Utrecht University

Faculty of science

Sard's theorem

Bachelor Thesis

Stijn van Opstal

Supervisors:

Dr. Guy Boyde
Mathematical Institute

Dr. Lennart Meier
Mathematical Institute

24 January 2025

Contents

1 Introduction

The goal of this thesis is to give the reader an exposition of the proof of Sard's theorem and an application of the theorem, the Whitney embedding theorem. The intended audience consists of bachelor students in mathematics roughly half-way through their bachelor degree at university level. We will assume basic knowledge of calculus, linear algebra and general topology and will briefly recall specifics when needed. We will give more detailed attention to more advanced topics, most notably the theory of differentiable manifolds, that are needed to understand Sard's theorem in the preliminaries.

We will start by illustrating Sard's theorem with an example consisting of a function $f : \mathbb{R} \to \mathbb{R}$. Then we will introduce some preliminaries needed to understand Sard's theorem, after which we will state and prove the full theorem. Then we will state some preliminaries for the Whitney embedding theorem and then state and prove this theorem.

2 Sard's theorem for real valued functions of a real variable

Sard's theorem is concerned with mathematical structures called 'manifolds'. To fully understand the theorem it is thus necessary to explore the theory behind these manifolds. But we can already start with one of the most basic forms of manifolds and functions between them, namely real valued functions of a real variable. We can state Sard's theorem for these functions without worrying about its generalisation to abstract manifolds yet.

When looking at real valued functions of a real variable, it is not difficult to produce a function whose derivative is equal to a constant on an interval of any desired length: just take the linear function with the desired slope. Likewise, if one desires a function that has constant derivative on an interval in its codomain the linear function with desired slope will also suffice, with the notable exception of a function whose derivative is zero on any given interval in its codomain. Producing a function that has zero derivative on an interval in its codomain is actually quite difficult, because in contrast to values where the derivative is non-zero, at values where the derivative is zero the function tends to 'stay the same' exactly because its derivative is zero.

Let us for the moment call the points in the domain of a function where its derivative is zero 'critical points' and the values in the codomain whose pre-image contains a critical point 'critical values'; we will give more general definitions that applies to functions between manifolds later on. If one starts looking for a function that has a critical value on each point in an interval in its codomain, one might start by considering one that oscillates in the interval and admits a critical value on isolated points in the interval. A simple example could be the function $\mathbb{R}_{\geq 1} \to [-1.1], x \mapsto \frac{\sin(\pi x)}{\lfloor x \rfloor}$, whose critical points lie at $x = n + \frac{1}{2}$ for each $n \in \mathbb{N}$ and whose value at those critical points is $\frac{1}{n}$. This gives us a function that has critical values at isolated points in the interval $[-1, 1]$, but this is still far from filling up the entire interval. Can we do better than this? Let us first note that in the given example all the critical points corresponding to the critical values are isolated in the domain, i.e. for each critical value $f(p)$ there is an interval $I_p = (p - \delta_p, p + \delta_p)$ with $\delta_p > 0$ such that the only critical point in I_p is p itself. The following argument shows that a function that has only this type of critical values will not suffice to produce a function whose set of critical values contains an entire interval: since every non-empty interval I_p contains a rational number and does not contain another critical point apart from p , we can map the set of critical points one-on-one to a subset of the rational numbers by mapping each critical point to the corresponding rational number in I_p. Since the set of rational numbers is countable, the set of critical points must also be countable and thus also the set of critical values. But if the set of critical values is countable, it can never be equal to an interval.

This does not fully answer the question of how large the set of critical values can be, as the set of critical points and critical values might not consist of just isolated points. Here Sard's theorem can be used to give limit the size of the set of critical values: it says that the measure of the set of critical values has to be zero: this in particular means that the set of critical values cannot contain an interval, since an interval has a measure greater than zero.

3 Preliminaries

3.1 Smooth manifolds and smooth functions on smooth manifolds

Sard's theorem is concerned with mathematical structures called 'smooth manifolds'. These smooth manifolds have the desired property that it is possible to perform calculations on them as one is familiar with doing on n-dimensional Euclidean space, such as in multivariable calculus. The object of study in multivariable calculus are functions from one m-dimensional space into another n-dimensional space and this study can be generalised to manifolds because, as we will see, manifolds locally look like n-dimensional Euclidean space.

We will start with the definition of an abstract topological manifold. We depart from the notion of a topological space, which allows us to talk about 'opens' (in particular, we view Euclidean space as a topological space with the associated standard topology) and we will remind the reader that a homeomorphism between two topological spaces is a bijection between these spaces that is continuous and whose inverse is also continuous and that its importance lies in the fact that two homeomorphic topological spaces are intrinsically 'the same' (from a topological viewpoint).

Definition 3.1. *A topological space M is called an m-dimensional (topological) manifold if there exists a countable number of opens $U_i \subseteq M$, $i \in I$ with I being countable, that cover M and are homeomorphic to an open subset of m-dimensional Euclidean space.*

We will call the opens that cover M together with the associated homeomorphisms $\phi_i : U_i \to \mathbb{R}^m$ 'coordinate charts' because they give us coordinates to work with on our abstract manifold M. We can roughly compare this to everyday topographical charts covering the earth in coordinates that allow us to measure distances etc. Some prototypical simple examples of topological manifolds are the circle (a one-dimensional manifold), the sphere (two-dimensional) or any n-sphere (n-dimensional) or the torus (two-dimensional). It is also important to note that Euclidean space itself and all of its open subsets are also manifolds (as they trivially meet the definition of a manifold)! For this last class of examples it is trivial to give concrete charts: any open subset U of n-dimensional Euclidean space is an n-dimensional topological manifold because the identity chart $Id : U \to \mathbb{R}^n : x \mapsto x$ covers all of U and is a homeomorphism to an open subset of $\mathbb{R}^n$ (namely U). For the circle, there are multiple collections of charts that can be constructed: via stereographic projection one can obtain two charts $S_n : S^1\backslash(0,1) \to \mathbb{R}$ and $S_s : S^1\backslash(0,-1) \to \mathbb{R}$ that together cover S^1 and are homeomorphisms between the circle and $\mathbb{R}$. Another possible choice would be projections from the upper and lower half-open circle onto the x-axis and left and right half-open circle onto the y-axis. Note that it is not possible to construct just one coordinate chart that covers all of S^1 as that would imply that S^1 is homeomorphic to an open subset of $\mathbb{R}$, which it is not.

Now to actually be able to perform calculus on a manifold, just a set of coordinate charts is not enough, because it is possible to construct a topological manifold with multiple charts for which there are functions that are smooth in one chart, but not in the other! We would like for all the charts to agree on whether a function on a manifold is smooth or not, and for this we impose an extra condition. For any two charts $\phi_1 : U_1 \to \mathbb{R}^n$ and $\phi_2 : U_2 \to \mathbb{R}^n$ for which U_1 and U_2 overlap it is possible to construct the change-of-coordinates map $c_{\phi_1}^{\phi_2} : \phi_1(U_1 \cap U_2) \subseteq \mathbb{R}^n \to \phi_2(U_1 \cap U_2) \subseteq \mathbb{R}^n : x \mapsto \phi_2(\phi_1^{-1}(x))$. Since this is a map between subsets of Euclidean space, we can impose the condition that this change-of-coordinates map is smooth, which we will call 'smooth compatibility'. If, for a collection of charts, all the change-of-coordinate maps are smooth, we call the collection of these charts a smooth atlas. The two different examples of collections of charts on the circle S^1 both give us a smooth atlas. In fact, all the charts from one atlas actually are smoothly compatible with all the charts of the other atlas, so we could join the to atlases to create a new smooth atlas. In fact, to formalize things we would actually like to work with atlases that contain all smooth charts that are constructible and are smoothly compatible with each other. These are called *maximal smooth atlases* or *smooth structures*. Fortunately, it turns out that for most practical calculations it is actually enough to check smoothness for a (non-maximal) smooth atlas contained in a maximal smooth atlas. A topological manifold together with a smooth structure is called a 'smooth manifold'.

We are now ready to introduce smooth functions between smooth manifolds, which is one of the main subjects of Sard's theorem. In general, if we consider functions from one topological manifold to another without considering smooth atlases, we can't say anything about smoothness of such a function because the (abstract) manifolds don't come with any structure of themselves that allows us to talk about concepts such as differentiability. With smooth atlases however, we can, and this is one of the main reasons for introducing smooth atlases. This leads us to the following definitions:

Definition 3.2. *Let $F : M \to N$ be a function from an m-dimensional smooth manifold M to an n-dimensional smooth manifold N. The representation of F with respect to two charts $\phi_1 : M \to \mathbb{R}^m$ and $\phi_2 : N \to \mathbb{R}^n$ from the smooth structure of M, respectively N, is defined as $F_{\phi_2}^{\phi_1} := \phi_2 \circ F \circ \phi_1^{-1}$.*

So the smooth charts in the smooth structures of M and N allow us to handle the function F as a function from a subset of $\mathbb{R}^m$ to a subset of $\mathbb{R}^n$. And for such functions we know what it means to be 'smooth'! This leads to the following definition:

Definition 3.3 (Smooth functions between manifolds)**.** *A function $F : M \to N$ from an m-dimensional smooth manifold M to an n-dimensional smooth manifold N is called 'smooth'* *if its representation with respect to each chart from the smooth structure of M and each chart of the smooth structure of N is smooth in the 'ordinary sense' (as a function of $\mathbb{R}^m$ to $\mathbb{R}^n$).*

So to check smoothness of a function $F : M \to N$ between smooth manifolds M and N, we need to check whether the representations of F with respect to all possible charts of the smooth structure of M and N are smooth (in the ordinary sense, as functions from $\mathbb{R}^m$ to $\mathbb{R}^n$). Fortunately, as we have remarked earlier, it is possible to show that actually just checking it for any (possibly non-maximal) smooth atlas contained in the smooth structures of M and N is sufficient.

3.2 Tangent spaces and differentials

When dealing with smooth (in the ordinary sense) functions from $\mathbb{R}^m$ to $\mathbb{R}^n$, we can use the total derivative of that function at a point to construct the plane that is tangent to the function at that point. When looking at manifolds that are embedded in say one or two dimensional Euclidean space, we can geometrically perform a similar construction, i.e. try to construct the plane that is tangent to the manifold at a certain point. For a manifold such as a circle this would amount to constructing the line that is tangent to the circle at any arbitrary point. In this section we will introduce 'tangent spaces', which are a generalisation of this idea. In particular, tangent spaces can be constructed for all smooth manifolds, even those that are not embedded in Euclidean space. When a manifold is smoothly embedded in an Euclidean space, the tangent space coincides with the usual, intuitive notion of a plane tangent to the manifold in the space it is embedded in.

We will start with the following definition:

Definition 3.4 (smooth curve)**.** *Let $\gamma : I \to M$ be a function from an open interval $I \subseteq \mathbb{R}$ endowed with the standard smooth structure to an m-dimensional smooth manifold M endowed with a smooth structure consisting of the charts $\{(U_i, \phi_i)\}$. We call γ a* smooth curve *when its representation γ_{ϕ_i} with respect to the identity chart on I and the chart ϕ_i on M is smooth as a function from $\mathbb{R}$ to $\mathbb{R}^m$ for each chart ϕ_i in the smooth structure of M.*

We will now specifically consider smooth curves that pass through a point $p \in M$, and for which $0 \in I$ and such that $\gamma(0) = p$ (this is mainly to simplify notation). For every chart ϕ_i we can take the derivative of $\gamma_{phi_i}(t)$ at $t = 0$, since it is simply a function of (a subset of) $\mathbb{R}$ to $\mathbb{R}^m$. The exact value of this derivative does however depend on the chart ϕ_i, so we cannot in general talk about the derivative of $\gamma(t)$ at $t = 0$. But what we cán say is that when two smooth curves γ_1 and γ_2 have equal derivative at $t = 0$ in one chart ϕ_1, i.e. $\frac{d\gamma_1^{\phi_1}}{dt}(0) = \frac{d\gamma_2^{\phi_1}}{dt}(0)$, they have equal derivative at $t = 0$ in any chart. This is due to the chain rule:

$$\begin{aligned} \frac{d\gamma_1^{\phi_i}}{t}(0) &= \frac{d}{dt}(c_{\phi_1}^{\phi_i} \circ \gamma_1^{\phi_1}(t))\Big|_{t=0} = (D_{\gamma_1^{\phi_1}(0)} c_{\phi_1}^{\phi_i}) \frac{d\gamma_1^{\phi_1}}{dt}(0) = \\ &(D_{\gamma_2^{\phi_1}(0)} c_{\phi_1}^{\phi_i}) \frac{d\gamma_2^{\phi_1}}{dt}(0) = \frac{d}{dt}(c_{\phi_1}^{\phi_i} \circ \gamma_2^{\phi_1}(t))\Big|_{t=0} = \frac{d\gamma_2^{\phi_i}}{t}(0). \end{aligned}$$

This means we can form an equivalence class of smooth curves that have the same derivative at $t = 0$, since the relation of having the same derivative at $t = 0$ is an equivalence relation. We will denote the equivalence class that any particular curve γ belongs to as $[\gamma]$, which we will call a *tangent vector*. As we mentioned, the value of the derivative of $\gamma_{phi_i}(t)$ at $t = 0$ depends on the chart ϕ; this means we can associate each tangent vector for each chart with a concrete vector in $\mathbb{R}^m$, namely via

$$[\gamma]^\phi := \frac{d\gamma^\phi}{dt}(0) \in \mathbb{R}^m$$

We call this the representation of the tangent vector with respect to the chart ϕ. As we just showed, these representations uniquely determine a tangent vector: whenever the representation of two tangent vectors $[\gamma_1]$ and $[\gamma_2]$ are equal for one chart, i.e. $[\gamma_1]^\phi = [\gamma_2]^\phi$, then they are equal for all charts and thus the tangent vectors themselves are equal, i.e. $[\gamma_1] = [\gamma_2]$. This also means we can define the sum of two tangent vectors and multiplication with a scalar of a tangent vector via their representations: for any chart we define $([\gamma_1]+[\gamma_2])^\phi = [\gamma_1]^\phi + [\gamma_2]^\phi$ and $(\lambda[\gamma_1])^\phi = \lambda[\gamma_1]^\phi$ for $\lambda \in \mathbb{R}$. This means the set of tangent vectors is a vector space, explaining the use of the term 'tangent vector'. We call the set of tangent vectors in a point $p \in M$ the *tangent space*, denoted by T_pM.

Now we can come to an (arguably quite natural) definition of what we call the differential of a smooth map $F : M \to N$ between smooth manifolds M and N:

Definition 3.5 (differential)**.** *The differential of a smooth map $F : M \to N$ between smooth manifolds M and N at a point $p \in M$, is a map $(dF)_p : T_pM \to T_{F(p)}N$ defined by $(dF)_p([\gamma]) := [F \circ \gamma]$.*

For this definition to make sense we do have to ensure that when $[\gamma] \in T_pM$ we have that $[F \circ \gamma] \in T_{F(p)}N$. First of all, $F \circ \gamma$ is a function from I to N. We then have:

$$(F \circ \gamma)^\chi = \chi \circ F \circ \gamma = \chi \circ F \circ \phi^{-1} \circ \phi \circ \gamma = F^\chi_\phi \circ \gamma^\phi,$$

so $F \circ \gamma$ is smooth with respect to an arbitrary chart χ. This means that $F \circ \gamma$ is a smooth curve. Finally, if $[\gamma_1] = [\gamma_2]$ then $[F \circ \gamma_1]^\chi = \frac{d(F\circ\gamma_1)^\chi}{dt}(0) = \frac{d(F^\chi_\phi \circ \gamma^\phi)}{dt}(0) = \frac{dF^\chi_\phi}{dt}(\gamma_1(0))[\gamma_1]^\phi = \frac{dF^\chi_\phi}{dt}(\gamma_2(0))[\gamma_2]^\phi = [F \circ \gamma_2]^\chi$, and so $(dF)_p$ is well defined.

3.3 Submersions and immersions

Sard's theorem concerns the size of the critical values of a function between manifolds. To introduce these, we start with the notion of a submersion for a smooth function between smooth manifolds (which is an adaptation of the regular definition for smooth functions between Euclidean spaces):

Definition 3.6. *A smooth function $F : M \to N$ between smooth manifolds M and N is a submersion in a point $p \in M$ if its differential at that point, $(dF)_p$, is surjective (as a linear map). We call the function itself a submersion, if it is a submersion at every $p \in M$.*

If $F : M \to N$ is a smooth function between manifolds, a point $p \in M$ is called a *regular point* of F if F is a submersion at p. If p is not a regular point, we call it a *critical point*. A *regular value* of F is a point $q \in N$ such that all points of the set $F^{-1}(q)$ are regular points. If a point $q \in N$ is not a regular value, we call it a *critical value*.

The importance of regular values lies in the following theorem:

Theorem 3.7. *Let $F : M \to N$ be a smooth function from a smooth manifold M of dimension m to a smooth manifold N of dimension n. If $q \in N$ is a regular value of F, then $F^{-1}(q)$ is a submanifold of M of dimension $m - n$.*

As with submersions, there is also a direct adaption of the notion of an immersion to smooth function between manifolds:

Definition 3.8. *A smooth function $F : M \to N$ between smooth manifolds M and N is a immersion in a point $p \in M$ if its differential at that point, $(dF)_p$, is injective (as a linear map). We call the function itself an immersion, if it is an immersion at every $p \in M$.*

3.4 Sets of measure zero

Sard's theorem involves the size of a specific subset of a smooth manifold, or more specifically its *measure*. Measure theory is a branch of mathematics that, among others, concerns itself with attributing a notion of 'size' to subsets of $\mathbb{R}^n$, called a measure of that set. For the more ordinary (geometric) subsets of $\mathbb{R}^n$ such as a square in $\mathbb{R}^2$ or a line in $\mathbb{R}$, it coincides with the intuitive idea of size of such an object, i.e. the surface area of the square or the length of the line. One may realise that this means that it is not possible to translate the notion of a measure to manifolds, as sets that have different measures, such as the intervals $[0, 1]$ and $[0, 2]$, are actually homeomorphic. So the measure of a set is not preserved by homeomorphisms. There is one exception to this however, namely for sets of measure zero. Since we are thus only concerned about sets of measure zero, we will give a definition of what it means to have measure zero without relying on concepts that belong to measure theory such as e.g. sigma-algebras and (Lebesgue) measures of arbitrary subsets of $\mathbb{R}^n$.

Definition 3.9. *A subset $A \subseteq \mathbb{R}^n$ is said to have n-dimensional measure zero if for every $\delta > 0$ it can be covered by a countable number of n-dimensional open balls such that the added volume of each ball is less then δ.*

Note that in the previous definition, we could have also chosen to cover A with open cubes instead of open balls. Now in the case of smooth manifolds we are working with smooth charts, so a natural way to extend the concept of measure zero is as follows:

Definition 3.10. *A subset $A \subseteq M$ of a smooth manifold is said to have measure zero if for every smooth chart (U, ϕ) the subset $\phi(A \cap U) \subseteq \mathbb{R}^n$ has n-dimensional measure zero.*

For the proof of Sard, we will also be using the following lemma as stated and proven by Lee [1]:

Lemma 3.11. *Let $A \subseteq \mathbb{R}^n$ be a compact subset such that the intersection of A with each hyperplane $c \times \mathbb{R}^{n-1}$ has $(n-1)$-dimensional measure zero. Then A has n-dimensional zero.*

3.5 Taylor polynomials

We will assume the reader is familiar with Taylor polynomials of real-valued functions of one real variable, but to reiterate, suitable functions $f : \mathbb{R} \to \mathbb{R}$ can be approximated at a point $c \in \mathbb{R}$ in their domain by a so-called Taylor polynomial of order n. This means that, in a neighbourhood of c, the function is equal to a polynomial of order n plus a 'remainder' or 'error' term that approaches zero as one gets closer to c:

$$f(x+c) = \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x+c)^k + R_c^n(x)$$

For functions $F : \mathbb{R}^m \to \mathbb{R}^n$ we can define the k-th Taylor polynomial as follows:

$$P_k(x) = F(a_i) + \sum_{j=1}^{k} \frac{1}{j!} \sum_{I:|I|=j} \partial_I F(a_i)(x-a_i)^I = F(a_i).$$

It is then possible to show that $F(x) - P_k(x) = R_k(x)$, where $R_k(x)$ is a remainder term defined by:

$$R_k(x) = \frac{1}{k!} \sum_{I:|I|=k+1} (x-a_i)^I \int_0^1 (1-t)^k \partial_I F(a_i + t(x-a_i))dt.$$

For details we refer to [2].

4 Sard's theorem

We are now ready to present Sard's theorem and its proof. The proof is modelled on the one given in the book of Lee [1].

Theorem 4.1 (Sard)**.** *Let $F : M \to N$ be a smooth function from a smooth manifold M of dimension m to a smooth manifold N of dimension n. Let C be the set of critical points in M of F. Then the set of critical values of F in N, $F(C)$, has n-dimensional measure zero.*

Proof. The proof will proceed by induction over the dimension of M. Throughout the proof we will be considering the representation of F with respect to a chart from the smooth structure of M, denoted $U \subseteq M$, and the smooth structure of N. As mentioned in the preliminaries, a point $c \in M$ is a critical point of F if and only if the representation of F with respect to any chart of the smooth structure of M containing c and any chart of the smooth structure of N containing $F(c)$ is not a submersion. So it suffices to show it for U.

The proof is divided into three steps that together will show that $F(C)$ has n-dimensional measure zero. For this, we will first define the following subsets of C: let $C_n \subseteq C$ for $n \in \mathbb{N}_{>0}$ be defined by

$$C_n := \{x \in M \mid \text{for } 1 \leq i \leq n \text{ all } i\text{-th partial derivatives of } F \text{ vanish at } x\}.$$

That C_n is indeed a subset of C can be seen by considering the fact that for all $x \in C_n$ all first partial derivatives of F vanish at x and thus the first row of (the matrix representation of) $(dF)_x$ consists only of zero's, making it of less than full rank and thus x a critical point. Note also that $C \supseteq C_1 \supseteq C_2 \supseteq \ldots$. Now the three steps are:

(1) Show that $F(C \backslash C_1)$ has measure zero.

(2) Show for all $k \in \mathbb{N}_{>0}$ that $F(C_k \backslash C_{k+1})$ has measure zero.

(3) Show for $k > \frac{m}{n} - 1$ that $F(C_k)$ has measure zero.

Note that while $F(C) = F(C\backslash C_1) \cup F(C_1)$ and $F(C_1) = F(C_1\backslash C_2) \cup F(C_2)$ and in general $F(C_k) = F(C_k\backslash C_{k+1}) \cup F(C_{k+1})$, we cannot say in general that $F(C)$ is equal to $F(C\backslash C_1) \cup \bigcup_{k\in\mathbb{N}} F(C_k \backslash F(C_{k+1})$ as there might be points $c \in C$ for which $F(c)$ does not belong to any $F(C_k\backslash C_{k+1})$, i.e. critical points of F at which all partial derivatives of F vanish. Thus we need step 3 to show that $F(C)$ has measure zero, namely by

$$F(C) = F(C\backslash C_1) \cup \bigcup_{k=1}^{\lceil \frac{m}{n} - 1 \rceil} F(C_k\backslash C_{k+1}) \cup F(C_{\lceil \frac{m}{n} \rceil})$$

and the fact that a finite union of sets of measure zero is itself of measure zero. Step 1 and 2 will make use of the induction hypothesis, i.e. that $F(C)$ has measure zero if M has dimension $m-1$ or less, to show that $F(C\backslash C_1)$ and $F(C_k\backslash C_{k+1})$ are of measure zero when M is of dimension m. Step 3 however will be done in a direct fashion.

First we must prove the base case: if $m = 0$, then the set of critical values $F(C)$ has measure zero. First assume that the dimension of N is also zero. Then N consists of a (countable) number of points and the differential of F at any point of M is a map of a zero-dimensional tangent space to another zero-dimensional tangent space and thus must be surjective. So there are no critical points of F in M and thus $F(C)$ is empty and of measure zero. Suppose then that the dimension of N is larger than zero. Then the whole image $F(M)$ consists of a countable number of points in N each having measure zero. Since the union of a countable number of points having measure zero is still of measure zero, the whole image $F(M)$ is of measure zero and thus in particular the set of critical values $F(C)$ is also.

Step 1: $F(C\backslash C_1)$ has measure zero

If $a \in C\backslash C_1$ then there must be at least one partial derivative of F that does not vanish in a, since $a \notin C_1$. This means that there is at least one component of F that has a first partial derivative that does not vanish in a, say $\frac{\partial F_i}{\partial x_j}(a) \neq 0$. By applying a coordinate transformation in M that exchanges the first coordinate with the i-th coordinate and a coordinate transformation in N that exchanges the first coordinate with the j-th coordinate, we can assume without loss of generality that $\frac{\partial F_1}{\partial x_1}(a) \neq 0$. By smoothness of F this means there is a neighbourhood V_a of a in U such that $\frac{\partial F_1}{\partial x_1}(x) \neq 0$ for all $x \in V_a$. We now consider the following coordinate transformation:

$$h : \mathbb{R}^m \to \mathbb{R}^m : (x_1, \ldots, x_m) \mapsto (F_1(x_1, \ldots, x_m), x_2, \ldots, x_m).$$

First of all note that since F is smooth, h is also smooth. Now to show that h is a proper coordinate transformation (i.e. a diffeomorphism from $\mathbb{R}^m$ to $\mathbb{R}^m$) we compute the determinant of the total derivative and show that it is non-zero:

$$D_x h = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x) & \frac{\partial F_1}{\partial x_2}(x) & \frac{\partial F_1}{\partial x_3}(x) & \dots & \frac{\partial F_1}{\partial x_m}(x) \\ \frac{\partial x_2}{\partial x_1}(x) & \frac{\partial x_2}{\partial x_2}(x) & \frac{\partial x_2}{\partial x_3}(x) & \dots & \frac{\partial x_2}{\partial x_m}(x) \\ \frac{\partial x_3}{\partial x_1}(x) & \frac{\partial x_3}{\partial x_2}(x) & \frac{\partial x_3}{\partial x_3}(x) & \dots & \frac{\partial x_3}{\partial x_m}(x) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_m}{\partial x_1}(x) & \frac{\partial x_m}{\partial x_2}(x) & \frac{\partial x_m}{\partial x_3}(x) & \dots & \frac{\partial x_m}{\partial x_m}(x) \end{pmatrix}$$
$$= \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x) & \frac{\partial F_1}{\partial x_2}(x) & \frac{\partial F_1}{\partial x_3}(x) & \dots & \frac{\partial F_1}{\partial x_m}(x) \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

and thus by standard rules for computing the determinant we have $\det(D_x h) = \frac{\partial F}{\partial x_1}(x) \neq 0$ for all $x \in V_a$. Since the determinant of $D_x h$ is non-zero, it is an isomorphism and thus by the inverse function theorem, h is a local diffeomorphism at each point $x \in V_a$ and thus on the whole of V_a.

Now since $h_1(x) = F_1(x)$ we have $F_1(h^{-1}(x)) = h_1(h^{-1}(x)) = x_1$ and thus $F \circ h^{-1}(x) = (F_1(h^{-1}(x)), F_2(h^{-1}(x)), \dots, F_n(h^{-1}(x))) = (x_1, F_2(h^{-1}(x)), \dots, F_n(h^{-1}(x)))$. That means that by applying the coordinate transformation h^{-1}, we get new coordinates in which we can write F as

$$F(u, v_1, \dots, v_{m-1}) = F(u, v) = (u, F_2(u, v), \dots, F_n(u, v)). \tag{1}$$

In these new coordinates the total derivative of F reads as:

$$D_{(u,v)} F = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v_1} & \frac{\partial F_2}{\partial v_2} & \dots & \frac{\partial F_2}{\partial v_{m-1}} \\ \frac{\partial F_3}{\partial u} & \frac{\partial F_3}{\partial v_1} & \frac{\partial F_3}{\partial v_2} & \dots & \frac{\partial F_3}{\partial v_{m-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial u} & \frac{\partial F_n}{\partial v_1} & \frac{\partial F_n}{\partial v_2} & \dots & \frac{\partial F_n}{\partial v_{m-1}} \end{pmatrix}$$

An important observation here is that for $(u, v) \in (C \backslash C_1) \cap V_a$, the matrix $D_{(u,v)} F$ cannot be of full rank, since (u, v) is a critical point. Due to the structure of the matrix, this can only happen if the $(n-1)\times(m-1)$-submatrix

$$\begin{pmatrix} \frac{\partial F_2}{\partial v_1} & \frac{\partial F_2}{\partial v_2} & \cdots & \frac{\partial F_2}{\partial v_{m-1}} \\ \frac{\partial F_3}{\partial v_1} & \frac{\partial F_3}{\partial v_2} & \cdots & \frac{\partial F_3}{\partial v_{m-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial v_1} & \frac{\partial F_2}{\partial v_2} & \cdots & \frac{\partial F_n}{\partial v_{m-1}} \end{pmatrix}, \tag{2}$$

evaluated at (u,v), is not of full rank. We will now use this observation to show that certain points of an appropriately defined function from $\mathbb{R}^{m-1}$ to $\mathbb{R}^{n-1}$ are critical points. For $c\in\mathbb{R}$ let $B_c := \{v\in\mathbb{R}^{m-1} : (c,v)\in V_a\}$ and define:

$$F_c : \mathbb{R}^{m-1} \supseteq B_c \to \mathbb{R}^{n-1} : v \mapsto (F_2(c,v),\ldots,F_n(c,v)).$$

Now the total derivative of this new function reads as:

$$D_vF_c = \begin{pmatrix} \frac{\partial F_2(c,v)}{\partial v_1} & \frac{\partial F_2(c,v)}{\partial v_2} & \cdots & \frac{\partial F_2(c,v)}{\partial v_{m-1}} \\ \frac{\partial F_3(c,v)}{\partial v_1} & \frac{\partial F_3(c,v)}{\partial v_2} & \cdots & \frac{\partial F_3(c,v)}{\partial v_{m-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n(c,v)}{\partial v_1} & \frac{\partial F_2(c,v)}{\partial v_2} & \cdots & \frac{\partial F_n(c,v)}{\partial v_{m-1}} \end{pmatrix}$$

which is exactly the matrix we found at (2), evaluated at (c,v). Now suppose that $w\in\mathbb{R}^n$ lies in the intersection between the hyperplane where $w_1 = c$ and $F((C\backslash C_1)\cap V_a)$. Then w is of the form $w = F(u,v) = (c, F_2(u,v),\ldots,F_n(u,v))$. Also, since $w\in F((C\backslash C_1)\cap V_a)\subseteq F(V_a)$, formula (1) gives us $u = c$ and thus $w = F(c,v) = (c,F_2(c,v),\ldots,F_n(c,v))$. Now by definition of F_c we get $w = (c, F_c(v))$. Also, since $(u,v) = (c,v)\in(C\backslash C_1)\cap V_a$ we know that the matrix in (2) is not of full rank and thus that D_vF_c is not of full rank. That means that $F_c(v)$ is a critical value of F_c. So any w that lies in the intersection between the hyperplane where $w_1 = c$ and $F((C\backslash C_1)\cap V_a)$ is of the form (c, w^*) with w^* a critical value of F_c.

Now we are ready to invoke the induction hypothesis to establish that the measure of the set of critical values of F_c is equal to zero, and thus the measure of all w that lie in the intersection between the hyperplane where $w_1 = c$ and $F((C\backslash C_1)\cap V_a)$ is also zero. When we combine this with lemma 3.11, we are able to conclude that $F((C\backslash C_1)\cap V_a)$ itself is of measure zero. For this, let us first note that F_c is a function from $B_c\subseteq\mathbb{R}^{m-1}$ to $\mathbb{R}^{n-1}$. In fact, since it is a smooth function from an $(m-1)$-dimensional to an $(n-1)$-dimensional manifold, by the induction hypothesis the $(n-1)$-dimensional measure of the set of critical values of F_c is equal to zero. This thus means that the measure of all w that lie in the intersection between the hyperplane where $w_1 = c$ and $F((C\backslash C_1)\cap V_a)$ must also be zero. Now by lemma 3.11 the measure of $F((C\backslash C_1)\cap V_a)$ must be zero.

Now all the V_a for every $a \in C\backslash C_1$ cover $C\backslash C_1$; this might be an uncountable cover (as $C\backslash C_1$ might contain an uncountable number of elements), but each V_a is a union of sets from a countable collection of opens since U is second countable, and thus all of $C\backslash C_1$ is cover by a countable we can shrink V_a such that it is a ball since $\mathbb{R}^m$ is Lindelöf, we can extract a countable subcover V. Now $F(C\backslash C_1) = \bigcup_{V_a \in V} F((C\backslash C_1) \cap V_a)$ and since for every V_a in V we have that the measure of $F((C\backslash C_1) \cap V_a)$ is zero and there are a countable number of V_a in V, the measure of $F(C\backslash C_1)$ must be zero.

Step 2: $F(C_k\backslash C_{k+1})$ has measure zero for all $k \in \mathbb{N}$

Let again a be an element of $C_k\backslash C_{k+1}$. Since a is not in C_{k+1} there must be at least one $(k+1)$-st partial derivative of F that does not vanish. This $(k+1)$-st derivative is the partial derivative of some k-th derivative of F. Let us now define $y : U \to \mathbb{R}$ as this k-th partial derivative (note that y is real-valued and we are more specifically defining it as the component of the k-th partial derivative of F that has a non-vanishing partial derivative in a). By applying an appropriate coordinate transformation that only interchanges the desired coordinates (i.e. the first and the one at which y has non-vanishing derivative) we can assume that it is the first partial derivative of y that does not vanish in a, thus $\frac{\partial y}{\partial x_1}(a) \neq 0$. Since y is real-valued and has a non vanishing partial derivative in a, the $(1 \times m)$matrix representation of its differential is of full rank (namely 1) and thus a is a regular value of y. Since y is smooth, there must be a neighbourhood around a, which we will again denote by V_a, where its first derivative does not vanish and thus where it is regular. By the level-set theorem, the zero set of y in V_a is a smooth hyper-surface, which we will denote by Y. Since by definition y is zero on all of $C_k\backslash C_{k+1}$ we have $V_a \cap (C_k\backslash C_{k+1}) \subseteq Y$. Now at any point $p \in V_a \cap (C_k\backslash C_{k+1})$ we have that d_pF is not surjective (because $p \in C_k \subseteq C$. If we now consider the function $F|_Y : Y \to \mathbb{R}^n$, then $d_pF|_Y = (d_pF)|_{T_pY}$ is also not surjective, since it is the restriction of linear map d_pF thas wasn't surjective in the first place. So any point $p \in V_a \cap (C_k\backslash C_{k+1})$ is a critical point of $F|_Y$ and so $F(V_a \cap (C_k\backslash C_{k+1}))$ is contained in the set of critical values of $F|_Y$. Here we can use the induction hypothesis again to establish that the measure of the set of critical values of $F|_Y$ is zero, and thus the measure of $F(V_a \cap (C_k\backslash C_{k+1}))$ is also zero. Now as in step 1, V_a is the union of opens from countable collection of opens that covers $F(C_k\backslash C_{k+1})$, proving that $F(C_k\backslash C_{k+1})$ also has measure zero.

Step 3: $F(C_k)$ has measure 0 for $k > \frac{m}{n} - 1$

This step is a direct proof that the measure of $F(C_k)$ is zero, for $k > \frac{m}{n} - 1$, attained by covering $F(C_k)$ with a countable union of balls whose added volume vanishes when decreasing the radius of the balls towards zero, thus showing that $F(C_k)$ has measure zero. We start by noting that we can cover U with a countable number of cubes such that for each $a_i \in C_k \cap U$ there is a cube E_i such that $a_i \in E_i$. Since we eventually want to cover $F(C_k)$ by a countable collection of balls, we are going to consider the volume of the image $F(E_i)$ of such a cube. We will find that $F(E_i)$ can be contained in a ball of a certain volume for each E_i, covering $F(C_k)$ with a countable number of balls. If we then reduce the side length of the cubes E_i, more cubes will be needed to cover C_k, resulting in more balls covering $F(C_k)$, but we will find that we can reduce the volume of those balls proportionally faster than the amount of balls necessary to cover $F(C_k)$. The crucial part that allows this is the following bound for values $F(x)$ when $x \in E_i$ that can be derived using Taylor's theorem:

$$|F(x) - F(a_i)| \leq A|x - a_i|^{k+1}. \tag{3}$$

where A is a constant.

To prove this bound, let us first note that since E_i is compact, all k+1-st partial derivatives of F are absolutely bounded on E_i, since they are continuous (since F is smooth). Let M be this bound. Since $a_i \in C_k$, the k-th order Taylor polynomial of F at a_i is equal to:

$$P_k(x) = F(a_i) + \sum_{j=1}^{k} \frac{1}{j!} \sum_{I:|I|=j} \partial_I F(a_i)(x - a_i)^I = F(a_i)$$

since all partial derivatives of order less then or equal to k vanish at a_i by definition of C_k. Thus:

$$|F(x) - F(a_i)| = |F(x) - P_k(x)| = |R_k(x)| \tag{4}$$

where R_k is the k-th remainder term at a_i given by:

$$R_k(x) = \frac{1}{k!} \sum_{I:|I|=k+1} (x - a_i)^I \int_0^1 (1-t)^k \partial_I F(a_i + t(x - a_i))dt.$$

Due to the absolute boundedness of the $(k+1)$-st partial derivatives of F by M, we can bound the absolute value of R_k at a_i by:

$$|R_k(x)| \leq \frac{M}{(k+1)!}|x - a_i|^{k+1}.$$

This combined with (4) gives us the desired bound as described in (3), with $A = \frac{M}{(k+1)!}$.

Let the side length of E_i be R. When $x \in E_i$, we know that since we also have $a_i \in E_i$, the distance between these two is at most the length of the diagonal of two opposite points of the cubes, thus $|x_i - a_i| \leq \sqrt{mR^m} = \sqrt{m}R$. Now by the bound in (3) we know that $|F(x) - F(a_i)| \leq A|x - a_i|^{k+1} \leq Am^{(k+1)/2}R^{k+1}$. Since we have a bound on the distance between $F(x)$ and $F(a_i)$ for any arbitrary point $x \in E_i$, this must mean that the image $F(E_i)$ is contained in a ball whose diameter is less than or equal to $Am^{(k+1)/2}R^{k+1}$. Now the volume of this ball is proportional to $C(Am^{(k+1)/2}R^{k+1})^n = A'R^{(k+1)n}$, with C and A' constants (with respect to R).

Now we can start to shrink our original cube E_i, and we do that by halving the length of its edge. We still want the cubes to cover $C_k \cap U$ (because we want $F(C_k \cap U)$ to be contained in the image of our cubes $F(E_i)$. When we halve the edge length of our cubes, we will need 2^m times the number of cubes to cover the original volume. This means that we will have 2^m times number of balls covering $F(C_k \cap U)$, but each ball having volume less than or equal to $A'(\frac{R}{2})^{(k+1)n}$. This means the total volume of all the balls covering $F(C_k \cap U)$ is less than or equal to $2^m A'(\frac{R}{2})^{(k+1)n} = A'R^{(k+1)n}2^{m-(k+1)n}$. That is $2^{m-(k+1)n}$ times the original volume before halving the cube length. This means that if $m - (k+1)n < 0$, we can decrease the volume of all the balls covering $F(C_k \cap U)$, while still maintaining a complete cover of

$F(C_k \cap U)$. Since we can do this an arbitrary amount of times, we can decrease the volume of the cover of $F(C_k \cap U)$ to arbitrarily close to zero. This shows that, under the condition that $m - (k+1)n < 0$, thus $k > \frac{m}{n} - 1$, the measure of $F(C_k \cap U)$ must be zero and thus that the measure of $F(C_k)$ is zero.

□

5 Preliminaries to the Whitney embedding theorem

5.1 Regular coordinate balls

A smooth manifold is covered by an atlas of smooth charts (U, ϕ). When the image of a smooth chart $\phi(U) \subseteq \mathbb{R}^n$ is an open ball, we call U a smooth coordinate ball, and we call it a regular coordinate ball when there is a smooth coordinate ball $U' \subseteq \overline{U}$ with associated map $\phi' : B' \to \mathbb{R}^n$ such that for some positive real numbers $r < r'$ we have:

$$\phi(U) = B_r(0), \quad \phi(\overline{U}) = \overline{B}_r(0) \quad \text{and} \quad \phi'(U') = B_{r'}(0),$$

where $B_r(0)$ denotes an open ball of radius r centered at zero. It can be shown that any smooth manifold admits a countable basis of regular coordinate balls; the main idea is that first assuming M has only one chart $\phi : M \to R \subseteq \mathbb{R}^n$, we can find a countable basis of R consisting of open balls B centered at rational points (since $R \subseteq \mathbb{R}^n$; since ϕ is a homeomorphism, the pre-images of these balls $\phi^{-1}(B)$ forms a countable basis of M and the restriction of ϕ to these balls produces the desired smooth charts. Then the case of a countable number of charts can be tackled by the fact that a countable collection of countable sets is again countable.

5.2 Smooth embeddings

The Whitney embedding theorem shows that any smooth manifold can be smoothly embedded in Euclidean space. A smooth embedding between manifolds is a smooth immersion that is also a topological embedding. The following theorem provides an important means to show that a function is actually an embedding:

Theorem 5.1. *Let $F : M \to N$ be a smooth function between smooth manifolds M and N such that it is an injective immersion. Then if any of the two following conditions hold, F is a smooth embedding: (1) F is a proper map, (2) M is compact.*

5.3 Smooth bump functions and partitions of unity

The proof of the Whitney embedding theorem uses the existence of smooth bump functions, which are defined as follows:

Definition 5.2. *Let $A \subseteq M$ be a closed subset of a smooth manifold M that is contained in an open subset U of M. A smooth bump function $\rho : M \to [0, 1]$ for A supported on U is a smooth function that is equal to 1 in A, and equal to zero outside of U.*

It can be shown that for every closed subset $A \subseteq M$ of a smooth manifold M and open subset $U \subseteq M$ containing A, there exists a smooth bump function for A and supported on U. With smooth bump functions, it is possible to extend a smooth function ϕ defined on an open subset $U \subseteq M$ to a smooth function $\tilde{\phi}$ such that $\tilde{\phi}(x) = \rho(x)\phi(x)$ on U and $\tilde{\phi}(x) = 0$ outside A. For details on smooth bump functions, we refer to Lee [1].

5.4 exhaustive functions and regular domains

6 Whitney embedding theorem

Here we will state and prove the Whitney embedding theorem:

Theorem 6.1. *Any m-dimensional smooth manifold M can be smoothly embedded into N-dimensional Euclidean space, for some $N \in \mathbb{N}$.*

Note that we will not prove the full theorem, which also gives a bound on the dimension of the Euclidean space in which the manifold is embedded, namely $2m+1$. With more work, this bound can even be lowered to $2m$. The proof is modelled on that give by Lee [1] and adapted to our needs; we will also use a lemma by Guillemin and Pollack [3].

Proof. The first part is to show that any compact m-dimensional manifold can be smoothly embedded into N-dimensional Euclidean space for some $N \in \mathbb{N}$. What we will prove here is actually a bit stronger, namely that any compact subset M of an arbitrary (thus possibly non-compact) manifold has a neighbourhood that can be smoothly embedded into N-dimensional Euclidean space for some $N \in \mathbb{N}$. We will not yet need Sard's theorem for this. Since we are only concerned here with proving it for some $N \in \mathbb{N}$, the argument might be characterized as somewhat brute-force: the idea is to cover M with a finite amount of coordinate balls B_i, $1 \leq i \leq I$, which is possible due to compactness, and then to use bump functions to construct a function F with domain M that has m components for each B_i (or each $i \in I$) and then add another I components which we will need to ensure F is injective. In total, M will thus be embedded in $\mathbb{R}^{mI+I}$.

To start, we reiterate that any smooth manifold can be covered by a countable set of regular coordinate balls and so any compact subset M of a smooth manifold can covered by this same countable set of regular coordinate balls. Since M is compact we can extract a finite subcover $\{B_1, \ldots, B_I\}$ consisting of a finite number of regular coordinate balls. For each regular coordinate ball B_i there exists a coordinate ball B'_i such that $\overline{B}_i \subseteq B'_i$ and a diffeomorphism $\phi_i : \overline{B}_i \to \mathbb{R}^m$ and a smooth bump function $\rho_i : M \to \mathbb{R}$ that is supported on B'_i, identically 1 on $\overline{B}_i$ and zero outside of B'_i. We now define the following function, which will be our desired embedding:

$$F : \bigcup_{i=1}^{I} \overline{B}_i \to \mathbb{R}^{mI+I} : p \mapsto (\rho_1(p)\phi_1(p), \ldots, \rho_I(p)\phi_I(p), \rho_1(p), \ldots, \rho_I(p)).$$

To show that it is an embedding, we show that it is a smooth injective immersion and then use the fact that a smooth injective immersion on a compact domain is actually a smooth embedding. First, to show that it is injective, suppose $F(p) = F(q)$ for $p, q \in \bigcup_{i=1}^{I} \overline{B}_i$.

There must be an i such that $p \in \overline{B}_i$ and thus $\rho_i(p) = 1$. Now since $F(p) = F(q)$, we have $\rho_i(q) = \rho_i(p) = 1$ and since ρ_i is only equal to 1 on $\overline{B}_i$, we have $q \in \overline{B}_i$. Thus both $p, q \in \overline{B}_i$ and since ϕ_i is injective on $B'_i \supseteq \overline{B}_i$ (because it's a diffeomorphism) and $\phi_i(p) = \phi_i(q)$ (because we supposed that $F(p) = F(q)$) it follows that $p = q$ and thus F is injective. The smoothness of F follows from the fact that its components are smooth, which is a consequence of the properties of smooth coordinate balls and smooth bump functions. Finally, to show that F is an immersion we consider the matrix representation of its derivative. For this we note that for any $p \in M$ we have $p \in B_i$, giving $\rho_i(p) = 1$, and thus the matrix representation of its derivative has the following structure:

$$(dF)_p = \begin{pmatrix} (d\rho_1\phi_1)_p \\ \vdots \\ (d\phi_i)_p \\ \vdots \\ (d\rho_I\phi_I)_p \\ (d\rho_1)_p \\ \vdots \\ (d\rho_I)_p \end{pmatrix},$$

where each $(d\rho_j\phi_j)_p$ for $j \neq i$ and $(d\phi_i)_p$ is an $(m \times m)$-matrix and $(d\rho_j)_p$ is a $(1 \times m)$-matrix. Now if we have two vectors $\mathbf{a}$ and $\mathbf{b}$, then $(dF)_p\mathbf{a} = (dF)_p\mathbf{b}$ implies

$$\begin{pmatrix} (d\rho_1\phi_1)_p \\ \vdots \\ (d\phi_i)_p \\ \vdots \\ (d\rho_I\phi_I)_p \\ (d\rho_1)_p \\ \vdots \\ (d\rho_I)_p \end{pmatrix} \mathbf{a} = \begin{pmatrix} (d\rho_1\phi_1)_p \\ \vdots \\ (d\phi_i)_p \\ \vdots \\ (d\rho_I\phi_I)_p \\ (d\rho_1)_p \\ \vdots \\ (d\rho_I)_p \end{pmatrix} \mathbf{b},$$

and thus in particular $(d\phi_i)_p\mathbf{a} = (d\phi_i)_p\mathbf{b}$. But this implies $\mathbf{a} = \mathbf{b}$ since ϕ is a diffeomorphism by definition and thus also an immersion. So we conclude that F is an immersion. With this, we have shown that F is a smooth injective immersion and since M is compact, it is a smooth embedding.

Now we need to treat the case where M is not compact. Since M is not compact, we cannot assume there is a finite cover by coordinate balls and thus we cannot construct a function F as we did in the compact case. The idea will be to partition M into a number of compact subsets using an exhaustion function and then to use the first part of the proof to construct a smooth embedding from the smooth embeddings of the compact subsets.

First we will state and prove the following lemma, which is part of the Whitney embedding theorem. It is necessary to prove that any non-compact smooth manifold can be embedded in Euclidean space, but is also of interest in its own right when applied to compact smooth manifolds. The proof of this lemma is by Guillemin and Pollack [3].

Lemma 6.2. *Let M be an m-dimensional smooth manifold that can be embedded in some Euclidean space $\mathbb{R}^K$. Then M admits a injective smooth immersion into $\mathbb{R}^{2m+1}$.*

Proof. We are going to show that if there is a smooth injective immersion $f : M \to \mathbb{R}^K$, with $K > 2m+1$ (which there is if M can be smoothly embedded in $\mathbb{R}^K$, then we can construct a projection onto a $(K-1)$-dimensional vector subspace of $\mathbb{R}^K$ (which is isomorphic to $\mathbb{R}^{K-1}$) which is an injective immersion. We will first define two maps: $h : M \times M \times \mathbb{R} \to \mathbb{R}^K : (x, y, t) \mapsto t(f(x) - f(y))$ and $g : T(M) \to \mathbb{R}^K : (x, v) \mapsto (df)_x v$, where $T(M) := \{(x, v) \in M \times \mathbb{R}^K : v \in T_x M\}$. Now note that since $K > 2m + 1$ the dimensions of the domains of h and g are larger then the codomain, implying that all values of the images of h and g are critical values and thus that the measure of the images of h and g must be zero in $\mathbb{R}^K$. This means there exists an $a \in \mathbb{R}^K$ which does not belong to the image of h and g. Note that $h(x, y, 0) = 0$ and $g(x, 0) = 0$, so $a \neq 0$. Now let $H = \{b \in \mathbb{R}^K : \langle a, b\rangle = 0\}$, meaning that H is the $(K-1)$-dimensional vector subspace of $\mathbb{R}^K$ consisting of all vectors of $\mathbb{R}^K$ that are orthogonal to a. Also, let $\pi : \mathbb{R}^K \to H$ be the orthogonal projection onto H. We now claim that $\pi \circ f$ is an injective immersion.

To show that $\pi \circ f$ is injective, suppose $\pi \circ f(x) = \pi \circ f(y)$. Since π is an orthogonal projection to H, $\pi(v) = \pi(w)$ means that $\pi(v - w) = \pi(v) - \pi(w) = 0$ and so $v - w = ta$ for some $t \in \mathbb{R}$. Thus $\pi \circ f(x) = \pi \circ f(y)$ means $f(x) - f(y) = ta$. Now if $x \neq y$, then $h(x, y, \frac{1}{t}) = \frac{1}{t}(f(x) - f(y)) = \frac{1}{t}ta = a$ which is contrary to the fact that a is not in the image of h. Thus we must have $x = y$ and so $\pi \circ f$ is injective.

To show that $\pi \circ f$ is an immersion at an arbitrary $x \in M$, we suppose that there is a vector $v \in T_x M$ such that $(d(\pi \circ f))_x(v) = 0$. Due to the chain rule we get $(d\pi)_{f(x)}(df)_x(v) = 0$ and since $(d\pi)_y = \pi$ (because π is linear) we get $\pi((df)_x(v)) = 0$. This again means that $(df)_x(v) = ta$ for some $t \in \mathbb{R}$. If $t = 0$, then $(df)_x(v) = 0$, but f is an immersion by definition, so the kernel of its differential as a linear map must only contain the zero vector. This means that t cannot be zero. Now we thus have $g(x, \frac{v}{t}) = (df)_x(\frac{v}{t}) = \frac{1}{t}(df)_x(v) = \frac{1}{t}ta = a$, which is again contrary to the fact that a is not in the image of g. So there is no non-zero vector v such that $(d(\pi \circ f))_x(v) = 0$, so $\pi \circ f$ is an immersion.

We have now shown that if M can be smoothly embedded in $\mathbb{R}^K$ for $K > 2m + 1$, there exists a smooth injective immersion of M into $\mathbb{R}^{K-1}$ and we can conclude, working inductively over K down to $K = 2m + 2$, that there is an injective immersion into $\mathbb{R}^{2m+1}$. □

We have already shown that any compact manifold M can be embedded in $\mathbb{R}^N$ for some $N \in \mathbb{N}$ and now know that there exists a smooth injective immersion from M into $\mathbb{R}^{2m+1}$. But since M is compact, this must be a smooth embedding and so any compact smooth manifold of dimension m can be smoothly embedded in $\mathbb{R}^{2m+1}$

First, let $f : M \to \mathbb{R}$ be a smooth exhaustion function. We are going to partition M into (overlapping) subsets using f. Define $D_0 = f^{-1}((-\infty, 1])$ and $D_i = f^{-1}([i, i + 1])$ for $i\mathbb{N}_{>0}$. Furthermore, define $E_0 = f^{-1}((-\infty, 1\frac{1}{2}))$ and $E_i = f^{-1}((i - \frac{1}{2}, i + 1\frac{1}{2}))$ for $i\mathbb{N}_{>0}$. Now we have in particular that E_i is open (since f is continuous), that D_i is compact (since f is proper) and that $D_i \subseteq E_i$. From the first part of the proof combined with lemma 6.2 we know that there is a smooth injective immersion $\phi_i : D_i^\star \to \mathbb{R}^{2n+1}$ for an open neighbourhood $D_i^\star \supseteq D_i$. Now let ρ_i be a smooth bump function equal to 1 on a neighbourhood of D_i and supported in E_i.

Now the following function $F : M \to \mathbb{R}^{2n+1} \times \mathbb{R}^{2n+1} \times \mathbb{R}$ will be our desired embedding:

$$F(p) = \left(\sum_{i \text{ even}} \rho_i(p)\phi_i(p), \sum_{i \text{ odd}} \rho_i(p)\phi_i(p), f(p) \right).$$

To show that it is a smooth embedding, we show that it is a smooth injective immersion that it is also proper and hence a smooth embedding.

To prove smoothness of F, first note that f is smooth by definition of being a smooth exhaustion function. Also ρ_i and ϕ_i are smooth for each $i \in \mathbb{N}$; this does not immediately mean that the first two components in the definition of F are smooth, since these are infinite sums of ρ_i and ϕ_i. Now for each $p \in M$, there is one $j \in \mathbb{N}$ such that $p \in D_j \subseteq E_j$. Since $E_k \cap E_l \neq$ only if $k = l-1$, $k = l$ or $k = l+1$ and ρ_i is supported on E_i, if j is even there will be no other even $j' \in \mathbb{N}$ such that $\rho_{j'} \neq 0$ and if j is odd there will be no other odd $j'' \in \mathbb{N}$ such that $\rho_{j''} \neq 0$ on a neighbourhood of p (since E_j is open). So in a neighbourhood of p at most one term of both sums will be non-zero, making them smooth on that neighbourhood. This means that F is locally smooth for each $p \in M$ and thus that F is smooth.

To show that F is proper, note that f is proper since it is a smooth exhaustion function and that the other two components of F are smooth (and thus continuous). Thus F is proper, since the product function $g \times h$ is proper if g is continuous and h is proper.

Now suppose $F(p) = F(q)$ for $p, q \in M$. There must be a $j \in \mathbb{N}$ such that $p \in D_j$. Now $F(p) = F(q)$ implies $f(p) = f(q)$. Since $p \in D_j$, $f(p) \in [j, j+1]$ and thus $f(q) \in [j, j+1]$ and so $q \in f^{-1}([j, j+1]) = D_j$. This means that $\rho_j(p) = 1 = \rho_i(q)$ and thus $\phi_i(q) = \rho_i(q)\phi_i(q) = \rho_i(p)\phi_i(p) = \phi_i(p)$ and since ϕ_i is injective we have $p = q$, meaning F is injective.

Finally suppose $p \in M$ and $p \in D_j$ with j even. Then we have just seen that $\sum_{i \text{ even}} \rho_i(p)\phi_i(p) = \phi_j(p)$, giving:

$$(dF)_p = \begin{pmatrix} (d\phi_j)_p \\ \star \\ (df)_p \end{pmatrix},$$

where the $\star$ denotes the differential of the second component. This means that if $(dF)_p\mathbf{a} = (dF)_p\mathbf{b}$ for two vectors $\mathbf{a}$ and $\mathbf{b}$, $(d\phi_j)_p\mathbf{a} = (d\phi_j)_p\mathbf{b}$ and since ϕ_j is an immersion, $\mathbf{a} = \mathbf{b}$. This shows that $(dF)_p$ is injective and so F is an immersion at p. Now if j is odd, we get:

$$(dF)_p = \begin{pmatrix} \star \\ (d\phi_j)_p \\ (df)_p \end{pmatrix},$$

with $\star$ denoting the differential of the first component of F. Via a similar argument we now get that if $(dF)_p\mathbf{a} = (dF)_p\mathbf{b}$ for two vectors $\mathbf{a}$ and $\mathbf{b}$, we get $(d\phi_j)_p\mathbf{a} = (d\phi_j)_p\mathbf{b}$ and thus $\mathbf{a} = \mathbf{b}$ since ϕ_j is an immersion. This gives us that F is an immersion at an arbitrary p and so F is an immersion. Altogether this gives us that F is the desired smooth embedding. □

References

[1] J. M. Lee, *Introduction to smooth manifolds* (Springer, New York, 2012).

[2] J. P. v. B. H. AuthorsJ J Duistermaat, Johan A C Kolk, *Multidimensional real analysis. I, Differentiation* (Cambridge University Press, Cambridge, 2004).

[3] V. Guillemin and A. Pollack, *Differential topology* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1974).