta < \varepsilon$, $c \in K$, and $x \in N$, with $x - c \in Z$ and $0 < |x - c| \le \delta/L$. (If there are no such c and x, then the claim is an automatic consequence of our conventions.) Let $\pi \colon [0, 1] \to N$ be the C^1 path $\pi(t) = \rho_N(c + t(x - c))$. For each t we have

$$\pi(t) - c = \big(\rho_N(c + t(x - c)) - c - t(x - c)\big) + t(x - c) \in Z,$$

so $\pi'(t) \in T_{\pi(t)} N \cap Z$. Evidently $|D\rho_N(x)y| \le L|y|$ for all $x \in \tilde{K}_\varepsilon$ and $y \in E$, so $|\pi'(t)| \le L|x - c|$ and $|\pi(t) - c| \le L|x - c| \le \delta$. Therefore

$$\begin{aligned}\big|Df(\pi(t))\pi'(t)/|\pi'(t)|\big| &\le \zeta(N, K, f, \kappa - 1, \delta)|\pi(t) - c|^{\kappa-1} \\ &\le L^{\kappa-1}\zeta(N, K, f, \kappa - 1, \delta)|x - c|^{\kappa-1}.\end{aligned}$$

We conclude that $\eta(N, K, f, \kappa, \delta/L) \le L^\kappa \zeta(N, K, f, \kappa - 1, \delta)$ because

$$\begin{aligned}|f(x) - f(c)| \le \int_0^1 |(f \circ \pi)'(t)|\, dt &\le \int_0^1 \big|Df(\pi(t))\pi'(t)/|\pi'(t)|\big| \cdot |\pi'(t)|\, dt \\ &\le L^\kappa \zeta(N, K, f, \kappa - 1, \delta)|x - c|^\kappa.\end{aligned}$$ ∎

While one would typically prove Taylor's theorem in a pleasantly structured setting, our lack of prior knowledge concerning the structure of $C_{f,r}$ requires us to work in a general and rather cumbersome structure. A $(k, \alpha+)$-*flattening configuration* is a k-tuple $((N_1, K_1), \ldots, (N_k, K_k))$ such that:

(a) Each N_ℓ is a full featured $C^{\ell,\alpha+}$ manifold, and $N_1 \subset \cdots \subset N_k$.

(b) Each K_ℓ is a compact subset of N_ℓ, and $K_1 \subset \cdots \subset K_k$.

(c) For each $\ell = 2, \dots, k$, if $\phi\colon N_\ell \to Y$ is a $C^{\ell,\alpha+}$ function that vanishes on K_ℓ, then $D\phi(x)z = 0$ for all $x \in K_{\ell-1}$ and $z \in T_x N_\ell \cap Z$.

Fix such a configuration $((N_1, K_1), \dots, (N_k, K_k))$. For each $z \in Z$ let $\varphi_z\colon N_k \to Y$ be the function $\varphi_z(x) = Df(x)(D\rho_{N_k}(x)z) = D(f \circ \rho_{N_k})(x)z$.

Lemma 3. *If $Df(c)z = 0$ for all $c \in K_k$ and $z \in T_c N_k \cap Z$ and $\eta(N_1, K_1, \varphi_z, k - 1 + \alpha) = 0$ for all $z \in Z$, then $\zeta(N_1, K_1, f, k - 1 + \alpha) = 0$.*

Proof. If $c \in K_k$ and $z \in Z$, then $D\rho_N(c)z \in Z$ (because $\mathrm{Id}_{A_N} - \rho_N$ maps into Z) and of course $D\rho_N(c)z \in T_c N_k$, so $\varphi_z(c) = 0$. Let $e_1, \dots, e_{m-r}$ be an orthonormal basis of Z. If $c \in K_1$, $x \in N_1$, $x \neq c$, $x - c \in Z$, $z \in T_x N_1 \cap Z$, and $|z| \leq 1$, then $D\rho_{N_k}(x)z = z$ because $T_x N_1 \subset T_x N_k$ and $\rho_{N_k}|_{N_k}$ is the identity, so

$$|Df(x)z| = |Df(x)(D\rho_{N_k}(x)z)| = |\sum_i \langle z, e_i\rangle \varphi_{e_i}(x)| \leq |z| \sum_i |\varphi_{e_i}(x) - \varphi_{e_i}(c)|.$$

Dividing this by $|x - c|^{k-1+\alpha}$ reveals that

$$\zeta(N_1, K_1, f, k - 1 + \alpha) \leq \sum_i \eta(N_1, K_1, \varphi_{e_i}, k - 1 + \alpha). \qquad \blacksquare$$

Proposition 3. *If $f|_{N_k}$ is $C^{k,\alpha+}$ with $Df(c)z = 0$ for all $c \in K_k$ and $z \in T_c N_k \cap Z$, then $\eta(N_1, K_1, f, k + \alpha) = 0$.*

Proof. In view of Lemma 2 it suffices to prove that $\zeta(N_1, K_1, f, k - 1 + \alpha) = 0$, which we do by induction on k. If $k = 1$ and $K_1' \subset N_1$ is a compact neighborhood of K_1, then, since $D(f \circ \rho_{N_1})$ is $C^{0,\alpha+}$, there is a nondecreasing continuous $\epsilon_{K'} : \mathbb{R} \to \mathbb{R}$ with $\epsilon_{K'}(0) = 0$ such that

$$|D(f \circ \rho_{N_1})(x) - D(f \circ \rho_{N_1})(c)| \leq \epsilon_{K'}(|x - c|)|x - c|^\alpha$$

for all $c \in K_1$ and $x \in K_1'$. If $c \in K_1$, $x \in K_1'$, $x - c \in Z$, and $z \in T_x N_1 \cap Z$ with $|z| \leq 1$, then, as above, $D\rho_{N_1}(x)z = z$ and $D\rho_{N_1}(c)z \in T_c N_1 \cap Z$, so

$$\begin{aligned} |Df(x)z| &= |Df(x)(D\rho_{N_1}(x)z) - Df(c)(D\rho_{N_1}(c)z)| \\ &\leq |D(f \circ \rho_{N_1})(x) - D(f \circ \rho_{N_1})(c)|. \end{aligned}$$

Suppose that $k \geq 2$ and we have proved the desired inequality with k replaced by $k - 1$. As a composition of $C^{k-1,\alpha+}$ functions, φ_z is $C^{k-1,\alpha+}$. Since $\varphi_z(c) = 0$ for all $c \in K_k$, $D\varphi_z(c)z = 0$ for all $c \in K_{k-1}$ and $z \in T_c N_k \cap Z$. The induction hypothesis (with φ_z in place of f) gives $\zeta(N_1, K_1, \varphi_z, k - 2 + \alpha) = 0$. Now Lemma 2 yields $\eta(N_1, K_1, \varphi_z, k - 1 + \alpha) = 0$, after which Lemma 3 gives the desired result. $\blacksquare$

7. MORSE DECOMPOSITION Let M be a full featured $C^{k,\alpha+}$ manifold, and let $C \subset M$ be closed. A $C^{k,\alpha+}$ *Morse decomposition* of C is a countable collection $\{((N_1^j, K_1^j), \dots, (N_k^j, K_k^j))\}$ of $(k, \alpha+)$-flattening configurations with $N_k^j \subset M$ for all j and $\bigcup_j K_1^j = C$. Let Q be the set of $x \in C$ such that $D\phi(x)z = 0$ for all $z \in T_x M \cap Z$ whenever $\phi\colon M \to Y$ is a $C^{k,\alpha+}$ function that vanishes on C. If $x \in C \setminus Q$, $\phi\colon M \to Y$ is a $C^{k,\alpha+}$ function that vanishes on C, $z \in T_x M \cap Z$, and $D\phi(x)z \neq 0$, then a neighborhood of x in C can be “peeled off” into a $(\mu_M - 1)$-dimensional submanifold of M be applying the implicit function theorem. This will allow a Morse decomposition to be constructed by a double induction on μ_M and k.

First we need to show that we can require the result of the application of the implicit function theorem to be full featured. One of these features can be fabricated from scratch.

Lemma 4. *If N is a $C^{k,\alpha+}$ manifold, $x_0 \in N$, and $T_{x_0}N + Z = E$, then there is an open neighborhood $N' \subset N$ of x_0 for which there is an open neighborhood $A \subset E$ of N' and a $C^{k,\alpha+}$ retraction $\rho\colon A \to N'$ such that $y - \rho(y) \in Z$ for all $y \in A$.*

Proof. Let Z' be an $(m - \mu_N)$-dimensional linear subspace of Z such that $T_{x_0}N + Z' = E$. The function $(x', z') \mapsto x' + z'$ from $N \times Z'$ to E is $C^{k,\alpha+}$, and its derivative at $(x_0, 0)$ is nonsingular, so the implicit function theorem gives a neighborhood B of $(x_0, 0)$ and a $C^{k,\alpha+}$ diffeomorphism between B and a neighborhood $A \subset E$ of x_0. Let the inverse of this diffeomorphism be $(\rho, \sigma)\colon A \to N \times Z'$, and let $N' = \{\, x \in N : (x, 0) \in B \,\}$. We now replace A with $\rho^{-1}(N')$ and ρ with its restriction to this set. ■

Lemma 5. *Let N be a full featured $C^{k,\alpha+}$ manifold. If $\phi\colon N \to Y$ is $C^{k,\alpha+}$, $x_0 \in \phi^{-1}(0)$, $z \in T_{x_0}N \cap Z$, and $D\phi(x_0)z \neq 0$, then there is a full featured $(\mu - 1)$-dimensional $C^{k,\alpha+}$ manifold N' that contains a neighborhood of x_0 in $\phi^{-1}(0)$.*

Proof. The consequence of the Hahn-Banach theorem used in the proof of Lemma 1 implies that there is a continuous linear $L\colon Y \to \mathbb{R}$ with $L(D\phi(x_0)z) \neq 0$. Let $\phi' = L \circ \phi$. Since $D\phi'(x_0)z = (L \circ D\phi(x_0))z \neq 0$ and $\phi^{-1}(0) \subset {\phi'}^{-1}(0)$, it suffices to prove the claim with ϕ replaced by ϕ', so we may assume that $Y = \mathbb{R}$.

Let $\mu_{N'} = \mu_N - 1$. Let $\xi_0 = \psi_N^{-1}(x_0)$ and $\zeta = D(\psi_N^{-1})(x_0)z$, and let $W_{N'}$ be a $\mu_{N'}$-dimensional linear subspace of W_N that does not contain ζ. Let $\tilde{\phi} = \phi \circ \psi_N\colon U_N \to \mathbb{R}$. Since $D\tilde{\phi}(x_0)\zeta = D\phi(x_0)z \neq 0$, the $C^{k,\alpha+}$ implicit function theorem gives an open $U_{N'} \subset W_{N'}$ and a $C^{k,\alpha+}$ function $g\colon U_{N'} \to \mathbb{R}$ such that $\{\, \xi + g(\xi)\zeta : \xi \in U_{N'} \,\}$ is a neighborhood of ξ_0 in $\tilde{\phi}^{-1}(0)$. Let $\psi_{N'}\colon U_{N'} \to N$ be the function

$$\psi_{N'}(\xi) = \psi_N(\xi + g(\xi)\zeta).$$

Let $N' = \psi_{N'}(U_{N'})$; of course N' is a neighborhood of x_0 in $\phi^{-1}(0)$. By construction $T_{x_0}N'$ and z span $T_{x_0}N$, and $T_{x_0}N$ and Z span E, so $T_{x_0}N'$ and Z span E. By continuity, if $U_{N'}$ is a small enough neighborhood of ξ_0, then $T_{x'}N'$ and Z span E for all $x' \in N'$. Thus N', $W_{N'}$, $U_{N'}$, and $\psi_{N'}$ satisfy (a) and (b).

Let $Z_{N'} = W_{N'} \cap Z_N$. By (i), $\zeta \in Z_N$, so $Z_{N'}$ has codimension 1 in Z_N. For $\xi \in U_{N'}$ and $\nu \in W_{N'}$ we are given that $D\psi_N(\xi + g(\xi)\zeta)\nu \in Z$ if and only if $\nu \in Z_N$, and $D\psi_{N'}(\xi)\nu = D\psi_N(\xi + g(\xi)\zeta)\nu + (Dg(\xi)\nu)z$, so $D\psi_{N'}(\xi)\nu \in Z$ if and only if $\nu \in Z_{N'}$. For $\xi, \xi' \in U_{N'}$ we are given that $\psi_{N'}(\xi) - \psi_{N'}(\xi') \in Z$ if and only if $(\xi + g(\xi)\zeta) - (\xi' + g(\xi')\zeta) \in Z_N$, and this is the case if and only if $\xi - \xi' \in Z_{N'}$.

Now Lemma 4 allows us to replace N' (with related adjustments to $U_{N'}$ and $\psi_{N'}$) with an open neighborhood of x_0 for which there is an open neighborhood $A_{N'} \subset E$ of N' and a $C^{k,\alpha+}$ retraction $\rho_{N'}\colon A_{N'} \to N'$ such that $y - \rho(y) \in Z$ for all $y \in A_{N'}$. ■

Proposition 4. *There is $C^{k,\alpha+}$ Morse decomposition of C.*

Proof. A collection of Morse decompositions of the elements of a countable cover of C by compact sets can be united to give a Morse decomposition of C, so we may assume that C itself is compact. Note that (c) of the definition of a $(k, \alpha+)$-flattening configuation holds vacuously when $k = 1$, and it also holds when N_ℓ is r-dimensional because $T_xN_\ell \cap Z = \{0\}$ for all $x \in N_\ell$. When $k = 1$, $\{((M, C))\}$ is a Morse decomposition of C. If $\mu_M = r$, then $Z_M = \{0\}$, so a suitable collection is given by $\{((M, K_1), \ldots, (M, K_k))\}$ where $K_1 = \cdots = K_k = C$. By induction it suffices to show that the claim holds for given $\mu_M > r$ and $k > 1$ if it has already been established with (μ_M, k) replaced by either $(\mu_M - 1, k)$ or $(\mu_M, k - 1)$.

Let Q be as above. If $x \in C \setminus Q$, $\phi\colon M \to Y$ is a $C^{k,\alpha+}$ function that vanishes on C, $z \in T_xM \cap Z$, and $D\phi(x)z \neq 0$, then Lemma 5 gives a full featured $(\mu_M - 1)$-dimensional $C^{k,\alpha+}$ manifold containing a neighborhood of x in $\phi^{-1}(0) \supset C$. Since $C \setminus Q$ is separable, there is a countable cover $\{N^i\}$ of $C \setminus Q$ by such manifolds. By induction on μ_M, for each i there a countable collection $\{((N_1^{i\ell}, K_1^{i\ell}), \ldots, (N_k^{i\ell}, K_k^{i\ell}))\}$ of $(k, \alpha+)$-flattening configurations with $\bigcup_\ell K_1^{i\ell} = N^i \cap C$.

Evidently Q is closed, hence compact. By induction on k there is a countable collection $\{((N_1^h, K_1^h), \ldots, (N_{k-1}^h, K_{k-1}^h))\}$ of $(k - 1, \alpha+)$-flattening configurations such that $N_{k-1}^h \subset M$ for all h and $\bigcup_h K_1^h = Q$. Each

$$((N_1^h, K_1^h), \ldots, (N_{k-1}^h, K_{k-1}^h), (M, C))$$

is a $(k, \alpha+)$-flattening configuration, so we have the satisfactory countable collection

$$\{((N_1^{i\ell}, K_1^{i\ell}), \ldots, (N_k^{i\ell}, K_k^{i\ell}))\} \cup \{((N_1^h, K_1^h), \ldots, (N_{k-1}^h, K_{k-1}^h), (M, C))\}. \quad \blacksquare$$

Proof of Theorem 3. As we explained in Section 4, we may assume that $C_{f,r} = \{\, c \in U : \ker Df(c) = Z \,\}$. Note that U itself is a full featured m-dimensional $C^{k,\alpha+}$ manifold if we take $\mu_N = m$, $W_U = E$, $U_U = U$, ψ_U the identity, $A_U = U$, ρ_U the identity, and $Z_U = Z$. Therefore Proposition 4 gives a $C^{k,\alpha+}$ Morse decomposition $\{((N_1^j, K_1^j), \ldots, (N_k^j, K_k^j))\}$ of $C_{f,r}$, for each j Proposition 3 gives $\eta(N_1^j, K_1^j, f, k + \alpha) = 0$, and according to whether $\mu_{N_1^j} = r$ or $\mu_{N_1^j} > r$, Proposition 1 or Proposition 2 implies that $f(S \cap K_1^j)$ is $d(s)$-null. ■

8. WHAT ELSE? If the reader continues on to [??], the first thing she will find is a much more comprehensive discussion of the history of the Morse-Sard theorem and related literature. The function f is $C^{k,\alpha}$ if it is C^k and, for each compact $K \subset U$, there is an $M_K > 0$ such that $|D^k f(x) - D^k f(y)| \le M_K |x - y|^\alpha$ for all $x, y \in K$. For this slightly weaker condition there is a slightly weaker result: if f is $C^{k,\alpha}$, $s \le r$, and $S \subset C_{f,r}$ is s-sigmafinite, then $f(S)$ is $d(s)$-sigmafinite. The proof parallels the argument here. There is also a special case: if f is $C^{k,\alpha}$, then $f(C_{f,r})$ itself is $d(m)$-null. The proof of this follows the methods used to prove Proposition 2, but is more sophisticated and applies advanced results from measure theory. Finally, there are brief descriptions of the beautiful examples constructed by Whitney, Dubovitskiĭ, and Federer to show that the degree-of-differentiability assumptions cannot be relaxed.

ACKNOWLEDGMENT. I am grateful for helpful correspondence with Mikhail Korobkov, Frank Morgan, and Ravi Vakil.

REFERENCES

1. G. Debreu. Economies with a finite set of equilibria. *Econometrica*, 38:387–392, 1970.
2. A. Ya. Dubovitskiĭ. On the set of degenerate points. *Izv. Akad. Nauk SSSR Ser. Mat.*, 31:27–36, 1967.
3. H. Federer. *Geometric Measure Theory*. Springer, New York, 1969.

4. J. C. Harsanyi. Oddness of the number of equilibrium points: A new proof. *Int. J. Game Theory*, 2:235–250, 1973.
5. M.W. Hirsch. *Differential Topology*. Springer-Verlag, New York, 1976.
6. J. Milnor. *Topology from the Differentiable Viewpoint*. University Press of Virginia, Charlottesville, 1965.
7. A.P. Morse. The behavior of a function on its critical set. *Ann. of Math.*, 40:62–70, 1939.
8. A. Sard. The measure of the critical values of differentiable maps. *Bull. Amer. Math. Soc.*, 48:883–897, 1942.
9. A. Sard. The equivalence of n-measure and Lebesgue in E_n. *Bull. Amer. Math. Soc.*, 49:758–759, 1943.
10. M. Shishikura. The boundary of the Mandelbroit set has Hausdorff dimension two. *Astérisque*, 7(222):389–405, 1994.

ANDREW MCLENNAN received his B.A. in mathematics from the University of Chicago, but then detoured, receiving a Ph.D. in economics from Princeton University. After stints at the University of Toronto and Cornell University, he endured 18 bitter winters at the University of Minnesota before moving to Australia, briefly at the University of Sydney, and then for the last decade at the University of Queensland. In addition to teaching and research in economic theory, and sometimes mathematics, he enjoys life with his partner Shino and son Sean, and noodling on the piano.
School of Economics, University of Queensland, St Lucia QLD 4067
a.mclennan@economics.uq.edu.au