Mathematics 241A
Introduction to Global Analysis

John Douglas Moore
Department of Mathematics
University of California
Santa Barbara, CA, USA 93106
e-mail: moore@math.ucsb.edu

Fall, 2010

Preface

These are slightly revised lecture notes for the graduate course, Topics in Geometry, given at UCSB during the fall quarter of 2009. We intend to include additional topics later. It might be helpful to start with an overview of the subject presented.

Morse theory might have developed in three main stages, although as events transpired, the three stages were actually intertwined.

The first stage should have been finite-dimensional Morse theory, which relates critical points of proper nonnegative functions on finite-dimensional manifolds to the topology of these manifolds. Indeed, the foundations for this were laid in Marston Morse's first landmark article on Morse theory [55], but he quickly turned his attention to problems from the calculus of variations, which ultimately became part of the infinite-dimensional theory. In subsequent developments, finite-dimensional Morse theory became a primary tool for studying the topology of finite-dimensional manifolds and has had many successes, including the celebrated h-cobordism theorem of Smale [51], which settled the generalized Poincaré conjecture in dimensions greater than five: any compact manifold of dimension at least five which is homotopy equivalent to a sphere must be homeomorphic to a sphere. Modern expositions of finite-dimensional Morse theory often construct a chain complex from the free abelian group generated by the critical points, the boundary being defined by orbits of the gradient flow which connect the critical points. The homology of this chain complex, called Morse homology, is isomorphic to the usual integer homology of the manifold; see [74] for example.

What might have been the second stage—the Morse theory of geodesics—formed the core of what Morse [56] called "the calculus of variations in the large." Morse was interested in studying the critical points of the length function or action function on the space of paths joining two points in a Riemannian manifold, the critical points being geodesics. His idea was to approximate the infinite-dimensional space of paths by a finite-dimensional manifold of very high dimension, and then apply finite-dimensional Morse theory to this approximation. As explained in Milnor's classical book on Morse theory [50], this approach produced many striking results in the theory of geodesics in Riemannian geometry, such as the theorem of Serre that any two points on a compact Riemannian manifold can be joined by infinitely many geodesics. It also provided the first proof of the Bott periodicity theorem from homotopy theory. One might regard

the Morse theory of geodesics as an application of topology to the study of ordinary differential equations, in particular, to those equations which like the equation for geodesics, arise from classical mechanics. Thus Morse theory arises from the very core of "applied mathematics."

Palais and Smale were able to provide an elegant reformulation of the Morse theory of geodesics in the language of infinite-dimensional Hilbert manifolds [62]. They showed that the action function on the infinite-dimensional manifold of paths satisfies "Condition C," a condition replacing "proper" in the finite-dimensional theory, which they showed is sufficient for the development of Morse theory in infinite dimensions. This became the standard approach to the Morse theory of geodesics during the last several decades of the twentieth century.

Future historians will most likely regard the third stage of Morse theory as encompassing many strands, but our viewpoint is to focus on techniques for applying Morse theory to nonlinear elliptic partial differential equations coming from the calculus of variations in which the domain is a two-dimensional compact surface. Morse himself hoped to apply the ideas of his theory to a central case—the partial differential equations for minimal surfaces in Euclidean space. The first steps in this direction were taken by Morse and Tompkins, as well as Shiffman, who established the theorem that if a simple closed curve in Euclidean space $\mathbb{R}^3$ bounds two stable minimal disks, it bounds a third, which is not stable. This provided a version of the so-called "mountain pass lemma" for minimal disks in Euclidean space. The results of Morse, Tompkins and Shiffman suggested that Morse inequalities should hold for minimal surfaces in Euclidean space with boundary constrained to lie on a given Jordan curve, and indeed, such inequalities were later established under appropriate hypotheses [41].

But the most natural extension of the Morse theory of geodesics to the realm of partial differential equations would be a Morse theory of two-dimensional minimal surfaces in a more general curved ambient Riemannian manifold M, instead of the ambient Euclidean space used in the classical theory of minimal surfaces. The generalization to a completely general ambient space requires new techniques. Unfortunately, if Σ is a connected compact surface, it becomes somewhat awkward to extend the finite-dimensional approximation procedure—so effective in the theory of geodesics—to the space of mappings $\mathrm{Map}(\Sigma, M)$ from Σ to M. One might hope that a better approach would be based upon the theory of infinite-dimensional manifolds, as developed by Palais and Smale, but a serious problem is encountered: the standard energy function

$$E : \mathrm{Map}(\Sigma, M) \to \mathbb{R},$$

used in the theory of harmonic maps and parametrized minimal surfaces, fails to satisfy the Condition C which Palais and Smale had used so effectively in their theory, when $\mathrm{Map}(\Sigma, M)$ is completed with respect to a norm strong enough to lie within the space of continuous functions.

To get around this difficulty, Sacks and Uhlenbeck introduced the α-energy [68], [69], a perturbation of the usual energy which does satisfy Condition C when $\mathrm{Map}(\Sigma, M)$ is completed with respect to a Banach space norm which

is both weak enough to satisfy condition C and strong enough to have the same homotopy type as the space of continuous maps from Σ to M. In simple terms, we could say that the α-energy lies within "Sobolev range." The α-energy approaches the usual energy (plus a constant) as the parameter α in the perturbation goes to one, and we can therefore say that the usual energy is "on the border of Sobolev range." Using the α-energy, Sacks and Uhlenbeck were able to establish many striking results in the theory of minimal surfaces in Riemannian manifolds, including the fact that any compact simply connected Riemannian manifold contains at least one nonconstant minimal two-sphere, which parallels the classical theorem of Fet and Lyusternik stating that any compact Riemannian manifold contains at least one smooth closed geodesic. But they also discovered the phenomenon of "bubbling" as $\alpha \to 1$, which prevents Morse inequalitites from holding for compact parametrized minimal surfaces in complete generality.

A somewhat different approach to existence of parametrized minimal surfaces in Riemannian manifolds was developed at about the same time by Schoen and Yau [72], using Morrey's solution to the Plateau problem in a Riemannian manifold and arguments based upon a "replacement procedure." Their approach yielded striking theorems relating positive scalar curvature to topology of three-manifolds, and was a step toward the first proof of the positive mass theorem of general relativity. The Schoen-Yau replacement procedure can also be used to obtain many of the existence results of Sacks and Uhlenbeck, and an alternate treatment of many of their theorems is provided by Jost [39], as well as by other authors, who use yet different techniques, including heat flow.

However, the approach via Sacks-Uhlenbeck perturbation seems to provide the strongest link with the global analysis approach to nonlinear partial differential equations, and the clearest insight into bubbling, the phenomenon mentioned above which is observed as the perturbation is turned off. Bubbling appears to be an essential component of any complete critical point theory of minimal surfaces within Riemannian manifolds, and this phenomenon also appears in the study of other nonlinear partial differential equations, such as the Yang-Mills equation on four-dimensional manifolds.

If one were to replace the Riemannian metric on the ambient manifold by a conformal equivalence class of such metrics, the theory of two-dimensional minimal surfaces would become part of a broader context—nonlinear partial differential equations which are conformally invariant and lie on the border of Sobolev range. These equations include the Yang-Mills equations on four-dimensional manifolds from the standard model for particle physics, the anti-self-dual equations used so effectively by Donaldson, the Seiberg-Witten equations, and Gromov's equations for pseudoholomorphic curves. A technology has been developed for studying these equations. First one needs to develop a transversality theory using Smale's generalization of Sard's theorem from finite-dimensional differential topology. This generally shows that in generic situtations solutions to the nonlinear partial differential equation form a finite-dimensional submanifold of an infinite-dimensional function space. The tangent space to this submanifold is studied via the linearization of the nonlinear partial differential equation

at a given solution; often, the dimension of the tangent space is obtained by application of the Atiyah-Singer index theorem, which reduces to the Riemann-Roch theorem in the case of parametrized minimal surfaces. Next one develops a suitable compactness theorem. Finally, one uses topological and geometric methods to derive important geometrical conclusions (for example, existence of minimal surfaces) or to construct differential topological invariants of manifolds (as in Seiberg-Witten theory).

As mentioned before, for minimal surfaces, bubbling implies that the most obvious extension of the Morse inequalities to minimal surfaces in Riemannian manifolds cannot hold, but also provides the framework for analyzing how the Morse inequalities fail.

Not only does Sacks-Uhlenbeck bubbling interfere with the Morse inequalities, but when minimizing area one must allow for variations not only over the space of functions but over the conformal structure on the surface, an element of Teichmüller space, and a sequence of harmonic maps may degenerate as the conformal structure moves to the boundary. Moreover, branched coverings of a given minimal surfaces count as new critical points within the space of functions, although they are not geometrically distinct from the covered surface. Finally, the energy is invariant under the action of the mapping class group, so the energy descends to a function on the quotient space, a space which has a more complicated topology than that of $\mathrm{Map}(\Sigma, M)$ when the genus of Σ is at least two. One might suspect that a procedure for constructing minimal surfaces that can fail in several different ways is too flawed to be of much use. However, we argue that a different perspective is more productive—since the minimax procedure for a given homology class does not always yield interesting minimal surfaces, one should divide homology classes into various categories depending upon which of the possible difficulties will likely arise.

It is our purpose here to provide some of the foundations for such a theory, a theory which promises important applications similar to those found in the Morse theory of geodesics.

It has long been known that that the extension of Morse theory to infinite-dimensional manifolds is not really necessary for the study of geodesics. Indeed, Bott expresses it this way in his beautiful survey article on Morse theory [9] in 1982: "I know of no aspect of the geodesic question where [the infinite-dimensional approach] is essential; however it clearly has some aesthetic advantages, and points the way for situations where finite dimensional approximations are not possible..." One could imagine constructing a finite-dimensional approximation suitable for studying the α-energy when $\alpha > 1$, but it would be far more awkward than the infinite-dimensional manifold of maps, and an analysis of how it breaks down as $\alpha \to 1$ seems to require a study of infinite-dimensional manifold pieces. Thus, for any partial Morse theory of minimal surfaces, in contrast to closed geodesics, calculus on infinite-dimensional manifolds appears to play an essential conceptual and simplifying role.

We assume the reader is familiar with the basics of finite-dimensional differential geometry, including geodesics, curvature, the tubular neighborhood theorem, and the second variation formula for geodesics. We will also assume

some familiarity with basic complex analysis, the foundations of Banach and Hilbert space theory, and the willingness to accept results from the linear theory of elliptic partial differential operators, in particular the theory of Fredholm operators on Sobolev spaces. All of these topics are treated very well in highly accessible sources, to which we can refer at the appropriate time.

We begin with an overview of global analysis on infinite-dimensional manifolds of maps. The second chapter reviews the theory of geodesics on Riemannian manifolds which owes much to the pioneering work of Bott, Gromoll, Klingenberg and Meyer. We then turn to minimal surfaces, providing a brief introduction to the key theorems of Sacks, Uhlenbeck, Meeks, Schoen and Yau which helped elucidate the topology of three-dimensional manifolds. A final chapter is planned that will cover the bumpy metric theorem of [52] and some of its immediate applications. Starred sections can be omitted without loss of continuity.

Doug Moore

Santa Barbara, CA, December, 2010

Contents

Chapter 1

Infinite-dimensional manifolds

1.1 A global setting for nonlinear DE's

Linear differential equations are often fruitfully studied via techniques from linear functional analysis, including the theory of Banach and Hilbert spaces. In contrast, the proper setting for an important class of nonlinear partial differential equations is a nonlinear version of functional analysis, which is based upon infinite-dimensional manifolds modeled on Banach and Hilbert spaces. The theory of such manifolds was developed by Eells, Palais and Smale among others in the 1950's and 1960's, and has proven to be extremely useful for understanding many of the nonlinear differential equations which arise in geometry, such as

1. the equation for geodesics in a Riemannian manifold,
2. the equation for harmonic maps from surfaces into a Riemannian manifold, or for minimal surfaces in a Riemannian manifold,
3. the equations for pseudoholomorphic curves in a symplectic manifold,
4. and the Seiberg-Witten equations.

In all of these examples, the solutions can be expressed as critical points of a real-valued function (often called the action or the energy) defined on an infinite-dimensional manifold, such as the function space manifolds described in the following pages. In favorable cases, a gradient (or pseudogradient) of the action or energy can then be used to locate critical points (solutions to the differential equations) via what is often called the "method of steepest descent."

For this procedure to work, the topology on the function space must be strong enough for the action or energy to be differentiable, yet weak enough to force convergence of a sequence which is tending toward a minimum (or to a minimax solution for a given constraint). The two conflicting conditions often

select a unique acceptable topology for the space of functions. In the most favorable circumstances, the topology is strong enough so that it lies within the space of continuous functions, a space which has been studied extensively by topologists. The function space is then often homotopy equivalent, that is equivalent in the sense of homotopy theory, to the space of continuous functions. This makes some existence questions within the theory of nonlinear differential equations accessible by topological methods.

In the case of the Seiberg-Witten equations, the logic is reversed. Instead of topology shedding light on existence of solutions to partial differential equations, it is the space of solutions to the Seiberg-Witten equations that enable one to distinguish between different smooth structures on smooth four-manifolds (as explained in [54]). This illustrates that at a very fundamental level, topology and nonlinear partial differential equations are closely related, and underscores the importance of developing a global theory of partial differential equations based upon the theory of infinite-dimensional manifolds.

1.2 Infinite-dimensional calculus

Our first topic is the theory of infinite-dimensional manifolds. We refer to the excellent presentations of Lang [43] or of Abraham, Marsden and Ratiu [2] for further elaboration of topics only briefly introduced in the following pages.

It was pointed out by Smale, Abraham, Lang and others in the 1960's that several variable calculus could be developed not just in finite-dimensional Euclidean spaces, but also with very little extra work within the context of infinite-dimensional Hilbert and Banach spaces. Most of the proofs of theorems are straightforward modifications of the proofs in $\mathbb{R}^n$, so we will go very rapidly over this basic material.

Definition. A *pre-Hilbert space* is a real vector space E together with a function $\langle \cdot, \cdot \rangle : E \times E \to \mathbb{R}$ which satisfies the following axioms:

1. $\langle x, y \rangle = \langle y, x \rangle$, for $x, y \in E$.
2. $\langle ax, y \rangle = a\langle x, y \rangle$, for $a \in \mathbb{R}$ and $x, y \in E$.
3. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$, for $x, y, z \in E$.
4. $\langle x, x \rangle \geq 0$, for $x \in E$, equality holding only when $x = 0$.

These axioms simply state that $\langle x, y \rangle$ is a positive-definite symmetric bilinear form on E.

Given such a pre-Hilbert space, we can define a map $\| \cdot \| : E \to \mathbb{R}$ by $\|x\| = \sqrt{\langle x, x \rangle}$. We can regard E as a metric space by defining the distance between elements x and y of E to be $d(x, y) = \|x - y\|$.

Definition. A *Hilbert space* is a pre-Hilbert space which is complete in terms of the metric d.

An example of a finite-dimensional Hilbert spaces is $\mathbb{R}^n$ with inner product $\langle \cdot, \cdot \rangle$ defined by

$$\langle (x_1, \ldots, x_n), (y_1, \ldots y_n) \rangle = x_1 y_1 + \cdots + x_n y_n.$$

An example of an infinite dimensional Hilbert space is the space $L^2([0,1], \mathbb{R})$ studied in basic analysis courses. To construct it, one starts with defining an inner product $\langle \cdot, \cdot \rangle$ on the space

$$C^\infty([0,1], \mathbb{R}) = \{ \ C^\infty \text{ functions } f : [0,1] \to \mathbb{R} \ \}$$

by

$$\langle \phi, \psi \rangle = \int_0^1 \phi(t)\psi(t)dt.$$

This inner product satisfies the first four of the above axioms but not the last one. We let $L^2([0,1], \mathbb{R})$ denote the equivalence classes of Cauchy sequences from $C^\infty([0,1], \mathbb{R})$, two Cauchy sequences $\{\phi_i\}$ and $\{\psi_i\}$ being equivalent if for each $\epsilon > 0$ there is a positive integer N such that

$$i, j > N \quad \Rightarrow \quad \|\phi_i - \psi_i\| < \epsilon. \tag{1.1}$$

Equivalence classes of sequences form a vector space, and we define an inner product $\langle \cdot, \cdot \rangle$ on $L^2([0,1], \mathbb{R})$ by

$$\langle \{\phi_i\}, \{\psi_i\} \rangle = \lim_{i, i \to \infty} \langle \phi_i, \psi_i \rangle.$$

We say that $L^2([0,1], \mathbb{R})$ is the *completion* of $C^\infty([0,1], \mathbb{R})$ with respect to $\langle \cdot, \cdot \rangle$. The process we have described is virtually the same as that often used to construct the real numbers from the rationals.

Definition. A *pre-Banach space* is a vector space E together with a function $\| \cdot \| : E \to \mathbb{R}$ which satisfies the following axioms:

1. $\|ax\| = |a|\|x\|$, when $a \in \mathbb{R}$ and $x \in E$.
2. $\|x + y\| \leq \|x\| + \|y\|$, when $x, y \in E$.
3. $\|x\| \geq 0$, for $x \in E$
4. $\|x\| = 0$ only if $x = 0$.

A function $\| \cdot \| : E \to \mathbb{R}$ which satisfies the first three axioms is called a *seminorm* on E. If, in addition, it satisfies the fourth axiom it is called a *norm*.

As in the case of Hilbert spaces, we can make E into a metric space by defining the distance between elements x and y of E to be $d(x, y) = \|x - y\|$.

Definition. A *Banach space* is a pre-Banach space which is complete in terms of the metric d.

Of course, every Hilbert space is a Banach space with norm $\| \cdot \|$ defined by $\|x\| = \sqrt{\langle x, x \rangle}$. Let

$$C^0([0,1], \mathbb{R}) = \{ \text{ continuous functions } f : [0,1] \to \mathbb{R} \ \},$$

and define

$$\|\cdot\| : C^0([0,1],\mathbb{R}) \to \mathbb{R} \quad \text{by} \quad \|f\| = \sup\{|f(t)| : t \in [0,1]\}.$$

Then $\|\cdot\|$ makes $C^0([0,1],\mathbb{R})$ into a Banach space. More generally, we can consider the space

$$C^k([0,1],\mathbb{R}) = \{ \text{ functions } f : [0,1] \to \mathbb{R} \text{ which have continuous derivatives up to order } k \},$$

a Banach space with respect to the norm

$$\|\cdot\|_k : C^k([0,1],\mathbb{R}) \to \mathbb{R} \quad \text{defined by} \quad \|f\|_k = \sup\left\{\sum_{i=0}^{k} |f^{(i)}(t)| : t \in [0,1]\right\}, \tag{1.2}$$

where $f^{(i)}(t)$ denotes the derivative of f of order i.

When $1 \le p < \infty$ and $p \ne 2$, the spaces $L^p([0,1],\mathbb{R})$ studied in basic analysis courses are Banach spaces which are not Hilbert spaces. To construct these spaces, one starts with defining a function $\|\cdot\|$ on the space $C^\infty([0,1],\mathbb{R})$ of C^∞ functions $f : [0,1] \to \mathbb{R}$ by

$$\|\phi\| = \left[\int_0^1 |\phi(t)|^p dt\right]^{1/p},$$

which is shown to be a norm by means of the Minkowski inequality. This agrees with the norm previously defined on $L^2([0,1],\mathbb{R})$ when $p = 2$. Just as in the construction of $L^2([0,1],\mathbb{R})$, one can use this norm to define Cauchy sequences, and let $L^p([0,1],\mathbb{R})$ be the set of equivalence classes of Cauchy sequences, where two Cauchy sequences $\{\phi_i\}$ and $\{\psi_i\}$ are equivalent if for each $\epsilon > 0$ there is a positive integer N such that (1.1) holds. The basic properties of L^p spaces are treated in standard references on functional analysis; thus for the Hölder and Minkowski inequalities, for example, one can refer to Theorem III.1 of [65].

Each Banach space E has a metric

$$d : E \times E \to \mathbb{R} \quad \text{defined by} \quad d(e_1, e_2) = \|e_1 - e_2\|,$$

and we say that a subset U of E is *open* if

$$p \in U \quad \Rightarrow \quad B_\epsilon(p) = \{q \in E : d(p.q) < \epsilon\} \subset U,$$

for some $\epsilon > 0$. A map $T : E_1 \to E_2$ between Banach spaces is *continuous* if $T^{-1}(U)$ is open for each open $U \subset E_2$, or equivalently, if there is a constant $c > 0$ such that

$$\|T(e_1)\|_2 \le c\|e_1\|_1, \qquad \text{for all } e_1 \in E_1,$$

where $\|\cdot\|_1$ and $\|\cdot\|_2$ are the norms on E_1 and E_2 respectively.

We let $L(E_1, E_2)$ be the space of continuous linear maps $T : E_1 \to E_2$, a Banach space in its own right under the norm

$$\|T\| = \sup\{\|T(e_1)\| : e_1 \in E_1, \|e_1\| = 1\}.$$

It is easily shown that $L(E_1, E_2)$ is the same as the space of linear maps from E_1 to E_2, which are bounded in terms of this norm. In particular, we can define the *dual* of a Banach space E to be $E^\star = L(E, \mathbb{R})$. We say that a Banach space is *reflexive* if $(E^\star)^\star$ is isomorphic to E. It is proven in analysis courses that $L^p([0,1], \mathbb{R})$ is reflexive when $1 < p < \infty$ but $L^1([0,1], \mathbb{R})$ and $C^0([0,1], \mathbb{R})$ are not.

Banach spaces and continuous linear maps form a category, as do Hilbert spaces and continuous linear maps. The subject functional analysis is concerned with properties of the categories of Hilbert spaces, Banach spaces and more general spaces of functions, and is one of the major tools in studying linear partial differential equations. Key theorems from the theory of Banach spaces include the Open Mapping Theorem, the Hahn-Banach Theorem and the Uniform Boundedness Theorem. These theorems are discussed in [65], [66] and [67]; we will need to use the statements of these theorems and their consequences.

The Open Mapping Theorem states that if $T : E_1 \to E_2$ is a continuous surjective map between Banach spaces, it takes open sets to open sets. Thus if T is a continuous bijection, its inverse is continuous. The Hahn-Banach Theorem implies that if e is a nonzero element of a Banach space E, then there is a continuous linear function $\lambda : E \to \mathbb{R}$ such that $\lambda(e) \neq 0$. A bilinear map $B : E_1 \times E_2 \to F$ is said to be *continuous* if there is a constant $c > 0$ such that

$$\|B(e_1, e_2)\| \leq c\|e_1\|_1\|e_2\|_1, \qquad \text{for all } e_1 \in E_1 \text{ and all } e_2 \in E_2.$$

One of the consequences of the Uniform Boundedness Theorem is that such a Bilinear map is continuous is and only if

$$B(\cdot, e_2) : E_1 \to F \quad \text{and} \quad B(e_1, \cdot) : E_1 \to F$$

are continuous for each $e_1 \in E_1$ and each $e_2 \in E_2$. The three theorems are almost trivial to prove for finite-dimensional Banach spaces, but the proofs are more subtle for infinite-dimensional Banach spaces.

We now turn to the question of how to develop differential calculus for functions defined on Banach spaces. It is actually the topology, or the equivalence class of norms on the Banach space, that is important for calculus, two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a linear space E being equivalent if there is a constant $c > 1$ such that

$$\frac{1}{c}\|x\|_1 < \|x\|_2 < c\|x\|_1, \qquad \text{for } x \in E.$$

Lang [43] calls a vector space E a *Banachable space* if it is endowed with an equivalence class of Banach space norms. However, most other authors do not use this term, and we will simply use the simpler term Banach space, it being understood however, that we may sometimes pass to an equivalent norm in the

middle of an argument, when it is the underlying vector space with its topology, the "topological vector space"—not the norm itself—that is important.

Definition. Suppose that E_1 and E_2 are Banach spaces, and that U is an open subset of E_1. A continuous map $f : U \to E_2$ is said to be *differentiable* at the point $x_0 \in U$ if there exists a continuous linear map $T : E_1 \to E_2$ such that

$$\lim_{\|h\|\to 0} \frac{\|f(x_0 + h) - f(x_0) - T(h)\|}{\|h\|} = 0,$$

where $\|\cdot\|$ denotes both the Banach space norms on E_1 and E_2. We will call T the *derivative* of f at x_0 and write $Df(x_0)$ for T. Note that the derivative can also be calculated from the formula

$$Df(x_0)h = \lim_{t\to 0} \frac{f(x_0 + th) - f(x)}{t}.$$

Just as in ordinary calculus, the derivative $Df(x_0)$ determines the *linearization* of f near x_0, which is the affine function

$$\tilde{f}(x) = f(x_0) + Df(x_0)(x - x_0)$$

which most closely approximates f near x_0.

If f is differentiable at every $x \in U$, a derivative $Df(x)$ is defined for each $x \in U$ and thus we have a set-theoretic map $Df : U \to L(E_1, E_2)$. If this map Df is continuous, we can also ask whether it is differentiable at $x_0 \in U$. This will be the case if there is a continuous linear map $T : E_1 \to L(E_1, E_2)$ such that

$$\lim_{\|h\|\to 0} \frac{\|Df(x_0 + h) - Df(x_0) - T(h)\|}{\|h\|} = 0,$$

in which case we write $D^2 f(x_0)$ for T and call $D^2 f(x_0)$ the second derivative of f at x_0. Note that

$$\begin{aligned} D^2 f(x_0) \in L(E_1, L(E_1, E_2)) &= L^2(E_1, E_2) \\ &= \{\text{continuous bilinear maps } T : E_1 \times E_1 \to E_2\}, \end{aligned}$$

which can also be made into a Banach space in an obvious way.

We say that a function $f : U \to E_2$, where U is an open subset of E_1, is

1. C^0 if it is continuous.
2. C^1 if it is continuous and differentiable at every $x \in U$, and $Df : U \to L(E_1, E_2)$ is continuous.
3. C^k for $k \geq 2$ if it is C^1 and $Df : U \to L(E_1, E_2)$ is C^{k-1}.
4. C^∞ or *smooth* if it is C^k for every nonnegative integer k.

With the above definition of differentiation, many of the arguments in several variable calculus can be transported without difficulty to the Banach space setting, as carried out in detail in [43] or [2]. For example, the Leibniz rule for differentiating a product carries over immediately to the infinite-dimensional setting:.

Proposition 1.2.1. *Suppose that $B : F_1 \times F_2 \to G$ is a continuous bilinear map between Banach spaces, that U is an open subsets of a Banach space E and*

$$f : U_1 \longrightarrow F_1, \qquad f_2 : U_2 \longrightarrow F_2$$

are C^1 maps. Then $e \mapsto g(e) = B(f_1(e), f_2(e))$: is a C^1 map, and

$$Dg(x_0)h = B(Df_1(x_0)h, f_2(e)) + B(f_1(x_0)Df_2(x_0)h).$$

Sketch of Proof: To simplify notation, we write

$$g(x) = B(f_1(x), f_2(x)) = f_1(x) \cdot f_2(x).$$

Then

$$g(x+h) - g(x) = (f_1(x+h) - f_1(x))f_2(x+h) + f_1(x)(f_2(x+h) - f_2(x)),$$

and hence

$$\frac{g(x+h) - g(x)}{\|h\|} = \frac{f_1(x+h) - f_1(x)}{\|h\|} f_2(x+h) + f_1(x)\frac{f_2(x+h) - f_2(x)}{\|h\|}.$$

The Proposition follows by taking the limit as $\|h\| \to 0$.

The following lemma is called the chain rule.

Proposition 1.2.2. *If U_1 and U_2 are open subsets of Banach spaces E_1 and E_2 and*

$$f : U_1 \longrightarrow U_2, \qquad g : U_2 \longrightarrow E_3$$

are C^1 maps, then so is $g \circ f : U_1 \to E_3$, and

$$D(g \circ f)(x_0) = Dg(f(x_0))Df(x_0), \qquad \text{for } x_0 \in U_1.$$

Sketch of Proof: We have

$$f(x_0 + h) = f(x_0) + Df(x_0)h + o(h),$$

where the symbol $o(h)$ stands for an element in E_2 such that $o(h)/\|h\| \to 0$ as $\|h\| \to 0$. Similarly,

$$g(f(x_0) + k) = g(f(x_0)) + Dg(f(x_0))(k) + o(k).$$

Setting $k = Df(x_0)h + o(h)$ yields

$$g(f(x_0 + h)) = g(f(x_0)) + Dg(f(x_0))(Df(x_0)h + o(h)) + o(k).$$

One checks without difficulty using continuity of g that an $o(k)$ expression, where k is a bounded linear function of h, is also $o(h)$, and hence

$$g(f(x_0 + h)) = g(f(x_0)) + Dg(f(x_0))Df(x_0)h + o(h),$$

which gives the desired conclusion.

By induction, one immediately shows that the composition of C^k maps is C^k and the composition of C^∞ maps is C^∞.

Example 1.2.3. We suppose that the domain of the function is the Banach space $E = L^p(S^1, \mathbb{R}^N)$, the completion of the space $C^\infty(S^1, \mathbb{R}^N)$ of smooth $\mathbb{R}^N$-valued functions on S^1 with respect to the L^p-norm

$$\|\phi\|_{L^p} = \left[\int_{S^1} |\phi(t)|^p dt\right]^{1/p}, \qquad \text{for } \phi \in L^p(S^1, \mathbb{R}^N).$$

Here S^1 is regarded as the quotient of the interval $[0, 1]$ obtained by identifying the points 0 and 1, and possesses the standard measure dt with respect to which S^1 has measure one. A useful tool for dealing with functions in the L^p spaces is the Hölder inequality which states: if $\phi \in L^p$, $\psi \in L^q$ and

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} \text{ where } p, q, r \geq 1, \quad \text{then} \quad \phi\psi \in L^r \text{ and } \|\phi\psi\|_{L^r} \leq \|\phi\|_{L^p}\|\psi\|_{L^q}.$$

Using this inequality and the chain rule, it is not difficult to show that when $p \geq 2$, the function

$$f : E \longrightarrow \mathbb{R} \qquad \text{defined by} \qquad f(\phi) = \int_{S^1} (1 + |\phi(t)|^2)^{p/2} dt$$

is C^2 and that its first and second derivatives are given by the formulae

$$Df(\phi)(\psi) = p \int_{S^1} (1 + |\phi(t)|^2)^{p/2-1} \phi(t) \cdot \psi(t) dt$$

and

$$\begin{aligned}(D^2 f)(\phi)(\psi_1, \psi_2) = p &\int_{S^1} (1 + |\phi(t)|^2)^{p/2-1} \psi_1(t) \cdot \psi_2(t) dt \\ &+ p(p-2) \int_{S^1} (1 + |\phi(t)|^2)^{p/2-2} (\phi(t) \cdot \psi_1(t))(\phi(t) \cdot \psi_2(t)) dt.\end{aligned}$$

Indeed, to carry this out, we apply the chain rule to $f = h \circ g$, where h is integration over S^1 and

$$g : E \to L^1(S^1, \mathbb{R}) \qquad \text{by} \qquad g(\phi) = (1 + |\phi(t)|^2)^{p/2}.$$

On the other hand, if f were C^3, one can verify that g would also be C^3, with third derivative given by the formula

$$(D^3g)(\psi_1,\psi_2,\psi_3) = (3p/2)(p/2-1)(1+|\phi|^2)^{p/2-1}\phi\psi_1\psi_2\psi_3 \\ + (3p/2)(p/2-1)(p/2-2)(1+|\phi|^2)^{p/2-2}\phi^3\psi_1\psi_2\psi_3,$$

and if $p<3$, for any smooth choice of ϕ, we could choose ψ_1, ψ_2 and ψ_3 in L^p such that the product $\psi_1\psi_2\psi_3$ is not in L^1. This implies that f cannot be C^3 when $2<p<3$.

Another familiar theorem from several variable calculus in finite dimension is the "equality of mixed partials." To state the infinite-dimensional version, we let E and F be Banach spaces, U an open subset of E. Suppose that $f: U \to F$ is a C^2 map. Then for $x_0 \in U$,

$$D^2f(x_0) \in L(E,L(E,F)) = L^2(E,F) \\ = \{\text{continuous bilinear maps } T: E\times E \to F\}.$$

Of course, a very important case is the one where $F=\mathbb{R}$, the base field of real numbers.

Proposition 1.2.4. *$D^2f(x_0)$ is symmetric; in other words,*

$$D^2f(x_0)(h,k) = D^2f(x_0)(k,h), \qquad \text{for } h,k \in E.$$

Sketch of proof: First note that by the Hahn-Banach theorem, it suffices to show that if $\lambda: F\to\mathbb{R}$ is any continuous linear functional, then $D^2(\lambda\circ f)(x_0) = \lambda\circ D^2f(x_0)$ is symmetric, because if $D^2f(x_0)$ is not symmetric, the same will be true of $\lambda\circ D^2f(x_0)$, for some linear function λ. This reduces the proof to the case where $F=\mathbb{R}$. Next note that via the chain rule,

$$f(x+h)-f(x) = \int_0^1 (Df)(x+th)h\,dt,$$

and by iteration,

$$f(x+h+k)-f(x+k)-f(x+h)+f(x) = \int_0^1 (Df)(x+th+k)h\,dt \\ = \int_0^1\int_0^1 (D(Df)(x+th+sk)(k))h\,ds\,dt.$$

Interchanging h and k yields

$$f(x+h+k)-f(x+h)-f(x+k)+f(x) = \int_0^1 (Df)(x+tk+h)k\,dt \\ = \int_0^1\int_0^1 (D(Df)(x+sk+th)(h))k\,ds\,dt.$$

Since the left-hand sides of the last two expressions are equal, so are the right-hand sides, and hence

$$\int_0^1 \int_0^1 [(D(Df)(x+th+sk)(k))h - (D(Df)(x+th+sk)(h))k]dsdt = 0.$$

Since $D(Df)$ is a continuous function, this can only happen if $D^2f(x)(k,h) = D^2f(x)(h,k)$, for all x, h and k, which is exactly what we needed to prove.

More generally, if $f : U \to F$ is a C^k-map, then

$$D^k f(x_0) = D(D^{k-1}f)(x_0) \in L(E, L^{k-1}(E,F)) = L^k(E,F),$$

and by an induction based on the previous lemma, we see that in fact

$$D^k f(x_0) \in L^k_s(E,F) = \{T \in L^k(E,F) : \ T \text{ is symmetric}\}.$$

By symmetric we mean that

$$T(h_{\sigma(1)}, \ldots, h_{\sigma(k)}) = T(h_1, \ldots, h_k),$$

for all permutations σ in the symmetric group S_k on k letters.

It is often useful to have an explicit formula for the higher derivatives of a composition. The following such formula comes from §3 of [1]. If p and k are positive integers with $k \leq p$ and $(i_1, \ldots, i_k)$ is a k-tuple of positive integers such that $i_1 + \cdots + i_k = p$ and $i_1 \leq i_2 \leq \cdots i_k$, we define integers $\sigma^p_k(i_1, \ldots, i_k)$ recursively by requiring that $\sigma^1_1(1) = 1$ and

$$\sigma^p_k(i_1, \ldots, i_k) = \delta^1_{i_1}\sigma^{p-1}_{k-1}(i_2, \ldots, i_k) + \sum_{j=1}^{k} \sigma^{p-1}_k(i_1, \ldots, i_j + 1, \ldots i_k),$$

where $\delta^1_{i_1}$ is 1 if $i_1 = 1$ and 0 otherwise.

Proposition 1.2.5. *Suppose that U and V are open subsets of Banach spaces E and F respectively, and that $f : U \to V$ and $g : V \to G$ are C^p maps, where G is a third Banach space. Then*

$$D^p(g \circ f) = \sum_{k=1}^{p} (D^k g \circ f) P^p_k(f),$$

where

$$P^p_k(f) = \sum \sigma^p_k(i_1, \ldots, i_k)(D^{i_1}f, \ldots, D^{i_k}f),$$

the sum being taken over all p-tuples of positive integers such that $i_1 + \cdots + i_k = p$ and $i_1 \leq i_2 \leq \cdots \leq i_k$.

The proof of Proposition 1.2.5 is by induction on p starting with the case $p = 1$, which is an immediate application of the chain rule. For $p \geq 2$, one obtains the

formula for σ_k^p and the expression for $D^p(g \circ f)$ by applying the chain rule and the Leibniz rule for differentiating a product.

In order to put Taylor's theorem in the Banach space setting, we need to define the integral of a continuous map $\gamma : [0,1] \to E$ into a Banach space E. The definition is particularly easy if E is a reflexive Banach space; in this case, we just set

$$\int_0^1 \gamma(t)dt = e, \quad \text{where} \quad \lambda(e) = \int_0^1 \lambda \circ \gamma(t)dt.$$

A definition of the integral for general Banach spaces can be found in [43].

Let U be an open subset of a Banach space E. Following [2], we define a *thickening* of U to be an open subset $\widetilde{U} \subset E \times E$ such that

1. $U \times \{0\} \subset \widetilde{U}$.
2. $(x,h) \in \widetilde{U} \Rightarrow x + th \in U$, for $t \in [0,1]$.
3. $(x,h) \in \widetilde{U} \Rightarrow x \in U$.

Proposition 1.2.6. (Taylor's Theorem.) *If a map $f : U \to F$ is of class C^r, there exist continuous maps*

$$\phi_k : U \to L_s^k(E,F), \quad \text{for } 1 \le k \le r \quad \text{and} \quad R : \widetilde{U} \to L_s^r(E,F),$$

where $\tilde{U}$ is a thickening of U, such that $R(x,0) = 0$ and

$$f(x+h) = f(x) + \phi_1(x)h + \frac{1}{2!}\phi_2(x)(h,h) + \cdots + \frac{1}{r!}\phi_r(x)(h,\cdots,h) + R(x,h)(h,\cdots,h).$$

Here $\phi_k(x) = (D^k f)(x)$, for $1 \le k \le r$.

To see this, we first reduce to the case where $F = \mathbb{R}$ by applying the Hahn-Banach Theorem, and then establish via induction,

$$\begin{aligned} f(x+h) &= f(x) + \int_0^1 (Df)(x+th)hdt \\ &= f(x) + (Df)(x)h + \int_0^1 [(Df)(x+th) - (Df)(x)]hdt \\ &= f(x) + (Df)(x)h + \int_0^1 \int_0^1 [(D^2 f)(x+sth)](h,h)tdsdt \\ &= f(x) + (Df)(x)h + \frac{1}{2!}(D^2 f)(x)(h,h) \\ &\quad + \int_0^1 \int_0^1 [(D^2 f)(x+sth) - D^2 f(x)](h,h)tdsdt \end{aligned}$$

Continuing in the same manner, we find that

$$f(x+h) = f(x) + (Df)(x)h + \frac{1}{2!}(D^2f)(x)(h,h) \\ + \cdots + \frac{1}{k!}(D^kf)(x)(h,\cdots,h) + R(x,h)(h,\cdots,h),$$

where $R(x,h) \in L^k_s(E,F)$ depends continuously on x and h and $R(x,0) = 0$.

We have seen that many of the basic results of differential calculus of several variables extend with little change to the infinite-dimensional context. The following theorem is somewhat deeper:

Theorem 1.2.7. (Inverse Function Theorem.) *If U_1 and U_2 are open subsets of Banach spaces E_1 and E_2 with $x_0 \in U_1$, and $f : U_1 \to U_2$ is a C^∞ map such that $Df(x_0) \in L(E_1, E_2)$ is invertible, then there are open neighborhoods V_1 of x_0 and V_2 of $f(x_0)$, and a C^∞ map $g : V_2 \to V_1$, such that*

$$f \circ g = id_{V_2} \qquad \text{and} \qquad g \circ f = id_{V_1}.$$

Moreover, $Dg(f(x)) = [Df(x)]^{-1}$, for $x \in V_1$.

Sketch of proof: We can assume without loss of generality that $x_0 = 0 \in U_1$ and $f(0) = 0 \in U_2$. We can assume, moreover, that $E_1 = E_2$ and $Df(0)$ is the identity map by replacing f by $Df(0)^{-1} \circ f$. Define $h : U_1 \to E_1$ by $h(x) = x - f(x)$. Then $Dh(0) = 0$, and by continuity of Dh there exists $\delta > 0$ such that

$$\|x\| \le \delta \quad \Rightarrow \quad x \in U_1 \quad \text{and} \quad \|dh(x)\| < \frac{1}{2}.$$

If $\|x\|, \|y\| < \delta$, then it follows from the chain rule that

$$\|h(x) - h(y)\| = \left\| \int_0^1 \frac{d}{dt}(h(tx + (1-t)y))dt \right\| \\ = \left\| \int_0^1 Dh(tx + (1-t)y)(x-y)dt \right\| \\ \le \left[\int_0^1 \|Dh(tx + (1-t)y)\|dt \right] \|x - y\| < \frac{1}{2}\|x - y\|.$$

More generally, if $\|y\| < \delta/2$, and we define the map h_y by $h_y(x) = h(x) + y$, then

$$\|x\| \le \delta \quad \Rightarrow \quad \|h_y(x)\| \le \|y\| + \|h(x)\| < \frac{\delta}{2} + \frac{1}{2}\|x\| \le \delta,$$

so h_y takes the closed ball of radius δ to itself and

$$\|h_y(x) - h_y(x')\| = \|h(x) - h(x')\| < \frac{1}{2}\|x - x'\|,$$

so h_y is a contraction. Thus by the well-known Contraction Lemma, given y with $\|y\| < \delta/2$, there is a unique fixed point x of h_y; that is, there is a unique x such that $\|x\| \le \delta$ and

$$h_y(x) = x \quad \Rightarrow \quad x - f(x) + y = x \quad \Rightarrow \quad f(p) = q.$$

Let

$$V_2 = \{y \in E_1 : \|y\| < \delta/2\} \quad \text{and} \quad V_1 = \{x \in E_1 : \|x\| < \delta, f(x) \in V_2\}$$

and define

$$g : V_2 \to V_1 \qquad \text{by} \qquad g(y) = x \in V_1 \quad \Leftrightarrow \quad f(x) = y.$$

Then g is a set-theoretic inverse to $f : V_1 \to V_2$.

To show that g is continuous, it suffices to show that $|x-x'| \le 2|f(x)-f(x')|$. But

$$\begin{aligned} |x - x'| &\le |(x - f(x)) - (x' - f(x'))| + |f(x) - f(x')| \\ &\le |h(x) - h(x')| + |f(x) - f(x')| \le \frac{1}{2}|x - x'| + |f(x) - f(x')|, \end{aligned}$$

which clearly implies the desired result.

To see that that g is C^1, we note first that if $x_0 \in V_1$,

$$f(x) - f(x_0) = Df(x_0)(x - x_0) + o(|x - x_0|).$$

If $y_0 = f(x_0)$ and $y = f(x)$, we can rewrite this equation as

$$\begin{aligned} y - y_0 &= Df(x_0)(g(y) - g(y_0)) + o(|x - x_0|), \\ &\text{or} \quad g(y) - g(y_0) = [Df(x_0)]^{-1}(y - y_0 + o(|x - x_0|)). \end{aligned}$$

The continuity argument shows that $[Df(x_0)]^{-1}(o(|x - x_0|))$ is $o(|y - y_0|)$, so g is C^1 with derivative $Dg(y) = (Df)^{-1}(g(y))$.

Finally, one uses "bootstrapping":

$$g \in C^1 \;\Rightarrow\; (Df)^{-1} \circ g \in C^1 \;\Rightarrow\; Dg \in C^1 \;\Rightarrow\; g \in C^2 \;\Rightarrow\; \cdots$$

to conclude that g is C^∞, and the theorem is proven. Later, we will see that the technique of bootstrapping used here is a fundamental tool for the study of nonlinear PDE's.

Before stating an important corollary of the Inverse Function Theorem, we point out one of the difficulties in dealing with Banach spaces. A closed linear subspace E_1 of a Banach space E is a Banach space in its own right, but there may not exist a complementary closed subspace E_2 such that E is linearly homeomorphic to $E_1 \oplus E_2$. We say that a subspace E_1 of a Banach space E is *split* if there does exist such a complement E_2. For example, any finite-dimensional subspace of a

Banach space is split as is any closed subspace of finite codimension. Moreover, any closed subspace of a Hilbert space is split, because the inner product can be used to define the orthogonal complement.

Corollary 1.2.8. *If U is an open subset of the Banach space E with $x \in U$, and $f : U \to F$ is a C^∞ map such that $Df(x) \in L(E, F)$ is surjective with split kernel, then there exists an open subset $V \subset U$ and a diffeomorphism $\phi : V_1 \times V_2 \to V$, where V_1 and V_2 are open subsets of Banach spaces E_1 and E_2 with $E = E_1 \oplus E_2$ and $E_2 \cong F$, such that $f \circ \phi$ is the projection on the second factor.*

Sketch of proof: We can assume without loss of generality that $x = 0$ and $f(0) = 0$. Let E_1 be the kernel of $Df(0)$ and since it splits, let E_2 be a complement such that $E = E_1 \oplus E_2$. Note that $Df(p)$ establishes an isomorphism from E_2 to F. Let

$$g : E = E_1 \oplus E_2 \to E = E_1 \oplus E_2 \qquad \text{by} \qquad g(x_1, x_2) = (x_1, f(x_1, x_2)).$$

One easily checks that $Dg(0)$ is invertible. Now apply the inverse function theorem to construct a smooth map

$$\phi : V_1 \times V_2 \to V \subset U \qquad \text{such that} \qquad g \circ \phi = \mathrm{id},$$

the identity map. The projection on E_2 is just $f \circ \phi$.

Remark. Notably absent from our examples of Banach spaces is the space

$$C^\infty([0,1], \mathbb{R}) = \{ \text{ functions } f : [0,1] \to \mathbb{R} \text{ which have continuous derivatives of all orders } \},$$

which one suspects should have importance for the theory of nonlinear partial differential equations. Unfortunately, the natural topology to use on this space is not defined by a single norm or seminorm, but by a countable collection of norms $\{\| \cdot \|_k : k \in \mathbb{Z}, k \geq 0\}$, where $\| \cdot \|_k$ is defined by (1.2).

Definition. A *Fréchet space* is a vector space E together with a countable collection of seminorms $\{\| \cdot \|_k : k \in \mathbb{Z}, k \geq 0\}$ satisfying the following axioms:

1. $\|x\|_k = 0$ for all k only if $x = 0$.
2. Suppose that $\{x_i\}$ is a sequence of elements from E. If for each $\epsilon > 0$, there is a N such that $i, j \geq N$ implies that $\|x_i - x_j\|_k < \epsilon$ for all k, then there is an element $x \in E$ such that $\|x_i - x\|_k$ converges to zero for all k.

Of course every Banach space is a Fréchet space, but $C^\infty([0,1], \mathbb{R})$ with the collection of norms defined above is a Fréchet space which is not a Banach space. Given a Fréchet space E with seminorms $\{\| \cdot \|_k : k \in \mathbb{Z}, k \geq 0\}$, we can define a distance function

$$d : E \times E \to \mathbb{R} \quad \text{by} \quad d(x, y) = \sum_{k=0}^{\infty} \frac{1}{2^k} \|x - y\|_k,$$

which makes E into a metric space. Thus we can talk of continuous maps $f : E_1 \to E_2$ from the Fréchet space E_1 to the Fréchet space E_2, and we have a category consisting of Fréchet spaces and continuous linear maps.

Definition. Suppose that E_1 and E_2 are Fréchet spaces, and that U is an open subset of E_1. A continuous map $f : U \to E_2$ is said to be *continuously differentiable* on U if there exists a continuous map $Df : U \times E_1 \to E_2$ such that

$$Df(x)y = \lim_{t \to 0} \frac{f(x+ty) - f(x)}{t},$$

where t ranges throughout $\mathbb{R} - \{0\}$, it being understood that the limit on the right-hand side exists for all $x \in U$ and all $y \in E_1$.

It is proven in Hamilton's survey article [32], Part I, 3.2.3 and 3.2.5 that if $f : U \to E_2$ is continuously differentiable, the map $y \mapsto Df(x)y$ is linear. Thus if E_1 and E_2 are Banach spaces the above definition agrees with the definition previously given.

We could develop much of the infinite-dimensional calculus and the theory of infinite-dimensional manifolds in the category of Fréchet spaces, and in fact this is carried out in [32], but the Inverse Function Theorem does not hold for Fréchet spaces. Moreover, in the existence theory for solutions to nonlinear partial differential equations, it is often convenient to first prove existence in a given Banach space and then prove regularity using bootstrapping, just as we did in the proof of the Inverse Function Theorem. This technique seems particularly well-adapted to Banach spaces. For these reasons, we prefer to think of $C^\infty([0,1], \mathbb{R})$ as the intersection of a "chain" of Banach spaces,

$$\cdots \subseteq C^{k+1}([0,1], \mathbb{R}) \subseteq C^k([0,1], \mathbb{R}) \subseteq C^{k-1}([0,1], \mathbb{R}) \subseteq \cdots \subseteq C^0([0,1], \mathbb{R}).$$

For proving theorems, we will usually work in the category of Banach spaces so that we can use theorems like the Inverse Function Theorem. However, the statements of theorems are sometimes more elegant when phrased in the category of Fréchet spaces.

1.3 Infinite-dimensional manifolds

Definition. Let E be a Banach space. A connected *smooth manifold* modeled on E is a connected Hausdorff space $\mathcal{M}$ together with a collection $\mathcal{A} = \{(U_\alpha, \phi_\alpha) : \alpha \in A\}$, where each U_α is an open subset of $\mathcal{M}$ and each

$$\phi_\alpha : U_\alpha \longrightarrow \phi_\alpha(U_\alpha) \subset E$$

is a homeomorphism such that

1. $\bigcup\{U_\alpha : \alpha \in A\} = \mathcal{M}$.
2. $\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$ is C^∞, for all $\alpha, \beta \in A$.

A nonconnected Hausdorff space is called a *smooth manifold* or a *Banach manifold* if each component is a connected smooth manifold modeled on some Banach space. (There is no harm in allowing different components to be modeled on different Banach spaces.) The smooth manifold is called a *Hilbert manifold* if each component is modeled on a Hilbert space.

We say that $\mathcal{A} = \{(U_\alpha, \phi_\alpha) : \alpha \in A\}$ is the *atlas* defining the smooth structure on $\mathcal{M}$, and each (U_α, ϕ_α) is one of the *charts* in the atlas.

Remark. We could define *Fréchet manifolds* by simply replacing the phrase "Banach space" by "Fréchet space" in the above definition.

Let $\mathcal{M}$ and $\mathcal{N}$ be smooth manifolds modeled on Banach spaces E and F respectively. Suppose that $\mathcal{M}$ and $\mathcal{N}$ have atlases $\mathcal{A} = \{(U_\alpha, \phi_\alpha) : \alpha \in A\}$ and $\mathcal{B} = \{(V_\beta, \psi_\beta) : \beta \in B\}$. Then a continuous map $F : \mathcal{M} \to \mathcal{N}$ is said to be *smooth* if $\psi_\beta \circ F \circ \phi_\alpha^{-1}$ is C^∞, where defined, for $\alpha \in A$ and $\beta \in B$. It follows from the chain rule that the composition of smooth maps is smooth. In this way we obtain a category whose objects are smooth manifolds modeled on Banach spaces and whose morphisms are smooth maps between such manifolds.

As in the case of finite-dimensional manifolds, a *diffeomorphism* is a smooth map between manifolds with smooth inverse. We will often identify two smooth manifolds if there is a diffeomorphism from one to the other. Later we will construct invariants (such as de Rham cohomology) that will often enable us to determine that two infinite-dimensional manifolds cannot be diffeomorphic.

Of course, the simplest example of a smooth manifold modeled on E is an open subset U of E in which the atlas is $\{(U, \mathrm{id}_U)\}$. However, the examples of most interest to us will be function spaces.

Example. Suppose that M^n is a complete Riemannian manifold of finite dimension n, which we can assume is isometrically imbedded in Euclidean space $\mathbb{R}^N$ by the celebrated Nash imbedding theorem [57]. Suppose that Σ is a compact smooth manifold of finite dimension m. Then

$$C^0(\Sigma, \mathbb{R}^N) = \{\text{continuous maps } f : \Sigma \to \mathbb{R}^N\}$$

is a Banach space, and we claim that the subspace

$$C^0(\Sigma, M) = \{\text{continuous maps } f : \Sigma \to M\} \subseteq C^0(\Sigma, \mathbb{R}^N)$$

is an infinite-dimensional Banach manifold.

To construct the charts on $C^0(\Sigma, M)$ we use the exponential map of M. Suppose that $f : \Sigma \to M$ is a smooth (C^∞) map, and consider the Banach space

$$C^0(f^*TM) = \{\text{continuous sections of } f^*TM\},$$

with the C^0-norm

$$\|X\| = \sup\{|X(p)| : p \in \Sigma\},$$

where the $|\cdot|$ on the right is length as defined by the Riemannian metric on M. For $\epsilon > 0$, we set

$$V_{f,\epsilon} = \{X \in C^0(f^*TM) : \|X\| < \epsilon\},$$

and we define

$$U_{f,\epsilon} = \{g \in C^0(\Sigma, M) : d_M(g(p), f(p)) < \epsilon \text{ for all } p \in \Sigma\},$$

where $d_M : M \times M \to \mathbb{R}$ is the distance function on M defined by the Riemannian metric. Then we define $\psi_{f,\epsilon} : V_{f,\epsilon} \to U_{f,\epsilon}$ by

$$\psi_{f,\epsilon}(X)(p) = \exp_{f(p)}(X(p)).$$

By the the proof of the standard tubular neighborhood theorem from finite-dimensional Riemannian geometry, one concludes that when $\epsilon > 0$ is sufficiently small, $\psi_{f,\epsilon}$ is a bijection, and we set $\phi_{f,\epsilon} = \psi_{f,\epsilon}^{-1}$. We set

$$\mathcal{A} = \{(U_{f,\epsilon}, \phi_{f,\epsilon}) : \ f : \Sigma \to M \text{ is a } C^\infty \text{ map and } \epsilon > 0 \text{ is small enough that } \psi_{f,\epsilon} \text{ is a homeomorphism } \}.$$

Since smooth maps $f : \Sigma \to M$ are dense in $C^0(\Sigma, M)$, the union of the elements of $\mathcal{A}$ cover $C^0(\Sigma, M)$.

Hence to verify that $C^0(\Sigma, M)$ is a smooth manifold, we need only check that $\phi_{f_2,\epsilon_2} \circ (\phi_{f_1,\epsilon_1})^{-1}$ is smooth where defined, when

$$(U_{f_1,\epsilon_1}, \phi_{f_1,\epsilon_1}), (U_{f_2,\epsilon_2}, \phi_{f_2,\epsilon_2}) \in \mathcal{A}.$$

Let U be the open subset of the total space of f_1^*TM on which $(\exp_{f_2(p)})^{-1} \circ \exp_{f_1(p)}$ is defined. We can assume that $U \cap T_{f_1(p)}M$ is convex with compact closure for each $p \in \Sigma$, and define

$$g : U \to (\text{total space of } f_2^*TM)$$

by

$$g(p, v) = (p, (\exp_{f_2(p)})^{-1} \circ \exp_{f_1(p)}(v)), \quad \text{for } p \in \Sigma,\ v \in T_{f_1(p)}M.$$

We can think of g as a family of smooth maps

$$p \mapsto \tilde{g}_p : (T_{f_1(p)}M \cap U) \to T_{f_2(p)}M,$$

and since Σ is compact and U has compact closure in each fiber, all derivatives $D^k(\tilde{g}_p)$ are bounded. Using the open neighborhood U of the zero-section, we define

$$\tilde{U} = \{X \in C^0(f_1^*TM) : X(\Sigma) \subseteq U\}$$

and define $\omega_g : \tilde{U} \to C^0(f_2^*TM)$ by $\omega_g(X) = g \circ X$.

To finish the proof that $C^0(\Sigma, M)$ is a smooth manifold, it suffices to show that ω_g is smooth. We can formalize this statement in a theorem, known as the ω-lemma:

Lemma 1.3.1. *Suppose that E_1 and E_2 are finite-dimensional vector bundles over the compact smooth manifold Σ and that U is a bounded open neighborhood of the zero section of E_1 whose restriction to each fiber of E_1 is convex.*

If $g : U \to$ (total space of E_2) is a smooth map which takes the fiber of E_1 over p to the fiber of E_2 over p (for each $p \in \Sigma$), and

$$\tilde{U} = \{\sigma \in C^0(E_1) : \sigma(\Sigma) \subseteq U\},$$

then the map $\omega_g : \tilde{U} \to C^0(E_2)$, defined by $\omega_g(\sigma) = g(\sigma)$, is smooth.

The first step is to show that ω_g is continuous; this is straightforward and we can safely leave it to the reader.

Suppose now that $\sigma, \eta, \sigma + \eta \in \tilde{U}$. It follows from Taylor's theorem that for each $p \in \Sigma$,

$$\tilde{g}_p((\sigma + \eta)(p)) = \tilde{g}_p(\sigma(p)) + D\tilde{g}_p(\sigma(p))\eta(p) + R(\sigma, \eta)(p)(\eta(p)),$$

where

$$R(\sigma, \eta)(p)(\eta(p)) = \left[\int_0^1 [D\tilde{g}_p((\sigma + t\eta)(p)) - D\tilde{g}_p(\sigma(p))]dt\right]\eta(p). \tag{1.3}$$

We can write this as

$$\omega_g(\sigma + \eta) = \omega_g(\sigma) + T(\sigma)(\eta) + R(\sigma, \eta)\eta,$$

where

$$T(\sigma)(\eta)(p) = D\tilde{g}_p(\sigma(p))\eta(p),$$

and $R(\sigma, \eta)\eta$ is the remainder term given by (1.3). Note that

$$\|T(\sigma)(\eta)\| = \sup_{p\in\Sigma} |D\tilde{g}_p(\sigma(p))\eta(p)| \leq \sup_{p\in\Sigma} |D\tilde{g}_p(\sigma(p))|\|\eta\|,$$

so $T(\sigma)$ is a continuous linear map from a neighborhood of 0 in $C^0(E_1)$ to $C^0(E_2)$. Thus T extends to a linear map from $C^0(E_1)$ to $C^0(E_2)$. We next estimate the error term $R(\sigma, \eta)\eta$; since

$$\left|\int_0^1 [D\tilde{g}_p((\sigma + t\eta)(p)) - D\tilde{g}_p(\sigma(p))]dt\right| \leq |D^2\tilde{g}_p(\sigma(p))||\eta(p)|,$$

we conclude that

$$\|R(\sigma, \eta)\eta\| \leq \sup_{p\in\Sigma} |D^2\tilde{g}_p(\sigma(p))\|\eta\|^2 = o(\eta).$$

This implies that ω_g has the derivative $D\omega_g(\sigma) = T(\sigma)$ at σ. We have defined a map

$$D\omega_g : \tilde{U} \longrightarrow L(C^0(E_1), C^0(E_2)), \tag{1.4}$$

and it is relatively straightforward to show that $D\omega_g$ is continuous, showing that ω_g is C^1.

We would like to extend this argument to higher derivatives, and for this we need to factor the derivative given by (1.4) as follows: Recalling that $g : U \to$ (total space of E_2), we define a corresponding map

$$\widetilde{Dg} : U \to (\text{total space of } L(E_1, E_2)) \quad \text{by setting} \quad \widetilde{Dg}(p) = D\tilde{g}_p.$$

(We can regard $\widetilde{Dg}$ as a "partial derivative" of g in which the point $p \in \Sigma$ is held fixed.) We can then define

$$\omega_{\widetilde{Dg}} : \tilde{U} \to C^0(\Sigma, L(E_1, E_2)) \quad \text{by} \quad \omega_{\widetilde{Dg}}(\sigma)(p) = D\tilde{g}_p(\sigma(p)).$$

Using the fact that $C^0(\Sigma, \mathbb{R})$ is a Banach algebra, we can show that the map

$$A : C^0(\Sigma, L(E_1, E_2)) \to L(C^0(\Sigma, E_1), C^0(\Sigma, E_2)) \\ \text{defined by} \quad A(T)(\sigma)(p) = T(p) \cdot \sigma(p),$$

is smooth, thus providing the desired factorization,

$$D\omega_g = A \cdot \omega_{\widetilde{Dg}}.$$

The argument presented in the preceding paragraph now shows that $\omega_{\widetilde{Dg}}$ is C^1, from which we conclude that $D\omega_g$ is C^1, and hence ω_g is C^2.

Next, we show that $\omega_{\widetilde{D^2 g}}$ is C^1, which implies that ω_g is C^3, and so forth. By induction, we establish that ω_g is C^k for all $k \in \mathbb{N}$, and hence ω_g is C^∞.

The above lemma has the following consequence:

Theorem 1.3.2. *If Σ and M are smooth manifolds with Σ compact, then $C^0(\Sigma, M)$ is a smooth manifold modeled on the Banach spaces*

$$C^0(f^*TM) = \{\textit{continuous sections of } f^*TM\},$$

where $f : \Sigma \to M$ is a smooth map. Moreover, if $g : M \to N$ is a C^∞ map, then the map

$$\omega_g : C^0(\Sigma, M) \to C^0(\Sigma, N) \quad \textit{defined by} \quad \omega_g(f) = g \circ f,$$

is smooth.

To prove the moreover part of the theorem, we need to show that if $f_1 : \Sigma \to M$ and $f_2 : \Sigma \to N$ are smooth, then

$$\phi_{f_2,\epsilon_2} \circ \omega_g \circ \phi^{-1}_{f_1,\epsilon_1} \quad \text{is smooth where defined.}$$

The proof of this is a straightforward application of Lemma 1.2.1.

A modification of the previous example is often quite useful. Let Σ be a compact smooth manifold of finite dimension m with boundary $\partial\Sigma$, M^n a complete

Riemannian manifold of finite dimension n and $f_0 : \partial\Sigma \to M$ a fixed smooth mapping. We claim that

$$C_0^0(\Sigma, M) = \{\text{continuous maps } f : \Sigma \to M : f|\partial\Sigma = f_0\}.$$

is a Banach manifold. The component containing f is modeled on the Banach space

$$C_0^0(f^*TM) = \{\text{continuous sections of } f^*TM \text{ which vanish on the boundary of } M\}.$$

The proof is a straightforward modification of the proof of Theorem 1.2.2.

Unfortunately, the manifolds $C^0(\Sigma, M)$ are not sufficient for constructing a global theory of partial differential equations. We need to be able to differentiate elements in our function spaces. Thus we need to start off with a somewhat stronger Banach space than the space $C^0(\Sigma, \mathbb{R})$ of continuous real-valued functions on Σ.

Thus for $k \in \mathbb{N}$, we are led to consider the space $C^k(\Sigma, \mathbb{R})$ of real-valued functions on Σ which have continuous derivatives up to order k, a Banach space with respect to the norm

$$\|f\|_{C^k} = \sup\{\|f\|(p) + \|Df\|(p) + \cdots + \|D^k f\|(p) : p \in \Sigma\}. \tag{1.5}$$

In fact, it is easily checked that if $f, g \in C^k(\Sigma, \mathbb{R})$, then

$$\|fg\|_{C^k} \leq \|f\|_{C^k}\|g\|_{C^k}, \tag{1.6}$$

so $C^k(\Sigma, \mathbb{R})$ is in fact a Banach algebra.

More generally, we can consider the space $C^k(\Sigma, \mathbb{R}^N)$ of $\mathbb{R}^N$-valued functions on Σ which have continuous derivatives up to order k, which is also a Banach space with norm defined by (1.5). The Banach algebra condition (1.6) ensures that we can define a continuous multiplication

$$\mu : C^k(\Sigma, L(\mathbb{R}^N, \mathbb{R}^M)) \times C^k(\Sigma, \mathbb{R}^N) \longrightarrow C^k(\Sigma, \mathbb{R}^M)$$

by simply multiplying in the range.

If M is an n-dimensional Riemannian manifold which isometrically imbedded in $\mathbb{R}^n$, we let

$$C^k(\Sigma, M) = \{f \in C^k(\Sigma, \mathbb{R}^N) : f(p) \in M \text{ for all } p \in \Sigma\ \}.$$

We claim that $C^k(\Sigma, M)$ is a smooth manifold.

The construction of the atlas is just like the construction for $C^0(\Sigma, M)$. For $\epsilon > 0$, we set

$$V_{f,\epsilon} = \{X \in C^k(f^*TM) : \|X\|_{C^0} < \epsilon\},$$

an open subset of $C^k(f^*TM)$, and we define

$$U_{f,\epsilon} = \{g \in C^0(\Sigma, M) : d_M(g(p), f(p)) < \epsilon \text{ for all } p \in \Sigma\}.$$

As before, we define $\psi_{f,\epsilon} : V_{f,\epsilon} \to U_{f,\epsilon}$ by

$$\psi_{f,\epsilon}(X)(p) = \exp_{f(p)}(X(p)).$$

When $\epsilon > 0$ is sufficiently small, $\psi_{f,\epsilon}$ is a bijection, and we set $\phi_{f,\epsilon} = \psi_{f,\epsilon}^{-1}$. As smooth atlas for $C^k(\Sigma, M)$, we take

$$\mathcal{A} = \{(U_{f,\epsilon}, \phi_{f,\epsilon}) : \ f : \Sigma \to M \text{ is a } C^\infty \text{ map and } \epsilon > 0 \text{ is small enough that } \psi_{f,\epsilon} \text{ is a homeomorphism } \}.$$

Just as before, since smooth maps $f : \Sigma \to M$ are dense in $C^k(\Sigma, M)$, the union of the elements of $\mathcal{A}$ cover $C^k(\Sigma, M)$. To finish the proof that $C^k(\Sigma, M)$ is an infinite-dimensional smooth manifold, we need to verify once again that $\phi_{f_2,\epsilon_2} \circ (\phi_{f_1,\epsilon_1})^{-1}$ is smooth where defined, when

$$(U_{f_1,\epsilon_1}, \phi_{f_1,\epsilon_1}), (U_{f_2,\epsilon_2}, \phi_{f_2,\epsilon_2}) \in \mathcal{A}.$$

But this follows from a corresponding ω-lemma:

Lemma 1.3.3. *Suppose that E_1 and E_2 are finite-dimensional vector bundles over the compact smooth manifold Σ and that U is a bounded open neighborhood of the zero section of E_1 whose restriction to each fiber of E_1 is convex. If $g : U \to$ (total space of E_2) is a smooth map which takes the fiber of E_1 over p to the fiber of E_2 over p (for each $p \in \Sigma$), and*

$$\tilde{U} = \{\sigma \in C^k(E_1) : \sigma(\Sigma) \subseteq U\},$$

then the map $\omega_g : \tilde{U} \to C^k(E_2)$, defined by $\omega_g(\sigma) = g(\sigma)$, is smooth.

The proof is virtually identical to the proof given for Lemma 1.2.1. The proof extends to C^k maps because the space $C^k(\Sigma, \mathbb{R})$ has two key properties:

1. It is a Banach algebra, and
2. there is be a continuous inclusion from $C^k(\Sigma, \mathbb{R})$ into the Banach algebra $C^0(\Sigma, \mathbb{R})$ of continuous functions.

Thus just as before, we can construct an important family of Banach manifolds:

Theorem 1.3.4. *If Σ and M are smooth manifolds with Σ compact, then for each $k \in \mathbb{N}$, $C^k(\Sigma, M)$ is a smooth manifold modeled on the Banach spaces*

$$C^k(f^*TM) = \{C^k \text{ sections of } f^*TM\},$$

whenever $f : \Sigma \to M$ is a smooth map. Moreover, if $g : M \to N$ is a C^∞ map, then the map

$$\omega_g : C^k(\Sigma, M) \to C^k(\Sigma, N) \quad \text{defined by} \quad \omega_g(f) = g \circ f,$$

is smooth.

To summarize, for each pair (Σ, M) of finite-dimensional smooth manifolds, with Σ compact, we have a chain of Banach manifolds,

$$\cdots \subseteq C^{k+1}(\Sigma, M) \subseteq C^k(\Sigma, M) \subseteq C^{k-1}(\Sigma, M) \subseteq \cdots \subseteq C^0(\Sigma, M).$$

The intersection of these manifolds is the space $C^\infty(\Sigma, M)$ of C^∞ maps from Σ to M, which could be made into a Fréchet manifold, but we will not enter into the details of that here.

We can now try to formulate calculus of variations problems in terms of the infinite-dimensional manifolds that we have constructed. Thus, for example, we can define the *action function*

$$J : C^1(S^1, M) \longrightarrow \mathbb{R} \qquad \text{by} \qquad J(\gamma) = \frac{1}{2} \int_{S^1} |\gamma'(t)|^2 dt,$$

and check without much difficulty that J is a smooth map. As we learned in elementary differential geometry courses, the "critical points" for J are the smooth closed geodesics on M.

Suppose that Σ is a compact two-dimensional Riemann surface. Thus we can imagine that Σ has a Riemannian metric, but we forget about the metric except for its conformal equivalence class, which we denote by ω. Suppose that

$$\{(U_\alpha, (x_\alpha, y_\alpha)) : \alpha \in A\}$$

is an atlas of isothermal coordinate charts on Σ, and let $\{\psi_\alpha : \alpha \in A\}$ be a partition of unity subordinate to $\{U_\alpha : \alpha \in A\}$. We can then define the *Dirichlet integral*

$$E_\omega : C^1(S^1, M) \longrightarrow \mathbb{R} \quad \text{by} \quad E_\omega(f) = \frac{1}{2} \int_\Sigma \sum_{\alpha \in A} \psi_\alpha \left[\left| \frac{\partial f}{\partial x_\alpha} \right|^2 + \left| \frac{\partial f}{\partial y_\alpha} \right|^2 \right] dx_\alpha dy_\alpha.$$

Once again it is relatively easy to check that E_ω is a smooth map on the infinite-dimensional manifold $C^1(\Sigma, M)$. Later we will see that the "critical points" for E_ω are harmonic maps.

More generally, suppose that Σ is an m-dimensional Riemannian manifold with Riemannian metric expressed in local coordinates $({}^1, \ldots x^m)$ on Σ by

$$h = \sum_{a,b=1}^m \eta_{ab} dx^a dx^b.$$

If $f : \Sigma \to M \subseteq \mathbb{R}^N$ is a smooth map and (η^{ab}) denotes the matrix inverse to (η_{ab}), we set

$$|df|^2 = \sum_{a,b=1}^m \eta^{ab} \frac{\partial X}{\partial x^a} \cdot \frac{\partial X}{\partial x^b} \quad \text{and} \quad dA = \sqrt{\det(\eta_{ab})} dx^1 \cdots dx^m.$$

We can then define the *Dirichlet integral*

$$E : C^1(\Sigma, M) \longrightarrow \mathbb{R} \qquad \text{by} \qquad E(f) = \frac{1}{2} \int_\Sigma |df|^2 dA,$$

which is once again a smooth real-valued function on the infinite-dimensional manifold $C^1(\Sigma, M)$. In the case where the domain is two-dimensional, choice of isothermal parameters leads to exactly the same integrand as before, so this generalizes the previous energy functions to higher dimensional domains.

1.4 The basic mapping spaces

For the study of geodesics, harmonic and minimal surfaces and pseudoholomorphic curves, as well as other nonlinear partial differential equations, we need a collection of function spaces with weak enough topologies that it is relatively easy to prove convergence of a sequence which is tending toward an infimum of energy on a given component. The infinite-dimensional manifolds that have proven to be most useful in this regard are those modeled on Sobolev spaces. In this section, we describe the simplest of these spaces.

If Σ is a compact Riemannian manifold, we can define an inner product $(\cdot, \cdot)$ on the space $C^\infty(\Sigma, \mathbb{R}))$ of smooth real-valued maps by

$$(f, g) = \int_\Sigma (fg + \langle Df, Dg\rangle) dA,$$

where the inner product $\langle \cdot, \cdot \rangle$ on the right is the usual inner product in the cotangent space defined by the Riemannian metric on Σ, and dA denotes the area or volume form on Σ. The inner product $(\cdot, \cdot)$ makes space $C^\infty(\Sigma, \mathbb{R}))$ of smooth functions into a pre-Hilbert space. Any pre-Hilbert space has a Hilbert space completion, the set of equivalence classes of Cauchy sequences, as described at the beginning of §1.2. The Hilbert space completion in our case is denoted by $L^2_1(\Sigma, \mathbb{R}))$, and is called the *Sobolev space* of L^2_1-functions on Σ.

A second important Sobolev space generalizes the L^p spaces studied in real analysis when $1 < p < \infty$. We start by defining a norm $\| \cdot \|_{L^p_1}$ on $C^\infty(\Sigma, \mathbb{R}))$ by

$$\left(\|f\|_{L^p_1}\right)^p = \int_\Sigma (|f|^p + |Df|^p) dA,$$

where $|Df|$ is the length calculated with respect to the Riemannian metric on Σ. This norm makes $C^\infty(\Sigma, \mathbb{R}))$ into a pre-Banach space. As before, we can construct the Banach space completion. This Banach space completion is denoted by $L^p_1(\Sigma, \mathbb{R}))$ and is called the *Sobolev space* of L^p_1-functions on Σ. Of course, when $p = 2$ this reduces to the previous example.

We can also define versions of these Sobolev spaces for higher numbers of derivatives. Thus we can define a norm $\| \cdot \|_{L^p_k}$ on $C^\infty(\Sigma, \mathbb{R}))$ by

$$\left(\|f\|_{L^p_k}\right)^p = \int_\Sigma (|f|^p + |Df|^p + \cdots + |Df^k|^p) dA,$$

and construct the completion with respect to this norm, which is denoted $L^p_k(\Sigma, \mathbb{R})$. The resulting space is a Banach space, and a Hilbert space when $p = 2$. We thus obtain a chain of Banach spaces,

$$\cdots \subseteq L^p_k(\Sigma, \mathbb{R}) \subset \cdots \subseteq L^p_1(\Sigma, \mathbb{R}) \subseteq L^p(\Sigma, \mathbb{R}),$$

and the intersection of all the spaces in the chain is just the space $C^\infty(\Sigma, \mathbb{R})$) of smooth functions. These spaces are essential for the modern theory of partial differential equations and they are compared by means of the Sobolev Lemma, which in general form states

$$L^p_k(\Sigma, \mathbb{R}) \subseteq C^l(\Sigma, \mathbb{R}) \qquad \text{for} \qquad p(k-l) > \dim(\Sigma), \tag{1.7}$$

where $\subseteq$ indicates continuous inclusion. A complement to the Sobolev Lemma states that for $pk > \dim(\Sigma)$, $L^p_k(\Sigma, \mathbb{R})$ is a Banach algebra in an appropriate norm defining the "Banachable" structure. Additional information on the Sobolev spaces is found in standard references, such as Evans [19].

For the case where $\Sigma = S^1$, where S^1 is the unit interval $[0, 1]$ with endpoints identified, the Sobolev Lemma is relatively easy to establish, and we do that here:

Lemma 1.4.1. *There is a continuous linear injection $i : L^2_1(S^1, \mathbb{R}) \to C^0(S^1, \mathbb{R})$ which extends the inclusion $C^\infty(S^1, \mathbb{R}) \subset C^0(S^1, \mathbb{R})$.*

Proof: We begin with the sequence of inequalities:

$$|f(t)| \leq |f(\tau)| + \int_t^\tau |f'(u)|du \leq |f(\tau)| + \int_{S^1} |f'(u)|du.$$

Averaging over τ and using the Cauchy-Schwarz inequality yields

$$\begin{aligned} |f(t)| \leq \left[\int_{S^1} |f(u)|du + \int_{S^1} |f'(u)|du\right] \\ \leq \left[\int_{S^1} [|f(u)|^2 + |f'(u)|^2]du\right]^{1/2} = (f, f)^{1/2}, \end{aligned}$$

where $(\cdot, \cdot)$ denotes the L^2_1 inner product. Taking the supremum over all t yields

$$\|f\|_{C^0} \leq \|f\|_{L^2_1}.$$

Thus a Cauchy sequence with respect to the L^2_1 inner product gets taken under the inclusion $C^\infty(S^1, \mathbb{R}) \subset C^0(S^1, \mathbb{R})$ to a Cauchy sequence with respect to the C^0-norm. By definition, an element of L^2_1 is an equivalence class of Cauchy sequences, and the map i is defined by sending this equivalence class to the limit of the C^0 Cauchy sequence. It is immediate that i is injective.

Lemma 1.4.2. *$L^2_1(S^1, \mathbb{R})$ is a Banach algebra; multiplication of functions is a continuous bilinear map*

$$L^2_1(S^1, \mathbb{R}) \times L^2_1(S^1, \mathbb{R}) \longrightarrow L^2_1(S^1, \mathbb{R}).$$

Proof: It follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}\|fg\|_{L^2_1}^2 = (fg, fg) &= \int_{S^1} [(fg)^2 + [(fg)']^2]dt = \int_{S^1} [(fg)^2 + (f'g + fg')^2]dt \\ &= \int_{S^1} [(fg)^2 + (f')^2g^2 + 2fgf'g' + f^2(g')^2]dt \\ &\leq \int_{S^1} \left[(fg)^2 + (f')^2g^2 + f^2(g')^2\right] dt + 2\|fg\|_{C^0} \int_{S^1} (f'g')dt \\ &\leq \|f\|_{C^0}^2\|g\|_{L^2_1}^2 + \|g\|_{C^0}^2\|f\|_{L^2_1}^2 + \|f\|_{C^0}^2\|g\|_{L^2_1}^2 + 2\|fg\|_{C^0}\|f\|_{L^2_1}\|g\|_{L^2_1}.\end{aligned}$$

Since the C^0 norm is less than the L^2_1, we find that

$$\|fg\|_{L^2_1}^2 \leq \|f\|_{L^2_1}^2\|g\|_{L^2_1}^2,$$

finishing the proof of the Lemma.

It follows from this Lemma that the multiplication map

$$L^2_1(S^1, \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)) \times L^2_1(S^1, \mathbb{R}^m) \longrightarrow L^2_1(S^1, \mathbb{R}^n)$$

is continuous.

Suppose now that M is a complete connected finite-dimensional Riemannian manifold isometrically imbedded as a proper submanifold of an ambient Euclidean space $\mathbb{R}^N$. We let

$$L^2_1(S^1, M) = \{\gamma \in L^2_1(S^1, \mathbb{R}^N) : \gamma(t) \in M \text{ for all } t \in S^1\},$$

which is a closed subspace of $L^2_1(S^1, \mathbb{R}^N)$ by Lemma 1.4.1.

We claim that $L^2_1(S^1, M)$ is an infinite-dimensional smooth manifold, the proof being just like the proof for $C^0(S^1, M)$. If $\gamma : S^1 \to M$ is a smooth curve, we let

$$\begin{aligned}V_{\gamma,\epsilon} &= \{X \in L^2_1(S^1, \gamma^*TM) : \text{the } L^2_1 \text{ norm of } X \text{ is } < \epsilon\}, \\ U_{\gamma,\epsilon} &= \{\lambda \in L^2_1(S^1, M) : d_M(\lambda(t), \gamma(t)) < \epsilon \text{ for all } t \in S^1\},\end{aligned}$$

and if $\epsilon > 0$ is sufficiently small, we define

$$\psi_{\gamma,\epsilon} : V_{\gamma,\epsilon} \to U_{\gamma,\epsilon} \quad \text{by} \quad \psi_{\gamma,\epsilon}(X)(p) = \exp_{\gamma(t)}(X(t)), \quad \phi_{\gamma,\epsilon} = \psi^{-1}_{\gamma,\epsilon}.$$

We then set

$$\begin{aligned}\mathcal{A} = \big\{(U_{\gamma,\epsilon}, \phi_{\gamma,\epsilon}) : \ &\gamma : S^1 \to M \text{ is a } C^\infty \text{ map and } \epsilon > 0 \text{ is} \\ &\text{small enough that } \psi_{\gamma,\epsilon} \text{ is a homeomorphism } \big\}.\end{aligned}$$

and prove that it is a smooth atlas by the same argument used to establish Lemma 1.3.1. Thus we obtain:

Theorem 1.4.3. *If M is a smooth manifold, then $L^2_1(S^1, M)$ is a smooth manifold modeled on the Banach spaces $L^2_1(\gamma^*TM)$ for $\gamma : S^1 \to M$ a smooth map. Moreover, if $g : M \to N$ is a C^∞ map, then the map*

$$\omega_g : L^2_1(S^1, M) \to L^2_1(S^1, N) \quad \text{defined by} \quad \omega_g(\gamma) = g \circ \gamma,$$

is also C^∞.

We are also interested in L^p_1-maps from a compact oriented surface Σ. It turns out that these are Hölder continuous in accordance with the following definition.

Definition. If Σ is a metric space, a map $f : \Sigma \to \mathbb{R}$ is said to be *Hölder continuous* with Hölder exponent $\gamma \in (0, 1]$ if there is a constant $C > 0$ such that

$$|f(p) - f(q)| \leq Cd(p, q)^\gamma \quad \text{for all } p, q \in \Sigma.$$

We let $C^{0,\gamma}(\Sigma, \mathbb{R})$ be the space of all functions $f : \Sigma \to \mathbb{R}$ which are Hölder continuous. If $f \in C^{0,\gamma}(\Sigma, \mathbb{R})$, we let

$$[f]_{C^{0,\gamma}} = \sup \left\{ \frac{f(p) - f(q)}{(d(p, q)^\gamma} : p, q \in \Sigma, p \neq q \right\}.$$

Lemma 1.4.4. *If Σ is a compact oriented surface and $p > 2$, there is a continuous linear injection $i : L^p_1(\Sigma, \mathbb{R}) \to C^{0,\gamma}(\Sigma, \mathbb{R})$, where $\gamma = 1 - 2/p$, which extends the inclusion $C^\infty(\Sigma, \mathbb{R}) \subset C^{0,\gamma}(\Sigma, \mathbb{R})$. Moreover, there is a constant $C > 0$ such that*

$$[f]_{C^{0,\gamma}} \leq C\|f\|_{L^p_1}.$$

A complete proof of this is given in Evans [19], §5.6.2. We only prove the weaker result that $L^p_1(\Sigma, \mathbb{R}) \subseteq C^{0,\gamma}(\Sigma, \mathbb{R})$. We begin by assuming that Σ is the torus with flat Riemannian metric expressed in terms of suitable conformal coordinates as $ds^2 = dx^2 + dy^2$. We consider a smooth function $f(x, y)$ on the disk $D(p, r_0)$ of radius r_0 about $p \in \Sigma$ defined in terms of Euclidean coordinates centered at p by $x^2 + y^2 \leq r_0^2$, then after shifting to polar coordinates (r, θ) defined by

$$x = r\cos\theta, \quad y = r\sin\theta,$$

we see that

$$|f(s, \theta) - f(p)| = \int_0^s \frac{\partial f}{\partial r}(r, \theta)dr \leq \int_0^s |Df|(r, \theta)dr,$$

and hence

$$\int_0^{2\pi} |f(s, \theta) - f(p)|d\theta \leq \int_0^{2\pi} \int_0^s |Df|(r, \theta)drd\theta \leq \int_{D(p,r_0)} \frac{|Df|}{r}dxdy.$$

Thus

$$\int_0^{2\pi} \int_0^{r_0} |f(s, \theta) - f(p)|sdsd\theta \leq \left[\int_0^{r_0} rdr\right] \left[\int_{D(p,r_0)} \frac{|Df|}{r}dxdy\right]$$

and hence

$$\int_{D(p,r_0)} |f(x,y) - f(p)| dxdy \leq \frac{r_0^2}{2} \int_{D(p,r_0)} \frac{|Df|}{r} dxdy.$$

It follows from the Hölder inequality that

$$\frac{r_0^2}{2} \int_{D(p,r_0)} \frac{|Df|}{r} dxdy \leq \frac{r_0^2}{2} \left[\int_{D(p,r_0)} |Df|^p dxdy \right]^{1/p} \left[\int_{D(p,r_0)} \frac{dxdy}{r^{p/(p-1)}} \right]^{(p-1)/p}$$

while direct integration yields

$$\begin{aligned} \int_{D(p,r_0)} \frac{dxdy}{r^{p/(p-1)}} &= \int_{D(p,r_0)} r^{p/(1-p)} dxdy \\ &= \int_0^{2\pi} \int_0^{r_0} r^{1/(1-p)} dr d\theta = \frac{2\pi(p-1)}{p-2} r_0^{(p-2)/(p-1)}. \end{aligned}$$

Thus

$$\int_{D(p,r_0)} |f(x,y) - f(p)| dxdy \leq \frac{r_0^2}{2} \left(\frac{2\pi(p-1)}{p-2} \right)^{(p-1)/p} r_0^{(p-2)/p} \|Df\|_{L^p}. \quad (1.8)$$

It follows from (1.8) and the Hölder inequality that

$$\begin{aligned} \pi r_0^2 |f(0)| &\leq \int_{D(p,r_0)} |f(x,y) - f(p)| dxdy + \int_{D(p,r_0)} |f(x,y)| dxdy \\ &\leq (\text{constant}) \|Df\|_{L^p} + (\text{constant}) \|f\|_{L^1} \\ &\leq (\text{constant}) \|Df\|_{L^p} + (\text{constant}) \|f\|_{L^p} (\text{area of } D(p,r_0))^{(p-1)/p} \\ &\leq (\text{constant}) \|f\|_{L_1^p}, \end{aligned}$$

which quickly yields the desired result when Σ is the flat torus.

If Σ is a more general Riemann surface, we can give Σ a Riemannian metric of constant curvature of constant curvature and total volume one. Choose $r_0 > 0$ less than the injectivity radius of this metric. A modification of the above argument can then be applied to a normal coordinate ball of radius r_0 showing that if $p \in \Sigma$, then

$$|f(p)| \leq (\text{constant}) \|f\|_{L_1^p}, \qquad \text{and hence} \qquad \|f\|_{C^0} \leq (\text{constant}) \|f\|_{L_1^p}.$$

(Note that changing the Riemannian metric on Σ merely replaces the L_1^p-norm by an equivalent norm, so adopting the constant curvature metric imposes no restriction.) Thus if a sequence $\{f_i\}$ of smooth functions on Σ converges to a limit in L_1^p-norm, $\{f_i\}$ converges also in C^0 norm to a unique limit function $f_\infty \in C^0$. Thus any element of $L_1^p(\Sigma, \mathbb{R})$ can be identified with a continuous function and the Lemma is proven.

Remark 1.4.5. The previous Sobolev Lemma for L_1^p maps from a surface begins to fail as p approaches 2 from above. The reason is that the highest order term in the L_1^2-norm is conformally invariant and hence invariant under dilations. In the case where D is the unit disk centered at the origin in $\mathbb{R}^2$ this highest order term is

$$\int_D |Df|^2 dxdy.$$

In contrast, the highest order term in the L_1^p-norm is not invariant under dilations. Given a smooth map $f : D \to \mathbb{R}^N$ which takes the boundary ∂D_ϵ to a point, we can define a dilated map $f_\epsilon : D_\epsilon \to \mathbb{R}^N$, where D_ϵ is the ball of radius $\epsilon > 0$ centered at the origin in $\mathbb{R}^2$, by

$$f_\epsilon(x,y) = f\left(\frac{x}{\epsilon}, \frac{y}{\epsilon}\right). \quad \text{Then} \quad |Df_\epsilon(x,y)| = \frac{1}{\epsilon}\left|Df\left(\frac{x}{\epsilon}, \frac{y}{\epsilon}\right)\right|.$$

and if $\tilde{x} = x/\epsilon$ and $\tilde{y} = y/\epsilon$ denote the coordinates corresponding to x and y on the unit disk D_1,

$$\int_{D_\epsilon} |Df_\epsilon|^p dxdy = \left(\frac{1}{\epsilon}\right)^{(p-2)} \int_{D_1} |Df|^{2\alpha} d\tilde{x}d\tilde{y}.$$

Thus as $\epsilon \to 0$, the L_1^p-norm of f_ϵ approaches infinity as long as $p > 2$. In particular, a bound on the L_1^2-norm does not imply a bound on the C^0-norm.

Lemma 1.4.6. *If Σ is a compact oriented surface and $p > 2$, $L_1^p(\Sigma, \mathbb{R})$ satisfies*

$$\|fg\|_{L_1^p} \leq 2\|f\|_{L_1^p}\|g\|_{L_1^p}.$$

Thus after passing to an equivalent norm, we can show that $L_1^p(\Sigma, \mathbb{R})$ is a Banach algebra.

Sketch of Proof: By the previous Lemma,

$$\|fg\|_{L^p} \leq \|f\|_{C^0}\|g\|_{L^p} + \|g\|_{C^0}\|f\|_{L^p} \leq \|f\|_{L_1^p}\|g\|_{L^p} + +\|g\|_{L_1^p}\|f\|_{L^p}$$

and

$$\|D(fg)\|_{L^p} \leq \|f\|_{C^0}\|Dg\|_{L^p} + \|g\|_{C^0}\|Df\|_{L^p} \leq \|f\|_{L_1^p}\|Dg\|_{L^p} + +\|g\|_{L_1^p}\|Df\|_{L^p}.$$

Adding these two inequalities yields the statement of the Lemma.

If Σ is a compact surface and $p > 2$, we let

$$L_1^p(\Sigma, M) = \{f \in L_1^p(\Sigma, \mathbb{R}^N) : f(p) \in M \text{ for all } p \in \Sigma\},$$

a closed subspace of $L_1^p(\Sigma, \mathbb{R}^N)$ by Lemma 1.4.3.

If Σ is a compact surface and $p > 2$, we claim that $L_1^p(\Sigma, M)$ is an infinite-dimensional smooth manifold. In this case, when $f : \Sigma \to M$ is a smooth curve, we let

$$V_{\gamma,\epsilon} = \{X \in L_1^p(\Sigma^* TM) : \text{the } L_1^p \text{ norm of } X \text{ is } < \epsilon\},$$
$$U_{\gamma,\epsilon} = \{g \in L_1^p(\Sigma, M) : d_M(f(p), g(p)) < \epsilon \text{ for all } p \in \Sigma\},$$

and if $\epsilon > 0$ is sufficiently small, we define

$$\psi_{\gamma,\epsilon} : V_{\gamma,\epsilon} \to U_{\gamma,\epsilon} \quad \text{by} \quad \psi_{\gamma,\epsilon}(X)(p) = \exp_{f(p)}(X(p)), \quad \phi_{\gamma,\epsilon} = \psi_{\gamma,\epsilon}^{-1}.$$

We then set

$$\mathcal{A} = \{(U_{\gamma,\epsilon}, \phi_{\gamma,\epsilon}) : \ f : \Sigma \to M \text{ is a } C^\infty \text{ map and } \epsilon > 0 \text{ is small enough that } \psi_{\gamma,\epsilon} \text{ is a homeomorphism } \}.$$

and once again prove that it is a smooth atlas by the same argument used to establish Lemma 1.3.1. Thus we obtain:

Theorem 1.4.7. *If Σ is a compact smooth surface and M is a smooth manifold, then for $p > 2$, $L_1^p(\Sigma, M)$ is a smooth manifold modeled on the Banach spaces $L_1^p(f^*TM)$ for $f : \Sigma \to M$ a smooth map. Moreover, if $g : M \to N$ is a C^∞ map, then the map*

$$\omega_g : L_1^p(\Sigma, M) \to L_1^p(\Sigma, N) \quad \textit{defined by} \quad \omega_g(\gamma) = g \circ \gamma,$$

is also C^∞.

In exactly the same way, we could show that if Σ is an m-dimensional smooth manifold and $p > m$, then $L_1^p(\Sigma, M)$ is an infinite-dimensional smooth manifold.

1.5 Homotopy type of the space of maps

A continuous map $f : X \to Y$ between topological spaces is said to be a *homotopy equivalence* if there is a continuous map $g : Y \to X$ such that $f \circ g$ and $g \circ f$ are both homotopic to the identity. The following theorem was proven quite early in the theory of manifolds of maps; see Eells [16] for the appropriate references.

Theorem 1.5.1. *Let M be a compact connected Riemannian manifold. Then the inclusions*

$$C^k(S^1, M) \subset L_1^2(S^1, M) \quad \textit{and} \quad C^k(S^1, M) \subset C^0(S^1, M)$$

are homotopy equivalences when $k \geq 1$.

The point of this Theorem is that from the point of view of homotopy theory, $L_1^2(S^1, M)$ is essentially the same as the space of continuous maps $C^0(S^1, M)$ with the compact open topology. This latter space has been extensively studied by topologists and much is known about its homotopy and homology groups, as we will see later.

The proof of the Theorem is an application of the theory of "smoothing operators."

For preparation, we suppose that $\phi : \mathbb{R} \to \mathbb{R}$ is a smooth map which vanishes outside $[-1, 1]$. Suppose, moreover, that

$$\phi \geq 0 \qquad \text{and} \qquad \int_{\mathbb{R}} \phi = 1.$$

For $\epsilon > 0$, let $\phi_\epsilon(t) = (1/\epsilon)\phi(t/\epsilon)$, so that

$$\text{supp}(\phi_\epsilon) \subset [-\epsilon, \epsilon] \qquad \text{and} \qquad \int_{\mathbb{R}} \phi_\epsilon = 1.$$

If $\gamma \in C^0(S^1, \mathbb{R}^N)$, we can regard γ as an element of $C^0(\mathbb{R}, \mathbb{R}^N)$ such that $\gamma(t+1) = \gamma(t)$ for all t, and we define $\phi_\epsilon * \gamma \in C^0(\mathbb{R}, \mathbb{R}^N)$ by

$$(\phi_\epsilon * \gamma)(t) = \int_{\mathbb{R}} \phi_\epsilon(t - \tau)\gamma(\tau)d\tau = \int_{\mathbb{R}} \phi_\epsilon(s)\gamma(t - s)ds.$$

It is immediately checked that $\phi_\epsilon * \gamma$ is C^∞ and

$$\frac{d^k}{dt^k}(\phi_\epsilon * \gamma)(t) = \frac{d^k}{dt^k}(\phi_\epsilon) * \gamma = \int_{\mathbb{R}} \left(\frac{d^k}{dt^k}\phi_\epsilon\right)(t - \tau)\gamma(\tau)d\tau.$$

Moreover, $(\phi_\epsilon * \gamma)(t+1) = (\phi_\epsilon * \gamma)(t)$, and hence $\phi_\epsilon * \gamma \in C^\infty(S^1, \mathbb{R}^N)$. We can thus define *smoothing operators*

$$S_\epsilon : C^0(S^1, \mathbb{R}^N) \to C^k(S^1, \mathbb{R}^N), \qquad S_\epsilon : L^2_1(S^1, \mathbb{R}^N) \to C^k(S^1, \mathbb{R}^N)$$

by $S_\epsilon(\gamma) = \phi_\epsilon * \gamma$. It is not difficult to show that the maps

$$S_\epsilon : C^0(S^1, \mathbb{R}^N) \to C^k(S^1, \mathbb{R}^N), \qquad S_\epsilon : L^2_1(S^1, \mathbb{R}^N) \to C^k(S^1, \mathbb{R}^N)$$

are continuous.

Proof of Theorem 1.5.1: Recall that we regard M as a submanifold of $\mathbb{R}^N$. Choose $\delta > 0$ so small that the exponential map

$$\exp : NM \to \mathbb{R}^N, \qquad \text{defined by} \qquad \exp(v) = p + v, \quad \text{for } v \in T_pM,$$

maps $NM_\delta = \{v \in NM : |v| < \delta\}$ diffeomorphically onto

$$M(\delta) = \{p \in \mathbb{R}^N : d(p, M) < \delta\}.$$

Then the "nearest point projection" map $r : M(\delta) \to M$, defined by $r(p+v) = p$ for $p \in M$, is a strong deformation retraction from $M(\delta)$ to M. To see this, we define

$$h : M(\delta) \times [0, 1] \to M(\delta) \quad \text{by} \quad h(p + v, t) = p + (1 - t)v,$$

and check that

1. $h(q, 0) = q$, for $q \in M(\delta)$,

2. $h(q,1) = r(q) \in M$, for $q \in M(\delta)$, and

3. $h(p,t) = p$, for $p \in M$.

We have a similar strong deformation retraction on the function space level. The ω-Lemma gives us a smooth map

$$\omega_h : L_1^2(S^1, M(\delta) \times [0,1]) \to L_1^2(S^1, M(\delta)) \quad \text{defined by} \quad \omega_h(\gamma) = h \circ \gamma.$$

We define

$$j : L_1^2(S^1, M(\delta)) \times [0,1] \to L_1^2(S^1, M(\delta) \times [0,1]) \quad \text{by} \quad j(\gamma,\tau)(t) = (\gamma(t), \tau)$$

and let $H = \omega_h \circ j$. Then

1. $H(\gamma, 0) = \gamma$, for $\gamma \in L_1^2(S^1, M(\delta))$,

2. $H(\gamma, 1) = r \circ \gamma \in L_1^2(S^1, M)$, for $\gamma \in L_1^2(S^1, M(\delta))$, and

3. $H(\gamma, t) = \gamma$, for $\gamma \in L_1^2(S^1, M)$.

We can therefore define a strong deformation retraction

$$R : L_1^2(S^1, M(\delta)) \longrightarrow L_1^2(S^1, M)$$

by $R(\gamma) = \omega_r(\gamma) = H(\gamma, 1)$. In a similar fashion, we can define a strong deformation retraction

$$R : C^k(S^1, M(\delta)) \longrightarrow C^0(S^1, M),$$

whenever $k \geq 0$.

Let $\varepsilon_k = 2^{-k}$ and let

$$\begin{aligned} C^0(S^1, M)^{\varepsilon_k} = \{\gamma \in C^0(S^1, M) : \ &\gamma \text{ maps the closed interval} \\ &[(m-1)2^{-k}, (m+1)2^{-k}] \text{ into the open ball} \\ &B(\gamma(m2^{-k}); \delta) \text{ for each integer } m \text{ such that } 0 \leq m \leq 2^k\}, \end{aligned}$$

an open set in the compact-open topology. They key point of this set is that when $|s-t| < 2^{-k}$ then the straight line from $\gamma(s)$ to $\gamma(t)$ in $\mathbb{R}^N$ lies entirely within $M(\delta)$ hence $S_\epsilon \star \gamma$ lies within $M(\delta)$ when $\epsilon \leq \epsilon_k$. Note that

$$C^0(S^1, M) = \bigcup_{k=1}^{\infty} C^0(S^1, M)^{\varepsilon_k}, \quad L_1^2(S^1, M)^{\varepsilon_{k+1}} \subset L_1^2(S^1, M)^{\varepsilon_k},$$

and we therefore say that $C^0(S^1, M)$ is a *monotone union* of the subspaces $C^0(S^1, M)^{\varepsilon_k}$. Similarly, we let

$$\begin{aligned} L_1^2(S^1, M)^{\varepsilon_k} &= L_1^2((S^1, M) \cap C^0(S^1, M)^{\varepsilon_k}, \\ C^k(S^1, M)^{\varepsilon_k} &= C^k((S^1, M) \cap C^0(S^1, M)^{\varepsilon_k}, \quad \text{when } k \geq 1, \end{aligned}$$

thereby expressing $L^2_1(S^1, M)$ and $C^k(S^1, M)$ as monotone unions for $k \geq 1$.

By analogous formulae, we define

$$C^0(S^1, M(\delta))^{\varepsilon_k}, \quad L^2_1(S^1, M(\delta))^{\varepsilon_k} \quad \text{and} \quad C^k(S^1, M(\delta))^{\varepsilon_k},$$

for $k \geq 1$. We can then define smoothing operators

$$S_{\varepsilon_k} : L^2_1(S^1, M)^{\varepsilon_k} \longrightarrow C^k(S^1, M(\delta))^{\varepsilon_k},$$

since

$$\gamma \in L^2_1(S^1, M)^{\varepsilon_k} \quad \Rightarrow \quad S_{\varepsilon_k} \star \gamma \in C^k(S^1, M(\delta))^{\varepsilon_k}.$$

We define s to the composition of

$$S_{\varepsilon_k} : L^2_1(S^1, M)^{\varepsilon_k} \to C^k(S^1, M(\delta))^{\varepsilon_k} \quad \text{and}$$
$$R : C^k(S^1, M(\delta))^{\varepsilon_k} \to C^k(S^1, M)^{\varepsilon_k}.$$

We claim that if $i : C^k(S^1, M)^{\varepsilon_k} \subset L^2_1(S^1, M)^{\varepsilon_k}$ is the inclusion, then

$$s \circ i \quad \text{and} \quad i \circ s$$

are homotopic to the identity. This is easy to verify. To get the homotopy from $s \circ i$ to the identity, we simply define

$$H_1 : C^k(S^1, M)^{\varepsilon_k} \times [0, 1] \to C^k(S^1, M)^{\varepsilon_k}$$
$$\text{by} \qquad H_1(\gamma, t) = R \circ (tS_{\varepsilon_k} + (1 - t)\mathrm{id})) \circ i(\gamma).$$

Similarly, to get the homotopy from $i \circ s$ to the identity, we define

$$H_2 : L^2_1(S^1, M)^{\varepsilon_k} \times [0, 1] \to L^2_1(S^1, M)^{\varepsilon_k}$$
$$\text{by} \qquad H_2(\gamma, t) = i \circ R \circ (tS_{\varepsilon_k} + (1 - t)\mathrm{id})(\gamma),$$

This shows that for each $k \in \mathbb{N}$, the inclusion

$$C^k(S^1, M)^{\varepsilon_k} \subset L^2_1(S^1, M)^{\varepsilon_k}$$

is a homotopy equivalence.

To finish the proof, we must take an appropriate limit as $k \to \infty$. Suppose that the metrizable space X is a monotone union of open subsets, by which we mean that we have a sequence of spaces

$$U_1 \subset U_2 \subset U_3 \subset \cdots \quad \text{such that} \quad X = \bigcup \{U_k : k \in \mathbb{N}\}.$$

Suppose, moreover, that we let

$$X^\star = (U_1 \times [1, 2]) \cup (U_2 \times [2, 3]) \cup (U_3 \times [3.4]) \cup \cdots$$

topologized as a subset of $X \times \mathbb{R}$. We say that X is the *homotopy direct limit* of the subspaces $\{U_k : k \in \mathbb{N}\}$ if the projection $p : X^\star \to X$ on the first factor

is a homotopy equivalence. If the subsets U_k are open, then the open cover $\{U_k : k \in \mathbb{N}\}$ has a C^0 subordinate partition of unity $\{\psi_k : k \in \mathbb{N}\}$. In this case, the map

$$f : X \to X^\star \quad \text{defined by} \quad f(x) = \left(x, \sum_{k=1}^{\infty} k\psi_k(x) \right)$$

is a homotopy inverse to p, showing that X is a homotopy direct limit in this case. Using this argument, one easily verifies that $C^k(S^1, M)$ is a homotopy direct limit of its subspaces $C^k(S^1, M)^{\varepsilon_k}$ and $L^2_1(S^1, M)$ is a homotopy direct limit of $L^2_1(S^1, M)^{\varepsilon_k}$.

Now we apply the following Lemma, which is just Theorem A from the Appendix to Milnor's book on Morse theory [50]:

Lemma 1.5.2. *Suppose that X is the homotopy direct limit of $\{U_k : k \in \mathbb{N}\}$ and that Y is the homotopy direct limit of $\{V_k : k \in \mathbb{N}\}$. If $f : X \to Y$ is a continuous map such that $f(U_k) \subseteq V_k$ and the restriction of f to U_k is a homotopy equivalence from U_k to V_k, then f itself is a homotopy equivalence.*

We refer the reader to Milnor for the proof of this Lemma. It implies that the inclusion $C^k(S^1, M) \subset L^2_1(S^1, M)$ is a homotopy equivalence when $k \geq 1$. In a similar manner, one verifies that the inclusion $C^k(S^1, M) \subset C^0(S^1, M)$ is a homotopy equivalence when $k \geq 1$.

Theorem 1.5.3. *Let M be a compact connected Riemannian manifold, Σ a compact connected Riemann surface, $p > 2$. Then the inclusions*

$$C^k(\Sigma, M) \subset L^p_1(\Sigma, M), \qquad C^k(\Sigma, M) \subset C^0(\Sigma, M)$$

are homotopy equivalences.

The proof is essentially the same as for the previous theorem, with 2 replaced by p, except for the definition of smoothing operators defined on Σ. The construction of such operators is a standard technique in the theory of partial differential equations. We describe only the case $\Sigma = T^2$ here, where $T^2 = \mathbb{C}/\Lambda$, Λ being a lattice in $\mathbb{C}$; the construction in this case is particularly transparent. (In the general case, the ideas are the same, but one constructs the smoothing operators by piecing together using a partition of unity on Σ.)

Note that an element $f \in L^p_1(T^2, \mathbb{R}^N)$ can be regarded as a map $f : \mathbb{C} \to \mathbb{R}^N$ such that $f(z + \lambda) = f(z)$ for $\lambda \in \Lambda$.

Suppose that $\phi : \mathbb{C} \to [0, \infty)$ is a smooth map which vanishes outside $D = \{z \in \mathbb{C} : |z| \leq 1\}$ such that

$$\int_{\mathbb{C}} \phi \; dxdy = 1,$$

where (x, y) are the standard coordinates on $\mathbb{C}$. Let $\phi_\epsilon(z) = (1/\epsilon^2)\phi(z/\epsilon)$, so that

$$\mathrm{supp}(\phi_\epsilon) \subset \{z \in \mathbb{C} : |z| \leq \epsilon\} \qquad \text{and} \qquad \int_{\mathbb{C}} \phi_\epsilon dxdy = 1.$$

If $f : \mathbb{C} \to M$ comes from an element $f \in L_1^p(T^2, \mathbb{R}^N)$, we define $\phi_\epsilon * f \in C^\infty(\mathbb{C}, \mathbb{R}^N)$ by

$$(\phi_\epsilon * \gamma)(z) = \int_{\mathbb{C}} \phi_\epsilon(z - w)\gamma(w)dw.$$

It is immediately checked that $(\phi_\epsilon * f)(z + \lambda) = (\phi_\epsilon * f)(z)$ for $\lambda \in \Lambda$, so $(\phi_\epsilon * f)$ can be identified with an element of $C^\infty(T^2, \mathbb{R}^N)$.

Thus we can define *smoothing operators*

$$S_\epsilon : C^0(T^2, \mathbb{R}^N) \to C^k(T^2, \mathbb{R}^N), \qquad S_\epsilon : L_1^p(T^2, \mathbb{R}^N) \to C^k(T^2, \mathbb{R}^N)$$

by $S_\epsilon(f) = \phi_\epsilon * f$. The proof of the Theorem for maps from $\Sigma = T^2$ now continues in exactly the same way as for maps from S^1.

Remark 1.5.4. It is interesting to consider Sobolev spaces of maps which do not lie in "Sobolev range." Thus we could consider

$$W_1^p(\Sigma, M) = \{f \in L_1^p(\Sigma, M) : \ f(p) \in M \text{ for almost all } p \in \Sigma \ \},$$

$$H_{1,S}^p(\Sigma, M) = (\text{Closure of } C^\infty(\Sigma, M) \text{ in } L_1^p(\Sigma, M)),$$

for $p \leq \dim \Sigma$. Although $H_{1,S}^p(\Sigma, M) \subseteq W_1^p(\Sigma, M)$ the inclusion is often strict, and neither space is in general homotopically equivalent to the space $C^0(\Sigma, M)$ of continuous maps. These Sobolev spaces have been extensively studied by Hang and Lin [33], among others, and applications to the theory of harmonic maps are described in the review article of Brezis [11]. When the dimension of Σ is at least three, harmonic maps from Σ to M are vastly more complicated than geodesics and harmonic surfaces; for example, they need not be smooth.

1.6 The α-Lemma*

We have seen that there is a covariant functor from finite-dimensional smooth manifolds and smooth maps to infinite-dimensional smooth manifolds and smooth maps,

$$M \mapsto C^k(\Sigma, M), \qquad g : M \to N \quad \mapsto \quad \omega_g : C^k(\Sigma, M) \to C^k(\Sigma, N),$$

where $\omega_g(f) = g \circ f$. One might hope that when M is a fixed smooth manifold, there is a similar contravariant functor from compact manifolds Σ to infinite-dimensional smooth manifolds $C^k(\Sigma, M)$ in which

$$f : \Sigma_1 \to \Sigma_2 \quad \mapsto \quad \alpha_f : C^k(\Sigma_2, M) \to C^k(\Sigma_1, M),$$

where $\alpha_f(g) = g \circ f$. However, it turns out that the maps α_f are no longer C^∞ smooth, but only C^k:

α-Lemma 1.6.1. *If M is a fixed smooth manifold, a smooth map $g : \Sigma_1 \to \Sigma_2$ between smooth compact manifolds induces a C^k map*

$$\alpha_g : C^k(\Sigma_2, M) \to C^k(\Sigma_1, M),$$

for each k.

A proof of this lemma can be found in [1].

We could try to put the α- and ω-Lemmas into a single theorem. This is partially accomplished by the following theorem, which we will not prove; it is stated in the survey article [16], which also includes numerous references:

Theorem 1.6.2. *If S, M and N are smooth finite-dimensional manifolds, then*

$$\Phi : C^k(S, M) \times C^{k+s}(M, N) \to C^k(S, N), \qquad \Phi(f, g) = g \circ f,$$

is C^s.

Remark. The loss of derivatives in the statement of Theorem 1.6.2 has far-reaching implications. For example, suppose that we want to construct examples of infinite-dimensional Banach Lie groups, which are defined just like ordinary Lie groups, except that of being finite-dimensional manifolds, they are infinite-dimensional Banach manifolds. We could start with a finite-dimensional Lie group G with identity e, smooth multiplication map

$$\mu : G \times G \to G, \qquad \mu(\sigma, \tau) = \sigma \cdot \tau$$

and smooth inverse map

$$\nu : G \to G, \qquad \nu(\sigma) = \sigma^{-1}.$$

We could then define the corresponding *loop group* $L^2_1(S^1, G)$ with identity the constant map to e, smooth multiplication

$$\omega_\mu : L^2_1(S^1, G) \times L^2_1(S^1, G) \to L^2_1(S^1, G), \qquad \omega_\mu(f, g) = f \cdot g,$$

where the dot on the right denotes multiplication within G, and smooth inverse

$$\omega_\nu : L^2_1(S^1, G) \to L^2_1(S^1, G), \qquad \omega_\nu(f)(p) = (f(p))^{-1}.$$

It is not difficult to check that $L^2_1(S^1, G)$ is in fact an infinite-dimensional Banach Lie group. A rich theory of these loop groups has been developed, although they are often modeled on Fréchet rather than Banach spaces.

On the other hand, although the space of C^k diffeomorphisms is a perfectly well-behaved topological group under composition, this composition fails to be smooth because of the loss of derivatives implicit in the statement of the Theorem 1.6.2. There seems to be no simple way to make the group of C^k diffeomorphisms into an infinite-dimensional Lie group modeled on a Banach space. This fact seems to interfere with potential applications of global analysis techniques to important nonlinear systems of PDE's, such as those governing incompressible fluids.

One consequence of Theorem 1.6.2 is a smoothness result for the evaluation map

$$\mathrm{ev} : C^k(\Sigma, M) \times M \to M \quad \text{defined by} \quad \mathrm{ev}(f, p) = f(p). \tag{1.9}$$

Lemma 1.6.3. *The map $ev : C^k(\Sigma, M) \times \Sigma \to M$, defined by (1.9).*

For a direct proof we refer to [2], page 99.

We can extend Theorem 1.6.2 and many of its consequences to the Sobolev manifolds $L^p_j(\Sigma_2, M)$ when p and k are large enough that $L^p_j(\Sigma_2, M) \subset C^k(\Sigma_2, M)$ in accordance with (1.7). For example, one consequence is:

Lemma 1.6.4. *If $L^p_j(\Sigma, M) \subset C^k(\Sigma, M)$, the map*

$$ev : L^p_j(\Sigma, M) \times \Sigma \to M \quad \textit{defined by} \quad ev(f, p) = f(p)$$

is C^k.

Note that it follows from this Lemma that

$$\text{ev} : L^2_k(S^1, M) \times \Sigma \to M \qquad \text{is } C^{k-1},$$

while if Σ is a surface,

$$\text{ev} : L^p_k(\Sigma, M) \times \Sigma \to M \qquad \text{is } C^{k-1},$$

when $p > 2$.

1.7 The tangent and cotangent bundles

Many constructions from the theory of finite-dimensional manifolds can be generalized to infinite-dimensional Banach or Hilbert manifolds. These include tensors of various ranks, vector fields and differential equations, connections, Riemannian metrics on Hilbert manifolds, Finsler metrics on Banach manifolds, differential forms and de Rham cohomology. Many of these constructions are carried out in great detail in the Lang's book [43] on infinite-dimensional manifolds. We provide a brief summary here.

We first extend familiar definitions of tangent and cotangent bundles to the infinite-dimensional context. Let $\mathcal{M}$ be an infinite-dimensional smooth manifold modeled on a Banach space E with smooth atlas $\{U_\alpha, \phi_\alpha) : \alpha \in A\}$. Consider the collection of triples (α, p, v), where $\alpha \in A$, $p \in U_\alpha$ and $v \in E$. On this collection of triples we define an equivalence relation $\sim$ by

$$(\alpha, p, v) \sim (\beta, q, w) \quad \Leftrightarrow \quad p = q \text{ and } w = D(\phi_\beta \circ \phi_\alpha^{-1})(\phi_\alpha(p))v.$$

The set of equivalence classes is called the *tangent bundle* of $\mathcal{M}$ and is denoted by $T\mathcal{M}$.

Let $[\alpha, p, v]$ denote the equivalence class of (α, p, v) and define

$$\pi : T\mathcal{M} \longrightarrow \mathcal{M} \quad \text{by} \quad \pi([\alpha, p, v]) = p.$$

Let $\tilde{U}_\alpha = \{[\alpha, p, v]; p \in U_\alpha, v \in E\}$, and define

$$\tilde{\phi}_\alpha : \tilde{U}_\alpha \longrightarrow E \oplus E \quad \text{by} \quad \tilde{\phi}_\alpha([\alpha, p, v]) = (\phi_\alpha(p), v).$$

Then $\{(\tilde{U}_\alpha, \tilde{\phi}_\alpha) : \alpha \in A\}$ is a smooth atlas on $T\mathcal{M}$ making $T\mathcal{M}$ into a smooth manifold modeled on the Banach space $E \oplus E$. If $p \in \mathcal{M}$, we let $T_p\mathcal{M} = \pi^{-1}(p)$, the fiber of the tangent bundle over p, and call $T_p\mathcal{M}$ the *tangent space* to $\mathcal{M}$ at p.

Just as in the finite-dimensional case, elements of $T_p\mathcal{M}$ are called *tangent vectors*. If $\gamma : (a, b) \to \mathcal{M}$ is a smooth curve, $t \in (a, b)$ and $\gamma(t) \in U_\alpha$, we define

$$\gamma'(t) \in T_p\mathcal{M} \qquad \text{by} \qquad \gamma'(t) = [\alpha, \gamma(t), D(\phi_\alpha \circ \gamma)(t) \cdot 1],$$

a tangent vector called the *velocity vector* to γ at t.

If $F : \mathcal{M} \to \mathcal{N}$ is a smooth map between manifolds with atlases $\{(U_\alpha, \phi_\alpha) : \alpha \in A\}$ and $\{(V_\beta, \psi_\beta) : \beta \in B\}$ and $p \in \mathcal{M}$, we can define the differential $(F_*)_p : T_p\mathcal{M} \to T_{F(p)}\mathcal{N}$ by

$$(F_*)_p([\alpha, p, v]) = (\beta, F(p), (D(\psi_\beta \circ F \circ \phi_\alpha^{-1})(\phi_\alpha(p))(v)],$$

where $p \in U_\alpha$ and $F(p) \in V_\beta$. Note that if $\gamma : (a, b) \to \mathcal{M}$ is a C^1 curve,

$$(F_*)_p(\gamma'(t)) = (F \circ \gamma)'(t), \qquad \text{for } t \in (a, b).$$

The differentials fit together to form a map of tangent bundles $F_* : T\mathcal{M} \to T\mathcal{N}$.

In a very similar way, we can also describe the *cotangent bundle* of $\mathcal{M}$. We consider a similar collection of triples (α, p, v^*), where $\alpha \in A$, $p \in U_\alpha$ and $v^* \in E^*$, where E^* is the Banach space dual to E. This time we choose the equivalence relation

$$(\alpha, p, v^*) \sim (\beta, q, w^*) \quad \Leftrightarrow \quad p = q \text{ and } v^* = [D(\phi_\beta \circ \phi_\alpha^{-1})(\phi_\alpha(p))]^* w^*,$$

where $(\cdot)^*$ denotes transpose map defined by

$$\left([D(\phi_\beta \circ \phi_\alpha^{-1})(\phi_\alpha(p))]^* w^*\right)(v) = w^*(D(\phi_\beta \circ \phi_\alpha^{-1})(\phi_\alpha(p))(v).$$

The cotangent bundle $T^*\mathcal{M}$ is the set of equivalence classes and is a smooth manifold modeled on the Banach space $E \oplus E^*$. Once again, we have a projection

$$\pi : T\mathcal{M} \longrightarrow \mathcal{M} \quad \text{defined by} \quad \pi([\alpha, p, v]) = p.$$

If $p \in \mathcal{M}$, the fiber $T_p^*\mathcal{M}$ of the cotangent bundle over p is called the *cotangent space* to $\mathcal{M}$ at p. A smooth map $F : \mathcal{M} \to \mathcal{N}$ induces a map in the opposite direction

$$(F^*)_p : T^*_{F(p)}\mathcal{N} \to T_p\mathcal{M} \qquad \text{by} \qquad (F^*)_p(w^*)(v) = w^*((F_*)_p(v)).$$

In terms of local coordinates, this can be written

$$(F^*)_p([\beta, F(p), w^*]) = [\alpha, p, (D(\psi_\beta \circ F \circ \phi_\alpha^{-1})(\phi_\alpha(p))^*(w^*)]. \tag{1.10}$$

We can also define the k-th *tensor power* and the k-th *exterior power* of the cotangent bundle. To do this, we start with the Banach space $L^k(E, \mathbb{R})$ of maps

$$T : E \times E \times \cdots E(k \text{ times}) \longrightarrow \mathbb{R}$$

which are linear in each variable, the so-called space of *k-linear maps*, or its subspace of alternating k-linear maps $L^k_a(E,\mathbb{R})$. An element $T \in L^k(E,\mathbb{R})$ is said to be *alternating* if

$$T(h_{\sigma(1)},\dots,h_{\sigma(k)}) = (\mathrm{sgn}(\sigma))T(h_1,\dots,h_k), \qquad \text{for all } \sigma \in S_k,$$

where S_k denotes the symmetric group on k letters and $\mathrm{sgn}(\sigma)$ denotes the sign of the permutation $\sigma \in S_k$. A continuous linear map $\Phi : E \to F$ between Banach spaces induces continuous linear maps

$$\Phi^* : L^k(F,\mathbb{R}) \to L^k(E,\mathbb{R}), \qquad \Phi^* : L^k_a(F,\mathbb{R}) \to L^k_a(E,\mathbb{R})$$

by means of the formula

$$(\Phi^* T)(v_1,\dots v_k) = T(\Phi(v_1),\dots,\Phi(v_k)).$$

To define the k-th tensor power of the cotangent bundle, we start with triples (α,p,T), where $\alpha \in A$, $p \in U_\alpha$ and $T \in L^k(E,\mathbb{R})$ and the equivalence relation

$$(\alpha,p,T_\alpha) \sim (\beta,q,T_\beta) \quad \Leftrightarrow \quad p = q \text{ and } T_\alpha = \Phi^* T_\beta,$$

where $\Phi = D(\phi_\beta \circ \phi_\alpha^{-1})(\phi_\alpha(p))$. The k-th tensor power of the cotangent bundle $\otimes^k T^*\mathcal{M}$ is the set of equivalence classes. The k-th exterior power is defined the same way, except that T is taken to lie in $L^k_a(E,\mathbb{R})$. The fibers $\otimes^k T^*_p\mathcal{M}$ and $\Lambda^k T^*_p\mathcal{M}$ are called the k-th *tensor power* and the *exterior power* of the cotangent space to $\mathcal{M}$ at p.

In the case where $\mathcal{M} = L^2_1(S^1,M)$, M being oriented, the tangent bundle has another description, namely

$$TL^2_1(S^1,M) = L^2_1(S^1,TM).$$

To see this, recall how we constructed the atlas on $L^2_1(S^1,M)$. If γ is a smooth element of $L^2_1(S^1,M)$, we let

$$U_{\gamma,\epsilon} = \{\lambda \in L^2_1(S^1,M) : d_M(\lambda(t),\gamma(t)) < \epsilon \text{ for all } t \in S^1\},$$

$$V_{\gamma,\epsilon} = \{X \in L^2_1(S^1,\mathbb{R}^n)) : \langle X(t),X(t)\rangle < \epsilon \text{ for all } t \in S^1\},$$

and define

$$\psi_{\gamma,\epsilon} : V_{\gamma,\epsilon} \longrightarrow U_{\gamma,\epsilon} \quad \text{by} \quad (\psi_{\gamma,\epsilon}(X))(t) = \exp_{\gamma(t)}(X(t)).$$

For ϵ sufficiently small, $\psi_{\gamma,\epsilon}$ is a bijection with inverse $\phi_{\gamma,\epsilon} : U_{\gamma,\epsilon} \to V_{\gamma,\epsilon}$, and

$$\{(U_{\gamma,\epsilon},\phi_{\gamma,\epsilon}) : \gamma \text{ is smooth and } \epsilon \text{ is sufficiently small}\}$$

is a smooth atlas for $L^2_1(S^1,M)$.

Now for each smooth γ, we can construct a lift $\tilde{\gamma} \in L^2_1(S^1,TM)$ by setting

$$\tilde{\gamma}(t) = 0_{\gamma(t)} \in T_{\gamma(t)}M,$$

and construct a corresponding chart on $L^2_1(S^1, TM)$. The remarkable fact is that we can choose the chart to be valid over all of

$$\tilde{U}_{\gamma,\epsilon} = \Omega_\pi^{-1}(U_{\gamma,\epsilon}) = \{X \in L^2_1(S^1, TM) : \pi \circ X \in U_{\gamma,\epsilon}\}.$$

Indeed, we can set

$$\tilde{V}_{\gamma,\epsilon} = V_{\gamma,\epsilon} \times \{ \text{ } L^2_1\text{-sections of } \gamma^* TM \text{ } \}$$

and define $\tilde{\psi}_{\gamma,\epsilon} : \tilde{V}_{\gamma,\epsilon} \to \tilde{U}_{\gamma,\epsilon}$ by

$$\tilde{\psi}_{\gamma,\epsilon}(X, Y) = (\exp_{\gamma(t)} X(t), (d(\exp_{\gamma(t)})_{X(t)} Y(t)).$$

Finally, define $\tilde{\phi}_{\gamma,\epsilon} : \tilde{U}_{\gamma,\epsilon} \to \tilde{V}_{\gamma,\epsilon}$ by $\tilde{\phi}_{\gamma,\epsilon} = \tilde{\psi}^{-1}_{\gamma,\epsilon}$. Then

$$(\tilde{\phi}_{\gamma_1,\epsilon} \circ \tilde{\phi}^{-1}_{\gamma_2,\epsilon})(X, Y) = ((\phi_{\gamma_1,\epsilon} \circ \phi^{-1}_{\gamma_2,\epsilon})(X), D(\phi_{\gamma_1,\epsilon} \circ \phi^{-1}_{\gamma_2,\epsilon})(X)(Y)).$$

Thus the charts transform exactly the way they should for the tangent bundle.

In a quite similar fashion, we can show that if Σ is a compact Riemann surface and $p > 2$,

$$TL^p_1(\Sigma, M) = L^p_1(\Sigma, TM).$$

It is important to observe that just as the imbedding $i : M \to \mathbb{R}^N$ induces an imbedding $\omega_i : L^2_1(S^1, M) \to L^2_1(S^1, \mathbb{R}^N)$, so the imbedding $i : TM \to T\mathbb{R}^N = \mathbb{R}^{2N}$ induces an imbedding

$$\omega_i : TL^2_1(S^1, TM) = L^2_1(S^1, TM) \longrightarrow L^2_1(S^1, T\mathbb{R}^N) = L^2_1(S^1, \mathbb{R}^{2N}),$$

allowing us to realize $TL^2_1(S^1, TM)$ as a subspace of a Banach space. Similarly, $TL^p_1(\Sigma, M)$ can be regarded as a subspace of a Banach space.

Note that the Hilbert space inner product allows us to identify the model space $L^2_1(S^1, \mathbb{R}^n)$ for $\mathcal{M} = L^2_1(S^1, M)$ with its dual. Using this fact, it is not difficult to verify that

$$T^* L^2_1(S^1, M) = L^2_1(S^1, T^*M), \qquad \otimes^k T^* L^2_1(S^1, M) = L^2_1(S^1, \otimes^k T^* M),$$

and

$$\Lambda^k T^* L^2_1(S^1, M) = L^2_1(S^1, \Lambda^k T^* M).$$

It is actually the "sections" of the tangent bundle and the exterior powers of the cotangent bundle that will be of most importance for us; these are called vector fields and differential forms, respectively.

Definition. A *smooth vector field* on $\mathcal{M}$ is a smooth map

$$\mathcal{X} : \mathcal{M} \longrightarrow T\mathcal{M} \qquad \text{such that} \qquad \pi \circ \mathcal{X} = \mathrm{id}_{\mathcal{M}}.$$

Note that if $\mathcal{X} : \mathcal{M} \to T\mathcal{M}$ is a smooth vector field and $f : \mathcal{M} \to \mathbb{R}$ is a smooth function, we can define the vector field $f\mathcal{X} : \mathcal{M} \to T\mathcal{M}$ by $(f\mathcal{X})(p) = f(p)\mathcal{X}(p)$.

This makes the space of smooth vector fields into a module over the ring of smooth real-valued functions.

Example. If M is a finite-dimensional Riemannian manifold and $X : M \to TM$ is a smooth vector field on M, then

$$\omega_X : L_1^2(S^1, M) \longrightarrow L_1^2(S^1, TM) = TL_1^2(S^1, M) \qquad \text{and}$$

$$\omega_X : L_1^p(\Sigma, M) \longrightarrow L_1^p(\Sigma, TM) = TL_1^2(\Sigma, M)$$

are smooth vector fields on $L_1^2(S^1, M)$ and $L_1^p(\Sigma, M)$.

1.8 Differential forms

For calculations on smooth manifolds, differential forms are often more convenient to use than general tensor fields. We now describe how some of the familiar operations on differential forms extend to infinte-dimensional manifolds.

Definition. A *smooth covariant tensor field* of rank k on $\mathcal{M}$ is a smooth map

$$\phi : \mathcal{M} \longrightarrow \otimes^k T^*\mathcal{M} \qquad \text{such that} \qquad \pi \circ \phi = \mathrm{id}_{\mathcal{M}}.$$

A *smooth differential form* of degree k on $\mathcal{M}$ (or a smooth k-form) is a smooth map

$$\phi : \mathcal{M} \longrightarrow \Lambda^k T^*\mathcal{M} \qquad \text{such that} \qquad \pi \circ \phi = \mathrm{id}_{\mathcal{M}}.$$

As in the case of vector fields, we can multiply covariant tensor fields or differential forms by functions.

An important example of differential one-form occurs when $f : \mathcal{M} \to \mathbb{R}$ is a smooth function. Then for the coordinate chart (U_α, ϕ_α), we have

$$D(f \circ \phi_\alpha^{-1}) : U_\alpha \to L(E, \mathbb{R}) = E^*, \tag{1.11}$$

where E is the model space of $\mathcal{M}$. The *differential* of f is the smooth one-form df such that

$$df(p) = [\alpha, p, D(f \circ \phi_\alpha^{-1})(p)], \qquad \text{for } p \in U_\alpha.$$

It is readily verified that the local representatives transform as they should under change of coordinates. Moreover it is easily checked that the differentials of functions at a point p generate the cotangent space. The Leibniz rule for differentiation implies that $d(fg) = gdf + fdg$.

Definition. A point $p \in \mathcal{M}$ is a *critical point* for the real-valued function $f : \mathcal{M} \to \mathbb{R}$ if $df(p) = 0$.

An important example to keep in mind is the real-valued function

$$J : L_1^2(S^1, M) \to M, \qquad J(\gamma) = J(\gamma) = \frac{1}{2} \int_{S^1} |\gamma'(t)|^2 dt,$$

when M is a Riemannian manifold. In this case, regularity theory will show that a critical point in this case is actually a C^∞ map, hence a smooth closed geodesic in M.

Using the notion of differential of a function, we can define the *directional derivative* of a function f in the direction of $\mathcal{X}$ by

$$\mathcal{X}(f)(p) = df(p)(\mathcal{X}(p)),$$

the right hand side being the dual pairing between cotangent and tangent spaces.

Lemma 1.8.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be smooth vector fields on the Banach manifold $\mathcal{M}$. Then there is a unique vector field $[\mathcal{X}, \mathcal{Y}]$ on $\mathcal{M}$ which satisfies the equation*

$$([\mathcal{X}, \mathcal{Y}](f))(p) = (\mathcal{X}\mathcal{Y}(f))(p) - (\mathcal{Y}\mathcal{X}(f))(p).$$

Sketch of proof: It suffices to show that the above formula is equivalent to an expression for $[\mathcal{X}, \mathcal{Y}]$ in terms of a local coordinate chart (U_α, ϕ_α). Suppose that

$$\tilde{\mathcal{X}}, \tilde{\mathcal{Y}} : U_\alpha \to E \quad \text{are defined by} \quad \mathcal{X}(p) = [p, \alpha, \tilde{\mathcal{X}}(p)], \quad \mathcal{Y}(p) = [p, \alpha, \tilde{\mathcal{Y}}(p)].$$

Using the chain rule, one can check that the Lie bracket must then be given by

$$[\mathcal{X}, \mathcal{Y}](p) = \left[p, \alpha, D\tilde{\mathcal{Y}}(p)\tilde{\mathcal{X}}(p) - D\tilde{\mathcal{X}}(p)\tilde{\mathcal{Y}}(p)\right].$$

The vector field $[\mathcal{X}, \mathcal{Y}]$ is known as the Lie bracket of $\mathcal{X}$ and $\mathcal{Y}$; it is easily verified that it satisfies the identity:

$$[f\mathcal{X}, g\mathcal{Y}] = fg[\mathcal{X}, \mathcal{Y}] + f\mathcal{X}(g)\mathcal{Y} - g\mathcal{Y}(f)\mathcal{X}.$$

We next note that differential forms are "functorial." If $F : \mathcal{M} \to \mathcal{N}$ is a smooth map and ϕ is a smooth differential form of degree k on $\mathcal{N}$, we can define a differential form $F^*\phi$ on $\mathcal{M}$ by

$$[F^*\phi](p) = F_p^*(\phi(\mathcal{F}(p));$$

it follows from (1.10) that $F^*\phi$ is smooth.

If $\mathcal{X}_1, \ldots, \mathcal{X}_k$ are smooth vector fields on $\mathcal{M}$ and ϕ is a smooth k-form on $\mathcal{M}$, then the smooth function

$$\phi(\mathcal{X}_1, \ldots, \mathcal{X}_k) : \mathcal{M} \longrightarrow \mathbb{R}$$

is defined fiberwise via the continuous $(k+1)$-linear map

$$L_a^k(E, \mathbb{R}) \times E \times \cdots \times E \longrightarrow \mathbb{R},$$

where E is the model space for $\mathcal{M}$.

Definition. If ϕ is a smooth k-form on $\mathcal{M}$ and ω is a smooth l-form on $\mathcal{M}$, the *wedge product* of ϕ and ω is the $(k+l)$-form on $\mathcal{M}$ defined by

$$(\phi \wedge \omega)(\mathcal{X}_1, \ldots, \mathcal{X}_{k+l}) = \frac{k!l!}{(k+l)!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}(\sigma)(\mathcal{X}_{\sigma(1)}, \ldots, \mathcal{X}_{\sigma(k+l)}).$$

Here S_{k+l} is the symmetric group on $k+l$ letters and $\operatorname{sgn}(\sigma)$ is the sign of the permutation $\sigma \in S_{k+l}$.

We remind the reader that some authors prefer to define the wedge product using the factor

$$\frac{1}{(k+l)!} \quad \text{instead of} \quad \frac{k!l!}{(k+l)!}.$$

With either convention, the wedge product is bilinear and associative, just as in the case of finite-dimensional manifolds, but not commutative. If ϕ is a k-form and ω an l-form,

$$\phi \wedge \omega = (-1)^{kl} \omega \wedge \phi.$$

Definition. If ϕ is a smooth k-form on $\mathcal{M}$ and $\mathcal{X}$ is a smooth vector field, the *interior product* $\iota_{\mathcal{X}}\phi$ is the smooth $(k-1)$-form on $\mathcal{M}$ defined by the formula

$$\iota_{\mathcal{X}}\phi(\mathcal{X}_2, \ldots, \mathcal{X}_k) = \phi(\mathcal{X}, \mathcal{X}_2, \ldots, \mathcal{X}_k),$$

whenever $\mathcal{X}_1, \ldots, \mathcal{X}_k$ are smooth vector fields on $\mathcal{M}$. (It is readily checked that there is a unique such differential form.) It is easily checked that

$$\iota_{\mathcal{X}}(\phi \wedge \psi) = (\iota_{\mathcal{X}}\phi) \wedge \psi + (-1)^{\deg(\phi)} \phi \wedge \iota_{\mathcal{X}}\psi.$$

Finally, the *exterior derivative* d is the collection of $\mathbb{R}$-linear maps from k-forms to $(k+1)$-forms which satisfy the following axioms, familiar from finite-dimensional differential topology:

1. If ω is a k-form, the value $d\omega(p)$ depends only on ω and its derivatives at p.
2. If f is a smooth real-valued function regarded as a differential 0-form, $d(f)$ is the differential of f defined before.
3. $d \circ d = 0$.
4. If ω is a k-form and ϕ is an l-form, then
$$d(\omega \wedge \phi) = (d\omega) \wedge \phi + (-1)^k \omega \wedge (d\phi).$$
5. If $F : \mathcal{N} \to \mathcal{M}$ is a smooth map, $F^* \circ d = d \circ F^*$ on differential forms.

Just as in the finite-dimensional case, one can prove:

Theorem 1.8.2. *There is a unique of linear maps of real vector spaces,*

$$d : \{\textit{differential } k\textit{-forms}\} \longrightarrow \{\textit{differential } (k+1)\textit{-forms}\},$$

which satisfy the five above axioms. Moreover, these linear maps satisfy the explicit formula

$$\begin{aligned} d\omega(\mathcal{X}_0, \ldots, \mathcal{X}_k) = &\sum (-1)^i \mathcal{X}_i \left(\omega(\mathcal{X}_0, \ldots, \widehat{\mathcal{X}_i}, \ldots, \mathcal{X}_k)\right) \\ &+ \sum_{i<j} (-1)^{i+j} \omega([\mathcal{X}_i, \mathcal{X}_j], \mathcal{X}_0, \ldots, \widehat{\mathcal{X}_i}, \ldots, \widehat{\mathcal{X}_j}, \ldots, \mathcal{X}_k), \quad (1.12) \end{aligned}$$

where the hats denote elements which are left out.

We sketch the proof under the assumption that the corresponding theorem for finite-dimensional manifolds has been established. Using the fifth axiom, we can reduce the proof of uniqueness to the case where the manifold is an open subset of the model space. Since (1.12) is linear over functions, it suffices to establish the formula in the case where $\mathcal{X}_0, \ldots \mathcal{X}_k$ are constant in terms of the local chart, in which case all brackets $[\mathcal{X}_i, \mathcal{X}_j]$ vanish. Thus it suffices to prove uniqueness when the vector fields $\mathcal{X}_0, \ldots \mathcal{X}_k$ are tangent to a $(k+1)$-dimensional affine subspace of the model space, and in this case (1.12) follows from the corresponding formula on this $(k+1)$-dimensional subspace, a finite-dimensional submanifold.

To prove the local existence, one merely defines the exterior derivatives by the formula (1.12) and check that they satisfy all of the axioms. It is simplest to verify the first four axioms by showing that if any of these axioms were to fail, it would have to fail already on finite-dimensional manifolds.

Finally, one notes that if operators satisfying the five axioms are defined and unique for any open subset of the model space, they are defined and unique for any set U_α in a smooth atlas $\{U_\alpha : \alpha \in A\}$ for $\mathcal{M}$. These operators must agree on overlaps $U_\alpha \cap U_\beta$, for $\alpha, \beta \in A$, and hence must fit together to form well-defined operators on $\mathcal{M}$, and the resulting operators must be unique because their restrictions to each U_α are unique.

Differential forms of degree k should be thought of as integrands for integration over compact oriented k-dimensional submanifolds. From this point of view, it is evident that they should be determined by their restrictions to finite-dimensional manifolds, where the above axioms for exterior derivative are familiar.

1.9 Riemannian and Finsler metrics

To construct critical points of functions such as the action or energy, we need to develop the "method of steepest descent" within the context of infinite-dimensional manifolds. And to find the direction of steepest descent, we are

led to seek a notion of gradient, which in the finite-dimensional context depends upon a Riemannian metric. Thus it is important to consider how to extend the notion of Riemannian metric to infinite-dimensional manifolds. It is to be expected that a somewhat stronger theory is possible for Hilbert manifolds than for Banach manifolds. Indeed, we will see that there is no fully satisfactory notion of Riemannian metric or gradient on Banach manifolds. We will need to make do with weaker notions of Finsler metrics and pseudogradients.

Suppose, therefore that $\mathcal{M}$ is a Hilbert manifold modeled on the Hilbert space E. If (U_α, ϕ_α) is a smooth chart on $\mathcal{M}$ and $\tilde{U}_\alpha$ is the subset of $T\mathcal{M}$ projecting to U_α, define

$$\varepsilon_\alpha : \tilde{U}_\alpha \longrightarrow E \qquad \text{by} \qquad \varepsilon_\alpha([\alpha, p, v]) = v.$$

The Hilbert space inner product $(\cdot, \cdot)$ pulls back via ε_α to $T_p\mathcal{M}$: we let

$$(\cdot, \cdot)_\alpha : T_p\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R} \qquad \text{by} \qquad (v, v)_\alpha = (\varepsilon_\alpha(v), \varepsilon_\alpha(w)).$$

Definition. A *Riemannian metric* on a Hilbert manifold $\mathcal{M}$ is a function which assigns to each $p \in \mathcal{M}$ an inner product

$$\langle \cdot, \cdot \rangle_p : T_p\mathcal{M} \times T_p\mathcal{M} \longrightarrow \mathbb{R}$$

such that:

1. There is some constant $c_p > 0$ such that

$$\frac{1}{c_p}(v, v)_\alpha < \langle v, v \rangle_p < c_p(v, v)_\alpha, \qquad \text{for all } v \in T_p\mathcal{M}.$$

   (Thus the topology induced by the Riemannian metric on $T_p\mathcal{M}$ agrees with the model space topology.)

2. $\langle \cdot, \cdot \rangle_p$ varies smoothly with p; in other words, $p \mapsto \langle \cdot, \cdot \rangle_p$ is a smooth covariant tensor field of rank two.

Example 1.9.1. Suppose that we have a proper isometric imbedding of the smooth Riemannian manifold M into $\mathbb{R}^N$. Then the Hilbert manifold $L^2_1(S^1, M)$ can be given a very natural Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ as follows: if $X, Y \in T_\gamma L^2_1(S^1, M)$, we can regard X and Y as maps $X, Y : S^1 \to \mathbb{R}^N$ such that $X(t) \in T_{\gamma(t)}M$ for each $t \in S^1$. We set

$$\langle X, Y \rangle_\gamma = \int_{S^1} [X(t) \cdot Y(t) + X'(t) \cdot Y'(t)] dt.$$

This can be regarded as the pullback of the "flat" Riemannian metric on the Hilbert space $L^2_1(S^1, \mathbb{R}^N)$, and it is smooth because the pullback of a smooth covariant tensor field via a smooth map is smooth.

Definition. If $\mathcal{M}$ is a Hilbert manifold with Riemannian metric $p \mapsto \langle \cdot, \cdot \rangle_p$, the *gradient* of a C^1 function $f : \mathcal{M} \to \mathbb{R}$ is the vector field $\mathrm{grad}(f)$ defined by

$$\langle \mathrm{grad}(f)(p), v \rangle_p = df_p(v), \qquad \text{for all } v \in T_p\mathcal{M}.$$

The idea behind the method of steepest descent for finding critical points of a nonnegative function f is to follow flowlines for the vector field $-\mathrm{grad}(f)$; in favorable cases, these flowlines will converge to a critical point for f.

In the case of Banach manifolds, the metrics best suited to our applications are not Riemannian, but Finsler. If (U_α, ϕ_α) is a smooth chart on a Banach manifold $\mathcal{M}$, $\tilde{U}_\alpha$ is the subset of $T\mathcal{M}$ projecting to U_α and

$$\varepsilon_\alpha : \tilde{U}_\alpha \longrightarrow E \qquad \text{by} \qquad \varepsilon_\alpha([\alpha, p, v]) = v$$

as before, the Banach space norm $\| \cdot \|$ on E pulls back via ε_α to $T_p\mathcal{M}$: we let

$$\| \cdot \|_\alpha : T_p\mathcal{M} \to \mathbb{R} \qquad \text{by} \qquad \|v\|_\alpha = \|\varepsilon_\alpha(v)\|.$$

Definition. A *Finsler metric* on a Banach manifold $\mathcal{M}$ is a function which assigns to each $p \in \mathcal{M}$ a norm

$$\| \cdot \|_p : T_p\mathcal{M} \longrightarrow \mathcal{R}$$

such that

1. There is some constant $c_p > 0$ such that

$$\frac{1}{c_p}\|v\|_\alpha < \|v\|_p < c_p\|v\|_\alpha, \qquad \text{for all } v \in T_p\mathcal{M}.$$

   (Thus the Finsler norm on $T_p\mathcal{M}$ is equivalent to the Banach space norm for the model space.)

2. $\| \cdot \|_p$ varies continuously with p.

Note that any Riemannian metric on a Hilbert manifold determines a Finsler metric: simply define

$$\| \cdot \|_p : T_p\mathcal{M} \longrightarrow \mathbb{R} \qquad \text{by} \qquad \|v\|_p = \sqrt{\langle v, v \rangle_p}.$$

Even in this special case, however, the norm $\| \cdot \|_p$ is only continuous, not smooth as a function on $T\mathcal{M}$.

The Riemannian metric on a Hilbert manifold establishes a norm-preserving vector bundle isomorphism from $T_p\mathcal{M}$ to $T_p^*\mathcal{M}$. We do not have such an isomorphism in the case of a Finsler metric on a Banach manifold, but the norm $\| \cdot \|_p$ on $T_p\mathcal{M}$ induces a dual norm (which we also denote by $\| \cdot \|_p$ for simplicity) on $T_p^*\mathcal{M}$:

$$\|\phi\|_p = \sup\{|\phi(v)| : v \in T_p\mathcal{M} \text{ and } \|v\|_p = 1\}.$$

Example 1.9.2. Suppose that we have a proper isometric imbedding of the smooth Riemannian manifold M into $\mathbb{R}^N$. Given a compact Riemann surface Σ and a real number $p > 2$, the Banach manifold $L^p_1(\Sigma, M)$ can be given a Finsler metric as follows: if $X \in T_f(L^p_1(\Sigma, M)$, we can regard X as a map $X : \Sigma \to \mathbb{R}^N$ such that $X(p) \in T_{f(p)}M$ for each $p \in \Sigma$. We then let $\|X\|_f$ be the L^p_1-norm of X as a mapping into Euclidean space. The Finsler metric $f \mapsto \|\cdot\|_f$ can be regarded as the pullback of the "flat" Finsler metric on the Banach space $L^p_1(\Sigma, \mathbb{R}^N)$.

If $\mathcal{M}$ is a connected Banach manifold with Finsler metric $\|\cdot\|$ and $\gamma : [0, 1] \to \mathcal{M}$ is a C^1 curve, we can define its length $L(\gamma)$ by

$$L(\gamma) = \int_0^1 \|\gamma'(t)\| dt,$$

where integration along a path can be defined as in [43], §1.4. We can then define a distance function $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ by

$$d(p, q) = \inf \left\{ L(\gamma) : \ \gamma : [0, 1] \to \mathcal{M} \text{ is a } C^1 \text{ path, } \gamma(0) = p, \gamma(1) = q \right\}. \quad (1.13)$$

Proposition 1.9.3. *Given a Finsler metric on a regular Banach manifold, the distance function d defined aboveis a metric in the metric space sense, and the metric topology agrees with the manifold topology.*

We apologize to the reader for not giving a complete proof of this Proposition, referring instead to [61] (see the appendix to §2).

However we will describe a proof of the Proposition for Examples 1.8.1 and 1.8.2. Note that it is quite easily verified that the distance function d satisfies

$$d(p, q) = d(q, p), \quad d(p, r) \leq d(p, q) + d(q, r), \quad \text{and} \quad d(p, p) = 0.$$

That only leaves the property $d(p, q) = 0 \Rightarrow p = q$.

Lemma 1.9.4. *In each of our two key examples, $L^2_1(S^1, M)$ and $L^p_1(\Sigma, M)$ with $p > 2$, d is a metric and the metric topology agrees with the manifold topology. Moreover, $L^2_1(S^1, M)$ and $L^p_1(\Sigma, M)$ are complete as metric spaces.*

Proof: Let us consider $L^p_1(\Sigma, M)$. If $f, g \in L^p_1(\Sigma, M)$ and $d(f, g) = 0$, then there exist arbitrarily short paths connecting f and g in $L^p_1(\Sigma, M)$. But a path connecting f and g in $L^p_1(\Sigma, M)$ also connects f and g in the ambient Banach space $E = L^p_1(\Sigma, \mathbb{R}^N)$, so there are arbitrarily short curves connecting f and g in the ambient Banach space. However, by a straightforward modification of the finite-dimensional argument, it is easily verified that if $\gamma : [0, 1] \to E$ is a C^1-map into a Banach space, then

$$\int_0^1 \|\gamma'(t)\| dt \geq \|\gamma(0) - \gamma(1)\|,$$

so two distinct points in E cannot be joined by curves of arbitrarily small length.

Since the map $\omega_i : L_1^p(\Sigma, M) \to L_1^p(\Sigma, \mathbb{R}^N)$ induced by the inclusion $i : M \to \mathbb{R}^N$ is an imbedding, it is now easy to verify that the metric topology agrees with the manifold topology. Finally, since $\omega_i : L_1^p(\Sigma, M) \to L_1^p(\Sigma, \mathbb{R}^N)$ is distance-decreasing, a Cauchy sequence $\{f_i\}$ in $L_1^p(\Sigma, M)$ is also a Cauchy sequence in $L_1^p(\Sigma, \mathbb{R}^N)$, and must therefore converge. Since $L_1^p(\Sigma, M)$ is a closed subset of $L_1^p(\Sigma, \mathbb{R}^N)$, we see that $L_1^p(\Sigma, M)$ must be complete as a metric space.

The case of $L_1^2(S^1, M)$ is treated in the same way.

1.10 Vector fields and ODE's

It is well-known that the global qualitative theory of systems of ordinary differential equations is best formulated within the language of vector fields on finite-dimensional manifolds. This theory, including the fundamental existence and uniqueness theorem for systems of ordinary differential equations, can be extended to infinite-dimensional manifolds. A detailed exposition of this extension is presented in Chapter IV of [43].

Definition. A C^1 curve $\gamma : (a, b) \to \mathcal{M}$ is called an *integral curve* for the vector field $\mathcal{X}$ if

$$\mathcal{X}(\gamma(t)) = \gamma'(t), \qquad \text{for } t \in (a, b), \tag{1.14}$$

where $\gamma'(t)$ is the velocity vector to γ at t.

Just as in the finite-dimensional case, a fundamental existence and uniqueness theorem states that given a smooth vector field $\mathcal{X}$ on $\mathcal{M}$, there is a unique integral curve for $\mathcal{X}$ which passes through any point of $\mathcal{M}$:

Theorem 1.10.1. (Existence and Uniqueness Theorem for Ordinary Differential Equations.) *Suppose that $\mathcal{X}$ is a C^1 vector field on $\mathcal{M}$ and $p \in \mathcal{M}$. Then there is an open neighborhood U of p, an $\epsilon > 0$ and a C^1 map*

$$\phi : (-\epsilon, \epsilon) \times U \longrightarrow \mathcal{M}$$

such that if $\phi_t(q) = \phi(t, q)$ for $t \in (-\epsilon, \epsilon)$ and $q \in U$, then

1. *each curve $t \mapsto \phi_t(q)$ is an integral curve for $\mathcal{X}$,*
2. *any integral curve for $\mathcal{X}$ which passes through U is of the form $t \mapsto \phi_t(q)$ for some $q \in U$,*
3. *ϕ_0 is the inclusion $U \subset \mathcal{M}$, and*
4. *$\phi_t \circ \phi_s = \phi_{t+s}$, whenever, both sides are defined.*

We will call $(-\epsilon, \epsilon) \times U$ a local flow box *for $\mathcal{X}$.*

Idea of proof (following Chapter IV of Lang [43]): The proof is based upon the Contraction Lemma, just like the proof of the Inverse Function Theorem.

We can replace the differential equation (1.14) by its local coordinate representation, and consider the initial value problem

$$\gamma'(t) = f(\gamma(t)), \quad \gamma(0) = q, \tag{1.15}$$

for $\gamma : (a, b) \to V$ and $f : V \to E$, where V is a suitable open subset of a Banach space E. We can assume that

$$\|f\| \leq K \quad \text{and} \quad \|Df\| \leq L$$

on V. Integrating both sides of (1.15) yields the equivalent integral equation

$$\gamma(t) = q + \int_0^t f(\gamma(u))du. \tag{1.16}$$

We can assume that the closed ball $\overline{B_{2\delta}(p)}$ of radius δ about p is contained in V and suppose that $q \in B_\delta(p)$, the open ball of radius δ about p. Let $I = [-\epsilon, \epsilon]$, where $\epsilon > 0$ will be chosen later, and let

$$X = \left\{\gamma : I \to U : \ \gamma \text{ is continuous, } \gamma(0) = q \text{ and } \gamma(I) \subset \overline{B_{2\delta}(p)}\right\}.$$

We can make X into a complete metric space by defining the distance function d by

$$d(\gamma_1, \gamma_2) = \sup\{\|\gamma_1(t) - \gamma_2(t)\| : t \in I\}.$$

If $\gamma \in X$, we set

$$T(\gamma)(t) = q + \int_0^t f(\gamma(u))du.$$

We choose ϵ so that $\epsilon < \delta/K$, and hence

$$\|T(\gamma)(t) - q\| \leq \epsilon K \leq \delta,$$

so $T(\gamma) \in X$. Finally, we note that

$$\begin{aligned} d\left(T(\gamma_1), T(\gamma_2)\right) &= \sup\left\{\|T(\gamma_1)(t) - T(\gamma_2)(t)\| : t \in I\right\} \\ &\leq \epsilon \sup\{\|f(\gamma_1(t)) - f(\gamma_2(t))\| : t \in I\} \\ &\leq \epsilon L \sup\{\|\gamma_1(t)\gamma_2(t)\| : t \in I\} = \epsilon L d(\gamma_1, \gamma_2). \end{aligned}$$

Thus by choosing ϵ so that $\epsilon L < 1$, we can ensure that $T : X \to X$ will be a contraction. Then, by the Contraction Lemma, T has a unique fixed point $\gamma_q \in X$, which must be a solution to the integral equation (1.16). This fixed point γ_q is C^1 and is the unique solution to the initial value problem (1.15).

Thus we can define

$$\phi : (-\epsilon, \epsilon) \times B_\delta(p) \to V \quad \text{by} \quad \phi(t, q) = \gamma_q(t).$$

It is relatively easy to check that properties 2, 3 and 4 of the Theorem hold and that ϕ is continuous. It is a little more challenging to check that ϕ is C^1, and for that we refer the reader to the excellent presentation in [43].

Once we have the Existence and Uniqueness Theorem, we can piece together the locally defined maps to form a map

$$\phi : V \longrightarrow \mathcal{M}, \quad \text{where } V \text{ is an open neighborhood of } \{0\} \times \mathcal{M} \text{ in } \mathbb{R} \times \mathcal{M}.$$

We say that the maps $\{\phi_t\}$ defined by $\phi_t(q) = \phi(t, q)$ form the *one-parameter group of local diffeomorphisms* of $\mathcal{M}$ which corresponds to the vector field $\mathcal{X}$.

1.11 Condition C

We want to apply the existence and uniqueness theorem from the preceding section to find critical points of a C^2 real-valued map $f : \mathcal{M} \to [0, \infty)$ via the method of steepest descent, where $\mathcal{M}$ is an infinite-dimensional manifold. In order to get this method to work, we need to assume that the function f assumes a "compactness condition" introduced by Palais and Smale [62]:

Definition. Suppose that $\mathcal{M}$ is a Banach manifold with a complete Finsler metric. (For example, $\mathcal{M}$ might be a Hilbert manifold with a complete Riemannian metric.) Then a C^2 function $f : \mathcal{M} \to [0, \infty)$ is said to satisfy *condition C* if whenever $\{p_i\}$ is a sequence in $\mathcal{M}$ such that

1. $f(p_i)$ is bounded and
2. $\|df(p_i)\|$ is not bounded away from zero,

then $\{p_i\}$ possesses a subsequence which converges to a critical point for f.

It is easiest to utilize this compactness condition in the case where $\mathcal{M}$ is a Hilbert manifold:

Theorem 1.11.1. *Suppose that $\mathcal{M}$ is a Hilbert manifold with a complete Riemannian metric $\langle \cdot, \cdot \rangle$. If $f : \mathcal{M} \to [0, \infty)$ is a C^2 function which satisfies condition C, $\mathcal{X} = -grad(f)$ and $\{\phi_t\}$ is the local one-parameter group of diffeomorphisms corresponding to $\mathcal{X}$, then*

1. *for each $p \in \mathcal{M}$, $\phi_t(p)$ is defined for all $t \geq 0$, and*
2. *there is a sequence $t_i \to \infty$ such that $\{\phi_{t_i}(p)\}$ converges to a critical point for f.*

The proof will be based upon a collection of inequalities. Suppose that $0 < t_1 < t_2$. Then

$$\begin{aligned} f(\phi_{t_1}(p)) - f(\phi_{t_2}(p)) &= -\int_{t_1}^{t_2} \frac{d}{dt} f(\phi_t(p)) dt \\ &= -\int_{t_1}^{t_2} df(\phi_t(p))(\mathcal{X}(\phi_t(p)) dt = \int_{t_1}^{t_2} \langle \text{grad}(f)(\phi_t(p)), \mathcal{X}(\phi_t(p)) \rangle dt \\ &= \int_{t_1}^{t_2} \|\text{grad}(f)(\phi_t(p))\|^2 dt = \int_{t_1}^{t_2} \|df(\phi_t(p))\|^2 dt. \quad (1.17) \end{aligned}$$

On the other hand, using the metric d on the Riemannian manifold $\mathcal{M}$, we have

$$d(\phi_{t_1}(p), \phi_{t_2}(p)) \leq \int_{t_1}^{t_2} \left\| \frac{d}{dt}(\phi_t(p)) \right\| dt$$
$$\leq \int_{t_1}^{t_2} \|\mathcal{X}(\phi_t(p))\| dt = \int_{t_1}^{t_2} \|df(\phi_t(p))\| dt. \quad (1.18)$$

Now we use the Cauchy-Schwarz inequality to obtain

$$d(\phi_{t_1}(p), \phi_{t_2}(p))^2 \leq \left[\int_{t_1}^{t_2} \|df(\phi_t(p))\| dt \right]^2$$
$$\leq (t_2 - t_1) \int_{t_1}^{t_2} \|df(\phi_t(p))\|^2 dt = (t_2 - t_1)(f(\phi_{t_1}(p)) - f(\phi_{t_2}(p))). \quad (1.19)$$

Let $\bar{t} = \sup\{t \in \mathbb{R} : \phi_t(p) \text{ is defined}\}$, and let $\{t_i\}$ be a sequence of real numbers $< \bar{t}$ such that $t_i \to \bar{t}$. It follows from (1.19) that $\{\phi_{t_i}(p)\}$ is a Cauchy sequence in $\mathcal{M}$. Since $(\mathcal{M}, d)$ is a complete metric space, $\phi_{t_i}(p) \to q$, for some $q \in \mathcal{M}$. But by the Existence and Uniqueness Theorem there is a flow box $(-\epsilon, \epsilon) \times U$ containing $(0, q)$. This implies that the curve $t \mapsto \phi_t(p)$ can be extended beyond $\bar{t}$, giving a contradiction. Thus we see that $\phi_t(p)$ is defined for all $t \geq 0$, and the first statement of the Theorem is proven.

Next, it follows from (1.17) that

$$\int_0^\infty \|\text{grad}(f)(\phi_t(p))\|^2 dt < \infty,$$

and hence there must exist a sequence $t_i \to \infty$ such that

$$\|df(\phi_{t_i}(p))\| = \|\text{grad}(f)(\phi_{t_i}(p))\| \to 0.$$

By Condition C, a subsequence of $\{\phi_{t_i}(p)\}$ converges to a critical point for f, finishing the proof of the Theorem.

Of course, we would like a version of the above Theorem to hold for the case of a C^2 function $f : \mathcal{M} \to \mathbb{R}$, where $\mathcal{M}$ is only a Banach manifold. However, it is not possible to define Riemannian metrics on $\mathcal{M}$ in this case, so we need a replacement for the notion of gradient, such as the following, similar to a definition suggested by Palais [59]:

Definition. Suppose that $f : \mathcal{M} \to \mathbb{R}$ is a C^2 function on a Banach manifold which has a Finsler metric and let U be an open subset of $\mathcal{M}$. A C^1 vector field $\mathcal{X} : U \to T\mathcal{M}$ is called a *pseudogradient* for f over U if there is a constant $\epsilon > 0$ such that for each $p \in U$,

1. $df_p(\mathcal{X}(p)) \geq \epsilon \|df_p\|^2$,
2. $\|\mathcal{X}(p)\| \leq (1/\epsilon)\|df_p\|$.

In both inequalities, $\|df_p\| = \sup\{|df_p(v)| : v \in T_p\mathcal{M} \text{ and } \|v\| \leq 1\}$, which is the dual norm on the cotangent space to $\mathcal{M}$ at p.

Of course, in the case of a Hilbert manifold, $\mathcal{X} = \text{grad}(f)$ satisfies both conditions in the definition with $\epsilon = 1$; in other words, a gradient on a Hilbert manifold is also a pseudogradient. The definition was set up so that the following Theorem would be true:

Theorem 1.11.2. *Suppose that $\mathcal{M}$ is a Banach manifold with a complete Finsler metric $\|\cdot\|$. Suppose that $f : \mathcal{M} \to [0, \infty)$ is a C^2 function which satisfies condition C. Let*

$$K = \{p \in \mathcal{M} : df(p) = 0\}$$

and let $U = \mathcal{M} - K$. If $\mathcal{X}$ is a pseudogradient for f on U, and $\{\phi_t\}$ is the local one-parameter group of diffeomorphisms corresponding to $-\mathcal{X}$, then

1. *for each $p \in \mathcal{M}$, $\phi_t(p)$ is defined for all $t \geq 0$, and*
2. *there is a sequence $t_i \to \infty$ such that $\{\phi_{t_i}(p)\}$ converges to a critical point for f.*

The proof is almost identical to the proof of Theorem 1.11.1 (except that we call the vector field $-\mathcal{X}$ instead of $\mathcal{X}$). Assuming that p is not a critical point for f and that $0 < t_1 < t_2$, we use the first condition in the definition of pseudogradient to replace (1.17) by the inequality

$$\begin{aligned} f(\phi_{t_1}(p)) - f(\phi_{t_2}(p)) &= -\int_{t_1}^{t_2} \frac{d}{dt} f(\phi_t(p))dt \\ &= \int_{t_1}^{t_2} df(\phi_t(p))(\mathcal{X}(\phi_t(p))dt \geq \epsilon \int_{t_1}^{t_2} \|df(\phi_t(p))\|^2 dt. \quad (1.20) \end{aligned}$$

We use the second condition in the definition of pseudogradient to replace (1.18) by

$$d(\phi_{t_1}(p), \phi_{t_2}(p)) \leq \int_{t_1}^{t_2} \|\mathcal{X}(\phi_t(p))\|dt \leq \frac{1}{\epsilon} \int_{t_1}^{t_2} \|df(\phi_t(p))\|dt. \quad (1.21)$$

Then we use the Cauchy-Schwarz inequality exactly as before to obtain

$$\begin{aligned} d(\phi_{t_1}(p), \phi_{t_2}(p))^2 &\leq \frac{1}{\epsilon^2} \left[\int_{t_1}^{t_2} \|df(\phi_t(p))\|dt\right]^2 \\ &\leq \frac{t_2 - t_1}{\epsilon^2} \int_{t_1}^{t_2} \|df(\phi_t(p))\|^2 dt \leq \frac{t_2 - t_1}{\epsilon^3}(f(\phi_{t_1}(p)) - f(\phi_{t_2}(p))). \quad (1.22) \end{aligned}$$

We can now use (1.22) instead of (1.19) to show that $\phi_t(p)$ is defined for all $t \geq 0$. Finally, it follows from (1.20) that

$$\int_0^\infty \|df(\phi_t(p))\|^2 dt < \infty,$$

which enables us to find a sequence $t_i \to \infty$ such that $\|df(\phi_{t_i}(p))\| \to 0$, and by condition C, a subsequence of $\{\phi_{t_i}(p)\}$ converges to a critical point for f.

The only problem now is to show that given a C^2 function $f : \mathcal{M} \to [0, \infty)$ on a Banach manifold, we can construct a corresponding pseudogradient on $U = \mathcal{M} - K$, where K is the set of critical points for f. The standard technique for constructing a pseudogradient consists of constructing pseudogradients over each open set of an open cover of U and then piecing these together using a partition of unity.

Let $\mathcal{M}$ be a metrizable infinite-dimensional smooth manifold modeled on a Banach space. According to a well-known theorem of Stone, $\mathcal{M}$ must be paracompact. That means that every open cover of $\mathcal{M}$ must have an open locally finite refinement.

Definition. Let $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ be an open cover of $\mathcal{M}$. A *partition of unity* subordinate to $\mathcal{U}$ is a collection $\{\psi_\alpha : \alpha \in A\}$ of continuous real-valued functions on $\mathcal{M}$ such that

1. $\psi_\alpha : \mathcal{M} \to [0, 1]$,
2. the support of ψ_α is a closed subset of U_α,
3. if $p \in \mathcal{M}$, there is an open neighborhood V of p which intersects the supports of only finitely many ψ_α, and
4. $\sum \psi_\alpha = 1$.

It is known (and proven in topology texts) that any open cover of a paracompact Hausdorff space possesses a subordinate continuous partition of unity.

Moreover, as proven in Lang [43], C^∞ Hilbert manifolds possess C^∞ partitions of unity. However, for Banach manifolds, we encounter a perhaps unexpected obstacle. Banach manifolds need not possess C^∞ partitions of unity. As pointed out in [16], for example, to construct C^k partitions of unity one needs to be able to construct nontrivial real-valued C^k functions on the model Banach space E with bounded support.

Fortunately, in the case where Σ is a Riemann surface and $p > 2$ the Banach manifold $L^p_1(\Sigma, M)$ does possess partitions of unity of class C^2. To see why, we notice that the function

$$f : L^p_1(\Sigma, \mathbb{R}) \longrightarrow \mathbb{R} \qquad \text{defined by} \qquad f(\phi) = \|\phi\|^p$$

is C^2, by an argument similar to that given in Example 1.2.3. Let $g : \mathbb{R} \to [0, 1]$ be a smooth function such that

1. $g(s) = 1$ when $|s| \leq 1$, and
2. $g(s) = 0$ when $|s| \geq 2$.

Then the map

$$f_\epsilon : L_1^p(\Sigma, \mathbb{R}^n) \to [0,1] \quad \text{defined by} \quad f_\epsilon(\phi) = g\left(\frac{2f(\phi)}{\epsilon}\right)$$

is a C^2 function which equals one on a small neighborhood of the origin and has support contained in the set $\{\phi \in L_1^p(\Sigma, \mathbb{R}^n) : \|\phi\| \leq \epsilon\}$. Using local coordinates, we can transport this function to $L_1^p(\Sigma, M)$ thereby obtaining a C^2 function $f : L_1^p(\Sigma, M) \to \mathbb{R}$ which is one in a neighborhood of a given point p and vanishes outside a given open neighborhood of p.

Using this function to start with, we can follow the familiar argument (such as given in Lang [43], Chapter II, §3) to construct C^2 partitions of unity subordinate to any open cover on the smooth manifold $L_1^p(\Sigma, M)$.

Lemma 1.11.3. *If $f : L_1^p(\Sigma, M) \to \mathbb{R}$ is a C^2 function, where $p \geq 2$, then f possesses a C^2 pseudogradient $\mathcal{X}$ which is tangent to every $L_k^p(\Sigma, M) \subset L_1^p(\Sigma, M)$, for $k \in \mathbb{N}$.*

Proof: Suppose that $0 < \epsilon < 1$. If p is not a critical point for f, we can choose a unit vector $u \in T_p\mathcal{M}$ such that $|df_p(u)| > \sqrt{\epsilon}\|df_p\|$; then

$$v = \sqrt{\epsilon}\|df_p\|u \qquad \text{satisfies} \qquad \|v\| \leq \|df_p\|, \quad df_p(v) \geq \epsilon\|df_p\|^2,$$

the two conditions in the definition of pseudogradient at the point p. We can extend v to a smooth vector field on some neighborhood of p which is a pseudogradient; for example, we could choose it to be constant in terms of some smooth chart. Thus we can construct a pseudogradient on an open neighborhood about any point p which is not in the set K of critical points of F. If $\mathcal{M}$ admits C^2 partitions of unity, one can piece together a C^2 pseudogradient on $\mathcal{M} - K$.

In the above construction, we can choose v to lie in the dense subspace $L_k^p(\Sigma, M)$ of $L_1^p(\Sigma, M)$ and the C^2 partition of unity on $L_1^p(\Sigma, M)$ can be chosen so that it restricts to a C^2 partition of unity on $L_k^p(\Sigma, M)$ for every $k \geq 1$. When this is done, the pseudogradient will be tangent to every $L_k^p(\Sigma, M)$, finishing the proof of the Lemma.

1.12 Topological constraints give critical points

Suppose now that $\mathcal{M}$ is a Banach manifold with a complete Finsler metric. Suppose that $f : \mathcal{M} \to [0, \infty)$ is a C^2 function and let

$$\mathcal{M}^a = \{p \in \mathcal{M} : f(p) \leq a\}.$$

Definition. An *ambient isotopy* of $\mathcal{M}$ is a smooth map $\Psi : \mathcal{M} \times [0,1] \to \mathcal{M}$ such that if $\psi_t : \mathcal{M} \to \mathcal{M}$ is defined by $\psi_t(p) = \Psi(p,t)$, then ψ_t is a diffeomorphism for each $t \in [0,1]$ and $\psi_0 = \mathrm{id}$.

Theorem 1.12.1. (Deformation Theorem) *Suppose that $\mathcal{M}$ is a Banach manifold with a complete Finsler metric. If $f : \mathcal{M} \to \mathbb{R}$ is a nonnegative C^2*

function satisfying condition C and there are no critical points p for f such that $a \leq f(p) \leq b$, then

1. *$\mathcal{M}^a$ is a strong deformation retract of $\mathcal{M}^b$.*
2. *there is a smooth ambient isotopy $\Psi = \{\psi_t : t \in [0,1]\}$ of $\mathcal{M}$ such that $\psi_1(\mathcal{M}^b) \subset \mathcal{M}^a$.*

Proof: It follows from Condition C that there is an $\epsilon > 0$ such that there are no critical points p for f such that $a - \epsilon < f(p) < b + \epsilon$. Using a partition of unity, we construct a smooth function $\eta : \mathcal{M} \to [0,1]$ such that

1. $\eta \equiv 1$ on $\{p \in M : a \leq f(p) \leq b\}$,
2. $\eta \equiv 0$ outside $\{p \in M : a - \epsilon < f(p) < b + \epsilon\}$.

Let $\mathcal{Y}$ be a pseudogradient for f on $U = \mathcal{M} - K$, where K is the critical locus of F and set $\mathcal{X} = -\eta\mathcal{Y}$.

Since f satisfies Condition C, there is a constant $k > 0$ such that $\|df\| \geq k$ on $\{p \in M : a \leq f(p) \leq b\}$. Indeed, if not, there would exist a sequence $\{p_i\}$ in $\{p \in M : a \leq f(p) \leq b\}$ such that $\|df(p_i)\| \to 0$. By condition C, a subsequence of $\{p_i\}$ would converge to a critical point for f in $\{p \in M : a \leq f(p) \leq b\}$, contradicting the hypothesis of the Theorem.

Theorem 1.10.2 shows that the one-parameter group $\{\phi_t : t \in \mathbb{R}\}$ of local diffeomorphisms determined by $\mathcal{X}$ is globally defined for all $t \geq 0$. We claim that if $p \in \mathcal{M}^b$, then $\phi_t(p) \in \mathcal{M}^a$ for $t > (b-a)/(\epsilon k)$. Indeed, if $\phi_t(p) \notin \mathcal{M}^a$, it follows from (1.20) that

$$f(p) - f(\phi_t(p)) \geq \epsilon \int_0^t \|df(\phi_\tau(p))\| d\tau \geq \epsilon k t,$$

and hence $\epsilon k t \leq b - a$, or equivalently, $t \leq (b-a)/(\epsilon k)$. We now set $\psi_t = \phi_{ct}$, where $c = 2(b-a)/(\epsilon k)$, and define

$$\Psi : \mathcal{M} \times [0,1] \to \mathcal{M} \qquad \text{by} \qquad \Psi(p,t) = \psi_t(p).$$

Then Ψ is an ambient isotopy such that $\psi_1(\mathcal{M}^b) \subset \mathcal{M}^a$.

This proves the second assertion of the Theorem. For the first, we let $\tau(p)$ be the first time t such that $\phi_t(p) \in \mathcal{M}^a$ and define $\Phi : \mathcal{M}^b \times [0,1] \to \mathcal{M}^b$ by

$$\Phi(p,t) = \begin{cases} \phi_{t(\tau(p))}, & \text{for } p \in \mathcal{M}^b - \mathcal{M}^a, \\ p, & \text{for } p \in \mathcal{M}^a. \end{cases}$$

Then Φ is a strong deformation retraction from $\mathcal{M}^b$ to $\mathcal{M}^a$, finishing the proof of the Theorem.

Definition. Let $\mathcal{F}$ be a family of subsets of $\mathcal{M}$. We say that $\mathcal{F}$ is *ambient isotopy invariant* if

$$A \in \mathcal{F} \qquad \Rightarrow \qquad \psi_1(A) \in \mathcal{F}$$

whenever $\{\psi_t : t \in [0,1]\}$ is an ambient isotopy of $\mathcal{M}$.

Theorem 1.12.2. (Minimax Theorem) *Suppose that $\mathcal{M}$ is a smooth manifold with a complete Finsler metric. Suppose, moreover, that $f : \mathcal{M} \to [0,\infty)$ is a C^2 function satisfying condition C and $\mathcal{F}$ is a nonempty family of subsets of $\mathcal{M}$ which is ambient isotopy invariant. Then*

$$\textit{Minimax}(f,\mathcal{F}) = \inf\{\sup\{f(p) : p \in A\} : A \in \mathcal{F}\}$$

is a critical value for f.

Proof: Let $c = \text{Minimax}(f,\mathcal{F})$. If c is not a critical value for f, then Condition C implies that there exists an $\epsilon > 0$ such that there are no critical points $p \in \mathcal{M}$ with $f(p) \in (c-\epsilon, c+\epsilon)$. But by definition of $\text{Minimax}(f,\mathcal{F})$, there exists an $A \in \mathcal{F}$ such that $A \subset \mathcal{M}^{c+\epsilon}$. Theorem 1.11.1 then gives a smooth isotopy $\{\psi_t : t \in [0,1]\}$ such that

$$\psi_1(\mathcal{M}^{c+\epsilon}) \subset \mathcal{M}^{c-\epsilon}, \quad \text{so} \quad \psi_1(A) \subset \mathcal{M}^{c-\epsilon}.$$

But then $\psi_1(A) \in \mathcal{F}$ showing that $\text{Minimax}(f,\mathcal{F}) \leq c-\epsilon$, a contradiction.

For example, we could let $\mathcal{M}_0$ be a component of $\mathcal{M}$ and let

$$\mathcal{F} = \{A \subseteq \mathcal{M} : A \subseteq \mathcal{M}_0\}.$$

Then $\mathcal{F}$ is ambient isotopy invariant and the previous Theorem implies:

Corollary 1.12.3. *Suppose that $\mathcal{M}$ is a smooth manifold with a complete Finsler metric, and that $f : \mathcal{M} \to [0,\infty)$ is a C^2 function satisfying condition C. Then f assumes its minimum value on each component of $\mathcal{M}$.*

We can also use invariants from algebraic topology to construct critical points. For example, suppose that M is simply connected and $[\alpha]$ is a nonzero element in $\pi_k(\mathcal{M})$, the k-th homotopy group of $\mathcal{M}$. In this case,

$$\mathcal{F}_{[\alpha]} = \{\ h(S^k) \text{ such that } h : S^k \to \mathcal{M} \text{ is a continuous map representing } [\alpha]\}$$

is ambient isotopy invariant, and hence there is a minimax critical point corresponding to the homotopy class $[\alpha]$. Alternatively, suppose that x is a nonzero element in $H_k(\mathcal{M};\mathbb{R})$, the singular homology group of $\mathcal{M}$ of degree k with real coefficients, and let

$$\begin{aligned}\mathcal{F}_x = \{&h(A) \text{ such that } A \text{ is a compact oriented manifold of dimension } k\\ &\text{with fundamental class } [\mu_A] \text{ and } h : A \to \mathcal{M} \text{ is a}\\ &\quad\text{continuous map with } h_*([\mu_A]) \text{ a nonzero multiple of } x\}.\end{aligned}$$

Once again $\mathcal{F}_x$ is ambient isotopy invariant, and one obtains a minimax critical point corresponding to the homology class x. For a third example, we suppose

that θ is a differential k-form on $\mathcal{M}$ such that $d\theta = 0$, and let

$$\mathcal{F}_\theta = \{\ h(A) \text{ such that } A \text{ is a compact oriented manifold of dimension } k \text{ and } h : A \to \mathcal{M} \text{ is a } C^1\text{-map such that } \int_A h^*\theta \neq 0 \Big\}.$$

It follows from the Homotopy Lemma to be proven in the next section that $\mathcal{F}_\theta$ is ambient isotopy invariant, so once again we obtain a minimax critical point corresponding to the differential form θ.

Theorem 1.12.4. *Suppose that K is the set of critical points for f whose critical value is Minimax$(f, \mathcal{F})$ and that U is an open neighborhood of K within $\mathcal{M}$. If the elements in $\mathcal{F}$ are compact, then there is an element $A \in \mathcal{F}$ such that for some $\epsilon > 0$,*

$$\phi_t(A) \subset \mathcal{M}^{c-\epsilon} \cup U, \qquad \text{for all } t \geq 0.$$

Proof: We follow an argument presented by Klingenberg [42]. First choose an open subset V of $\mathcal{M}$ such that $K \subset V \subset U$ and the distance from V to $\mathcal{M} - U$ is a positive number δ. It follows from condition C that the norm of df is bounded below on $\mathcal{M} - V$—otherwise, a sequence of points in $\mathcal{M} - V$ would converge to a critical point which would not lie in K. Moreover, any orbit of $-\mathcal{X}$ which enters V can only leave U if it travels for a distance δ in $U - V$, but then it follows from (1.21) that the value of f must decrease by at least δ. Thus if we set $\epsilon = \delta/2$, there will exist an element $A \in \mathcal{F}$ such that $A \subset \mathcal{M}^{c+\epsilon}$, and when T is sufficiently large, every orbit starting in A will have either passed below level $c - \epsilon$ or entered V at least once in time T. Hence $A' = \phi_T(A)$ is an element of $\mathcal{F}$ and $\phi_t(A') \subset \mathcal{M}^{c-\epsilon} \cup U$ for all $t \geq 0$.

1.13 de Rham cohomology

Once one has C^2 partitions of unity on Banach manifolds, it is relatively straightforward to extend de Rham cohomology to Banach manifolds. Indeed, C^2 partitions of unity will enable us to piece together C^1 differential forms with C^1 exterior derivatives, elements of the vector space

$$\Omega^k(\mathcal{M}) = \{C^1 \text{ differential forms } \omega \text{ on } \mathcal{M} \text{ of degree } k : d\omega \in C^1\}.$$

We make the $\Omega^k(\mathcal{M})$'s into a cochain complex in which the differential is the exterior derivative

$$d : \Omega^k(\mathcal{M}) \longrightarrow \Omega^{k+1}(\mathcal{M}).$$

We say that an element $\omega \in \Omega^k(\mathcal{M})$ is *closed* if $d\omega = 0$ and *exact* if $\omega = d\theta$ for some $\theta \in \Omega^{k-1}(\mathcal{M})$. Since $d \circ d = 0$, every exact k-form is closed. The quotient space

$$H^k_{dR}(\mathcal{M}; \mathbb{R}) = \frac{\text{closed elements of } \Omega^k(\mathcal{M})}{\text{exact elements of } \Omega^k(\mathcal{M})}.$$

is called the *de Rham cohomology* of $\mathcal{M}$. If $\omega \in \Omega^k(\mathcal{M})$ is closed, we let $[\omega]$ denote its cohomology class in $H^k_{dR}(\mathcal{M};\mathbb{R})$. In the terminology of algebraic topology, the de Rham cohomology is the cohomology of the cochain complex

$$\cdots \to \Omega^{k-1}(\mathcal{M}) \to \Omega^k(\mathcal{M}) \to \Omega^{k+1}(\mathcal{M}) \to \cdots . \tag{1.23}$$

Note that de Rham cohomology has a *cup product* defined by

$$[\omega] \cup [\phi] = [\omega \wedge \phi],$$

which makes the direct sum

$$H^*_{dR}(\mathcal{M};\mathbb{R}) = \sum_{k=0}^{\infty} H^k_{dR}(\mathcal{M};\mathbb{R})$$

into a graded commutative algebra over the ring of smooth real-valued functions on M. Moreover, the cup product behaves well under smooth maps: If $F : \mathcal{M} \to \mathcal{N}$ is a smooth map, the linear map F^* on differential forms induces a linear map

$$F^* : H^k_{dR}(\mathcal{N};\mathbb{R}) \longrightarrow H^k_{dR}(\mathcal{M};\mathbb{R}) \quad \text{such that} \quad F^*([\omega] \cup [\phi]) = F^*[\omega] \cup F^*[\phi].$$

Moreover, the identity map on $\mathcal{M}$ induces the identity on de Rham cohomology and if $F : \mathcal{M} \to \mathcal{N}$ and $G : \mathcal{N} \to \mathcal{P}$ are smooth maps, then $(G \circ F)^* = F^* \circ G^*$, so

$$\mathcal{M} \mapsto H^k(\mathcal{M};\mathbb{R}), \qquad (F : \mathcal{M} \to \mathcal{N}) \mapsto (F^* : H^k_{dR}(\mathcal{N};\mathbb{R}) \to H^k_{dR}(\mathcal{M};\mathbb{R}))$$

is a contravariant functor from the category of smooth manifolds and smooth maps to the category of real vector spaces and linear maps.

Lemma 1.13.1. (Poincaré Lemma.) *If U is a convex open subset of a Banach space E, or more generally any contractible open subset of E, then the de Rham cohomology of U is trivial:*

$$H^k_{dR}(U;\mathbb{R}) \cong \begin{cases} \mathbb{R} \text{ if } k = 0, \\ 0 \text{ if } k \neq 0. \end{cases}$$

One can modify the proof that is used in the finite-dimensional case. We only sketch the key ideas—the reader should refer to Lang [43] for details. Since the inclusion from a point into the convex set U is a homotopy equivalence, the Poincaré Lemma is an immediate consequence of

Lemma 1.13.2. (Homotopy Lemma.) *Smoothly homotopic maps $F, G : \mathcal{M} \to \mathcal{N}$ induce the same map on cohomology,*

$$F^* = G^* : H^k_{dR}(\mathcal{N};\mathbb{R}) \longrightarrow H^k_{dR}(\mathcal{M};\mathbb{R}).$$

On the other hand via functoriality, this follows from the special case of the Homotopy Lemma for the inclusion maps

$$i_0, i_1 : \mathcal{M} \longrightarrow [0,1] \times \mathcal{M}, \qquad i_0(p) = (0,p), \quad i_1(p) = (1,p).$$

Indeed, if $H : [0,1] \times \mathcal{M} \to \mathcal{N}$ is a smooth homotopy from F to G, then by definition of homotopy, $F = H \circ i_0$ and $G = H \circ i_1$, so

$$i_0^* = i_1^* \Rightarrow F^* = i_0^* \circ H^* = i_1^* \circ H^* = G^*.$$

This special case, however, can be established by integrating over the fiber of the projection on the second factor $[0,1] \times \mathcal{M} \to \mathcal{M}$. More precisely, let t be the standard coordinate on $[0,1]$, T the vector field tangent to the fiber of $[0,1] \times \mathcal{M}$ such that $dt(T) = 1$. We then define integration over the fiber

$$\pi_* : \Omega^k([0,1] \times \mathcal{M}) \to \Omega^{k-1}(\mathcal{M}) \quad \text{by} \quad \pi_*(\omega)(p) = \int_0^1 (\iota_T \omega)(t,p) dt.$$

Here the interior product $(\iota_T\omega)(t,p)$ is an element of $\Lambda^k T^*_{(t,p)}([0,1] \times \mathcal{M})$ and the integration is possible because the exterior power at (t,p) is canonically isomorphic to $\Lambda^k T^*_{(0,p)}([0,1] \times \mathcal{M})$. The key to proving that $i_0^* = i_1^*$ in cohomology is the "cochain homotopy" formula

$$i_1^*\omega - i_0^*\omega = d(\pi_*(\omega)) + \pi_*(d\omega).$$

This formula can be verified by using naturality to reduce the proof to the finite-dimensional case, and then calculating in local coordinates just as in the familiar finite-dimensional treatment found in [10]. Note that

$$d\omega = 0 \Rightarrow i_1^*\omega - i_0^*\omega = d(\pi_*(\omega)) \Rightarrow [i_1^*\omega] = [i_0^*\omega],$$

and hence on the cohomology level $i_0^* = i_1^*$. This finishes our sketch of the proof of the Homotopy Lemma and the Poincaré Lemma.

Remark 1.13.3. If $\mathcal{N}$ is a submanifold of $\mathcal{M}$ with inclusion map $i : \mathcal{N} \to \mathcal{M}$, we let

$$\Omega^k(\mathcal{M}, \mathcal{N}) = \ker(i^* : \Omega^k(\mathcal{M}) \to \Omega^k(\mathcal{N})),$$

and note that the exterior derivative makes this into the k-th cochain group of a cochain complex $\Omega^*(\mathcal{M}, \mathcal{N})$. The cohomology of this complex is called the *relative de Rham cohomology* of the pair $(\mathcal{M}, \mathcal{N})$ and is denoted by $H^k_{dR}(\mathcal{M}, \mathcal{N}; \mathbb{R})$.

The short exact sequence of cochain complexes

$$0 \to \Omega^*(\mathcal{M}, \mathcal{N}) \to \Omega^*(\mathcal{M}) \to \Omega^*(\mathcal{N}) \to 0. \tag{1.24}$$

yields a long exact sequence via the "snake lemma" (Theorem 2.16 in [35]) from algebraic topology:

$$\begin{aligned} \cdots \to H^k_{dR}(\mathcal{M}, \mathcal{N}; \mathbb{R}) \to H^k_{dR}(\mathcal{M}; \mathbb{R}) \to H^k_{dR}(\mathcal{N}; \mathbb{R}) \\ \to H^{k+1}_{dR}(\mathcal{M}, \mathcal{N}; \mathbb{R}) \to H^{k+1}_{dR}(\mathcal{M}; \mathbb{R}) \to \cdots . \end{aligned}$$

This is very useful for calculating de Rham cohomology.

For us, one of the primary uses of differential forms will be to calculate cohomology of infinite-dimensional manifolds. It is important to realize that the de Rham cohomology is the same as the singular or Čech cohomology with real coefficients that is studied in algebraic topology:

Theorem 1.13.4. (de Rham Theorem.) *Suppose that either $\mathcal{M} = L^p_k(\Sigma, M)$ where $pk > dim(\Sigma)$ or $\mathcal{M}$ is finite-dimensional. Suppose, moreover that $\mathcal{M}$ admits C^2 partitions of unity. Then the cohomology of the cochain complex $\Omega^*(\mathcal{M})$ is isomorphic to the Čech cohomology of $\mathcal{M}$.*

The proof (due to André Weil) is via the zig-zag construction described in the excellent text on de Rham theory by Bott and Tu [10], which we follow closely.

In the case where $\mathcal{M}$ is finite-dimensional, we let $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ be a locally finite open cover of $\mathcal{M}$ by sets which are geodesically convex with respect to a Riemannian metric on $\mathcal{M}$. If $\mathcal{M} = L^p_k(\Sigma, M)$ where $pk > \dim(\Sigma)$, we construct an open cover $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ in which each open set U_α is of the form

$$U_\alpha = \{g \in L^p_k(\Sigma, M) : \|g - f\|_{C^0} < \delta\}.$$

The Sobolev Lemma guarantees the existence of such an open cover. We choose δ so small that if $p, q \in M$ and $d(p, q) < 2\delta$, then there is a unique minimizing geodesic

$$\gamma_{p,q} : [0, 1] \to M \quad \text{such that} \quad \gamma_{p,q}(0) = p, \quad \gamma_{p,q}(1) = q,$$

and moreover this geodesic depends smoothly on p and q. (Then any two points in a δ-ball about a given point can be connected by a unique such geodesic.) If g_1 and g_2 are two elements of U_α, we can then define a path

$$\Gamma_{g_1,g_2} : [0, 1] \to L^p_k(\Sigma, M) \quad \text{by} \quad \Gamma_{g_1,g_2}(t)(p) = \gamma_{g_1(p),g_2(p)}(t).$$

It is easily checked in either case that the intersection of any collection of elements from $\mathcal{U}$ is contractible. Readers familiar with cohomology theory will remember that the Čech cohomology of such an open covering is the same as the Čech cohomology of $\mathcal{M}$ (and we do not need to take direct limits). Such an open cover is often called a "good" cover or "Leray" cover.

We now construct a double complex $K^{*,*}$ in which the (p, q)-element is

$$K^{p,q} = \check{C}^p(\mathcal{U}, \Omega^q),$$

which is defined to be the space of functions ω which assign to each distinct ordered $(p + 1)$-tuple $(\alpha_0, \ldots, \alpha_p)$ of indices such that $U_{\alpha_0} \cap \cdots \cap U_{\alpha_p} \neq 0$ an element

$$\omega_{\alpha_0 \cdots \alpha_p} \in \Omega^q(U_{\alpha_0} \cap \cdots \cap U_{\alpha_p})$$

in such a way that if the order of elements in a sequence is permuted, $\omega_{\alpha_0,\ldots,\alpha_p}$ changes by the change of the permutation; thus

$$\omega_{\alpha_0\alpha_1} = -\omega_{\alpha_1\alpha_0}, \quad \omega_{\alpha\alpha} = 0, \qquad \text{and so forth.}$$

We have two differentials on the double complex, the exterior derivative

$$d : \check{\mathrm{C}}^p(\mathcal{U}, \Omega^q) \to \check{\mathrm{C}}^p(\mathcal{U}, \Omega^{q+1}) \quad \text{defined by} \quad (d\omega)_{\alpha_0 \cdots \alpha_p} = d\omega_{\alpha_0 \cdots \alpha_p},$$

and the Čech differential

$$\delta : \check{\mathrm{C}}^p(\mathcal{U}, \Omega^q) \to \check{\mathrm{C}}^{p+1}(\mathcal{U}, \Omega^q)$$

defined by

$$(\delta\omega)_{\alpha_0 \cdots \alpha_{p+1}} = \sum_{i=0}^{p+1} (-1)^i \omega_{\alpha_0 \cdots \hat{\alpha}_i \cdots \alpha_{p+1}},$$

the forms on the right being restricted to the intersection.

The first differential is exact except when $q = 0$ by the Poincaré Lemma, while in the case $q = 0$ we find that

$$[\text{Kernel of } d : \check{\mathrm{C}}^p(\mathcal{U}, \Omega^0) \to \check{\mathrm{C}}^p(\mathcal{U}, \Omega^1)] = \check{\mathrm{C}}^p(\mathcal{U}; \mathbb{R}),$$

the space of Čech cocycles for the covering $\mathcal{U}$ on $\mathcal{M}$. The Čech cohomology of the cover $\mathcal{U}$ is by definition the cohomology of the cochain complex

$$\cdots \to \check{\mathrm{C}}^{p-1}(\mathcal{U}; \mathbb{R}) \to \check{\mathrm{C}}^p(\mathcal{U}; \mathbb{R}) \to \check{\mathrm{C}}^{p+1}(\mathcal{U}; \mathbb{R}) \to \cdots \tag{1.25}$$

and is denoted by $\check{\mathrm{H}}^p(\mathcal{M}; \mathbb{R})$.

The second differential is exact except when $p = 0$ and it is at this point that the C^2 partition of unity $\{\psi_\alpha : \alpha \in A\}$ subordinate to $\mathcal{U}$ is used. Indeed, given a δ-cocycle $\omega \in \check{\mathrm{C}}^p(\mathcal{U}, \Omega^q)$, we set

$$\tau_{\alpha_0 \cdots \alpha_{p-1}} = \sum_\alpha \psi_\alpha \omega_{\alpha\alpha_0 \cdots \alpha_{p-1}} \in \check{\mathrm{C}}^{p-1}(\mathcal{U}, \Omega^q),$$

noting that since ψ_α is C^2 we stay in the class of C^1 forms with C^1 exterior derivatives. Then

$$(\delta\tau)_{\alpha_0 \cdots \alpha_p} = \sum_{i,\alpha} (-1)^i \psi_\alpha \omega_{\alpha\alpha_0 \cdots \hat{\alpha}_i \cdots \alpha_p}$$

and it follows from the fact that $\delta\omega = 0$ that

$$\sum_{i=1}^{p} (-1)^i \omega_{\alpha\alpha_0 \cdots \hat{\alpha}_i \cdots \alpha_p} = \omega_{\alpha_0 \cdots \alpha_p}.$$

Hence

$$(\delta\tau)_{\alpha_0 \cdots \alpha_p} = \left(\sum_\alpha \psi_\alpha \right) \omega_{\alpha_0 \cdots \alpha_p} = \omega_{\alpha_0 \cdots \alpha_p},$$

establishing exactness. When $p = 0$, we find that

$$\text{Kernel of } \delta : \check{\mathrm{C}}^0(\mathcal{U}, \Omega^q) \to \check{\mathrm{C}}^1(\mathcal{U}, \Omega^q) = \Omega^q(\mathcal{M}),$$

the space of smooth q-forms on $\mathcal{M}$.

We can summarize the previous discussion by stating that the rows and columns in the following commutative diagram are exact:

$$\begin{array}{ccccccccccc}
 & & & & \cdot & & \cdot & & \cdot & & \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & \Omega^2(\mathcal{M}) & \to & \check{\mathrm{C}}^0(\mathcal{U},\Omega^2) & \to & \check{\mathrm{C}}^1(\mathcal{U},\Omega^2) & \to & \check{\mathrm{C}}^2(\mathcal{U},\Omega^2) & \to \cdots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & \Omega^1(\mathcal{M}) & \to & \check{\mathrm{C}}^0(\mathcal{U},\Omega^1) & \to & \check{\mathrm{C}}^1(\mathcal{U},\Omega^1) & \to & \check{\mathrm{C}}^2(\mathcal{U},\Omega^1) & \to \cdots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & \Omega^0(\mathcal{M}) & \to & \check{\mathrm{C}}^0(\mathcal{U},\Omega^0) & \to & \check{\mathrm{C}}^1(\mathcal{U},\Omega^0) & \to & \check{\mathrm{C}}^2(\mathcal{U},\Omega^0) & \to \cdots \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & \check{\mathrm{C}}^0(\mathcal{U},\mathbb{R}) & & \check{\mathrm{C}}^1(\mathcal{U},\mathbb{R}) & & \check{\mathrm{C}}^2(\mathcal{U},\mathbb{R}) & & \\
 & & & & \uparrow & & \uparrow & & \uparrow & & \\
 & & & & 0 & & 0 & & 0 & &
\end{array}$$

The remainder of the proof uses this diagram. Given a de Rham cohomology class $[\omega] \in H^p_{dR}(\mathcal{M};\mathbb{R})$ with p-form representative ω we construct a corresponding cohomology class $s([\omega])$ in the Čech cohomology $\check{\mathrm{H}}^p(\mathcal{M};\mathbb{R})$ as follows: The differential form defines an element $\omega^{0p} \in \check{\mathrm{C}}^0(\mathcal{U},\Omega^p)$ by simply restricting ω to the sets in the cover. It is readily checked that ω^{0p} is closed with respect to the *total differential* $D = \delta + (-1)^p d$ on the double complex $K^{*,*} = \check{\mathrm{C}}^*(\mathcal{U},\Omega^*)$. Using the Poincaré Lemma, we construct an element $\omega^{0,p-1} \in \check{\mathrm{C}}^0(\mathcal{U},\Omega^{p-1})$ such that $d\omega^{0,p-1} = \omega^{0p}$. Let $\omega^{1,p-1} = \delta\omega^{0,p-1}$ and observe that $d\omega^{1,p-1} = 0$ and $\omega^{1,p-1}$ is cohomologous to ω^{0p} with respect to D. Using the Poincaré Lemma again, we construct an element $\omega^{1,p-2} \in \check{\mathrm{C}}^1(\mathcal{U},\Omega^{p-2})$ such that $d\omega^{1,p-2} = \omega^{1,p-1}$. Let $\omega^{2,p-2} = \delta\omega^{1,p-2}$ and note that $\omega^{2,p-2}$ is cohomologous to ω^{0p} with respect to D. Continue in this fashion until we reach a D-cocycle $\omega^{p0} \in \check{\mathrm{C}}^p(\mathcal{U},\Omega^0)$ which is cohomologous to ω^{0p}. Since $d\omega^{p0} = 0$, each function $\omega^{p0}_{\alpha_0\cdots\alpha_p}$ is constant, and thus ω^{p0} determines a Čech cocycle $s(\omega)$ whose cohomology class is $s([\omega])$.

By the usual diagram chasing, the cohomology class obtained is independent of choices made. Moreover, reversing the zig-zag construction described in the preceding paragraph yields an inverse to s. This finishes our sketch of the proof of de Rham's theorem; for more details, one can consult [10], Chapter 2.

Remark 1.13.5. The proof shows that the cohomologies of the two cochain complexes (1.23) and (1.25) are isomorphic. It follows that the cohomology of the cochain complex (1.25) is independent of the choice of good cover. On the other hand, in the case where $\mathcal{M}$ has C^∞ partitions of unity the argument can be repeated with C^∞ differential forms to show that the de Rham cohomology is the same whether calculated with C^∞ forms of C^1 forms with C^1 exterior derivatives.

Remark 1.13.6. This remark assumes some familiarity with singular cohomol-

ogy, as treated in Chapter 3 of [35]. One could replace the double complex the occurs in the proof by

$$K_s^{p,q} = C_s^p(\mathcal{U}, \Omega^q),$$

which is defined to be the space of functions ω which assign to each distinct ordered $(p+1)$-tuple $(\alpha_0, \ldots, \alpha_p)$ of indices such that $U_{\alpha_0} \cap \cdots \cap U_{\alpha_p} \neq 0$ an element $s_{\alpha_0 \cdots \alpha_p}$ in the space of singular cochains within $U_{\alpha_0} \cap \cdots \cap U_{\alpha_p}$ with coefficients in $\mathbb{R}$. The above proof can then be modified to give an isomorphism from singular cohomology to the Čech cohomology of $\mathcal{M}$. The argument can also be modified so that it applies to relative cohomology. Thus readers familiar with standard cohomology theory can rest assured that de Rham cohomology gives exactly the same results as the cohomology they have studied in algebraic topology courses.

Remark 1.13.7. Suppose we take an arbitrary cover of M, not necessarily a good cover. Then the rows in the above diagram are still exact, even though the columns are not. For example, we can take two open sets U and V such that $\mathcal{M} = U \cup V$. Then the above diagram collapses to a short exact sequence of de Rham complexes

$$0 \to \Omega^*(M) \to \Omega^*(U) \oplus \Omega^*(V) \to \Omega^*(U \cap V) \to 0.$$

By the "snake lemma" from algebraic topology, we get a long exact sequence

$$\begin{aligned} \cdots \to H_{dR}^k(M;\mathbb{R}) \to H_{dR}^k(U;\mathbb{R}) \oplus H_{dR}^k(V;\mathbb{R}) \to H_{dR}^k(U \cap V;\mathbb{R}) \\ \to H_{dR}^{k+1}(M;\mathbb{R}) \to H_{dR}^{k+1}(U;\mathbb{R}) \oplus H_{dR}^{k+1}(V;\mathbb{R}) \to \cdots \end{aligned}$$

which is called the *Mayer-Vietoris sequence*. The Mayer-Vietoris sequence, together with the Homotopy Lemma, is very helpful in computing de Rham cohomology.

Chapter 2

Morse Theory of Geodesics

2.1 Geodesics

Our next goal is to explain how critical point theory on infinite-dimensional manifolds can be used to used to produce periodic solutions to a class of important nonlinear ordinary differential equations, the equations for geodesics in Riemannian manifolds.

The geodesic equation is a generalization of the simplest second-order linear ordinary differential equation—the equation of a particle moving with zero acceleration in Euclidean space. This asks for a vector-valued function

$$\gamma : (a, b) \longrightarrow \mathbb{R}^N \qquad \text{such that} \qquad \gamma''(t) = 0,$$

and its solutions are the constant speed straight lines. The simplest way to make this differential equation nonlinear is to consider a proper submanifold M of $\mathbb{R}^N$ with the induced Riemannian metric, and ask for a function

$$\gamma : (a, b) \longrightarrow M \subset \mathbb{R}^N \qquad \text{such that} \qquad (\gamma''(t))^T = 0,$$

where $(\cdot)^T$ denotes projection into the tangent space of M. In simple terms, we are asking for the curves which are as straight as possible subject to the constraint that they lie within M. In the terminology of differential geometry, the tangential projection of the ordinary derivative is known as the covariant derivative, and one often writes

$$(\gamma''(t))^T = \nabla_{\gamma'}\gamma'(t) \qquad \text{or} \qquad (\gamma''(t))^T = \frac{D\gamma'}{dt}(t),$$

where ∇ and D are two commonly used notations for the covariant derivative. The smooth maps γ which satisfy the equation $\nabla_{\gamma'}\gamma'(t) = 0$ are called the smooth *geodesics* in M.

We can put the geodesic equation into the more general context of simple mechanical systems: We let $(M, \langle \cdot, \cdot \rangle)$ be a Riemannian manifold, which we can

assume has a proper isometric imbedding $\iota : M^n \to \mathbb{R}^N$ into Euclidean space; such an imbedding is provided by the Nash imbedding theorem. In addition, we consider a smooth function $\phi : M \to \mathbb{R}$, to be called the *potential*. The triple $(M, \langle \cdot, \cdot \rangle, \phi)$ is called a *simple mechanical system*. For a simple mechanical system, "Newton's equation of motion" is

$$\nabla_{\gamma'} \gamma'(t) = -(\text{grad } \phi)(\gamma'(t)).$$

The left-hand side can be interpreted as the acceleration of a moving particle of unit mass, while the right-hand side is the force (per unit mass) produced by the potential ϕ.

As we mentioned before, the idea behind the calculus of variations is to regard solutions to ordinary differential equations—such as the geodesic equation or Newton's equation of motion—as critical points of functions defined on infinite-dimensional manifolds. In the case of geodesics, the infinite-dimensional manifold is

$$L_1^2([0,1], M) = \{\gamma \in L_1^2([0,1], \mathbb{R}^N) : \ \gamma(t) \in M, \text{ for all } t \in [0,1] \ \},$$

where M is a proper submanifold of $\mathbb{R}^N$, or one of the many useful subspaces of $L_1^2([0,1], M)$. These include the free loop space

$$L_1^2(S^1, M) = \{\gamma \in L_1^2([0,1], M) : \gamma(0) = \gamma(1)\},$$

the space of paths from p to q,

$$\Omega(M, p, q) = \{\gamma \in L_1^2([0,1], M) : \gamma(0) = p, \gamma(1) = q\},$$

where p and q are points of M, and the space of paths from S_1 to S_2,

$$\Omega(M, S_0, S_1) = \{\gamma \in L_1^2([0,1], M) : \gamma(0) \in S_0, \gamma(1) \in S_1\},$$

when S_0 and S_1 are compact imbedded submanifolds of M.

The first step towards formulating the equations of geodesics within critical point theory is to define the Euclidean action

$$J_{\mathbb{R}^N} : L_1^2([0,1], \mathbb{R}^N) \to \mathbb{R} \qquad \text{by} \qquad J_{\mathbb{R}^N}(\gamma) = \frac{1}{2} \int_0^1 \gamma'(t) \cdot \gamma'(t) dt,$$

the dot denoting the Euclidean dot product. The map $J_{\mathbb{R}^N}$ is clearly smooth, being the restriction of a continuous bilinear map

$$(\gamma, \lambda) \mapsto \frac{1}{2} \int_{S^1} \gamma'(t) \cdot \lambda'(t) dt$$

to the diagonal.

Recall from §1.4 that the isometric imbedding $\iota : M \to \mathbb{R}^N$ induces by composition a map

$$\omega_\iota : L_1^2([0,1], M) \to L_1^2([0,1], \mathbb{R}^N)$$

which is also a smooth imbedding. Moreover, the Hilbert space structure on the spaces $L^2_1([0,1],\mathbb{R}^N)$ induces Riemannian metrics on $L^2_1([0,1],M)$ and its subspaces $L^2_1(S^1,M)$ and $\Omega(M,p,q)$. We define

$$J_M : L^2_1([0,1],M) \to \mathbb{R} \qquad \text{by} \qquad J_M = J_{\mathbb{R}^N} \circ \omega_\iota,$$

a map which is clearly smooth, being the composition of smooth maps. In addition, if $\phi : M \to \mathbb{R}$ is a smooth function, the map

$$\gamma \in L^2_1([0,1],M) \quad \mapsto \quad \int_0^1 \phi(\gamma(t))dt$$

is also smooth, since it is the composition of

$$\omega_\phi : L^2_1([0,1],M) \to L^2_1([0,1],\mathbb{R})$$

with integration, integration being a continuous linear map. Finally, we define the *action* for the simple mechanical system $(M,\langle\cdot,\cdot\rangle,\phi)$ to be the function

$$J_{M,\phi} : L^2_1([0,1],M) \to \mathbb{R}, \qquad J_{M,\phi}(\gamma) = \int_0^1 \left[\frac{1}{2}\langle\gamma'(t),\gamma'(t)\rangle - \phi(\gamma(t))\right] dt.$$

By restriction, we also get smooth maps

$$J_{M,\phi} : L^2_1(S^1,M) \to \mathbb{R}, \qquad J_{M,\phi} : \Omega(M,p,q) \to \mathbb{R}.$$

With these preparations out of the way, we can now state *Hamilton's principle of least action*: the motion of a simple mechanical system is described by a critical point for the action function $J_{M,\phi}$ on $L^2_1(S^1,M)$ or $\Omega(M,p,q)$.

Focusing first on the case of free loops, the case needed for studying periodic motion, we are led to ask: what is the differential

$$(dJ_{M,\phi})_\gamma : T_\gamma(L^2_1(S^1,M)) \longrightarrow \mathbb{R}?$$

We will assume that γ is smooth and that $V \in T_\gamma(L^2_1(S^1,M))$ lies in the space of C^∞ sections of γ^*TM.

To calculate the differential, we suppose that $\alpha : S^1 \times (-\epsilon,\epsilon) \to M$ is a smooth one parameter family of curves with $\alpha(t,0) = \gamma(t)$ and $D_2\alpha(t,0) = V(t)$, where D_2 denote the partial derivative with respect to the second slot. Then

$$\omega_\alpha : L^2_1(S^1, S^1 \times (-\epsilon,\epsilon)) \longrightarrow L^2_1(S^1,M)$$

is smooth. Let

$$\bar{\mu} : (-\epsilon,\epsilon) \to L^2_1(S^1, S^1 \times (-\epsilon,\epsilon)) \qquad \text{by} \qquad \bar{\mu}(\tau)(t) = (t,\tau),$$

for $t \in S^1$. Clearly, $\bar{\mu}$ is smooth and hence $\bar{\alpha} = \omega_\alpha \circ \bar{\mu}$ is a smooth curve in $L^2_1(S^1,M)$ with $\bar{\alpha}'(0) = X$. A straightforward calculation now shows that

$$dJ_{M,\phi}(\gamma)(V) = \frac{d}{dt}(J_{M,\phi} \circ \bar{\alpha})\Big|_{\tau=0} = \int_{S^1} [\langle\gamma'(t), \nabla_{\gamma'}V\rangle - d\phi(\gamma(t))(V(t))]dt,$$

where $\nabla_{\gamma'} V$ is the directional derivative of V in the direction of γ' projected into the tangent space, otherwise known as the covariant derivative of V. Since γ is assumed to be smooth, we can integrate by parts to obtain

$$dJ_{M,\phi}(\gamma)(V) = -\int_{S^1} \langle \nabla_{\gamma'}\gamma'(t) + (\text{grad } \phi)(\gamma'(t)), V(t)\rangle dt, \tag{2.1}$$

the boundary terms cancelling by periodicity. Since X can be an arbitrary smooth vector field along γ, a critical point for $J_{M,\phi}$ must be a periodic solution to Newton's equation of motion,

$$\nabla_{\gamma'}\gamma'(t) = -(\text{grad } \phi)(\gamma'(t)). \tag{2.2}$$

Thus Hamilton's principle implies that the motion of a simple mechanical system should be represented by solutions to Newton's equation of motion. In the case where $\phi = 0$, a critical point for the action is simply a smooth closed geodesic.

In a quite similar fashion, one finds that critical points to $J_{M,\phi} : \Omega(M; p, q) \to \mathbb{R}$ are solutions γ to (2.2) such that $\gamma(0) = p$ and $\gamma(1) = q$. On the other hand, if one calculates the derivative on the larger space $\Omega(M, S_0, S_1)$, the integration by parts gives additional boundary terms and we must replace (2.1) by

$$dJ_{M,\phi}(\gamma)(V) = -\int_{S^1} \langle \nabla_{\gamma'}\gamma'(t) + (\text{grad } \phi)(\gamma'(t)), V(t)\rangle dt \\ + \langle \gamma'(1), V(1)\rangle - \langle \gamma'(0), V(0)\rangle.$$

This more complicated formula must be used when considering critical points for $dJ_{M,\phi} : \Omega(M, S_0, S_1) \to \mathbb{R}$, and in this case $V(0)$ and $V(1)$ are constrained to be tangent to S_0 and S_1. The first variation formula implies that critical points are solutions to (2.2) which are perpendicular to S_0 and S_1.

2.2 Condition C for the action

In order to apply the method of steepest descent to calculus of the variations problems described in the preceding section, we need the topology on the space of maps to satisfy two conditions:

1. It must be strong enough so that $J_{M,\phi}$ is continuous.
2. It must be weak enough so that suitable sequences $\{\gamma_i\}$ (such as sequences such that $J_{M,\phi}(\gamma_i)$ tends to a minimum or minimax value subject to some constraint) will converge in the topology.

If we chose the topology to be too strong, this would make it difficult to establish convergence of $\{\gamma_i\}$. The two conflicting requirements single out L^2_1 as the appropriate space to use when studying critical points of the action $J_{M,\phi}$. indeed, we next show that when M is compact, the action functions

$$J_M, J_{M,\phi} : L^2_1(S^1, M) \longrightarrow \mathbb{R}$$

satisfy condition C, thereby making available the minimax principle from §1.11. Moreover, we will show that the critical points of J are actually C^∞ curves.

Theorem 2.2.1. *If $(M, \langle\cdot,\cdot\rangle)$ is a compact Riemannian manifold, the function $J_M : L_1^2(S^1, M) \to \mathbb{R}$, defined by*

$$J_M(\gamma) = \frac{1}{2}\int_{S^1} \langle \gamma'(t), \gamma'(t)\rangle dt,$$

satisfies Condition C: if $\{\gamma_i\}$ is a sequence in $L_1^2(S^1, M)$ such that

$$J_M(\gamma_i) \text{ is bounded} \qquad \text{and} \qquad \|dJ_M(\gamma_i)\| \to 0, \tag{2.3}$$

then it possesses a subsequence which converges to a critical point for J_M.

Proof: We start by recalling the proof of the Sobolev Lemma from §1.4. Suppose that $\{\gamma_i\}$ is a sequence of elements of $L_1^2(S^1, M)$ satisfying (2.3). We can regard each γ_i as an element of the space $L_1^2(S^1, \mathbb{R}^N)$. Then for $t_1 < t_2$,

$$|\gamma_i(t_1) - \gamma_i(t_2)| \le \int_{t_1}^{t_2} |\gamma_i'(t)|dt \le \sqrt{t_2 - t_1}\left[\int_{t_1}^{t_2} |\gamma_i'(t)|^2 dt\right]^{1/2}$$
$$\le \sqrt{t_2 - t_1}\sqrt{2J_{\mathbb{R}^N}(\gamma_i)}.$$

Since $J_M(\gamma_i)$ is bounded, we see that $\{\gamma_i\}$ is equicontinuous. Since γ_i takes values in the compact submanifold $M \subset \mathbb{R}^N$, $\{\gamma_i\}$ is also uniformly bounded. It therefore follows from Arzela's theorem or Ascoli's theorem ([66], page 179) that a subsequence of $\{\gamma_i\}$ will converge uniformly to a continuous map $\gamma_\infty : S^1 \to M$. For simplicity, we continue to denote the subsequence by $\{\gamma_i\}$. To finish the proof, we need to show that

$$\|dJ_M(\gamma_i)\| \to 0 \quad \Rightarrow \quad \{\gamma_i\} \text{ is a Cauchy sequence in } L_1^2(S^1, M).$$

To understand $|dJ_M(\gamma)\|$ we need to be able to compare an element of $T_\gamma L_1^2(S^1, \mathbb{R}^N)$ with its projection into $T_\gamma L_1^2(S^1, M)$. Recalling that M is a submanifold of $\mathbb{R}^N$ with inclusion $i : M \to \mathbb{R}^N$, we let P denote the vector bundle map

$$P : i^*T(\mathbb{R}^N) \longrightarrow TM$$

which projects onto the tangential component. By the ω-Lemma, we have a smooth map

$$\omega_P : L_1^2(S^1, i^*T(\mathbb{R}^N)) \longrightarrow L_1^2(S^1, TM)$$

which is also a vector bundle map.

Lemma 2.2.2. *If $\langle\cdot,\cdot\rangle$ denotes the Riemannian metric on either $L_1^2(S^1, \mathbb{R}^N)$ or $L_1^2(S^1, M)$ and $X \in L_1^2(S^1, i^*T(\mathbb{R}^N))$ with $\omega_\pi(X) = \gamma$, then*

$$\langle \omega_P(X), \omega_P(X)\rangle \le [1 + CJ_M(\gamma)]\langle X, X\rangle,$$

where C is some constant which depends on M but is independent of γ.

Proof: We can regard P as a section of the bundle $i^*(\mathrm{End}(\mathbb{R}^N))$ over M and it possesses a differential dP which is a section of $T^*M \otimes [i^*(\mathrm{End}(\mathbb{R}^N))]$. Let

$$C_1 = \sup\{\|dP(v)\| : \ v \text{ is a unit-length element of } TM \ \},$$

a finite constant since M is compact. Then it follows from the Leibniz rule that

$$(\omega_P(X))(t) = P_{\gamma(t)}X(t) \quad \Rightarrow \quad (\omega_P(X))'(t) = dP(\gamma'(t))(X(t)) + P_{\gamma(t)}X'(t).$$

Recall that

$$\langle \omega_P(X), \omega_P(X)\rangle = \int_{S^1} \left[\|(\omega_P(X))'(t)\|^2 + \|(\omega_P(X))(t)\|^2\right] dt,$$

where $\|\cdot\|$ denotes the norm in $\mathbb{R}^N$, and hence

$$\begin{aligned}\langle \omega_P(X), \omega_P(X)\rangle &= \int_{S^1} [\|dP(\gamma'(t))(X(t)) + P_{\gamma(t)}X'(t)\|^2 + \|P_{\gamma(t)}(X)(t)\|^2]dt \\ &\leq \int_{S^1} [\|dP(\gamma'(t))(X(t))\|^2 + \|P_{\gamma(t)}X'(t)\|^2 + \|P_{\gamma(t)}(X))(t)\|^2]dt \\ &\leq \int_{S^1} [C_1\|\gamma'(t)\|^2\|(X(t)\|^2 + \|X'(t)\|^2 + \|X(t)\|^2]dt.\end{aligned}$$

Thus

$$\begin{aligned}\langle \omega_P(X), \omega_P(X)\rangle \leq C_1 \sup\{\|(X(t)\|^2 : t \in S^1\} \int_{S^1} \|\gamma'(t)\|^2 dt + \langle X, X\rangle \\ \leq (CJ_M(\gamma) + 1)\langle X, X\rangle,\end{aligned}$$

for some positive constant C. This finishes the proof of the lemma.

Lemma 2.2.3. *If $X \in L^2_1(S^1, i^*T(\mathbb{R}^N))$ and $\omega_\pi(X) = \gamma$, where*

$$\omega_\pi : L^2_1(S^1, i^*T(\mathbb{R}^N)) \to L^2_1(S^1, M)$$

is the projection induced by $\pi : i^*T(\mathbb{R}^N) \to M$, then

$$dJ_M(\gamma)(\omega_P(X)) = \int_{S^1} [\gamma'(t) \cdot X'(t) + \alpha(\gamma'(t), \gamma'(t)) \cdot X(t)]dt, \tag{2.4}$$

where $\alpha : TM \times TM \to NM$ is the second fundamental form of M in $\mathbb{R}^N$ and the dot products are taken in the ambient Euclidean space $\mathbb{R}^N$.

Proof: Note that it suffices to establish (2.4) in the case where γ and X are smooth because both sides are continuous in the L^2_1-topology. It follows from (2.1) with $\phi = 0$ that

$$dJ_M(\gamma)(\omega_P(X)) = -\int_{S^1} \gamma''(t) \cdot P_{\gamma(t)}(X(t))dt.$$

But since $\alpha(\gamma'(t),\gamma'(t))$ is the normal component of $\gamma''(t)$,

$$\begin{aligned}\gamma''(t)\cdot X(t) &= P_{\gamma(t)}(\gamma''(t))\cdot X(t)+\alpha(\gamma'(t),\gamma'(t))\cdot X(t)\\ &= \gamma''(t)\cdot P_{\gamma(t)}(X(t))+\alpha(\gamma'(t),\gamma'(t))\cdot X(t),\end{aligned}$$

and hence

$$dJ_M(\gamma)(\omega_P(X)) = \int_{S^1}[-\gamma''(t)\cdot X(t)+\alpha(\gamma'(t),\gamma'(t))\cdot X(t)]dt.$$

An integration by parts yields the claim.

Returning to the proof of the theorem, we let $\{\gamma_i\}$ be a sequence satisfying (2.3) which converges uniformly to a continuous map $\gamma_\infty : S^1 \to M$. Since $J_M(\gamma_i)$ is bounded, $\langle\gamma_i,\gamma_i\rangle = \|\gamma_i\|^2$ is bounded, and hence

$$\|\gamma_i-\gamma_j\|^2 \le (\|\gamma_i\|+\|\gamma_j\|)^2$$

is also bounded as $i,j\to\infty$. Lemma 2.2.2 implies that $\|(\omega_P)_{\gamma_i}(\gamma_i-\gamma_j)\|$ is bounded as well. Since $\|dJ_M(\gamma_i)\|\to 0$, for every $\epsilon>0$ there exists $N\in\mathbb{N}$ such that

$$|dJ_M(\gamma_i)((\omega_P)_{\gamma_i}(\gamma_i-\gamma_j)) - dJ_M(\gamma_j)((\omega_P)_{\gamma_j}(\gamma_i-\gamma_j))| < \epsilon \qquad (2.5)$$

for $i,j>N$. Here $(\omega_P)_{\gamma_i}(\gamma_i-\gamma_j)$ lies in the tangent space to $L_1^2(S^1,M)$ at γ_i. It follows from (2.5) and the explicit formula (2.4) for dJ_M that for every $\epsilon>0$ there exists $N\in\mathbb{N}$ such that

$$\begin{aligned}&\int_{S^1}[\gamma_i'(t)\cdot(\gamma_i'(t)-\gamma_j'(t))+\alpha(\gamma_i'(t),\gamma_i'(t))\cdot(\gamma_i(t)-\gamma_j(t))]dt\\ &\qquad -\int_{S^1}[\gamma_j'(t)\cdot(\gamma_i'(t)-\gamma_j'(t))+\alpha(\gamma_j'(t),\gamma_j'(t))\cdot(\gamma_i(t)-\gamma_j(t))]dt<\epsilon,\end{aligned}$$

for $i,j>N$. But

$$|\alpha(\gamma_i'(t),\gamma_i'(t))| \le (\text{constant})|\gamma_i'(t)|^2,$$

the constant is a bound for the norm of the second fundamental form of M and hence depends only on the submanifold M. Since $|\gamma_i(t)-\gamma_j(t)|<1/n$ for i and j sufficiently large,

$$\left|\int_{S^1}\alpha(\gamma_i'(t),\gamma_i'(t))\cdot(\gamma_i(t)-\gamma_j(t))dt\right| \le (\text{constant})J_M(\gamma_i)\frac{1}{n}\to 0$$

as $i,j\to\infty$, a similar implication holding when i is replaced by j. Hence for every $\epsilon>0$ there exists $N\in\mathbb{N}$ such that

$$\left|\int_{S^1}(\gamma_i'(t)-\gamma_j'(t))\cdot(\gamma_i'(t)-\gamma_j'(t))dt\right|<\epsilon,$$

for $i,j>N$. This, together with the C^0 convergence of $\{\gamma_i\}$ implies that $\{\gamma_i\}$ is a Cauchy sequence in $L_1^2(S^1,\mathbb{R}^N)$. Since $L_1^2(S^1,\mathbb{R}^N)$ is complete, $\{\gamma_i\}$ has a

limit in $L_1^2(S^1, \mathbb{R}^N)$, which must of course be γ_∞. Thus $\gamma_i \to \gamma_\infty \in L_1^2(S^1, M)$, and by continuity of dJ_M, γ_∞ must be a critical point for J_M. This concludes the proof of Theorem 2.2.1.

Theorem 2.2.4. *Suppose that $(M, \langle \cdot, \cdot \rangle)$ is a compact Riemannian manifold, $\phi : M \to \mathbb{R}$ is a smooth function and $\psi : S^1 \to \mathbb{R}^N$ is an L_1^2 map. The real-valued functions $J_{M,\phi}$ and $J_{M,\psi}$ on $L_1^2(S^1, M)$, defined by*

$$J_{M,\phi}(\gamma) = \frac{1}{2} \int_{S^1} [\langle \gamma'(t), \gamma'(t) \rangle - \phi(\gamma(t))] \, dt \qquad \text{and}$$

$$J_{M,\psi}(\gamma) = \frac{1}{2} \int_{S^1} [\langle \gamma'(t), \gamma'(t) \rangle + \gamma(t) \cdot \psi(t)] \, dt,$$

satisfy Condition C.

Sketch of proof: Suppose that $\{\gamma_i\}$ is a sequence in $L_1^2(S^1, M)$ such that $J_{M,\phi}(\gamma_i)$ is bounded and $\|dJ_{M,\phi}(\gamma_i)\| \to 0$. Then $J_M(\gamma_i)$ is also bounded and just as before, $\{\gamma_i\}$ is uniformly bounded and equicontinuous and thus by the Arzela-Ascoli theorem, possesses a subsequence which converges uniformly to a continuous map γ_∞.

We can now mimic the preceding proof up to equation (2.5). At this point, we need to account for an extra term in $dJ_{M,\phi}$, namely

$$- \int_{S^1} d\phi(\gamma_i(t))(P_{\gamma_i(t)}(\gamma_i(t) - \gamma_j(t)))dt.$$

However, this term goes to zero, since $\{\gamma_i(t)\}$ is Cauchy in C^0. Similarly, there is an extra term in $J_{M,\psi}$ which goes to zero. With these minor changes, the argument proceeds to the desired conclusion exactly as before.

For the statement of the next theorem, we suppose that M is a complete Riemannian manifold which has a proper isometric immersion into some Euclidean space $\mathbb{R}^N$.

Theorem 2.2.5. *Suppose that $(M, \langle \cdot, \cdot \rangle)$ is a compact Riemannian manifold and $\phi : M \to \mathbb{R}$ is a smooth function. The real-valued functions*

$$J_{M,\phi} : \Omega(M, p, q) \to \mathbb{R} \quad \text{and} \quad J_{M,\phi} : \Omega(M, S_1, S_2) \to \mathbb{R},$$

defined by

$$J_{M,\phi}(\gamma) = \frac{1}{2} \int_0^1 [\langle \gamma'(t), \gamma'(t) \rangle - \phi(\gamma(t))] \, dt,$$

satisfy condition C.

Proof: A straightforward modification of preceding arguments.

Theorem 2.2.6. *if the path γ within $L_1^2(S^1, M)$, $\Omega(M, p, q)$ or $\Omega(M, S_1, S_2)$ is a critical point for J_M or $J_{M,\phi}$, then γ is C^∞. Moreover, if $\psi \in L_k^2(S^1, \mathbb{R}^N)$, then any critical point for $J_{M,\psi}$ lies in $L_{k+2}^2(S^1, M)$.*

Proof: We prove only the free loop space cases. If γ is a critical point for J_M, it follows from (2.4) that

$$\int_{S^1} \gamma'(t) \cdot X'(t) dt = -\int_{S^1} \alpha(\gamma'(t), \gamma'(t)) \cdot X(t) dt,$$

for all $X \in T_\gamma L_1^2(S^1, \mathbb{R}^N)$. Thus in the sense of distributions,

$$\gamma''(t) = \alpha(\gamma'(t), \gamma'(t)). \tag{2.6}$$

Note that

$$\gamma \in L_1^2 \quad \Rightarrow \quad \gamma' \in L^2 \quad \Rightarrow \quad \alpha(\gamma'(t), \gamma'(t)) \in L^1.$$

Thus it follows from (2.6) and standard theorems of analysis that γ is C^1. Now we use the technique of "elliptic bootstrapping":

$$\gamma \in C^1 \ \Rightarrow \ \text{RHS of (2.6) is } C^1 \ \Rightarrow \ \gamma \in C^2 \ \Rightarrow \ \text{RHS of (2.6) is } C^2 \ \Rightarrow \ \cdots .$$

By induction, we see that γ is C^∞ and Theorem 2 is established for J_M.

The proofs for $J_{M,\phi}$ and $J_{M,\psi}$ are the same except that (2.6) is replaced by

$$\gamma''(t) = \alpha(\gamma'(t), \gamma'(t)) - (\text{grad } \phi)(\gamma'(t)) \qquad \text{or} \qquad \gamma''(t) = \alpha(\gamma'(t), \gamma'(t)) + (\psi(t))^T,$$

where $(\cdot)^T$ is the tangential component.

2.3 Existence of smooth closed geodesics

In studying closed geodesics on a Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$, we let $\mathcal{M} = L_1^2(S^1, M)$, an infinite-dimensional Hilbert manifold with a complete Riemannian metric, and define the action

$$J : \mathcal{M} \longrightarrow \mathbb{R} \qquad \text{by} \qquad J(\gamma) = \frac{1}{2} \int_{S^1} \langle \gamma'(t), \gamma'(t) \rangle dt,$$

a function which satisfies condition C by Theorem 2.2.1. We have seen that $\mathcal{M} = L_1^2(S^1, M)$ is homotopy equivalent to $C^0(S^1, M)$, the free loop space studied by topologists.

Theorem 2.3.1. *Let $(M, \langle \cdot, \cdot \rangle)$ be a compact connected Riemannian manifold and let $[S^1, M]$ denote the set of free homotopy classes of continuous maps from S^1 to M. If $\alpha \in [S^1, M]$ and*

$$L_{1,\alpha}^2(S^1, M) = \{\gamma \in L_1^2(S^1, M) : \textit{the free homotopy class of } \gamma \textit{ is } \alpha\},$$

then J assumes its minimum on $L_{1,\alpha}^2(S^1, M)$. If $\alpha \neq 0$, this minimum is achieved at a nonconstant smooth closed geodesic.

Proof: This is a direct consequence of Theorem 2.2.1 and Corollary 1.12.3.

The preceding theorem shows a nonsimply connected compact manifold always possesses a smooth closed geodesic. To treat the simply connected case, we need the following theorem due to Lusternik and Fet:

Theorem 2.3.2. *If $(M, \langle \cdot, \cdot \rangle)$ is a compact simply connected Riemannian manifold, M contains a nonconstant smooth closed geodesic.*

Before proving this theorem, we recall some concepts we need from homotopy theory. A *fibration* is a continuous map $f : E \to B$ which has the homotopy lifting property; this means that if the continuous map $\tilde{g} : Y \to E$ has the property that its projection into the base $f \circ \tilde{g} : Y \to M$ can be extended to a homotopy

$$H : Y \times [0,1] \to B \qquad \text{with} \qquad H(y,0) = f \circ \tilde{g}(y), \quad \text{for } y \in I^n,$$

then this homotopy H can be lifted to

$$\tilde{H} : Y \times [0,1] \to E \qquad \text{such that} \qquad \tilde{H}(y,0) = \tilde{g}(y) \quad \text{and} \quad f \circ \tilde{H} = H.$$

The key facts about fibrations are that the fibers $E_p = f^{-1}(p)$, for $p \in B$, are homotopy equivalent to each other, and the map f induces a long exact sequence of homotopy groups

$$\cdots \to \pi_k(E_p) \to \pi_k(E) \to \pi_k(B) \to \pi_{k-1}(E_p) \to \cdots .$$

The reader can refer to [10], §§16 and 17 or [35], Theorem 4.41 for proofs of these and related facts. In fact, the only thing needed for the exact sequence is that f be a *weak fibration*, which means that it has the homotopy lifting property for the case where $Y = [0,1]^n$, the n-cube, for all choices of n.

A key example concerns the path space

$$\mathcal{P} = \{ \text{ continuous maps } \gamma : [0,1] \to M \text{ such that } \gamma(0) = p \},$$

where p is some choice of base point in M. In this case, the map

$$\pi : \mathcal{P} \to M, \qquad \pi(\gamma) = \gamma(1)$$

is a fibration. Indeed, given

$$\tilde{g} : Y \to \mathcal{P} \quad \text{and} \quad H : Y \times [0,1] \to M \quad \text{with} \quad H(y,0) = \pi \circ \tilde{g}(y),$$

we can define the lift $\tilde{H} : Y \times [0,1] \to \mathcal{P}$ by

$$\tilde{H}(y,s)(t) = \begin{cases} \tilde{g}(y)\left(\frac{t}{1-(s/2)}\right), & \text{for } 0 \le t \le 1-(s/2), \\ H(y, 2t-2+s), & \text{for } 1-(s/2) \le t \le 1. \end{cases}$$

One readily checks that $\tilde{H}$ is continuous and has the desired properties. Indeed, when $t = 1-(s/2)$,

$$\tilde{g}(y)\left(\frac{t}{1-(s/2)}\right) = \tilde{g}(y)(1) = H(y,0) = H(y, 2t-2+s),$$

so the two pieces of the function fit together continuously, while when we set $t = 1$, we find that $\tilde{H}(y, s)(1) = H(y, 2 - 2 + s) = H(y, s)$, so $\tilde{H}$ is indeed a lift H.

The fiber over the base point p of this fibration is

$$\Omega_p = \{ \text{ continuous maps } \gamma : [0, 1] \to M \text{ such that } \gamma(0) = p = \gamma(1) \}$$

and is known as the *pointed loop space*. Its homotopy groups can be computed by the long exact sequence of the fibration,

$$\cdots \to \pi_k(\Omega_p) \to \pi_k(\mathcal{P}) \to \pi_k(M) \to \pi_{k-1}(\Omega_p) \to \cdots .$$

Since $\mathcal{P}$ is contractible via the homotopy

$$H : \mathcal{P} \times [0, 1] \to \mathcal{P}, \quad \text{where} \quad H(\gamma, s)(t) = \gamma((1 - s)t),$$

we see that $\pi_k(\mathcal{P}) = 0$ for all k and this long exact sequence collapses to yield

$$\pi_k(\Omega_p) \cong \pi_{k+1}(M).$$

A second example is the free loop space

$$C^0(S^1, M) = \{ \text{ continuous maps } \gamma : [0, 1] \to M \text{ such that } \gamma(0) = \gamma(1) \},$$

which is the total space of a fibration

$$\text{ev} : C^0(S^1, M) \to M \qquad \text{defined by} \qquad \text{ev}(\gamma) = \gamma(0).$$

The fiber over p in this case is Ω_p once again, so we obtain a long exact sequence

$$\cdots \to \pi_k(\Omega_p) \to \pi_k(C^0(S^1, M)) \to \pi_k(M) \to \pi_{k-1}(\Omega_p) \to \cdots .$$

In this case, however, $\text{ev}_* : \pi_k(C^0(S^1, M)) \to \pi_k(M)$ possesses a right inverse

$$i_* : \pi_k(M) \to \pi_k(C^0(S^1, M)) \quad \text{induced by the map} \quad i : M \to C^0(S^1, M)$$

which takes the point $p \in M$ to the constant loop located at p. Hence the exact sequence for ev splits, and we conclude that

$$\pi_k(C^0(S^1, M)) \cong \pi_k(M) \oplus \pi_k(\Omega_p) \cong \pi_k(M) \oplus \pi_{k+1}(M).$$

Thus we see that the homotopy groups of the free loop space $C^0(S^1, M)$ are quite easily determined from the homotopy groups of M.

Proof of Theorem 2.3.2: Since $\pi_1(M) = 0$, M is orientable and hence $H_n(M; \mathbb{Z}) \neq 0$. Let q be the smallest positive integer such that $H_q(M, \mathbb{Z}) \neq 0$. It follows from the Hurewicz isomorphism theorem that

$$\pi_i(M) = 0, \quad \text{for } 0 < i < q, \quad \text{and} \quad \pi_q(M) \cong H_q(M, \mathbb{Z}) \neq 0.$$

It follows from Theorem 1.5.1 that $\mathcal{M} = L_1^2(S^1, M)$ is homotopy equivalent to $C^0(S^1, M)$. Hence

$$\pi_k(\mathcal{M}) \cong \pi_k(M) \oplus \pi_k(\mathcal{M}_p) \cong \pi_k(M) \oplus \pi_{k+1}(M).$$

In particular,

$$\pi_{q-1}(\mathcal{M}) \cong \pi_q(M) \neq 0.$$

Since M is simply connected, $q \geq 2$ and $\pi_{q-1}(\mathcal{M}) \cong \pi_q(M)$ is abelian. Moreover, we can identify $\pi_{q-1}(\mathcal{M})$ with $[S^{q-1}, \mathcal{M}]$, the space of free homotopy classes of maps from S^{q-1} to $\mathcal{M}$. Choose a nonzero element $\alpha \in [S^{q-1}, \mathcal{M}]$. Let

$$\mathcal{F} = \{g(S^{q-1}) \text{ such that } g : S^{q-1} \to \mathcal{M} \text{ is a continuous map in } [\alpha]\}.$$

Clearly, $\mathcal{F}$ is an ambient isotopy invariant family of sets. Hence Minimax$(J, \mathcal{F})$ is a critical value for J. We need only verify that Minimax$(J, \mathcal{F}) \geq \epsilon$ for some $\epsilon > 0$. Note that by the Cauchy-Schwarz inequality, $J^{-1}([0, \epsilon])$ consists of curves of length $\leq \sqrt{2\epsilon}$.

Recall that M is isometrically imbedded in an ambient Euclidean space $\mathbb{R}^N$. If $p \in M$, let N_pM denote the normal space to M in $\mathbb{R}^N$ and let νM be the disjoint union of all the normal spaces N_pM, for $p \in M$, the total space of a smooth vector bundle over M, called the normal bundle. Let $\nu M(\delta)$ denote the union of all the balls of radius $\delta > 0$ in $N - pM$, for $p \in M$, and let let $M(\delta)$ denote the open δ-neighborhood of M in $\mathbb{R}^N$. If $\delta > 0$ is sufficiently small, one can prove that the exponential map

$$\exp : (\nu M)(\delta) \longrightarrow M(\delta),$$

is a diffeomorphism; this is the content of the tubular neighborhood theorem from differential topology ([36], Chapter 4, §5). Thus if δ is sufficiently small, M is a strong deformation retract of $M(\delta)$ and hence M is homotopy equivalent to $M(\delta)$.

For $\epsilon > 0$ sufficiently small, any closed curve γ on M such that $J(\gamma) < \epsilon$ and hence of length $< \sqrt{2\epsilon}$ (by the Cauchy-Schwarz inequality) can be contracted to the point $\gamma(0)$ in $\mathbb{R}^N$ without leaving $M(\delta)$. Thus if $g(S^{q-1}) \subset J^{-1}([0, \epsilon])$, then g is homotopic to a smooth map

$$\tilde{g} : S^{q-1} \to M_0, \quad \text{where} \quad M_0 = \{\gamma \in \mathcal{M} : \ \gamma \text{ is constant } \}.$$

But $\pi_{q-1}(M_0) = \pi_{q-1}(M) = 0$, and hence g is homotopic to a constant, contradicting the definition of $\mathcal{F}$. Hence Minimax$(J, \mathcal{F}) \geq \epsilon$, and M must possess a nonconstant smooth closed geodesic.

Remark 2.3.3. The technique we have used to prove the Lusternik-Fet Theorem, based upon the minimax principle, is often called *Lusternik-Schnirelmann theory*.

2.4 Second variation

A defect in the minimax principle is that quite different topological constraints can in fact lead to the same minimax critical points. We need a more refined theory that can analyze the contributions to the topology made by each individual critical point. The contributions of "nondegenerate" critical points are the easiest to analyze. In this section, we take the first step towards understanding nondegenerate critical points by investigating the structure of the Hessian at a critical point.

Let $\mathcal{M}$ be a Banach manifold and let $f : \mathcal{M} \to \mathbb{R}$ be a smooth map. If $p \in \mathcal{M}$ is a critical point for f, then the *Hessian* of f at p is the symmetric bilinear map $d^2 f(p) : T_p\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R}$ defined in terms of a chart (U_α, ϕ_α) by

$$d^2 f(p)([\alpha, p, v], [\alpha, p, w]) = D^2(f \circ \phi_\alpha^{-1})(\phi_\alpha(p))(v, w).$$

It is straightforward to show that this is independent of choice of chart. Indeed, if $\alpha : (-\epsilon, \epsilon) \to \mathcal{M}$ is a smooth curve with $\alpha(0) = p$ and $\alpha'(0) = V$, then a short calculation shows that

$$d^2 f(p)(V, V) = \left.\frac{d^2}{dt^2}(f \circ \alpha)\right|_{t=0},$$

an expression which is clearly independent of choice of chart. Of course, once one knows $d^2 f(p)(V, V)$, one can determine $d^2 f(p)(V, W)$ for arbitrary $V, W \in T_p\mathcal{M}$ by the polarization identity:

$$d^2 f(p)(V, W) = \frac{1}{4}[d^2 f(p)(V + W, V + W) - d^2 f(p)(V - W, V - W)].$$

Definition. The *Morse index* of a critical point p for f is the maximal dimension of a linear subspace of $T_p\mathcal{M}$ on which $d^2 f(p)$ restricts to a negative definite symmetric bilinear form. The *nullity* of the critical point p is

$$\dim\{\ V \in T_p\mathcal{M} \text{ such that } d^2 f(p)(V, W) = 0 \text{ for all } W \in T_p\mathcal{M}\ \}.$$

We say that the critical point is *stable* if its Morse index is zero. In some sense, the index measures the extent to which a critical point fails to be stable.

We would like to calculate the Hessian at critical points for the simple mechanical systems we described in §2.1, where we take $\mathcal{M}$ to be a space of L^2_1 maps. In the case where $\mathcal{M} = L^2_1(S^1, M)$, it is proven in elementary differential geometry texts that the Hessian of the action J is given by the *second variation formula* or *index formula*:

Proposition 2.4.1. *If $(M, \langle\cdot,\cdot\rangle)$ is a compact Riemannian manifold and $J : L^2_1(S^1, M) \to \mathbb{R}$ is the usual action, then*

$$d^2 J(\gamma)(V, W) = \int_{S^1} [\langle \nabla_{\gamma'(t)} V, \nabla_{\gamma'(t)} W\rangle - \langle R(V, \gamma'(t))\gamma'(t), W\rangle] dt, \qquad (2.7)$$

for $V, W \in T_\gamma \mathcal{M}$.

Let us review how to establish (2.7). We consider a smooth variation of γ which has its support within a given coordinate chart (U, t) on S^1. Recall that such a variation is a smooth family of maps $u \mapsto \gamma_u$ in $L^2_1(S^1, M)$ and let

$$\alpha(t, u) = \gamma_u(t), \qquad V(t) = \frac{\partial \alpha}{\partial u}(t) \in T_{\gamma(t)} M.$$

Differentiating twice, we obtain

$$\begin{aligned} d^2(J)(\gamma)(V,V) = \frac{d^2}{du^2}(J(\gamma_u))\bigg|_{u=0} &= \int_{S^1} \frac{\partial}{\partial u} \left\langle \frac{\partial \alpha}{\partial t}, \frac{D}{\partial u}\frac{\partial \alpha}{\partial t} \right\rangle dt \bigg|_{u=0} \\ &= \int_{S^1} \left[\left\langle \frac{D}{\partial u}\frac{\partial \alpha}{\partial t}, \frac{D}{\partial u}\frac{\partial \alpha}{\partial t} \right\rangle + \left\langle \frac{\partial \alpha}{\partial t}, \frac{D^2}{\partial u^2}\frac{\partial \alpha}{\partial t} \right\rangle \right] dt \bigg|_{u=0}, \end{aligned}$$

where D denotes the covariant derivative of the Levi-Civita connection on the ambient Riemannian manifold M. Thus

$$d^2(J)(\gamma)(V,V) = \int_{S^1} \left[\left\langle \frac{DV}{\partial t}, \frac{DV}{\partial t} \right\rangle + \left\langle \frac{\partial \gamma}{\partial t}, \frac{D}{\partial u}\frac{DV}{\partial t} \right\rangle \right] dt$$

Using the definition of the Riemann-Christoffel curvature tensor R, we obtain

$$d^2 J_\gamma(V,V) = \int_{S^1} \left[\left\langle \frac{DV}{\partial t}, \frac{DV}{\partial t} \right\rangle + \langle \gamma'(t), R(V, \gamma'(t))V \rangle + \left\langle \gamma'(t), \frac{D}{\partial t}\frac{DV}{\partial u} \right\rangle \right] dt.$$

An integration by parts and use of the fact that γ satisfies the Euler-Lagrange equation eliminates the last term, thereby yielding (2.7).

We can write the formula for second variation as

$$d^2 J(\gamma)(V,W) = \int_\Sigma \langle L(V), W \rangle dt,$$

where L_γ is the Jacobi operator, defined by

$$L(V) = -\left[\frac{D}{\partial t} \circ \frac{DV}{\partial t} + R(V, \gamma'(t))\gamma'(t) \right]. \tag{2.8}$$

An element $V \in T_\gamma \mathcal{M}$ is called a *Jacobi field* along γ if $L_\gamma(V) = 0$. Note that if $\gamma \in C^\infty(S^1, M)$, then the Jacobi operator yields a continuous linear operator

$$L : L^2_k(\gamma^* TM)) \to L^2_{k-2}(\gamma^* TM)),$$

for all $k \geq 1$. Symmetry of $d^2 J$ implies that the Jacobi operator satisfies

$$\int_{S^1} \langle L(V), W \rangle dt = \int_{S^1} \langle V, L(W) \rangle dt, \quad \text{for all } V, W \in L^2_{k+2}(\gamma^* TM)),$$

and hence we say that it is *formally self-adjoint*. In studying the Hessian of J it is also useful to consider, for $\lambda \in \mathbb{C}$, the closely related operator

$$L_\lambda = L - \lambda\iota : L^2_k(\gamma^*TM)) \to L^2_{k-2}(\gamma^*TM)), \tag{2.9}$$

where $\iota : L^2_k(\gamma^*TM)) \to L^2_{k-2}(\gamma^*TM))$ is the inclusion. We say that λ is an *eigenvalue* for L if the kernel of L_λ is nonzero.

There is a similar second variation formulae for $J : \mathcal{M} \to \mathbb{R}$ when $\mathcal{M} = \Omega(M; p, q)$. In this case, we obtain

$$d^2J(\gamma)(V,V) = \int_0^1 \left[\left\langle \frac{DV}{\partial t}, \frac{DV}{\partial t} \right\rangle + \langle \gamma'(t), R(V, \gamma'(t))V \rangle \right] dt,$$

except that now V is an element of

$$T_\gamma\Omega(M; p, q) = \{V \in L^2_1([0,1], TM) : \omega_\pi(V) = \gamma, V(0) = 0 = V(1)\}.$$

In this case an integration by parts shows that

$$d^2J(\gamma)(V,W) = \int_0^1 \langle L(V), W\rangle dt, \quad \text{for} \quad V, W \in T_\gamma\Omega(M; p, q), \tag{2.10}$$

where L is the differential operator defined once again by (2.8). If we set

$$L^2_{1,0}(\gamma^*TM) = \{X \in L^2_k(\gamma^*TM)) : X(0) = X(1) = 0\},$$

then the Jacobi operator defines a continuous linear map

$$L : L^2_{1,0}(\gamma^*TM)) \cap L^2_k(\gamma^*TM)) \to L^2_{k-2}(\gamma^*TM)),$$

for all $k \geq 1$. Once again, we can define, for $\lambda \in \mathbb{C}$, the closely related operator

$$L : L^2_{1,0}(\gamma^*TM)) \cap L^2_k(\gamma^*TM)) \to L^2_{k-2}(\gamma^*TM)), \tag{2.11}$$

where ι is the inclusion.

The Jacobi operators L are special cases of formally self-adjoint elliptic differential operators and there is a well-developed classical theory for the eigenvalue problem for such operators. In order to describe this theory, we need the notion of Fredholm operator:

Definition. A linear operator $L : E \to F$ between Banach spaces is said to be *Fredholm* if

1. L has finite-dimensional kernel,
2. L has closed range, and
3. L has finite-dimensional cokernel, where the cokernel of L is the quotient space $F/(\text{Range}(L))$.

The *Fredholm index* of a Fredholm operator L is defined by the formula

$$\text{Fredholm index of } L = \dim(\text{Kernel of } L) - \dim(\text{Cokernel of } L).$$

Note that a linear operator between finite-dimensional Banach spaces, say $L : \mathbb{R}^n \to \mathbb{R}^p$, is always Fredholm, and its Fredholm index is just the difference in dimensions $n - p$.

Theorem 2.4.2 *For every $\lambda \in \mathbb{C}$ and every integer $k \geq 1$, the operators L_λ defined by (2.9) and (2.11) are Fredholm maps of Fredholm index zero. Moreover,*

1. *for each $\lambda \in \mathbb{C}$ the eigenspace $W_\lambda = Ker(L_\lambda)$ is finite-dimensional.*
2. *all the elements of W_λ are C^∞,*
3. *if W_λ is empty, then L_λ possesses an inverse G_λ, which is called a* Green's operator,
4. *if W_λ is nonempty, that is, λ is an eigenvalue, then $\lambda \in \mathbb{R}$, and*
5. *the eigenvalues can be arranged in a sequence*

$$\lambda_1 < \lambda_2 < \cdots < \lambda_i < \cdots \qquad \text{with } \lambda_i \to \infty,$$

and only finitely many of the eigenvalues are negative.

This is a special case of the basic theorem for second-order elliptic operators proven in basic courses on linear PDE's. It is worked out in various cases in Chapter 5 of [79], Chapter 3 of [44] and in [19].

2.5 Morse nondegenerate critical points

If $\mathcal{M}$ is a Hilbert manifold with a Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ and $f : \mathcal{M} \to \mathbb{R}$ is a C^2 map with a critical point p, it follows from the Riesz representation theorem that there is a continuous map $A : T_p\mathcal{M} \to T_p\mathcal{M}$ such that

$$d^2 f(p)(V, W) = \langle\langle A(V), W \rangle\rangle, \quad \text{for} \quad V, W \in T_p\mathcal{M}, \tag{2.12}$$

Definition. Suppose that $\mathcal{M}$ is a Hilbert manifold and that $f : \mathcal{M} \to \mathbb{R}$ is a C^2 map. A critical point p for f is *Morse nondegenerate* if for some choice (and hence every choice) of Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ the continuous map $A : T_p\mathcal{M} \to T_p\mathcal{M}$ satisfying (2.12) is an isomorphism; otherwise, it is *Morse degenerate*.

Example 2.5.1. A Morse nondegenerate critical point has zero nullity, but the converse is not true in general. Indeed, we can let $\mathcal{M} = L^2[0, 1]$ with the inner product

$$\langle \phi, \psi \rangle = \int_0^1 \phi(t)\psi(t)dt,$$

and define

$$f : L^2[0,1] \to \mathbb{R} \quad \text{by} \quad f(\phi) = \int_0^1 t\phi(t)^2 dt.$$

Then $\phi = 0$ is a critical point for f and

$$d^2 f(0)(\phi, \psi) = \langle A(\phi), \psi \rangle, \quad \text{where} \quad A(\phi(t)) = t\phi(t).$$

Then A is injective but the range of A does not include any constant functions, so it is not surjective. Thus the critical point 0 has zero nullity but is not Morse nondegenerate.

In the variational problems we have been considering, however, the map A is Fredholm and zero nullity does indeed imply nondegeneracy.

To be specific, let's focus on the case where M is a complete Riemannian manifold, properly and isometrically imbedded in $\mathbb{R}^N$ and $\mathcal{M} = \Omega(M, p, q)$. For Riemannian metric on $\mathcal{M}$, we choose the intrinsic Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ defined by the formula

$$\langle\langle V, W \rangle\rangle_\gamma = \int_0^1 \left[\langle V(t), W(t) \rangle + \left\langle \frac{DV}{\partial t}(t), \frac{DW}{\partial t}(t) \right\rangle \right] dt, \tag{2.13}$$

for $V, W \in T_\gamma \Omega(M, p, q)$, where D is the Levi-Civita connection on M. An integration by parts shows that

$$\langle\langle V, W \rangle\rangle_\gamma = \int_0^1 \langle P(V), W \rangle dt, \quad \text{where} \quad P = -\frac{D}{dt} \circ \frac{D}{dt} + \mathrm{id},$$

a formally self-adjoint second order elliptic partial differential operator. Since the Hilbert space inner product is positive definite, it follows from Theorem 2.4.2 that

$$P : L^2_{1,0}(\gamma^* TM) \to L^2_{-1}(\gamma^* TM)$$

has an inverse Green's operator G.

Now consider the action function $J : \Omega(M, p, q) \to \mathbb{R}$. It follows from the second variation formula (2.10) that

$$d^2 J(\gamma)(V, W) = \int_0^1 \langle L(V), W \rangle dt = \langle\langle G \circ L(V), W \rangle\rangle = \langle\langle A(V), W \rangle\rangle,$$

where A is a Fredholm operator of Fredholm index zero. Thus if γ is a critical point for J with zero nullity, then γ is a Morse nondegenerate critical point.

Definition. We say that $f : \mathcal{M} \to \mathbb{R}$ is a *Morse function* if all of its critical points are Morse nondegenerate.

Theorem 2.5.2. *Suppose that $(M, \langle \cdot, \cdot \rangle)$ is a complete Riemannian manifold and $p \in M$. For almost all choices of $q \in M$, the function*

$$J : \Omega(M; p, q) \longrightarrow [0, \infty)$$

is a Morse function.

The proof follows from the fact that there are nonzero Jacobi fields along $\gamma \in \Omega(M; p, q)$ vanishing at the endpoints if and only if q is a critical value for $\exp_p$. Moreover, it follows from Sard's Theorem (which is proven in [36], Chapter 3) that the set of regular values for $\exp_p$ form a set of full measure in M. Thus for almost all choices of q, all geodesics in $\Omega(M; p, q)$ will be Morse nondegenerate.

In the case where M is a compact Riemannian manifold and $\mathcal{M} = L_1^2(S^1, M)$, we utilize the similar Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ defined by

$$\langle\langle V, W \rangle\rangle_\gamma = \int_{S^1} \left[\langle V(t), W(t) \rangle + \left\langle \frac{DV}{\partial t}(t), \frac{DW}{\partial t}(t) \right\rangle \right] dt,$$

for $V, W \in T_\gamma L_1^2(S^1, M)$). Once again, we can integrate by parts to obtain

$$\langle\langle V, W \rangle\rangle_\gamma = \int_{S^1} \langle P(V), W \rangle dt, \quad \text{where} \quad P = -\frac{D}{dt} \circ \frac{D}{dt} + \mathrm{id},$$

and P has a Green's operator inverse G. The operator $A = G \circ L$ satisfying

$$d^2 J(p)(V, W) = \langle\langle A(V), W \rangle\rangle, \quad \text{for} \quad V, W \in T_p L_1^2(S^1, M)$$

is once again a Fredholm operator of Fredholm index zero.

For the space of free loops, however, $J : L_1^2(S^1, M) \to \mathbb{R}$ is never a Morse function. The reason is that the action J is invariant under a continuous right action of the circle group S^1 on $L_1^2(S^1, M)$,

$$\phi : L_1^2(S^1, M) \times S^1 \longrightarrow L_1^2(S^1, M), \quad \phi(\gamma, s) = \gamma_s, \quad \text{where} \quad \gamma_s(t) = \gamma(t + s),$$

and therefore whenever any point is critical for J, so is the entire S^1-orbit. The action J is also invariant under an extension of this action to an $O(2)$-action

$$\phi : L_1^2(S^1, M) \times O(2) \longrightarrow L_1^2(S^1, M),$$

the reflections in $O(2)$ changing the orientation of the geodesic critical points. However, after we describe Smale's extension of Sard's theorem to infinite-dimensional manifolds in the next section, the theory of Fredholm maps will allow us to perturb J to a Morse function.

We have seen that we needed to choose the L_1^2 topology on the space of maps in order for Condition C to hold, but we also know that for the variational problems coming from simple mechanical systems, the critical points are automatically C^∞. For many purposes it is convenient to have extra derivatives, and hence to work in the space of L_k^2 maps, for $k > 1$. This is admissible so long as we don't utilize convergence results that rely upon condition C.

For example, we can take $\mathcal{M} = L_k^2(S^1, M)$ with underlying Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$, and consider a function $f = J : \mathcal{M} \to \mathbb{R}$ of the form

$$J(\gamma) = \int_\Sigma \mathcal{L}(\gamma) dt, \quad \text{where} \quad \mathcal{L} : L_k^2(S^1, M) \to L^1(S^1, M)$$

is a sufficiently well-behaved function, called the *Lagrangian* of the variational problem. We assume that we can differentiate to obtain

$$(dJ)(\gamma)(V) = \int_{S^1} \langle F(\gamma), V\rangle dA, \tag{2.14}$$

for some function F. The equation $F(\gamma) = 0$ is called the *Euler-Lagrange equation* for the variational problem, and F is called the *Euler-Lagrange operator*. For example, in the case of a simple mechanical system $(M, \langle\cdot,\cdot\rangle, \phi)$, we take

$$\mathcal{L}(\gamma) = \frac{1}{2}\langle\gamma'(t), \gamma'(t)\rangle - \phi(\gamma(t)),$$

and the corresponding Euler-Lagrange operator is just

$$F(\gamma) = -\nabla'_\gamma\gamma' - (\text{grad } \phi)(\gamma'(t)),$$

a nonlinear second-order differential operator.

To have an appropriate range for the nonlinear differential operator F it is convenient to utilize pullback bundles. For $k \in \mathbb{N}$, $k > 2$, we can regard $L^2_{k-2}(S^1, TM)$ as the total space of a smooth vector bundle over $L^2_{k-2}(S^1, M)$, and let $\tilde{L}^2_{k-2}(S^1, TM)$ denote the total space of the pullback bundle via the inclusion ι to $L^2_k(S^1, M)$, so

$$\tilde{L}^2_{k-2}(S^1, TM) = \{(\gamma, V) \in L^2_k(S^1, M) \times L^2_{k-2}(S^1, TM) : X \in L^2_{k-2}(\gamma^*TM)\}. \tag{2.15}$$

However, note that the explicit construction (2.15) makes sense even if $k = 2$ or $k = 1$. Thus the Euler-Lagrange map F can be regarded as a map

$$F : L^2_k(S^1, M) \to \tilde{L}^2_{k-2}(S^1, TM),$$

for all $k \in \mathbb{N}$, which is differentiable when k is large. If γ is a critical point for J, we can differentiate to obtain a formula for the Hessian,

$$d^2J(\gamma)(V, W) = \int_\Sigma \langle(\pi_V \circ DF)(\gamma)(V), W\rangle dA, \tag{2.16}$$

where DF denotes the derivative with respect to $\gamma \in L^2_k(\Sigma, M)$ and π_V is the vertical projection into the fiber $L^2_{k-2}(\gamma^*TM)$. Of course, $L = \pi_V \circ (DF)(\gamma)$ is just the Jacobi operator at γ. The formula shows that we can regard the Jacobi operator L as the linearization of the Euler-Lagrange operator F at a solution γ to the Euler-Lagrange equation.

Recall that the vector field $\mathcal{X} = -\text{grad}(J)$ on $L^2_1(S^1, M)$ is defined by the equation

$$\langle\langle\mathcal{X}(\gamma), V\rangle\rangle = -dJ(\gamma)(V), \quad \text{for } V \in T_\gamma L^2_1(S^1, M). \tag{2.17}$$

Thus the linearization of $-\mathcal{X}$ at a critical point $\gamma \in L^2_k(S^1, M)$ (when $k > 2$) is a smoothed version of the Jacobi operator. Indeed, since $F(\gamma) = 0$, it follows from (2.14) that

$$-\pi_V \circ D\mathcal{X}(\gamma) = G \circ \pi_V \circ DF(\gamma) = G \circ L = A,$$

where A is a restriction of the Fredholm operator satisfying (2.12).

Remark 2.5.3. The vector field $\mathcal{X} = -\text{grad}(J)$ is tangent to each of the subpaces $L^2_k(S^1, M) \subseteq L^2_1(S^1, M)$. Indeed, as mentioned before,

$$\langle\langle V, W\rangle\rangle_\gamma = \int_{S^1} \langle P_\gamma(V), W\rangle dt, \quad \text{where} \quad P_\gamma = -\frac{D}{dt} \circ \frac{D}{dt} + \text{id}.$$

As γ varies, the differential operators P_γ fit together to form a vector bundle map

$$P : L^2_k(S^1, TM) \to \tilde{L}^2_{k-2}(S^1, TM),$$

where $\tilde{L}^2_{k-2}(S^1, TM)$ is the total space of a vector bundle over $L^2_k(S^1, M)$ as described in (2.15). The vector bundle map P has a "Green's operator" inverse

$$G : \tilde{L}^2_{k-2}(S^1, TM) \to L^2_k(S^1, TM),$$

a smoothing operator which increases the number of derivatives by two. If

$$F : L^2_k(S^1, M) \to \tilde{L}^2_{k-2}(S^1, M)$$

is the Euler-Lagrange map for J_ψ, then (2.17) implies that

$$\mathcal{X} = -G \circ F : L^2_k(S^1, M) \to L^2_{k-2}(S^1, TM).$$

2.6 The Sard-Smale Theorem

Sard's theorem, so useful in understanding the transversality theory of finite-dimensional manifolds, possesses an extension to infinite-dimensional manifolds that is due to Smale. The standard approach to constructing Morse functions on infinite-dimensional manifolds is based upon the Sard-Smale theorem.

Recall that a point $q \in \mathcal{M}_2$ is called a *regular value* for the smooth map $f : \mathcal{M}_1 \to \mathcal{M}_2$ if

$$p \in f^{-1}(q) \quad \Rightarrow \quad df_p \text{ is onto;}$$

otherwise it is called a *critical value*. Any finite-dimensional manifold has a measure which defines volume integrals on open subsets of M; indeed, if D is an open subset of a coordinate chart $(U, (x_1, \ldots x_n))$ on M, we can set

$$\text{Volume of } D = \int_D \sqrt{g} dx_1 \cdots dx_n.$$

Here g is the determinant of the component matrix of the metric tensor, and it is standard to check that the integral for volume is independent of coordinate chart chosen. Using the Riesz representation theorem from analysis, one can then define measurable subsets of M and speak of sets of measure zero.

Brown-Sard Theorem 2.6.1. *Suppose that $f : M_1 \to M_2$ is a C^k map between finite-dimensional manifolds, where*

$$k > 0 \qquad \textit{and} \qquad k > \dim(M_1) - \dim(M_2).$$

Then the subset of M_2 consisting of the critical values of f has measure zero.

A proof of this theorem can be found in [36], Chapter 3. The strange differentiability condition cannot be weakened to $k \geq \dim(\mathcal{M}_1) - \dim(\mathcal{M}_2)$. However, this finite-dimensional theorem actually does have a slight improvement that was found by Bates [6]. Let $C^{k,1}$ denote the class of functions which are C^k and moreover have k-th order derivatives which are Lipshitz. Bates showed that if $f : M_1 \to M_2$ is $C^{k,1}$, where

$$k > 0 \quad \text{and} \quad k \geq \dim(M_1) - \dim(M_2),$$

then the set of critical values of f has measure zero.

Suppose now that $\mathcal{M}_1$ and $\mathcal{M}_2$ are manifolds modeled on Banach spaces E_1 and E_2 respectively. A C^k map $f : \mathcal{M}_1 \to \mathcal{M}_2$, where $k \geq 1$, is said to be a *Fredholm map* if its linearization

$$(f_*)_p : T_p\mathcal{M}_1 \longrightarrow T_{f(p)}\mathcal{M}_2$$

is a Fredholm operator, for each $p \in \mathcal{M}_1$. The *Fredholm index* of a Fredholm map $f : \mathcal{M}_1 \to \mathcal{M}_2$ is the Fredholm index of $(f_*)_p$ for any $p \in \mathcal{M}_1$, this being constant if $\mathcal{M}_1$ is connected. Indeed, it can be proven that the Fredholm index is a continuous function from the space of Fredholm operators to the integers.

Note that if $\mathcal{M}_1$ and $\mathcal{M}_2$ are finite-dimensional, any C^1 map $f : \mathcal{M}_1 \to \mathcal{M}_2$ is Fredholm with Fredholm index $\dim(\mathcal{M}_1) - \dim(\mathcal{M}_2)$.

Thus the Fredholm index can be used to replace one of the hypotheses in the Brown-Sard Theorem. We also need to replace the notion of "measure zero," since we have no completely satisfactory notion of measure in infinite dimensions possessing all of the nice properties of the standard measure on submanifolds of finite-dimensional Euclidean space. We say that a subset of $\mathcal{M}_2$ is *residual* or *generic* if it is a countable intersection of open dense subsets of $\mathcal{M}_2$. With these preparations out of the way, we can now state Smale's version of Sard's theorem [78]:

Sard-Smale Theorem 2.6.2. *Suppose that $f : \mathcal{M}_1 \to \mathcal{M}_2$ is a C^k Fredholm map between separable Banach manifolds, where*

$$k > 0 \qquad \textit{and} \qquad k > \textit{the Fredholm index of } f.$$

Then the set of regular values of f is residual.

The main idea of Smale's proof is to reduce to the finite-dimensional case. We first need to show that a Fredholm map is locally proper:

Lemma 2.6.3. *If $f : \mathcal{M}_1 \to \mathcal{M}_2$ is a Fredholm map and $p \in \mathcal{M}_1$, then there is an open neighborhood U of p such that the restriction of f to the closure $\overline{U}$ of U is proper and closed.*

To prove the Lemma, we use the "local representative theorem" for f near p, a direct corollary of the inverse function theorem. According to this theorem,

there exist direct sum decompositions of the model spaces for $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$E_1 = \tilde{E} \oplus G_1, \qquad E_2 = \tilde{E} \oplus G_2,$$

where G_1 and G_2 are both finite-dimensional, and local charts (U, ϕ) on $\mathcal{M}_1$ and (V, ψ) on $\mathcal{M}_2$ centered at p and $f(p)$ respectively, such that

$$\psi \circ f \circ \phi^{-1} : \phi(U \cap F^{-1}(V)) \longrightarrow E_2$$

is of the special form

$$\psi \circ f \circ \phi^{-1}(u, v) = (u, \eta(u, v)), \qquad \text{for} \qquad u \in \tilde{E}, \quad v \in G_1,$$

where $D\eta(0, 0) = 0$. We denote $\psi \circ f \circ \phi^{-1}$ simply by f.

Suppose now that D_1 and D_2 are closed balls of radius $\epsilon > 0$ in $\tilde{E}$ and G_1 respectively. To show that $f|(D_1 \times D_2)$ is proper, we suppose that $p_i = (u_i, v_i) \in D_1 \times D_2$ and $f(p_i) = q_i \to q \in V$. We need to show that $\{p_i\}$ has a convergent subsequence. Since D_2 is compact, we can assume that $v_i \to v \in D_2$. On the other hand,

$$f(u_i, v_i) = (u_i, \eta(u_i, v_i)) \to q \qquad \Rightarrow \qquad u_i \to \text{ some } u \in D_1.$$

Hence $p_i \to (u, v) \in D_1 \times D_2$. It follows that $f|(D_1 \times D_2)$ is proper.

To prove that $f|(D_1 \times D_2)$ is closed, we suppose that $p_i = (u_i, v_i) \in D_1 \times D_2$ converges to $p = (u, v)$. Then u_i converges to some point u and $\eta(u_i, v_i)$ has a subsequence which converges to some point w. Then $f(u_i, v_i)$ converges to (u, w).

Returning to the proof of the theorem, we note that since $\mathcal{M}_1$ is separable, the Lemma implies that it can be covered by a countable collection of open sets U_i such that $f|\overline{U}_i$ is proper. Thus we can reduce the proof to the case where f is proper and closed. But this implies that the set of critical values is closed. Thus in this case the regular values form an open set, and it suffices to prove that the regular values are dense. In other words, it suffices to show that any open subset of $\mathcal{M}_2$ contains a regular value.

It suffices to show that there is a regular value in a product neighborhood U_i constructed as in the proof of the lemma. It follows from the finite-dimensional version of Sard's theorem that for each fixed choice of u_0, there is a regular value v_0 for

$$v \mapsto \eta_i(u_0, v),$$

and (u_0, v_0) is then a regular value, proving the theorem.

Note that by Corollary 1.2.8, if $f : \mathcal{M}_1 \to \mathcal{M}_2$ is a C^k Fredholm map between separable Banach manifolds of Fredholm index m, where $m < k$, then whenever q is a regular value of f, $f^{-1}(q)$ is a submanifold of of $\mathcal{M}_1$ of dimension m.

2.7 Existence of Morse functions

We now consider an important application of the Sard-Smale theorem, the perturbation of a given function to a Morse function.

In the case of the action integral, the circle group S^1 acts on the free loop space $L^2_1(S^1, M)$ preserving the action J, and this ensures that all critical points for J must be Morse degenerate, and in fact that all nonconstant multiples of γ' are Jacobi fields along γ. However, there is a simple perturbation of the action integral whose critical points are all Morse nondegenerate, and this perturbation is therefore a Morse function. The perturbation is obtained by making a suitable choice of $\psi \in C^k(S^1, \mathbb{R}^N)$ for k sufficiently large, and setting

$$J_\psi : L^2_1(S^1, M) \to \mathbb{R} \quad \text{by} \quad J_\psi(\gamma) = \frac{1}{2}\int_{S^1} \|\gamma'(t)\|^2 dt + \int_{S^1} \gamma(t) \cdot \psi(t),$$

where the dot denotes the usual dot product in the ambient Euclidean space. In this section, we will show that for most choices of ψ, J_ψ is indeed a Morse function. Recall that a subset of a complete metric space is *residual* if it is the countable intersection of open dense sets.

Theorem 2.7.1. *For a residual set of $\psi \in L^2_1(S^1, \mathbb{R}^N)$, the function*

$$J_\psi : L^2_1(S^1, M) \to \mathbb{R}$$

is a Morse function; that is, all of its critical points are nondegenerate.

Note. Since critical points of J_ψ are automatically elements of $L^2_3(S^1, M)$ when $\psi \in L^2_1(S^1, \mathbb{R}^N)$, it suffices to show that $J_\psi : L^2_3(S^1, M) \to \mathbb{R}$ is a Morse function.

Proof: It is readily verified that γ is a critical point for J_ψ if and only if

$$dJ_\psi(X) = 0, \qquad \text{for all } X \in T_\gamma(L^2_3(S^1, M)).$$

We can write

$$dJ_\psi(\gamma)(X) = \int_\Sigma \langle F(\gamma, \psi), X\rangle dA,$$

where $F(\gamma, \psi) = 0$ is the Euler-Lagrange equation for J_ψ, so that γ is a critical point for J_ψ if and only if $F(\gamma, \psi) = 0$.

A direct calculation shows that the Euler-Lagrange map

$$F : L^2_3(S^1, M) \times L^2_1(S^1, \mathbb{R}^N) \to L^2_1(S^1, TM) \quad \text{is given by}$$
$$F(\gamma, \psi) = -\nabla_{\gamma'}\gamma' + \psi^T,$$

where ψ^T denotes the orthogonal projection of ψ into the tangent space of M. We can regard F as a vector field on $L^2_3(S^1, M)$, depending on a parameter in $L^2_1(S^1, \mathbb{R}^N)$, which loses two derivatives, and hence takes its values in $L^2_1(S^1, TM)$:

$$-\nabla_{\gamma'}\gamma' + \psi^T \in \{X \in L^2_1(S^1, TM) : \omega_\pi \circ X = \gamma\}.$$

Lemma 2.7.2. *F is transverse to the zero-section of the vector bundle*

$$\omega_\pi : L_1^2(S^1, TM) \longrightarrow L_1^2(S^1, M).$$

Proof of Lemma: Taking the partial derivative with respect to the second variable, we obtain $\pi_V \circ (D_2F)(\gamma,\psi)(\eta) = -\eta^T$, where π_V denotes projection into the vertical tangent space at $(\gamma, 0) \in L_1^2(S^1, TM)$. The formula shows that $\pi_V \circ (D_2F)(\psi,\gamma)$ maps onto the fiber of ω_π over γ. Hence if γ is a critical point for J_ψ,

$$\begin{aligned}(\text{Range of } \pi_V \circ (D_2F)(\psi,\gamma)) + (\text{Tangent space to zero-section}) \\ = \text{Tangent space to } L_1^2(S^1, TM),\end{aligned}$$

which means that F is transverse to the zero-section.

Just as in the finite-dimensional case, it follows from Corollary 1.2.8 that the inverse image of a split submersion of a submanifold is itself a submanifold, so $F^{-1}(\text{zero-section})$ is a submanifold $\mathcal{S}$ of $L_3^2(S^1, M) \times L_1^2(S^1, \mathbb{R}^N)$. We can also describe this submanifold as the solution set

$$\mathcal{S} = \{(\gamma,\psi) \in L_3^2(S^1, M) \times L_1^2(S^1, \mathbb{R}^N) : \nabla_{\gamma'}\gamma' - \psi^T = 0\}.$$

Lemma 2.7.3. *The projection on the second factor $\pi : \mathcal{S} \to L_1^2(S^1, \mathbb{R}^N)$ is a Fredholm map of Fredholm index zero.*

We begin by determining the tangent space to $\mathcal{S}$, obtaining

$$\begin{aligned}T_{(\gamma,\psi)}\mathcal{S} = \{(X,\eta) \in TL_3^2(S^1, M) \times L_1^2(S^1, \mathbb{R}^N) : \\ \pi_V \circ D_1F(\gamma,\psi)X + \pi_V \circ D_2F(\gamma,\psi)\eta = 0\},\end{aligned}$$

or equivalently,

$$T_{(\gamma,\psi)}\mathcal{S} = \{(X,\eta) \in TL_3^2(S^1, M) \times L_1^2(S^1, \mathbb{R}^N) : L_{\gamma,\psi}(X) = -\eta^T\}, \qquad (2.18)$$

where

$$L_{\gamma,\psi} = \pi_V \circ D_1F(\gamma,\psi) : T_\gamma L_3^2(S^1, M) \to T_\gamma L_1^2(S^1, M)$$

is the Jacobi operator for J_ψ determined by the formula

$$d^2J_\psi(\gamma)(X,Y) = \int_\Sigma \langle L_{\gamma,\psi}(X), Y\rangle dA, \qquad \text{for } X, Y \in T_\gamma L_3^2(S^1, M).$$

Let $B : TM \times TM \to NM$ denote the second fundamental form of M in $\mathbb{R}^N$ and define for $n \in N_pM$,

$$A_n : TM \to TM \qquad \text{by} \qquad \langle A_n(x), y\rangle = B(x,y) \cdot n.$$

Then a calculation (similar to that used to prove Proposition 2.4.1) shows that the Jacobi operator is given by the formula

$$L_{\gamma,\psi}(X) = -\nabla_{\gamma'}\nabla_{\gamma'}(X) - R(X,\gamma')\gamma' - A_{\psi(t)^\perp}(X). \tag{2.19}$$

One readily verifies that $L_{\gamma,\psi}$ is a Fredholm operator of Fredholm index zero.

We next calculate the kernel and image of the map $d\pi_{(\gamma,\psi)} : T_{\gamma,\psi}\mathcal{S} \to L_1^2(S^1,\mathbb{R}^N)$ and after a short derivation, we obtain the results:

$$\text{Kernel of } d\pi_{(\gamma,\psi)} = \{(X,\eta) \in T_{(\gamma,\psi)}\mathcal{S} : \eta = 0, L_\gamma(X) = 0\},$$

$$\text{Range of } d\pi_{(\gamma,\psi)} = \{L_{\gamma,\psi}(X) + \chi^\perp : X \in T_\gamma L_3^2(S^1,M), \chi \in L_1^2(S^1,\mathbb{R}^N)\},$$

where $\chi^\perp$ denotes the orthogonal projection of ψ into the normal space to M. Note that the kernel of $d\pi_{(\gamma,\psi)}$ is isomorphic to the kernel of $L_{\gamma,\psi}$ while the range of $d\pi_{(\gamma,\psi)}$ is a subspace of $L_1^2(S^1,\mathbb{R}^N)$ which has the same codimension as $\text{Range}(L_{\gamma,\psi}) \subset T_\gamma L_1^2(S^1,M)$.

Thus $d\pi_{(\gamma,\psi)}$ is a Fredholm operator with the same Fredholm index as $L_{\gamma,\psi}$, namely zero, finishing the proof of Lemma 2.6.3.

According to the Sard-Smale Theorem 2.6.2, there is a countable intersection of open dense subsets of $\psi \in L_2^2(S^1,\mathbb{R}^N)$ consisting of regular values of π. But if ψ is a regular value for π, then $\text{Kernel}(L_{\gamma,\psi}) = 0$ at each critical point, so all critical points are Morse nondegenerate, and J_ψ is a Morse function, establishing Theorem 2.7.1.

Definition. Suppose that $\mathcal{M}$ is a Banach manifold and $f : \mathcal{M} \to \mathbb{R}$ is a C^2 function. A compact finite-dimensional submanifold N of $\mathcal{M}$ is said to be a *nondegenerate critical submanifold* for f if

1. every $p \in N$ is a critical point for f.
2. if $p \in N$, then T_pN is the set of $V \in T_p\mathcal{M}$ such that $d^2f(p)(V,W) = 0$, for all $W \in T_p\mathcal{M}$.

The *Morse index* of a connected nondegenerate critical submanifold N is the index of any critical point $p \in N$.

Proposition 2.7.4. *If M is a compact Riemannian manifold and*

$$M_0 = \{\gamma \in L_1^2(S^1,M) : \ \gamma \text{ is constant } \},$$

then M_0 is a nondegenerate critical submanifold of $L_1^2(S^1,M)$ of Morse index zero. Moreover, for some $\epsilon > 0$, the open neighborhood

$$U_\epsilon = \{\gamma \in L_1^2(S^1,M) : |\gamma - M_0|_{C^0} < \epsilon\}$$

has M_0 as a strong deformation retract and contains no critical points for the action J other than the elements of M_0.

To prove this, first note that if γ_p denotes the constant loop at the point $p \in M$, then

$$T_{\gamma_p} L_1^2(S^1, M) \cong L_1^2(S^1, T_p M)$$

and the tangent subspace $T_{\gamma_p} M_0$ consists of the constant maps into $T_p M$. On the other hand, the normal space to M_0 at γ_p can be identified with

$$N_{\gamma_p} M_0 = \left\{ V \in L_1^2(S^1, T_p M) : \int_{S^1} V(t) dt = 0 \right\}.$$

According to the second variation formula (2.7),

$$d^2 J(\gamma_p)(V, W) = \int_{S^1} \langle V'(t), W'(t) \rangle dt, \quad \text{for } V, W \in L_1^2(S^1, T_p M),$$

from which we conclude that $d^2 J(\gamma_p)(V, W) = 0$ if V is constant, that is, if V lies in $T_{\gamma_p} M_0$. On the other hand, it is readily verified that $d^2 J(\gamma_p)$ is positive definite on $N_{\gamma_p} M_0$, so M_0 is indeed a nondegenerate critical submanifold of Morse index zero.

Let NM_0 denote the total space of the normal bundle of M_0 and for $\epsilon > 0$, let

$$V_\epsilon = \{ \, V \in N_{\gamma_p} M_0 \text{ for some } p \in M \text{ such that } |V|_{C^0} < \epsilon \, \},$$

and define

$$\phi : V_\epsilon \to U_\epsilon \quad \text{by} \quad \phi(V) = \exp_p(V).$$

For $\epsilon > 0$ sufficiently small, we can define

$$H : U_\epsilon \times [0, 1] \to U_\epsilon \quad \text{by} \quad H(\exp_p(V), t) = \exp_p(tV).$$

Then H is smooth, $H(\gamma, 1) = \gamma$ for $\gamma \in U_\epsilon$, $H(\gamma, t) = \gamma$ for $\gamma \in M_0$ and all $t \in [0, 1]$, so $\gamma \mapsto H(\gamma, 0)$ is the desired deformation retraction from U_ϵ to M_0 . Finally, using a Taylor expansion about points in M_0, one verifies that $U_\epsilon - M_0$ contains no critical points for J.

We can now alter the definition of J_ψ slightly. Suppose that $\eta : \mathbb{R} \to [0, 1]$ is a smooth function such that

$$\eta(t) = \begin{cases} 0, & \text{for } t \leq \epsilon^2/4, \\ 1, & \text{for } t \geq \epsilon^2/2. \end{cases}$$

Note that since $L(\gamma)^2 \leq 2J(\gamma)$, $\eta(J(\gamma)) = 1$ outside U_ϵ. Given a smooth map $\psi : S^1 \to \mathbb{R}^N$, we consider the function J_ψ defined by

$$J_\psi(\gamma) = J(\gamma) + \eta(J(\gamma)) \int_{S^1} (\gamma \cdot \psi) dt. \tag{2.20}$$

It follows from Theorem 2.7.1 that if ψ is generic and sufficiently small, then $J_\psi \geq 0$ and $M_0 = J_\psi^{-1}(0)$ is a critical submanifold of Morse index zero. Moreover, all other critical points of J_ψ are Morse nondegenerate, so the restriction of J_ψ to $L_1^2(S^1, M) - M_0$ is a Morse function.

2.8 Bumpy metrics for smooth closed geodesics*

With additional work, one can show that if M is a compact manifold, then for generic choice of Riemannian metric on M, all nonconstant smooth closed geodesics have the property that their only Jacobi fields are those generated by the S^1 action on $L^2_1(S^1, M)$. This is the "Bumpy Metric Theorem" of Abraham, and a proof due to Anosov is found in [4]; an alternate proof can be found in [7]. We will present another proof in this section, based upon Bott's theory of iterated smooth closed geodesics [8]. The proof is long so some readers may want to skip it on first reading.

We first prove a simpler version for the case where the geodesic is prime, that is, not a nontrivial cover of a geodesic of smaller length:

Theorem 2.8.1. *If M is a compact manifold, then for generic choice of Riemannian metric on M, all nonconstant prime smooth closed geodesics have the property that their only Jacobi fields are those generated by the S^1 action on $L^2_1(S^1, M)$.*

Proof of Theorem 2.8.1: The techniques are exactly the same as those used before. Note first that since critical points for the action function J are smooth, we are not restricted to working in $L^2_1(S^1, M)$, but can work for example in a Sobolev space of more highly differentiable functions, such as $L^2_2(S^1, M)$, which is also a Hilbert manifold.

We need a preliminary step in the argument to deal with the fact that J is G-equivariant, where $G = S^1$. If N is a compact codimension one submanifold of M with boundary ∂N, we let

$$\mathcal{U}(N) = \{ \text{ nonconstant } \gamma \in L^2_k(S^1, M) \ : \ \gamma \text{ does not intersect } \partial N \text{ and has transversal intersection with the interior of } N\},$$

an open subset of $L^2_2(S^1, M)$. We cover $L^2_2(S^1, M)$ with a countable collection $\mathcal{U}_i = \mathcal{U}(N_i)$ of open sets corresponding to a countable collection of codimension two submanifolds N_i. If 0 is a choice of origin in S^1, we let

$$\mathcal{F}_i = \{\gamma \in \mathcal{U}(N_i) : \gamma(0) \in N_i\},$$

a C^1 submanifold of $L^2_2(S^1, M)$ by the smoothness theorems in §1.6. Note that $\mathcal{F}_i$ meets each nonconstant G-orbit in $\mathcal{U}(N_i)$ in a finite number of points.

Let $\mathcal{M}$ denote the manifold of L^2_ℓ Riemannian metrics on M, an open subset of the space of L^2_ℓ sections of the second symmetric power of T^*M, where ℓ is a large integer. We claim that

$$\mathcal{S} = \{(\gamma, g) \in \mathcal{F}_i \times \mathcal{M} : \gamma \text{ is a geodesic for the metric } g\}$$

is a smooth submanifold of $\mathcal{F}_i \times \mathcal{M}$. To see this, we let $L^2_*(S^1, TM)$ denote the vector bundle over $\mathcal{F}_i$ whose fiber at each $f \in L^2_2(S^1, M)$ is $L^2(S^1, \gamma^*TM)$. It follows from multiplication theorems from the theory of Sobolev spaces ($L^2_2 \cdot L^2 \subset$

L^2) that $L^2_*(S^1, TM)$ does in fact have the structure of a smooth vector bundle over $\mathcal{F}_i$. We next define a C^1 map

$$F : \mathcal{F}_i \times \mathcal{M} \longrightarrow L^2_*(S^1, TM) \quad \text{by} \quad F(\gamma, g) = \nabla^g_{\gamma'}\gamma',$$

where ∇^g is the Levi-Civita connection on TM determined by the metric g on M. Let $\mathcal{Z}$ denote the image of the zero section of $L^2_*(\Sigma, TM)$ and note that $\mathcal{S} = F^{-1}(\mathcal{Z})$. Our goal is to show that F is as transverse to $\mathcal{Z}$ as possible.

The first derivative of the action is given by the formula

$$dJ(\gamma)(X) = \int_{S^1} \langle F(\gamma, g), X\rangle dt.$$

Differentiating once again gives us the Hessian at a critical point,

$$d^2J(\gamma)(X, Y) = \int_{S^1} \langle D_1F(\gamma, g)(X), Y\rangle dt = \int_{S^1} \langle L_\gamma(X), Y\rangle dt, \tag{2.21}$$

where D_1F denotes derivative with respect to the first variable

$$\{\gamma \in L^2_2(S^1, M) : \gamma(0) \in N_i\},$$

and

$$L_\gamma : \{X \in L^2_2(S^1, \gamma^*TM) : X(0) \in T_{\gamma(0)}N_i\} \longrightarrow L^2(S^1, \gamma^*TM) \tag{2.22}$$

is the Jacobi operator. Note that at a zero (γ, g) of F the tangent space to $L^2_*(S^1, TM)$ can be divided into a direct sum

$$T_\gamma(L^2_*(S^1, TM)) = H \oplus V,$$

where H is horizontal (tangent to the zero section) and V is vertical (tangent to the fiber). If π_V denote the projection onto V along H, it follows from (2.21) that

$$\pi_V \circ (D_1F)_{(\gamma,g)}(X) = L_\gamma(X). \tag{2.23}$$

Because it is a Jacobi field, the section $T(t) = \gamma'(t)$ is not in the image of L_γ. However, we claim that

$$\{\text{the image of } \pi_V \circ DF_{\gamma,g}\} \oplus (\text{span of } T(t)) = L^2(S^1, \gamma^*TM), \tag{2.24}$$

or equivalently, that $\mathcal{V}$ is spanned by $T(t)$, where

$$\mathcal{V} = \{L^2\text{-sections } X \text{ of } \gamma^*TM : X \text{ is } \perp \text{ to the image of } \pi_V \circ DF_{\gamma,g}\}.$$

To show this, we need to calculate $\pi_V \circ (D_2F)$, where D_2F is the partial derivative with respect to the second variable $g \in \mathcal{M}$.

Suppose that γ is a nonconstant geodesic, which will be C^ℓ when the metric is L^2_ℓ. Choose a point $t \in S^1$ which possesses a neighborhood U such that γ imbeds U into some open set $W \subset M$ on which *Fermi coordinates* $(x_1, \ldots, x_n)$ are defined. Such coordinates satisfy the following conditions:

1. γ is described by the equations $x_2 = \cdots = x_n = 0$,
2. $x_1 \circ \gamma = t$,
3. the metric g takes the form $\sum g_{ij} du^i du^j$, such that on $\gamma(U)$, $g_{ij} = \delta_{ij}$, for $1 \leq i, j \leq n$.

In terms of these local coordinates, the equation for geodesics becomes

$$\frac{d^2 x_k}{dt^2} + \sum_{i,j} \Gamma^k_{ij} \frac{dx_i}{dt} \frac{dx_j}{dt} = 0.$$

The expression on the left side of this equation is called the acceleration of the path.

A perturbation in the metric $h \in T_g \mathcal{M}$ with compact support in W can be written in the form $h = \sum h_{ij} dx_i dx_j$. Under this perturbation, the only piece of the acceleration that changes is the Christoffel symbol

$$\Gamma^k_{ij} = \frac{1}{2} \sum g^{kl} \left(\frac{\partial g_{il}}{\partial x_j} + \frac{\partial g_{jl}}{\partial x_i} - \frac{\partial g_{ij}}{\partial x_l} \right),$$

and if $\dot{\Gamma}^k_{ij}$ denotes the derivative of Γ^k_{ij} in the direction of the perturbation,

$$\dot{\Gamma}^k_{ij} = \frac{1}{2} \left(\frac{\partial h_{ik}}{\partial x_j} + \frac{\partial h_{jk}}{\partial x_i} - \frac{\partial h_{ij}}{\partial x_k} \right).$$

We then find that

$$\pi_V \circ (D_2 F)_{(\gamma, g)}(h) = \sum_{i,j,k=1}^{n} \dot{\Gamma}^k_{ij} \frac{dx_i}{dt} \frac{dx_j}{dt} \frac{\partial}{\partial x_k} = \sum_{k=1}^{n} \dot{\Gamma}^k_{11} \frac{\partial}{\partial x_k}. \tag{2.25}$$

We can select the perturbation so that

$$h_{11} = x_2 \phi, \qquad h_{ij} = 0, \quad \text{for other choices of indices } i, j,$$

where ϕ has compact support in W and $2 \leq r \leq n$; we then see that

$$\dot{\Gamma}^r_{11}(t, 0, \ldots, 0) = -\frac{1}{2} \frac{\partial}{\partial x_2}(h_{11})(t, 0, \ldots, 0) = -\frac{1}{2} \phi(t, 0, \ldots, 0).$$

Thus the fiber projection of the partial derivative of F with respect to the second variable (in $\mathcal{M}$) is given by the expression

$$\pi_V \circ (D_2 F)_{f,g}(h) = -\frac{1}{2} \sum_{r=2}^{n} \dot{\Gamma}^r_{11} \frac{\partial}{\partial x_r} = -\frac{1}{2} \phi \frac{\partial}{\partial x_2}.$$

In this manner, we can show that any vector field of the form

$$\sum_{r=2}^{n} \phi^r \frac{\partial}{\partial x_r}$$

where ϕ^r is a smooth function with compact support, lies in the image of $\pi_V \circ D_2 F_{f,g}$, and hence elements of $\mathcal{V}$ must be tangent to γ in U.

Since a dense open subset of points of Σ can be covered by sets of the form U, we see that elements of $\mathcal{V}$ must be tangent to γ over all of S^1. On the other hand, they must also be image of L_γ restricted to complement of the span of $T(t)$ by (2.23). Thus we obtain the desired result (2.24).

The remainder of the proof is similar to the proof of Theorem 2.7.1. Note that the tangent space to $\mathcal{S}$ is given by the formula

$$T_{\gamma,g}\mathcal{S} = \{(X,h) \in T_\gamma \mathcal{F}_i \times T_g \mathcal{M} : L_\gamma(X) + \pi_V \circ (D_2 F)(h) = 0\},$$

which makes it easy to analyze the kernel and range of the map $d\pi_{(\gamma,g)} : T_{(\gamma,g)}\mathcal{S} \to T_g\mathcal{M}$. Indeed,

$$\text{Kernel of } d\pi_{(\gamma,g)} = \{(X,h) \in T_\gamma \mathcal{F}_i \times T_g \mathcal{M} : h = 0, L_\gamma(X) = 0\},$$

while

$$\begin{aligned}\text{Range of } d\pi_{(\gamma,g)} = \{h \in T_g\mathcal{M};\ &\text{there exists an element } X \in T_\gamma L_2^2(S^1, M) \\ &\text{such that } X(0) \in T_{\gamma(0)} N_i \text{ and } L_\gamma(X) = -\pi_V \circ (D_2 F)(h)\}.\end{aligned}$$

Thus the kernel of $d\pi_{(\gamma,g)}$ is isomorphic to the kernel of L_γ. On the other hand, if h is an element of $T_g\mathcal{M}$ such that

$$\pi_V \circ (D_2 F)(h) = 0, \qquad \text{then} \qquad (0,h) \in T_{(\gamma,g)}\mathcal{S},$$

and hence h lies in the range of $d\pi_{(\gamma,g)}$. It follows that complement to the range of $d\pi_{(\gamma,g)}$ must inject to a complement to the range of L_γ, and in particular, any such complement is finite-dimensional, so $d\pi_{(\gamma,g)}$ is a Fredholm map, and the dimension of the cokernel of $d\pi_{(\gamma,g)}$ is no larger than the dimension of the cokernel of L_γ. By the earlier transversality argument, $dF_{\gamma,g}$ maps surjectively onto a complement to the one-dimensional space generated by the Jacobi field $T(t)$, so the dimension of the cokernel of $d\pi_{(\gamma,g)}$ is actually one less than the dimension of the cokernel of L_γ.

Thus the Fredholm index of $d\pi_{(\gamma,\psi)}$ one more than the Fredholm index of the map L_γ of (2.22), which is one because the restriction $X(0) \in T_{\gamma(0)} N_i$ cuts down the kernel by one. It follows that the Fredholm index of $d\pi$ itself is zero. Thus we can use the Sard-Smale Theorem to show that a countable intersection of open dense subsets of $\mathcal{M}$ consist of regular values for π. If g_0 is such a "generic metric," all of the prime geodesics will be Morse nondegenerate, finishing the proof of the Bumpy Metric Theorem for prime geodesics.

Implication of the above Theorem: It follows from Theorem 2.8.1, together with condition C of Palais and Smale, that for a Riemannian metric belonging to a residual subset, the number of S^1-orbits of nonconstant prime geodesics of length less than a given bound is finite.

We will next extend the argument from the previous theorem to cover the case in which geodesics are not necessarily prime, thereby obtaining:

Bumpy Metric Theorem 2.8.2. *If M is a compact manifold, then for generic choice of Riemannian metric on M, all nonconstant smooth closed geodesics have the property that their only Jacobi fields are those generated by the S^1 action on $L^2_1(S^1, M)$.*

Of course a geodesic which is not prime covers a prime minimal geodesic and our strategy is to study this underlying prime geodesic.

We make the following definitions following Bott [8]. Suppose that $\gamma : \mathbb{R} \to M$ is a smooth geodesic with $\gamma(t + 2\pi) = \gamma(t)$. If $q \in \mathbb{Z}$ and $z \in S^1 \subset \mathbb{C}$, we let

$$\mathcal{V}^\infty_{q,z} = \{ \text{ smooth sections } X \text{ of } \gamma^* TM \otimes \mathbb{C} \ : X(t + 2\pi q) = zX(t)\},$$

and define an inner product

$$\langle \cdot, \cdot \rangle_q : \mathcal{V}^\infty_{q,z} \times \mathcal{V}^\infty_{q,z} \longrightarrow \mathbb{C}$$

by

$$\langle X, \bar{Y} \rangle_q = \int_0^{2\pi q} \left[\left\langle \frac{DX}{\partial t}, \frac{D\bar{Y}}{\partial t} \right\rangle + \langle X, \bar{Y} \rangle \right] dt,$$

where D denotes the covariant derivative defined by the Levi-Civita connection on M. Let $\mathcal{V}_{q,z}$ denote the completion of $\mathcal{V}^\infty_{q,z}$ with respect to $\langle \cdot, \cdot \rangle_q$ and define

$$I_q(\cdot, \cdot) : \mathcal{V}_{q,z} \times \mathcal{V}_{q,z} \longrightarrow \mathbb{C}$$

by

$$I_q(X, \bar{Y}) = \int_0^{2\pi q} \left[\left\langle \frac{DX}{\partial t}, \frac{D\bar{Y}}{\partial t} \right\rangle - \langle R(X, \gamma')\gamma', \bar{Y} \rangle \right] dt,$$

which is of course the restriction of $d^2 J(\gamma)$ to $\mathcal{V}_{q,z}$.

Lemma 2.8.3. *The inclusion*

$$\sum_{z^q = 1} \mathcal{V}_{1,z} \subset \mathcal{V}_{q,1} \tag{2.26}$$

is an isomorphism.

Since the inclusion is clearly injective, it suffices to show that it is surjective. If $X \in \mathcal{V}_{q,1}$ and z is a primitive q-th root of unity, we let

$$X_z(t) = \frac{1}{q} \sum_{j=0}^{q-1} z^{-j} X(t + 2\pi j).$$

Then

$$\begin{aligned} X_z(t + 2\pi) &= \frac{1}{q} \sum_{j=0}^{q-1} z^{-j} X(t + 2\pi(j + 1)) \\ &= \frac{z}{q} \sum_{j=0}^{q-1} z^{-(j+1)} X(t + 2\pi(j + 1)) = zX_z(t), \end{aligned}$$

so $X_z \in \mathcal{V}_{1,z}$. Moreover,

$$X = \sum_{z^q=1} X_z,$$

so the inclusion (2.26) is indeed an isomorphism, proving the Lemma.

Note that if $X \in \mathcal{V}_{1,z_1}$ and $Y \in \mathcal{V}_{1,z_2}$, where z_1 and z_2 are q-th roots of unity, then

$$\begin{aligned} I_q(X,\bar{Y}) = &\int_0^{2\pi} \left[\left\langle \frac{DX}{\partial t}, \frac{D\bar{Y}}{\partial t} \right\rangle - \langle R(X,\gamma')\gamma', \bar{Y}\rangle \right] dt \\ &+ \int_{2\pi}^{4\pi} z_1\bar{z}_2 \left[\left\langle \frac{DX}{\partial t}, \frac{D\bar{Y}}{\partial t} \right\rangle - \langle R(X,\gamma')\gamma', \bar{Y}\rangle \right] dt \\ &+ \cdots + \int_{2\pi(q-1)}^{2\pi q} z_1^{q-1}\bar{z}_2^{q-1} \left[\left\langle \frac{DX}{\partial t}, \frac{D\bar{Y}}{\partial t} \right\rangle - \langle R(X,\gamma')\gamma', \bar{Y}\rangle \right] dt \\ &= \left(\sum_{j=0}^{q-1} z_1^j \bar{z}_2^j \right) \int_0^{2\pi} \left[\left\langle \frac{DX}{\partial t}, \frac{D\bar{Y}}{\partial t} \right\rangle - \langle R(X,\gamma')\gamma', \bar{Y}\rangle \right] dt. \end{aligned}$$

Thus we see that

$$I_q(X,\bar{Y}) = \begin{cases} qI_1(X\bar{Y}), & \text{if } z_1 = z_2, \\ 0, & \text{if } z_1 \neq z_2, \end{cases}$$

and hence the direct sum decomposition $\sum_{z^q=1} \mathcal{V}_{1,z}$ of $\mathcal{V}_{q,1}$ is orthogonal with respect to the index form I_q. Let $N(z)$ denote the *nullity* of the index form I_1 restricted to $\mathcal{V}_{1,z}$,

$$N(z) = \dim_{\mathbb{C}}\{X \in \mathcal{V}_{1,z} : I_1(X,\bar{Y}) = 0 \text{ for all } Y \in \mathcal{V}_{1,z}\}.$$

The above discussion proves the following lemma due to Bott [8], which plays a key role in his analysis of the relationship between the index and nullity of a prime smooth closed geodesic and the index of nullity of its multiple covers:

Lemma 2.8.4. *Let γ^q denote the q-fold iterate of the smooth closed geodesics γ, so $\gamma^q(t) = \gamma(qt)$. Then*

$$\textit{Nullity of } \gamma^q = \sum_{z^q=1} N(z).$$

We now return to the proof of the bumpy metric theorem itself. It suffices to consider geodesics whose length is less than a given bound. We know that there are only finitely many prime geodesics with length below this bound and that if the metric is perturbed by a sufficiently small amount, no new prime geodesics will be introduced. Our strategy is to perturb the metric in a neighborhood of a given geodesic in such a way that the geodesic is preserved.

As in the proof of the preceding Theorem, we construct a perturbation of the Riemannian metric on M of a specific form. Once again, we choose a point

$t \in S^1$ and a neighborhood U containing t such that γ imbeds U into some open set $W \subset M$ on which Fermi coordinates $(x_1, \ldots, x_n)$ are defined, coordinates such that:

1. γ is described by the equations $x_2 = \cdots = x_n = 0$,
2. $x_1 \circ \gamma = t$,
3. the metric g takes the form $\sum g_{ij} du^i du^j$, such that on $\gamma(U)$, $g_{ij} = \delta_{ij}$, for $1 \leq i, j \leq n$.

Following Klingenberg ([42], Proposition 3.3.7), we construct a perturbation of the metric $\dot{g} = \sum \dot{g}_{ij} dx_i dx_j$ such that

$$\dot{g}_{11}(x_1, \ldots x_n) = \sum_{r,s=2}^{n} x_r x_s \alpha_{rs}(x_1, x_2, \ldots, x_n), \quad \dot{g}_{ij} = 0 \quad \text{if} \quad (i,j) \neq (1,1).$$

Here α_{rs} are smooth functions which vanish outside a small tubular neighborhood of γ. A straightforward calculation shows that the resulting changes in the Christoffel symbols

$$\dot{\Gamma}_{kij} = \frac{1}{2} \left(\frac{\partial \dot{g}_{ik}}{\partial u_j} + \frac{\partial \dot{g}_{ik}}{\partial u_j} - \frac{\partial \dot{g}_{ik}}{\partial u_j} \right)$$

vanish unless at least two of the indices are 1, and if $2 \leq r, s \leq n$,

$$\dot{\Gamma}_{r11} = -\sum_s u_s \alpha_{rs}, \quad \dot{\Gamma}_{1r1} = \sum_s u_s \alpha_{rs}.$$

The corresponding changes in curvature components are given by the formulae

$$\dot{R}_{lijk} = \frac{\partial}{\partial u_i} \left(\dot{\Gamma}_{ljk} \right) - \frac{\partial}{\partial u_j} \left(\dot{\Gamma}_{lik} \right) + \sum_m \dot{\Gamma}_{lim} \Gamma^m_{jk} + \sum_m \Gamma_{lim} \dot{\Gamma}^m_{jk} - \sum_m \dot{\Gamma}_{ljm} \Gamma^m_{ik} - \sum_m \Gamma_{ljm} \dot{\Gamma}^m_{ik}.$$

But along $f_0(\Sigma_0)$ all the $\dot{\Gamma}_{kij}$'s and $\dot{\Gamma}^k_{ij}$'s must vanish, so the resulting change in the curvature R_{1r1s} along γ is given by

$$\dot{R}_{1r1s}(x_1, 0, \ldots, 0) = \alpha_{rs}(x_1, 0, \ldots, 0).$$

The Jacobi operator L_γ on normal sections can be expressed in components as follows: If

$$X = \sum_{r=2}^{n} f_r \frac{\partial}{\partial x_r}, \quad \text{then} \quad L_\gamma(X) = -\sum_{r=2}^{n} \left[\frac{d^2 f_r}{dx_1^2} + \sum_{s=2}^{n} R_{1r1s} f_s \right] \frac{\partial}{\partial x_r}.$$

We can consider the family of formally self-adjoint second order differential operators T of the form

$$T(X) = -\sum_{r=2}^{n} \left[\frac{d^2 f_r}{dx_1^2} + \sum_{s=2}^{n} T_{rs} f_s \right] \frac{\partial}{\partial x_r}.$$

For each such operator and each q-th root of unity $z \in S^1 \subset \mathbb{C}$, we can define the nullity

$$N(z) = \dim_{\mathbb{C}}\{X \in \mathcal{V}_{1,z} : L(X) = 0\}.$$

For an open dense set of such operators T, $N(z) = 0$. The argument in the preceding paragraph shows that we can perturb the Jacobi equation along any geodesic so that $N(z) = 0$. Since there are a finite number of roots of unity corresponding to covers γ^q of a fixed prime geodesic which have length less than a given bound, we can ensure that for generic metrics all such geodesics γ^q will be Morse nondegenerate. This finally proves the Bumpy Metric Theorem for geodesics.

2.9 Adding handles

Suppose that $\mathcal{M}$ is a smooth manifold modeled on a Hilbert space with a complete Riemannian metric, and that $f : \mathcal{M} \to \mathbb{R}$ is a C^2 Morse function satisfying condition C. From the Deformation Theorem 1.11.1, we know that the topology of

$$\mathcal{M}^a = \{p \in \mathcal{M} : f(p) \leq a\}$$

does not change as a increases unless a passes a critical value for f. Our goal now is to understand what happens to the topology of $\mathcal{M}^a$ when a passes a critical value c for f such that $f^{-1}(c)$ contains a finite number of critical points, all Morse nondegenerate. In this section, we carry our that analysis for the case of Hilbert manifolds.

Theorem 2.9.1. *Suppose that $\mathcal{M}$ is a Hilbert manifold with a complete Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ and that $f : \mathcal{M} \to [0, \infty)$ is a smooth function satisfying condition C. If the interval $[a, b]$ contains a single critical value c for f, there is exactly one critical point p for f such that $f(p) = c$ and this critical point is Morse nondegenerate of Morse index λ, then $\mathcal{M}^b$ is homotopy equivalent to $\mathcal{M}^a$ with a handle of index λ attached.*

To prove this, we choose a coordinate chart (U, ϕ) about the nondegenerate critical point p with $\phi(p) = 0 \in E$, where E is the model space for $\mathcal{M}$, and let $\bar{f} = f \circ \phi^{-1}$, a smooth function on the open subset $\phi(U) \subset E$ which has a single nondegenerate critical point of index λ at the origin. Expanding the smooth function $\bar{f}$ in a Taylor series yields

$$\bar{f}(x) = c + \frac{1}{2} d^2 \bar{f}(0)(x, x) + R_1(x), \qquad \text{where } \|R_1(x)\| \leq \epsilon_1 \|x\|^2, \tag{2.27}$$

for $x \in \phi(U)$, where ϵ_1 is a positive constant which can be made arbitrarily small by making U small. We can identify $T_p\mathcal{M}$ with E and since p is Morse nondegenerate, the self-adjoint bounded linear operator A on E determined by

$$d^2 \bar{f}(0)(x, y) = \langle\langle Ax, y \rangle\rangle, \qquad \text{for } x, y \in E,$$

where $\langle\langle \cdot, \cdot \rangle\rangle$ is the Riemannian metric, is invertible, so its spectrum is a closed subset of the real axis, bounded away from zero. Corresponding to the restriction of the vector field $\mathcal{X} = \text{grad}(f)$ to U is a vector field on $\phi(U)$ with local representative $\bar{\mathcal{X}}$ which has the Taylor series expansion

$$\bar{\mathcal{X}}(x) = Ax + R_2(x), \qquad \text{where } \|R_2(x)\| \leq \epsilon_2 \|x\|, \tag{2.28}$$

for $x \in \phi(U)$, where ϵ_2 is an arbitrarily positive constant. The vector field $\bar{\mathcal{X}}$ on $\phi(U)$ is closely approximated by its linearization $\mathcal{X}_A$ which has local representative $\mathcal{X}_A(x) = Ax$, and one can check that this linearization is a pseudo-gradient (as defined in § 1.11) when ϵ_2 is sufficiently small.

If the critical point p has finite index, the negative part of the spectrum is discrete and consists of finitely many eigenvalues. In this case, we let E_- be the subspace of E generated by the negative eigenvalues of A and let E_+ be the closed orthogonal complement to E_- in E. In general, the spectral theorem allows us to divide E into a direct sum $E = E_+ \oplus E_-$, each summand being preserved by A, and hence we can think of A as dividing into a "block matrix"

$$A = \begin{pmatrix} A_- & 0 \\ 0 & A_+ \end{pmatrix}$$

with A_- and A_+ self-adjoint invertible operators on the subspaces E_- and E_+ respectively. The spectrum of A_- lies on the negative real axis, while the spectrum of A_+ lies on the positive real axis. Thus if we suppose that x_- and x_+ are the orthogonal projections of x into E_- and E_+ respectively, so $x = x_- + x_+$, the system of differential equations represented by the linearization $-\mathcal{X}_A$ is

$$\begin{cases} dx_-/dt = -A_- x_-, \\ dx_+/dt = -A_- x_+. \end{cases}$$

This is just a linear system with constant coefficients, which has $\{\tilde{\phi}_t : t \in \mathbb{R}\}$ as its one parameter group of diffeomorphisms, where

$$\tilde{\phi}_t(x_-, x_+) = (\exp(-tA_- x_-), \exp(-tA_+ x_+)). \tag{2.29}$$

The fluid flow for $-\mathcal{X}_A$ described by $\{\tilde{\phi}_t : t \in \mathbb{R}\}$ is expanding on the subspace E_-, contracting on E_+ and closely approximates the fluid flow for $\bar{\mathcal{X}}$.

Let V be an open subset of U such that $p \in V \subseteq \bar{V} \subseteq U$.

Lemma 2.9.2. *There exist $r_1 > 0$ and $s_1 > 0$ such that*

1. *$D_-(r) \times D_+(s) \subseteq V$ and $\mathcal{X}_A$ is transverse to $D_-(r) \times \partial D_+(s)$ when $r \leq r_1$ and $s \leq s_1$,*
2. *$\bar{f}(\partial D_-(r_1) \times D_+(s_1)) \leq c - \epsilon$, for some $\epsilon > 0$, and*
3. *$\bar{f}^{-1}(c - \epsilon)$ is transverse to $D_-(r_1) \times x_+$ for $x_+ \in D_+(s_1)$.*

The first condition follows immediately from the explicit form (2.29) of the fluid flow for $\mathcal{X}_A$.

To prove the second claim, we use the Taylor expansion (2.27) to conclude that

$$\bar{f}(x) = c + \frac{1}{2}\langle\langle A_- x_-, x_-\rangle\rangle + \frac{1}{2}\langle\langle A_+ x_+, x_+\rangle\rangle + R_1(x)$$
$$\leq c - \frac{1}{2\|A_-^{-1}\|}\|x_-\|^2 + + \frac{1}{2}\|A_+\|\|x_+\|^2 + \epsilon_1\|x\|^2.$$

We can choose ϵ_1 smaller than $(1/(4\|A_-\|^2)$. Thus if we choose r_1 small and then s_1 much smaller, we can arrange that the second claim holds. The third claim is proven in a similar fashion, and we leave it as an exercise for the reader.

In accordance with the terminology of geometric topology, we call $\partial D_-(r_1)$ and $D_-(r_1)$ the *descending sphere* and *descending disk* of the critical point p; $\partial D_-(r_1)$ is a sphere of dimension $\lambda - 1$. We also have an *ascending sphere* $\partial D_+(s_1)$ and an *ascending disk* $D_+(s_1)$, both of which are infinite-dimensional.

Returning now to the proof of the theorem, we let $\eta : \mathcal{M} \to [0, 1]$ be a smooth function such that $\eta(q) = 1$ for $q \in \mathcal{M} - U$, $\eta(q) = 0$ for $q \in V$ and let

$$\mathcal{X} = \eta \text{ grad}(f) + (1 - \eta)\ \phi_*^{-1}\mathcal{X}_A.$$

Finally, we let $\{\phi_t : t \in \mathbb{R}\}$ be the one-parameter group of local diffeomorphisms corresponding to the vector field $-\mathcal{X}$. For $q \in \mathcal{M}^b - \mathcal{M}^{c-\epsilon}$, let $\tau(q)$ denote the first time t such that

$$\phi_t(q) \in \mathcal{M}^{c-\epsilon} \cup \phi^{-1}(D_-(r_1) \times \partial D_+(s_1)).$$

Note that $\tau(q)$ is finite by the argument for the Theorem 1.10.1 and the transversality conditions of Lemma 2.8.4 show that $\tau(q)$ depends continuously on q. Let $h_t : \mathcal{M}^b \to \mathcal{M}^b$ by $h_t(q) = \phi_{t\tau(x)}(q)$. Then clearly h_0 is the identity map and

$$h_1(\mathcal{M}^b) \subset \mathcal{M}^{c-\epsilon} \cup \phi^{-1}(D_-(r_1) \times \partial D_+(s_1)).$$

In fact, we easily check that h_t gives a deformation retraction from $\mathcal{M}^b$ to $\mathcal{M}^{c-\epsilon} \cup \phi^{-1}(D_-(r_1) \times \partial D_+(s_1))$, which is homotopy equivalent to $\mathcal{M}^{c-\epsilon}$ with a handle of index λ attached. By the Deformation Theorem 1.11.1 this has the homotopy type of $\mathcal{M}^a$ with a handle of index λ attached.

Remark 2.9.3. In the proof, the Riemannian metric on $\mathcal{M}$ is used only to construct the continuous isomorphism $A : T_p\mathcal{M} \to T_p\mathcal{M}$ and the gradient vector field $\mathcal{X}$ away from the critical point p. This suggests that it might be convenient to formulate the notion of "gradient-like" vector field, more flexible than an ordinary gradient, a concept utilized by Milnor for finite-dimensional manifolds in [51]:

Definition. Suppose that $f : \mathcal{M} \to \mathbb{R}$ is a C^2 Morse function on a Hilbert manifold $\mathcal{M}$ with a complete Riemannian metric, and let $K \subseteq \mathcal{M}$ be the critical

locus of f. We say that a C^2 vector field $\mathcal{X}$ on $\mathcal{M}$ is *gradient-like* for f if $\mathcal{M}$ possesses an open cover $\{U, V\}$ such that

1. the restriction of f to $\mathcal{M} - K$ is a pseudogradient for f,
2. the open set U divides into connected components U_p, one for each $p \in K$, such that U_p is the domain of a coordinate system ϕ_p such that $\phi_p(p) = 0$, and
3. $(\phi_p)_*(\mathcal{X}|U_p) = \mathcal{X}_A$, where $\mathcal{X}_A$ has local representative

$$\mathcal{X}_A(x) = Ax, \quad \text{for } x \in \phi_p(U), \tag{2.30}$$

and $A : T_p\mathcal{M} \to T_p\mathcal{M}$ is invertible and self-adjoint with respect to some Hilbert space inner product $\langle\langle \cdot, \cdot \rangle\rangle$ on $T_p\mathcal{M}$.

If $\{\eta_V\} \cup \{\eta_p : p \in K\}$ is a C^k partition of unity with respect to the open cover $\{V\} \cup \{U_p : p \in K\}$, and $\mathcal{X}_V$ is a pseudogradient for f on V, then

$$\mathcal{X} = \eta_V \mathcal{X}_V + \sum_{p \in K} \eta_p [(\phi_p^{-1})_* \mathcal{X}_A]$$

is a C^k gradient-like vector field for f on $\mathcal{M}$, so long as the open neighborhoods U_p are chosen sufficiently small. Even if the Morse function f is only C^2 (so its gradient is only C^1), we can construct C^k gradient-like vector fields for f on $\mathcal{M} - K$ so long as $\mathcal{M}$ admits C^k partitions of unity.

Note that it was a gradient-like vector field for f that we utilized in the proof of Theorem 2.8.1. Since pseudogradients form an open subset of the space of vector fields, gradient-like vector fields give considerable flexibility in constructions involving flows corresponding to Morse functions, as we will see when constructing the Morse-Witten chain complex for a Morse function f in the next section.

We can extend Theorem 2.9.1 to the case of several Morse nondegenerate critical points:

Theorem 2.9.4. *Suppose that $\mathcal{M}$ is a Hilbert manifold with a complete Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ and that $f : \mathcal{M} \to [0, \infty)$ is a smooth function satisfying condition C. If the interval $[a, b]$ contains a single critical value c for f and there are finitely many critical point $p_1, \ldots, p_k$ for f such that $f(p_i) = c$ which are Morse nondegenerate with finite Morse indices $\lambda_1, \ldots \lambda_k$, then $\mathcal{M}^b$ is homotopy equivalent to $\mathcal{M}^a$ with handles of index $\lambda_1, \ldots, \lambda_k$ attached.*

The proof is a straightforward extension of the proof of Theorem 2.8.1.

Remark 2.9.5. Theorems 2.9.1 and 2.9.4 both hold for critical points of infinite Morse index, although attaching handles of infinite index turns out to be invisible from the homotopy theory point of view. Moreover, in the problems we have been considering from the calculus of variations, the Morse index of

critical points is always finite, so the handles constructed via these theorems are finite-dimensional.

The Bumpy Metric Theorem 2.8.2 provides motivation for extending Theorems 2.9.1 and 2.9.4 to the case of nondegenerate critical submanifolds. If $N \subset \mathcal{M}$ is a compact nondegenerate critical submanifold of finite Morse index λ, the normal bundle νN to N is defined via the Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ on $\mathcal{M}$; its fiber at $p \in N$ is

$$(\nu N)_p = \{V \in T_p\mathcal{M} : \langle\langle V, W \rangle\rangle = 0 \text{ for all } W \in T_pN\}.$$

The normal bundle has a finite-dimensional subbundle $\nu_- N$ whose fiber at p is generated by eigenvectors corresponding to the negative eigenvalues of $A : T_p\mathcal{M} \to T_p\mathcal{M}$, the rank of the bundle $\nu_- N$ being the Morse index of N. Let

$$D(\nu_- N) = \{V \in \nu_- N : \|V\| \leq 1\}, \qquad S(\nu_- N) = \{V \in \nu_- N : \|V\| = 1\},$$

the unit disk and unit sphere bundles in the negative normal bundle $\nu_- N$.

Theorem 2.9.6. *Suppose that $\mathcal{M}$ is a Hilbert manifold with a complete Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ and that $f : \mathcal{M} \to [0, \infty)$ is a smooth function satisfying condition C. If $[a, b]$ contains a single critical value c for f, the set of critical of points with value c forming a nondegenerate critical submanifold N of finite Morse index λ, then $\mathcal{M}^b$ is homotopy equivalent to $\mathcal{M}^a$ with the disk bundle $D(\nu_- N)$ attached to $\mathcal{M}^a$ along $S(\nu_- N)$.*

The proof is a relatively straightforward modification of the proof for Theorem 2.9.1. Note that we can allow N to have more than one component.

2.10 Morse inequalities

Suppose now that $\mathcal{M}$ is a Hilbert manifold with a complete Riemannian metric $\langle\langle \cdot, \cdot \rangle\rangle$ and that $f : \mathcal{M} \to [0, \infty)$ is a Morse function satisfying condition C, all critical points having finite Morse index. For $a \in \mathbb{R}$, we let

$$\mathcal{M}^a = \{p \in \mathcal{M} : f(p) \leq a\}.$$

Condition C implies that each $\mathcal{M}^a$ has only finitely many nondegenerate critical points. We would like to investigate how the topology of Ω^a changes as a increases. There are two ways of tracking the changes in topology, one is via the Morse inequalities which we discuss next, the other via the Morse-Witten chain complex for f which will be treated in § 2.11.

For each nonnegative integer λ, we let μ_λ^a denote the number of critical points for f within $\mathcal{M}^a$ of Morse index λ and for a given choice of field F (such as $\mathbb{R}$ or $\mathbb{Z}_2$), we let β_λ^a be the dimension of $H^i(\mathcal{M}^a; F)$. Then the *weak Morse inequalities* state that

$$\mu_\lambda^a \geq \beta_\lambda^a. \tag{2.31}$$

Instead of proving these directly, we will prove a stronger version of the Morse inequalities in terms of the *Morse polynomial* for f and *Poincaré polynomial* for $\mathcal{M}^a$, defined respectively by

$$\mathcal{M}^a(t) = \sum_{\lambda=0}^{\infty} \mu_\lambda^a t^\lambda, \qquad \mathcal{P}^a(t) = \sum_{\lambda=1}^{\infty} \beta_\lambda^a t^\lambda.$$

Then (2.31) is an immediate consequence of:

Theorem 2.10.1. (Morse inequalities) *$\mathcal{M}$ is a Hilbert manifold with a complete Riemannian metric and that $f : \mathcal{M} \to [0, \infty)$ is a Morse function satisfying condition C. Then there is a polynomial*

$$\mathcal{Q}^a(t) = \sum_{\lambda=0}^{\infty} q_\lambda^a t^\lambda \quad \text{with } q_\lambda^a \geq 0 \text{ such that} \quad \mathcal{M}^a(t) - \mathcal{P}^a(t) = (t+1)\mathcal{Q}^a(t). \tag{2.32}$$

We make the simplifying assumption that there is only critical point for each critical value, so that we can apply Theorem 2.9.1 instead of Theorem 2.9.4. This can be arranged quite easily by simply perturbing f. Our strategy is to apply an induction on a.

To start the induction, we note that when we set $a < 0$, $\mathcal{M}^a$ is empty and hence $\beta_\lambda^a = \mu_\lambda^a = 0$. We can therefore set $\mathcal{Q}^a(t) = 0$ in this case.

For the inductive step, observe first that if the interval $[a, b]$ contains a single critical value c corresponding to a Morse nondegenerate critical point of Morse index λ, then by Theorem 2.9.1,

$$H_k(\mathcal{M}^b, \mathcal{M}^a; F) \cong \begin{cases} F, & \text{if } k = \lambda, \\ 0, & \text{if } k \neq \lambda. \end{cases}$$

Thus it follows from the exact sequence

$$\cdots \to H_k(\mathcal{M}^a; \mathbb{Z}) \to H_k(\mathcal{M}^b; \mathbb{Z}) \to H_k(\mathcal{M}^b, \mathcal{M}^a; \mathbb{Z}) \to \cdots$$

that either

$$\mathcal{P}^b(t) = \mathcal{P}^a(t) + t^\lambda \quad \text{or} \quad \mathcal{P}^b(t) = \mathcal{P}^a(t) - t^{\lambda-1},$$

and one can check that the two cases depend on whether the descending sphere (pushed down into $\mathcal{M}^a$) bounds or not. In the former case, the descending disk can be completed to a cycle representing a new generator of λ-dimensional homology, while in the latter it yields a new relation in $(\lambda-1)$-dimensional homology. Since $\mathcal{M}^b(t) = \mathcal{M}^a(t) + t^\lambda$ in either case, we see that

$$(\mathcal{M}^b(t) - \mathcal{P}^b(t)) - (\mathcal{M}^a(t) - \mathcal{P}^a(t))$$

is either 0 or $t^{\lambda-1}(t+1)$. Thus assuming (2.32) for a, we can arrange that

$$\mathcal{M}^b(t) = \mathcal{P}^b(t) + (t+1)\mathcal{Q}^b(t)$$

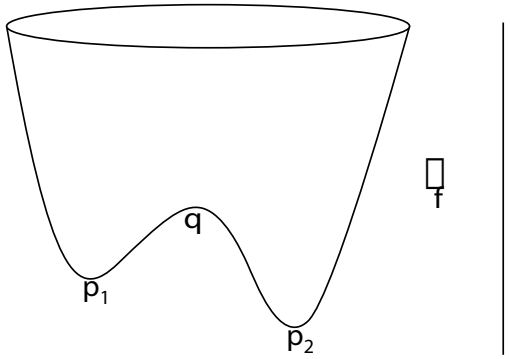

Figure 2.1: If a smooth proper Morse function $f : \mathbb{R}^2 \to [0, \infty)$ has two local minima at p_1 and p_2, it must have an additional critical point (say at q) of Morse index one. This consequence of the Morse inequalities is known as the "mountain pass lemma."

also holds by setting $\mathcal{Q}^b(t) = \mathcal{Q}^a(t)$ or $\mathcal{Q}^b(t) = \mathcal{Q}^a(t) + t^{\lambda-1}$. This finishes the inductive step, and the theorem follows.

If $\mathcal{M}$ has only finitely many critical points of each index, we can let μ_λ denote the number of critical points for f of Morse index λ and let β_λ be the dimension of $H^i(\mathcal{M}; F)$. If we define the *Morse series* for f and *Poincaré series* for $\mathcal{M}$ by

$$\mathcal{M}(t) = \sum_{\lambda=0}^{\infty} \mu_\lambda t^\lambda, \qquad \mathcal{P}(t) = \sum_{\lambda=0}^{\infty} \beta_\lambda t^\lambda,$$

it follows from Theorem 2.9.1 that there is a polynomial

$$\mathcal{Q}(t) = \sum_{\lambda=0}^{\infty} q_\lambda t^\lambda \quad \text{with } q_\lambda \geq 0 \text{ such that} \quad \mathcal{M}(t) - \mathcal{P}(t) = (t+1)\mathcal{Q}(t). \quad (2.33)$$

This is a statement of the Morse inequalities for $f : \mathcal{M} \to [0, \infty)$.

Corollary 2.10.2. (Lacunary Principle) *If $\mu^a_\lambda \neq 0 \Rightarrow \mu^a_{\lambda-1} = 0 = \mu^a_{\lambda+1}$, then $\mu^a_\lambda = \beta^a_\lambda$, for every nonnegative integer λ.*

Indeed, by the weak Morse inequalities, $\beta^a_{\lambda-1} = 0 = \beta^a_{\lambda+1}$, and hence the coefficients of $\lambda - 1$ and $\lambda + 1$ in $(t+1)\mathcal{Q}^a(t)$ must be zero. From this we conclude that $\mathcal{Q}^a(t) \equiv 0$.

Similarly, we can let $a \to \infty$ and obtain the lacunary principle for f on $\mathcal{M}$: If $\mu_\lambda \neq 0 \Rightarrow \mu_{\lambda-1} = 0 = \mu_{\lambda+1}$, then $\mu_\lambda = \beta_\lambda$, for every nonnegative integer λ.

Example 2.10.3. Suppose that $M = \mathbb{S}^n$, the n-dimensional sphere with metric of constant curvature one, where $n \geq 3$, and that p and q are points in M which are not antipodal. Then there is exactly one geodesic from p to 1 of Morse index $k(n-1)$, for each $k \in \mathbb{N} \cup \{0\}$, and hence

$$\mathcal{M}(t) = 1 + t^{n-1} + t^{2(n-1)} + \cdots + t^{k(n-1)} + \cdots .$$

From the "feedback" implicit in the Morse inequalities (2.33), we see that $\mathcal{M}(t) = \mathcal{P}(t)$, so

$$\mathcal{P}(t) = 1 + t^{n-1} + t^{2(n-1)} + \cdots + t^{k(n-1)} + \cdots ,$$

and we conclude that for any field F,

$$H_m(\Omega(S^n, p, q); F) \cong \begin{cases} F, & \text{if } m = k(n-1) \text{ or } k \in \mathbb{N} \cup \{0\}, \\ 0, & \text{otherwise.} \end{cases}$$

Thus if $\langle \cdot, \cdot \rangle$ is *any* Riemannian metric on $\mathbb{S}^n$, then for generic choice of p and q in $\mathbb{S}^n$, there will be infinitely many geodesics from p to q, at least one being of index $k(n-1)$ for each $k \in \mathbb{N} \cup \{0\}$.

We would also like Morse inequalities for the case where $\mathcal{M} = L_1^2(S^1, M)$, where M is a compact Riemannian manifold. If we give M a "generic" Riemannian metric, Theorem 2.7.5 implies that all nonconstant critical points for the action $J : L_1^2(S^1, M) \to \mathbb{R}$ lie on one-dimensional nondegenerate critical submanifolds. We can suppose that the metric is chosen so that only one $O(2)$-orbit of critical points lies at each critical level.

Thus suppose that the interval $[a, b]$ contains a unique critical value c with $J^{-1}(c)$ containing a unique $O(2)$-orbit of geodesics which comprise a nondegenerate critical submanifold N. Then Theorem 2.8.6 gives an isomorphism on cohomology

$$H_k(\mathcal{M}^b, \mathcal{M}^a; \mathbb{Z}) \cong H_k(D(\nu_- N), S(\nu_- N); \mathbb{Z}).$$

It follows from the Thom isomorphism theorem (with twisted coefficients) that if the Morse index of N is λ, then

$$H_k(D(\nu_- N), S(\nu_- N); \mathbb{Z}) \cong H_{k-\lambda}(N; \mathbb{Z} \otimes \theta_-),$$

where θ_- is the orientation bundle of N. If the bundle $\nu_- N$ is orientable, twisted homology reduces to usual homology, and hence

$$H_k(D(\nu_- N), S(\nu_- N); \mathbb{Z}) \cong H_{k-\lambda}(N; \mathbb{Z}).$$

If N is a nondegenerate critical submanifold for J, we let

$$\mathcal{P}_N(t) = \sum_\lambda [\dim H_\lambda(N; \theta_- \otimes F)] t^\lambda.$$

In the case where the normal bundle is orientable, this is the Poincaré polynomial of N. If $F = \mathbb{R}$ and N is a nondegenerate $O(2)$-orbit of geodesics, then

$$\mathcal{P}_N(t) = 2(1 + t) \qquad \text{or} \qquad \mathcal{P}_N(t) = 0,$$

depending upon whether the normal bundle is orientable or not, while if $F = \mathbb{Z}_2$, we always have $\mathcal{P}_N(t) = 2(1+t)$. The *Morse polynomial* of J on $\mathcal{M}^a$ is

$$\mathcal{M}^a(t) = \sum \{ t^{\lambda_N} \mathcal{P}_N(t) : \ N \text{ is a nondegenerate critical submanifold of Morse index } \lambda_N \text{ with } J(N) \leq a \ \}.$$

As in the case of a Morse function, the Morse series is related to the Poincaré polynomial of $\mathcal{M}^a$,

$$\mathcal{P}^a(t) = \sum_{\lambda=0}^{\infty} \dim H_\lambda(\mathcal{M}^a; F) t^\lambda,$$

by the Morse inequalities, which state that there is a polynomial

$$\mathcal{Q}^a(t) = \sum_{i=1}^{\infty} q_\lambda^a t^\lambda \quad \text{with } q_\lambda \geq 0 \text{ such that} \qquad \mathcal{M}^a(t) = \mathcal{P}^a(t) + (t+1)\mathcal{Q}^a(t). \tag{2.34}$$

If we know the homology (or cohomology) of $L_1^2(S^1, M)^a$, the Morse inequalities (2.34) enable us to estimate the number of smooth closed geodesics M must have with action $\leq a$ when M is given a generic metric.

2.11 The Morse-Witten complex

The Morse inequalities do not usually completely determine the integer homology $H_*(\mathcal{M}^b; \mathbb{Z})$. The additional information we would need for an inductive determination of $H_*(\mathcal{M}^b; \mathbb{Z})$ is the boundary map in the long exact sequence

$$\cdots \to H_*(\mathcal{M}^a; \mathbb{Z}) \to H_*(\mathcal{M}^b; \mathbb{Z}) \to H_*(\mathcal{M}^b, \mathcal{M}^a; \mathbb{Z}) \to \cdots .$$

However, implicit in the writings of Thom, Smale, Milnor and others, is a procedure for calculating the boundary map by determining the trajectories between critical points. This results in a chain complex $(C_*(f, \mathcal{X}), \partial)$ that depends on the function f and on a gradient-like vector field $\mathcal{X}$ used to calculate trajectories between critical points, a chain complex which calculates the homology of the manifold $\mathcal{M}^b$. This chain complex is often called the *Morse-Witten chain complex*, because of the fact that Witten gave an important quantum-mechanical interpretation of the boundary operator [83]. We merely sketch the ideas of the construction here; a much more complete treatment can be found in [74] and Chapter 6 of [40].

For the statement of the next theorem, we assume the reader is familiar with the notion of CW complex; background on this topic can be found in Chapter 0 of [35].

Theorem 2.11.1. *If $f : \mathcal{M} \to [0, \infty)$ is a Morse function on a complete Hilbert manifold that satisfies condition C and all of its critical points have finite index, then for each $a \in \mathbb{R}$, $\mathcal{M}^a$ has the homotopy type of a finite CW complex with one cell of dimension λ for each critical point of index λ.*

Sketch of proof: This is a consequence of Theorem 2.9.4 and it is an analog of Theorem 3.5 in [50]. In fact, one could prove that $\mathcal{M}$ itself has the homotopy type of a CW complex, but a rigorous proof would require a limiting procedure as $a \to \infty$. Such a procedure is given in Appendix A of [50].

Thus the homology of $\mathcal{M}$ should be just the cellular homology of the resulting CW complex, as described in books on algebraic topology; see, for example, [35], pages 137-141. This raises the problem of describing the cells and attaching maps, and finding an algorithm for calculating this homology.

Such an algorithm is based upon choice of a gradient-like vector field $\mathcal{X}$ for the Morse function $f : \mathcal{M} \to \mathbb{R}$. Let $\{\phi_t : t \in \mathbb{R}\}$ be the one-parameter group of local diffeomorphisms corresponding to $-\mathcal{X}$.

Definition. The *unstable manifold* $W_p(f, \mathcal{X})$ of a critical point $p \in \mathcal{M}$ consists of the images of all trajectories $t \mapsto \phi_t(q)$ which start at p in the remote past, in other words, such that $\phi_t(q) \to p$ as $t \to -\infty$. Similarly, the *stable manifold* $W_p^*(f, \mathcal{X})$ of a critical point p consists of the images of trajectories $t \mapsto \phi_t(q)$ such that $\phi_t(q) \to p$ as $t \to \infty$.

Lemma 2.11.2. *The unstable and stable manifolds $W_p(f, \mathcal{X})$ and $W_p^*(f, \mathcal{X})$ are in fact submanifolds of $\mathcal{M}$.*

Sketch of proof: Let us consider the case of the unstable manifold $W_p(f, \mathcal{X})$. First one uses the explicit description (2.30) of the gradient-like vector field $\mathcal{X}$ near p to show that for $\epsilon > 0$ sufficiently small,

$$W_p(f, \mathcal{X})^\epsilon = \phi_p^{-1}\left(\{x \in E_- : \|x\| < \epsilon\}\right)$$

is part of the unstable manifold. Then one notes that

$$W_p(f, \mathcal{X}) = \bigcup \{\phi_t(W_p(f, \mathcal{X})^\epsilon) : t \geq 0\}.$$

The properties of smooth flows corresponding to vector fields then show that $W_p(f, \mathcal{X})$ is a smooth manifold diffeomorphic to an open cell. The proof for $W_p^*(f, \mathcal{X})$ is similar, starting with

$$W_p^*(f, \mathcal{X})^\epsilon = \phi_p^{-1}\left(\{x \in E_+ : \|x\| < \epsilon\}\right).$$

It follows from Condition C and the explicit form of $\mathcal{X}$ that any orbit $q \mapsto \phi_t(q)$ converges to some critical point as $t \to \infty$ (although it may come close to several other critical points first).

Lemma 2.11.3. *We can adjust the gradient-like vector field for f so that if λ_p is the index of p and λ_q is the index of q,*

1. *$\lambda_p \leq \lambda_q \Rightarrow W_p(\mathcal{X}) \cap W_q^*(\mathcal{X})$ is empty, while*
2. *$\lambda_p > \lambda_q \Rightarrow W_p(\mathcal{X}) \cap W_q^*(\mathcal{X})$ is a submanifold of dimension $\lambda_p - \lambda_q$, when $\mathcal{X}$ is sufficiently smooth.*

Sketch of proof: If $W_p(\mathcal{X}) \cap W_q^*(\mathcal{X})$ is nonempty, then $f(p) > f(q)$ and we can choose a regular value c for f such that $f(p) > c > f(q)$. Then $\mathcal{N} = f^{-1}(c)$ is a codimension one submanifold of $\mathcal{M}$, and

$$\dim\,(\mathcal{N} \cap W_p(\mathcal{X})) = \lambda_p - 1, \qquad \text{codim}\,(\mathcal{N} \cap W_q^*(\mathcal{X})) = \lambda_q.$$

Moreover, if we choose c sufficiently close to the Morse nondegenerate critical point p, we see that $S = \mathcal{N} \cap W_p(\mathcal{X})$ is a $(\lambda_p - 1)$-dimensional sphere and lies in an open tubular neighborhood U of $\mathcal{N}$ with diffeomorphism $\psi : U \to S \times V$, where V is an open ball in a Banach space, with $\psi(S) = S \times \{0\}$. Let

$$N = U \cap W_q^*(\mathcal{X}), \quad \text{and consider} \quad g = \pi \circ \psi : N \to V,$$

where π is the projection on the second factor. We can check that g is a Fredholm map with Fredholm index $\lambda_p - 1 - \lambda_q$, and

$$\psi^{-1}(S \times \{x\}) \cap N = \{s \in N : g(s) = x\}.$$

Assuming that $\mathcal{X}$ (and hence N) is sufficiently smooth, we can use the Sard-Smale Theorem 2.6.2 to choose a regular value x for g. Finally, we can choose x as close as we want to 0, and construct an isotopy from a neighborhood of S to itself which carries $S \times \{0\}$ to $S \times \{x\}$. Since the conditions defining pseudogradient are open, we can replace $\mathcal{X}$ by a new gradient-like vector field so that

$$\mathcal{N} \cap W_p(\mathcal{X}) \quad \text{and} \quad \mathcal{N} \cap W_q^*(\mathcal{X})$$

have transverse intersection. If $\lambda_p \leq \lambda_q$, the dimension of $\mathcal{N} \cap W_p(\mathcal{X})$ is less than the codimension of $\mathcal{N} \cap W_q^*(\mathcal{X})$, so $W_p(\mathcal{X}) \cap W_q^*(\mathcal{X})$ is empty. If $\lambda_p > \lambda_q$, then $\mathcal{N} \cap W_p(\mathcal{X}) \cap W_q^*(\mathcal{X})$ is a submanifold of dimension $\lambda_p - \lambda_q - 1$.

In particular, if $\lambda_p = \lambda_q + 1$, then $\mathcal{N} \cap W_p(\mathcal{X}) \cap W_q^*(\mathcal{X})$ consists of a finite number of points and $W_p(\mathcal{X}) \cap W_q^*(\mathcal{X})$ consists of a finite collection of smooth curves from p to q.

We can now orient the unstable manifolds and define the *Morse-Witten complex* of the nonnegative Morse function f. We let $C_{\lambda+1}(f, \mathcal{X})$ be the free $\mathbb{Z}$-module generated by the critical points $p_{\lambda+1,1}, p_{\lambda+1,2} \ldots$ of f of index k and let ∂ be the $\mathbb{Z}$-module homomorphism

$$\partial : C_{\lambda+1}(f, \mathcal{X}) \longrightarrow C_\lambda(f, \mathcal{X})$$

defined by

$$\partial(p_{\lambda+1,j}) = \sum_q a_{jq} p_{\lambda,q}, \tag{2.35}$$

where $a_{jq} \in \mathbb{Z}$ is the oriented number of trajectories from $p_{\lambda+1,j}$ to $p_{\lambda,q}$. The sign of a trajectory $\gamma : (-\infty, \infty) \to M$ from p to q is determined as follows: First one orients each unstable manifold for f. Then one constructs a trivial vector bundle E along γ which is transverse to $TW_q^*(\mathcal{X})$ and of complementary dimension such that the fiber $E_{\gamma(t)}$ approaches a hyperplane in the negative

eigenspace $T_p\mathcal{M}_-$ for $d^2f(p)$ as $t \to -\infty$, and the positive eigenspace for $d^2f(q)$ as $t \to \infty$. Using this bundle we translate the oriented negative eigenspace $T_qW_q(\mathcal{X})$ for $d^2f(q)$ back to p. If $T_q\mathcal{M}_-$ denotes the resulting hyperplane in $T_pW_p(\mathcal{X})$, then

$$T_pW_p(\mathcal{X}) = T_q\mathcal{M}_- \oplus T,$$

where T is the oriented line in the direction of the trajectory leaving p. We assign the positive sign to the trajectory γ if the orientation of $T_pW_p(\mathcal{X})$ agrees with the direct sum orientation, the minus sign otherwise.

The next theorem says that $(C_*(f,\mathcal{X}),\partial)$ is a chain complex:

Theorem 2.11.4. *Suppose that $f : \mathcal{M} \to \mathbb{R}$ is a nonnegative Morse function on a complete Hilbert manifold that satisfies condition C and all of the critical points of f have finite index. Then $\partial\circ\partial = 0$ and the homology (with integer coefficients) of the Morse-Witten complex $(C_*(f,\mathcal{X}),\partial)$ is just the usual homology $H_*(\mathcal{M};\mathbb{Z})$.*

Sketch of proof: It suffices to prove this for the restriction of f to $\mathcal{M}^a$ which has only finitely many critical points, and we henceforth use the notation $\mathcal{M}$ for $\mathcal{M}^a$. After generic choice of $\mathcal{X}$, follows from Lemma 2.10.3 that unstable critical points of index λ intersect only stable manifolds of critical points of index $< \lambda$. Thus if we let $\mathcal{M}^{(\lambda)}$ denote the union of the unstable manifolds in $\mathcal{M}$ of index $\leq \lambda$, we obtain an increasing filtration

$$\cdots \subset \mathcal{M}^{(\lambda-1)} \subset \mathcal{M}^{(\lambda)} \subset \mathcal{M}^{(\lambda+1)} \subset \cdots$$

of $\mathcal{M}$ by closed subsets. Since the unstable manifolds of index λ are in one-to-one correspondence with generators of

$$H_*(\mathcal{M}^{(\lambda)}, \mathcal{M}^{(\lambda-1)};\mathbb{Z}),$$

we can set $C_*(f,\mathcal{X})$ equal to this homology group. If we let ∂' denote the boundary homomorphism in the exact sequence of the triple

$$(\mathcal{M}^{(\lambda)}, \mathcal{M}^{(\lambda-1)}, \mathcal{M}^{(\lambda-2)}),$$

one can use the usual argument from the theory of cellular homology to show that $\partial'\circ\partial' = 0$ and the cohomology of the Morse complex is the standard homology of $\mathcal{M}$ (as in the proof of Theorem 2.35 in [35]). Thus it is relatively easy to see that we do indeed get a chain complex, and we have reduced the proof to the verification that ∂' is the same as the homomorphism ∂ defined by (3.44).

For this, one can follow arguments used by Milnor ([51]) to prove the h-cobordism theorem of Smale. Note we can adjust the values of the Morse function f while simultaneously multiplying the gradient-like vector field $\mathcal{X}$ by a positive function $\eta : \mathcal{M}^a \to (0,\infty)$. In this way, we can replace the Morse function f by "self-indexing" function f^* which satisfies the conditions $f^\star \geq -1$ and $f^\star(p) = \lambda$, whenever p is a critical point of index λ, by following the proof

of Theorem 4.8 in [51], without changing critical points or the boundary maps in the Morse-Witten complex. After this is done, $\mathcal{M}^{(\lambda)}$ is a strong deformation retract of

$$\mathcal{M}^{\lambda+1/2} = \left\{ q \in \mathcal{M} : f^*(q) \leq \lambda + \frac{1}{2} \right\}.$$

To complete the sketch of the proof, we need only verify the following lemma:

Lemma 2.11.5. *The Witten boundary ∂ agrees with the boundary ∂' defined by the CW decomposition.*

The proof is a straightforward modification of the proof of Theorem 7.4 in [51], to which we refer for details. Here we give only a very superficial description of how the argument goes. The key idea is that if

$$\mathcal{N} = (f^*)^{-1}(\lambda + 1/2),$$

then the CW boundary ∂ can be regarded as a composition

$$H_{\lambda+1}(\mathcal{M}^{\lambda+3/2}, \mathcal{M}^{\lambda+1/2}; \mathbb{Z}) \longrightarrow H_\lambda(\mathcal{N}; \mathbb{Z}) \longrightarrow H_\lambda(\mathcal{M}^{\lambda+1/2}, \mathcal{M}^{\lambda-1/2}; \mathbb{Z})$$

where the first map takes the homology class corresponding to the $(\lambda+1)$-handle for a critical point $p_{\lambda+1,j}$ to the homology class of its boundary sphere $S_{\lambda+1,j} \subseteq \mathcal{N}$, and the second map is induced by inclusion. On the cochain level, we have a corresponding factorization of the coboundary

$$H^\lambda(\mathcal{M}^{\lambda+1/2}, \mathcal{M}^{\lambda-1/2}; \mathbb{R}) \longrightarrow H^\lambda(\mathcal{N}; \mathbb{R}) \longrightarrow H^{\lambda+1}(\mathcal{M}^{\lambda+3/2}, \mathcal{M}^{\lambda+1/2}; \mathbb{R}),$$

where we use real coefficients so that we can represent cycles by differential forms. Corresponding to a critical point $p_{\lambda,q}$ we can define a "Thom form" $\theta_{\lambda,q}$ (as described in [10]) which represents a cohomology class

$$[\theta_{\lambda,q}] \in H^\lambda(\mathcal{M}^{\lambda+1/2}, \mathcal{M}^{\lambda-1/2}; \mathbb{R})$$

such that

$$\int_{W_{p_{\lambda,r}}} \theta_{\lambda,q} = \delta_{rq} = \begin{cases} 1 & \text{if } r = q, \\ 0 & \text{if } r \neq q. \end{cases}$$

We can think of $\theta_{\lambda,q}$ as Poincarë dual to the class represented by the inifinite-dimensional stable manifold for $p_{\lambda,q}$. One now checks that

$$a_{jq} = \int_{S_{\lambda+1,j}} \theta_{\lambda,q} \in \mathbb{Z}$$

is both the integer appearing in the Witten boundary (3.44) and the integer appearing in the formula for the CW boundary.

Corollary 2.11.6. (Lacunary Principle) *If $\mu^a_\lambda \neq 0 \Rightarrow \mu^a_{\lambda-1} = 0 = \mu^a_{\lambda+1}$, then the boundary ∂ in the Morse-Witten chain complex is zero.*

The proof is immediate.

Note that applying Corollary 2.11.6 to Example 2.10.3. yields the integer homology of the loop space of $\mathbb{S}^n$ when $n \geq 3$:

$$H_m(\Omega(S^n, p, q); \mathbb{Z}) \cong \begin{cases} \mathbb{Z}, & \text{if } m = k(n-1) \text{ or } k \in \mathbb{N} \cup \{0\}, \\ 0, & \text{otherwise.} \end{cases}$$

In spite of the fact that the Morse-Witten complex gives stronger results, the Morse inequalities are often quite useful, since in many cases it is difficult to calculate the boundary operator ∂.

Chapter 3

Harmonic and minimal surfaces

3.1 The energy of a smooth map

Suppose now that Σ is a compact smooth Riemannian manifold of dimension m. In terms of local coordinates $(u^1, \ldots, u^m)$ on Σ we can write the Riemannian metric and area or volume element on Σ as

$$\sum_{a,b=1}^{m} \eta_{ab} du^a \otimes du^b \quad \text{and} \quad dA = \sqrt{\eta} du^1 \cdots du^m,$$

where η denotes the determinant of the matrix (η_{ab}). If M is a second Riemannian manifold of dimension n isometrically imbedded in Euclidean space $\mathbb{R}^N$ and $f : \Sigma \to M \subseteq \mathbb{R}^N$ is a smooth map, the *energy density* of f at a given point is given in terms of local coordinates by the formula

$$e(f) = \frac{1}{2}|df|^2, \quad \text{where} \quad |df|^2 = \sum_{a,b=1}^{m} \eta^{ab} \frac{\partial f}{\partial u^a} \cdot \frac{\partial f}{\partial u^b},$$

where (η^{ab}) denotes the matrix inverse to (η_{ab}) and the dot denotes the Euclidean dot product in $\mathbb{R}^N$. The energy of a smooth map $f : \Sigma \to M \subseteq \mathbb{R}^N$ is given by the *Dirichlet integral*

$$E(f) = \int_\Sigma e(f) dA = \frac{1}{2} \int_\Sigma \sum_{a,b=1}^{m} \eta^{ab} \frac{\partial f}{\partial u^a} \cdot \frac{\partial f}{\partial u^b} \sqrt{\eta} du^1 \cdots du^m, \tag{3.1}$$

the integrand being independent of choice of local coordinates. Note that if Σ is one-dimensional, the energy reduces to the action J that we studied in the previous chapter.

The energy defines a smooth map

$$E : C^2(\Sigma, M) \longrightarrow \mathbb{R}.$$

To find the critical points of E, we consider a variation of a given map f which has its support within a coordinate chart $(U, (u^1, \ldots, u^m))$ on Σ. Such a variation is simply a smooth family of maps $t \mapsto f(t)$ in $C^2(\Sigma, M)$ such that $f(0) = f$ and $f(t)(p) = f(p)$ for all t, when $p \in \Sigma - U$. In terms of the local coordinates, we set

$$\alpha(u^1, \ldots, u^m, t) = f(t)(u^1, \ldots, u^m),$$

and a straightforward calculation shows that

$$\begin{aligned}\frac{d}{dt}(E(f(t))\Big|_{t=0} &= \int_\Sigma \left[\sum_{a,b=1}^m \sqrt{\eta}\eta^{ab}\frac{\partial^2\alpha}{\partial t \partial u^a}\cdot\frac{\partial\alpha}{\partial u^b}\right] du^1\cdots du^m\Bigg|_{t=0} \\ &= -\int_\Sigma \frac{\partial\alpha}{\partial t}\cdot\left[\sum_{a,b=1}^m \frac{\partial}{\partial u^a}\left(\sqrt{\eta}\eta^{ab}\frac{\partial\alpha}{\partial u^b}\right)\right] du^1\cdots du^m\Bigg|_{t=0}.\end{aligned}$$

We can evaluate at $t = 0$, setting

$$V(u^1, \ldots u^m) = \frac{\partial\alpha}{\partial t}(u^1, \ldots u^m, 0),$$

to obtain the *first variation formula*,

$$dE(f)(V) = -\int_\Sigma V\cdot\left[\frac{1}{\sqrt{\eta}}\sum_{a,b=1}^m \frac{\partial}{\partial u^a}\left(\sqrt{\eta}\eta^{ab}\frac{\partial f}{\partial u^b}\right)\right] dA. \tag{3.2}$$

If f is a critical point for the energy E, then $dE(f)(V) = 0$ for all such variations V, and f must satisfy the partial differential equation

$$\left[\sum_{a,b=1}^m \frac{1}{\sqrt{\eta}}\frac{\partial}{\partial u^a}\left(\sqrt{\eta}\eta^{ab}\frac{\partial f}{\partial u^b}\right)\right]^\top = 0, \tag{3.3}$$

where $(\cdot)^\top$ denotes projection into the tangent space to M. Maps $f \in C^2(\Sigma, M)$ which satisfy equation (3.3) are called *harmonic maps*. Just like the equation for geodesics, the equation for harmonic maps is nonlinear because of the projection into the tangent space. In fact, we can rewrite (3.3) in terms of the Levi-Civita connection D on M as

$$\frac{1}{\sqrt{\eta}}\sum_{a,b=1}^m \frac{D}{\partial u^a}\left(\sqrt{\eta}\eta^{ab}\frac{\partial f}{\partial u^b}\right) = 0,$$

or alternatively, in terms of local coordinates $(x^, \ldots x^n)$ on M, we can write the equation of harmonic maps as

$$\frac{1}{\sqrt{\eta}}\sum_{a,b=1}^m \frac{\partial}{\partial u^a}\left(\sqrt{\eta}\eta^{ab}\frac{\partial x^i}{\partial u^b}\right) + \sum \Gamma^i_{jk}\eta^{ab}\frac{\partial x^j}{\partial u^a}\frac{\partial x^k}{\partial u^b} = 0,$$

where the Γ^i_{jk}'s are the Christoffel symbols.

Harmonic maps between Riemannian manifolds were introduced by Eells and Sampson [18] in 1964, who used the heat flow approach to establish the following theorem:

Theorem 3.1.1 (Eells and Sampson). *If M is a compact connected Riemannian manifold with nonpositive sectional curvatures and Σ is a compact Riemannian manifold, then every component of $C^2(\Sigma, M)$ contains a minimal energy representative, which is a smooth harmonic map.*

A very nice proof of this theorem using heat flow can be found at the beginning of Chapter 5 of [45]. A fews years after Eells and Sampson, Hartman [34] established a uniqueness result:

Theorem 3.1.2 (Hartman). *If Σ and M are compact connected Riemannian manifolds and M has negative sectional curvatures, then any component of $C^2(\Sigma, M)$ contains at most one harmonic map.*

One should note, however, that if M has nonpositive sectional curvature, the Hadamard-Cartan Theorem asserts that the exponential map $\exp_p : T_pM \to M$ is a smooth covering at any point $p \in M$, so M has Euclidean space as its universal cover, and all of its higher homotopy groups are zero. In this case, M is a $K(\pi, 1)$ and its topology is relatively simple.

The Eells-Sampson Theorem does not hold without the curvature assumption. Indeed, it fails spectacularly when the dimension of Σ is ≥ 3, and attempts to minimize energy within a given homotopy class of maps can lead to elements of $L^2_1(\Sigma, M)$ which have quite bad singularities, as discussed for example in [45].

Later we will prove the above theorems in the case where Σ has dimension two, as part of a general theory that will give existence results even when we make no assumptions on the curvature of the range M. In the case where Σ has dimension two, the energy is conformally invariant, and the theory of harmonic maps simplifies considerably due to the existence of isothermal coordinates on Σ. A nice proof of existence of isothermal charts, based upon Hodge theory, can be found in Chapter 5, §10 of [79]. According to this existence theorem, any two-dimensional Riemannian manifold possesses an atlas $\{(U_\alpha, (u_\alpha, v_\alpha)) : \alpha \in A\}$ consisting of isothermal charts, charts such that the Riemannian metric on Σ takes the form

$$ds^2 = \lambda_\alpha^2 (du_\alpha \otimes du_\alpha + dv_\alpha \otimes dv_\alpha),$$

where each λ_α is a positive smooth real-valued function on U_α. If $\{\psi_\alpha : \alpha \in A\}$ is a partition of unity subordinate to this atlas, then the energy of a C^2 map $f : \Sigma \to M$ is given by the formula

$$E(f) = \frac{1}{2} \sum_\alpha \int_\Sigma \psi_\alpha \left[\left\langle \frac{\partial f}{\partial u_\alpha}, \frac{\partial f}{\partial u_\alpha} \right\rangle + \left\langle \frac{\partial f}{\partial v_\alpha}, \frac{\partial f}{\partial v_\alpha} \right\rangle \right] du_\alpha dv_\alpha, \tag{3.4}$$

where $\langle \cdot, \cdot \rangle$ is the Riemannian metric on M induced by the dot product on $\mathbb{R}^N$,

and harmonic maps are simply the maps

$$f : \Sigma \longrightarrow M \qquad \text{such that} \qquad \left(\frac{\partial^2 f}{\partial u_\alpha^2} + \frac{\partial^2 f}{\partial v_\alpha^2}\right)^T = 0. \tag{3.5}$$

If one takes two coherently oriented isothermal charts $(U_\alpha, (u_\alpha, v_\alpha))$ and $(U_\beta, (u_\beta, v_\beta))$, one finds by a short calculation that on the intersection $U_\alpha \cap U_\beta$,

$$\frac{\partial u_\alpha}{\partial u_\beta} = \frac{\partial v_\alpha}{\partial v_\beta}, \qquad \frac{\partial u_\alpha}{\partial v_\beta} = -\frac{\partial v_\alpha}{\partial u_\beta}.$$

These are just the Cauchy-Riemann equations which state that $w_\alpha = u_\alpha + iv_\alpha$ is a holomorphic function of $w_\beta = u_\beta + iv_\beta$. Thus if Σ is an oriented surface, the atlas $\{(U_\alpha, z_\alpha) : \alpha \in A\}$ of positively oriented isothermal charts makes Σ into a one-dimensional complex manifold, otherwise known as a *Riemann surface*. In this way, Riemannian metrics on Σ are divided into equivalence classes, depending upon which Riemann surface structure is defined by the isothermal charts.

It is often convenient to choose a canonical metric on Σ within the conformal equivalence class determined by a Riemann surface structure. The uniformization theorem from Riemann surface theory states that any compact Riemann surface has as its universal cover one of three simply connected Riemann surfaces,

$$S^2 = \mathbb{C} \cup \{\infty\}, \qquad \mathbb{C} \qquad \text{or} \qquad D = \{z \in \mathbb{C} : |z| < 1\}.$$

The remarkable fact is that each of these model spaces possesses a Riemannian metric of constant Gaussian curvature compatible with its conformal structure, and the deck transformations of the universal cover of the compact Riemann surface Σ are isometries with respect to this Riemannian metric. If Σ is closed, we can normalize this metric by assuming that it has total area one, and it is then unique up to diffeomorphism. It follows from the Gauss-Bonnet formula that the sign of the curvature is positive, zero or negative when Σ has genus zero, one, or at least two, respectively.

In terms of the complex partial differential operators

$$\frac{\partial}{\partial w} = \frac{1}{2}\left(\frac{\partial}{\partial u} - \sqrt{-1}\frac{\partial}{\partial v}\right), \qquad \frac{\partial}{\partial \bar{w}} = \frac{1}{2}\left(\frac{\partial}{\partial u} + \sqrt{-1}\frac{\partial}{\partial v}\right)$$

we can write (3.3) as

$$\frac{D}{\partial \bar{w}}\left(\frac{\partial f}{\partial w}\right) = \left(\frac{\partial^2 f}{\partial w \partial \bar{w}}\right)^T = 0. \tag{3.6}$$

Note that we can regard

$$\frac{\partial f}{\partial w} \quad \text{as a section of} \quad \mathbf{E} = f^* TM \otimes \mathbb{C},$$

a complex vector bundle over the Riemann surface Σ. The vector bundle $\mathbf{E}$ has a Hermitian metric defined by

$$v, w \in T_pM \otimes \mathbb{C} \quad \mapsto \quad \langle v, \bar{w} \rangle,$$

where $\bar{w}$ is the conjugate of w, and the Levi-Civita connection D on f^*TM extends complex linearly to a metric connection on $\mathbf{E}$. The following theorem states that this connection gives $\mathbf{E}$ a canonical holomorphic structure, thereby making possible a remarkable relationship between harmonic surfaces and complex analysis:

Theorem 3.1.3 (Koszul and Malgrange). *If $\mathbf{E}$ is a complex vector bundle with Hermitian metric over a Riemann surface Σ and D is a metric connection on $\mathbf{E}$, then there is a unique holomorphic structure on $\mathbf{E}$ such that if σ is a section of $\mathbf{E}$,*

$$\sigma \text{ is holomorphic} \qquad \Leftrightarrow \qquad \frac{D\sigma}{\partial \bar{w}} = 0,$$

whenever w is a complex coordinate on Σ.

We will not prove this theorem here. An extension of this theorem is proven as Theorem 5.1 in Atiyah, Hitchin and Singer [5] using the Newlander-Nirenberg theorem on integrability of almost complex structures. Another proof is given in Donaldson and Kronheimer [14], Theorem 2.1.53.

Thus we see that the equation for harmonic maps from oriented surfaces can be expressed quite simply in terms of Riemann surface theory: Equation (3.6) asserts that a map $f : \Sigma \to M$ is harmonic if and only if in terms of any complex coordinate w on Σ, the section

$$\frac{\partial f}{\partial w} \quad \text{of} \quad \mathbf{E} = f^*TM \otimes \mathbb{C}$$

is holomorphic with respect to the holomorphic structure on $\mathbf{E}$ which is provided by the Koszul-Malgrange Theorem.

The locally defined holomorphic section

$$\frac{\partial f}{\partial w} \quad \text{may have isolated singularities, but} \quad \left[\frac{\partial f}{\partial w}\right] : \Sigma \to \mathbb{P}(\mathbf{E})$$

extends to all of Σ, where $\mathbb{P}(\mathbf{E})$ denotes the bundle of projective spaces of fibers of $\mathbf{E} = f^*TM \otimes \mathbb{C}$. Under change of complex coordinate

$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial w} = (\text{complex-valued function})\frac{\partial f}{\partial z},$$

and hence the various locally defined complex derivatives of f determine a complex line subbundle $\mathbf{L}$ of $\mathbf{E}$, which is holomorphic by the Koszul-Malgrange Theorem. The line bundle $\mathbf{L}$ is isomorphic to the tangent bundle to Σ if and only if f has no branch points, in accordance with the following definition:

Definition. A point $p \in \Sigma$ is a *branch point* for the harmonic map $f : \Sigma \to M$ if $(\partial f/\partial z)(p) = 0$, where z is any complex coordinate near p.

If p is a branch point for f, and z is a local complex coordinate defined on a small open neighborhood U of p with $z(p) = 0$, then we can write $(\partial f/\partial z) = z^\nu g$ for some positive integer ν, where g is a section of $\mathbf{L}$ over U such that $g(p) \neq 0$. We call ν the *order* or *multiplicity* of the branch point. If we let $w = z^{\nu+1}$, then $dw = (\nu+1)z^\nu dz$ and we can define a section

$$\frac{\partial}{\partial w} \quad \text{of } \mathbf{L} \text{ by} \quad \frac{\partial}{\partial w} = \frac{1}{(\nu+1)z^\nu}\frac{\partial}{\partial z}.$$

Thus we see that the restriction of L to U is obtained from the holomorphic tangent bundle $T\Sigma|(U - \{p\})$ by pasting it to a trivial bundle $U \times \mathbb{C}$ over U by means of the transition function

$$g_{0p} = \frac{1}{(\nu+1)z^\nu} : U - \{p\}) \longrightarrow \mathbb{C} - \{0\}).$$

From the trivial bundle over $\Sigma - \{p\})$ and this transition function one can construct the *point bundle* ζ_p^ν over Σ which has first Chern number $c_1(\zeta_p^\nu)([\Sigma]) = \nu$. If f has a single branch point p of branching order ν then $\mathbf{L} = T\Sigma \otimes \zeta_p^\nu$.

In general, the *divisor* of the harmonic map f is the finite sum

$$(f) = \nu_1 p_1 + \cdots + \nu_n p_m,$$

where $p_1, \ldots, p_m$ are the branch points of f and $\nu_1, \ldots, \nu_m$ are their branching orders. Then the above discussion shows that

$$\mathbf{L} = T\Sigma \otimes \zeta_{p_1}^{\nu_1} \otimes \cdots \otimes \zeta_{p_m}^{\nu_m}.$$

If ν denotes the total branching order of f, the total number of branch points of f, counted with multiplicity, then the first Chern number $\langle c_1(\mathbf{L}), [\Sigma]\rangle$ of the line bundle $\mathbf{L}$ (also known as the degree of $\mathbf{L}$ in many books on Riemann surfaces) is determined by the formula

$$\langle c_1(\mathbf{L}), [\Sigma]\rangle = 2 - 2g + \nu \tag{3.7}$$

where g is the genus of Σ.

Suppose, for example, that $h : \Sigma_2 \to M$ is a smooth harmonic map without branch points and that $g : \Sigma_1 \to \Sigma_2$ is a nontrivial holomorphic branched cover. (Thus, if $\Sigma_2 = S^2$, the Riemann sphere, we can regard g as a meromorphic function on Σ_1.) Then the composition $f = h \circ g : \Sigma_1 \to M$ is also harmonic; it is called a *branched cover* of the harmonic map $h : \Sigma_2 \to M$. We say that the harmonic map f is *prime* if it is not a nontrivial branched cover of another harmonic map. Note that if $f : \Sigma \to M$ is a k-fold branched cover of a harmonic sphere $h : S^2 \to M$ which is free of branch points, then the line bundle $\mathbf{L}$ for f satisfies

$$\langle c_1(\mathbf{L}), [\Sigma]\rangle = 2k.$$

3.2 Minimal surfaces

The theory of minimal surfaces in a Riemannian manifold M is concerned with critical points of the *area function*

$$A : \mathrm{Map}(\Sigma, M) \longrightarrow \mathbb{R} \qquad \text{defined by} \qquad A(f) = \int_\Sigma \left| \frac{\partial f}{\partial u} \wedge \frac{\partial f}{\partial v} \right| du dv,$$

which we will see is closely related to critical points for the energy function E defined by (3.4). In contrast to E, the integrand for A is independent of the choice of parametrization, and in particular does not depend upon the choice of a metric or conformal structure.

Proposition 3.2.1. *If Σ is given a conformal structure ω and $f \in Map(\Sigma, M)$, then $E(f) \geq A(f)$, with equality holding if and only if f satisfies the conditions*

$$\left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial u} \right\rangle = \left\langle \frac{\partial f}{\partial v}, \frac{\partial f}{\partial v} \right\rangle, \qquad \left\langle \frac{\partial f}{\partial u}, \frac{\partial f}{\partial v} \right\rangle = 0, \tag{3.8}$$

when $w = u + iv$ is any choice of complex chart, or equivalently,

$$\left\langle \frac{\partial f}{\partial w}, \frac{\partial f}{\partial w} \right\rangle = 0, \tag{3.9}$$

and $\langle \cdot, \cdot \rangle$ denotes the Riemannian metric on the ambient manifold M.

Proof: We utilize two well-known algebraic identities for vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^N$:

$$|\mathbf{v} \wedge \mathbf{w}|^2 + |\mathbf{v} \cdot \mathbf{w}|^2 = |\mathbf{v}|^2 |\mathbf{w}|^2, \qquad |\mathbf{v}||\mathbf{w}| \leq \frac{1}{2} \left(|\mathbf{v}|^2 + |\mathbf{w}|^2 \right),$$

with equality holding only if $|\mathbf{v}| = |\mathbf{w}|$. Using these two facts, we find that

$$\left| \frac{\partial f}{\partial u} \wedge \frac{\partial f}{\partial v} \right| \leq \left| \frac{\partial f}{\partial u} \right| \left| \frac{\partial f}{\partial u} \right| \leq \frac{1}{2} \left(\left| \frac{\partial f}{\partial u} \right|^2 + \left| \frac{\partial f}{\partial v} \right|^2 \right), \tag{3.10}$$

with equality holding if and only if (3.8) holds. This proves the proposition.

Definition. We say that a map $f : \Sigma \to M$ is *weakly conformal* with respect to a Riemann surface structure ω on Σ if it satisfies (3.9) when $w = u + iv$ is a complex coordinate for ω.

Weak conformality of an harmonic map $f : \Sigma \to M$ is expressed by the vanishing of the *Hopf differential*, which is the quadratic differential

$$\Omega(f) = \left\langle \frac{\partial f}{\partial w}, \frac{\partial f}{\partial w} \right\rangle dw \otimes dw. \tag{3.11}$$

It follows immediately from (3.6) that

$$\frac{\partial}{\partial \bar{w}} \left\langle \frac{\partial f}{\partial w}, \frac{\partial f}{\partial w} \right\rangle = 2 \left\langle \frac{D}{\partial \bar{w}} \frac{\partial f}{\partial w}, \frac{\partial f}{\partial w} \right\rangle = 0,$$

so the Hopf differential of a harmonic map is indeed a holomorphic quadratic differential, as studied in Riemann surface theory.

Theorem 3.2.2. *A weakly conformal harmonic map $f : \Sigma \to M$ is a critical point for the area function A.*

Proof: If f is weakly conformal and harmonic, and $t \mapsto f(t)$ is a smooth family of maps with $f(0) = f$, then

$$E(f(t)) \geq A(f(t)) \quad \text{and} \quad E(f(0)) = A(f(0))$$
$$\Rightarrow \qquad \frac{d}{dt} E(f(t))\Big|_{t=0} = \frac{d}{dt} A(f(t))\Big|_{t=0},$$

so f is critical for the area function.

Theorem 3.2.2 provides motivation for the following definition of parametrized minimal surface:

Definition. A *parametrized minimal surface* is a weakly conformal harmonic map $f : \Sigma \to M$.

Note that if f is an immersion, we could give Σ the Riemannian metric which makes f an isometric immersion. This gives Σ a Riemann surface structure which makes f is weakly conformal, and with this conformal structure, f is harmonic if and only if it is minimal.

More generally, if $f : \Sigma \to M$ is a weakly conformal map, then the complex line bundle $\mathbf{L}$ we described in the preceding section is "isotropic," that is, $\langle \mathbf{L}, \mathbf{L} \rangle = 0$. This constrains the possible singularities of f; the differential df_p at any point $p \in \Sigma$ has rank two if f is an immersion in a neighborhood of p, or rank zero if p is a branch point, but never has rank one.

We would like to consider minimal surfaces as critical points for a variational problem which involves the Dirichlet integral and the conformal structure, rather than the area function. In order to do this, we must allow the conformal structure on Σ to vary, as well as the map $f \in C^2(\Sigma, M)$. In this and the following section, we suppose that Σ is a compact oriented connected surface without boundary.

The simplest case is that where $\Sigma = S^2$, which has a unique conformal structure by the uniformization theorem. We can regard $\Sigma = S^2$ as $\mathbb{C} \cup \{\infty\}$ and take the atlas defined by the standard coordinate z on $\mathbb{C}$ and the coordinate $w = 1/z$ on $S^2 - \{0\}$. Then

$$\frac{\partial f}{\partial z} = \frac{\partial f}{\partial w} \frac{\partial w}{\partial z} = -\frac{1}{z^2} \frac{\partial f}{\partial w},$$

and

$$\frac{\partial f}{\partial w} \text{ bounded near } \infty \quad \Rightarrow \quad \frac{\partial f}{\partial z} \to 0 \text{ like } 1/z^2 \text{ as } z \to \infty. \tag{3.12}$$

By the Koszul-Malgrange Theorem 3.1.3, we can regard $\partial f / \partial z$ as a holomorphic section on $S^2 - \{\infty\}$ with a removeable singularity at ∞, and by the removeable

singularity theorem from complex analysis, $\partial f/\partial z$ extends to a holomorphic vector field on S^2. Thus when calculating energy on the Riemann sphere, we are fully justified in using a single coordinate $z = x + iy$ on $\mathbb{C} = S^2 - \{\infty\}$, and the energy can be expressed by the improper integral

$$E(f) = \frac{1}{2}\int_{\mathbb{C}} \left[\left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial x} \right\rangle + \left\langle \frac{\partial f}{\partial y}, \frac{\partial f}{\partial y} \right\rangle \right] dxdy.$$

Conversely, if $f : \mathbb{C} \to M$ is a harmonic map of finite energy, in other words, such that the above integral is finite, a removeable singularity theorem of Sacks and Uhlenbeck (Theorem 3.6 in [68]) will show that f extends to a harmonic map from S^2 into M.

Proposition 3.2.3. *A harmonic map $f : S^2 \to M$ is automatically weakly conformal, and hence a parametrized minimal surface.*

To prove this, we let $z = x + iy$ be the standard complex coordinate on $\mathbb{C} = S^2 - \{\infty\}$. Then it follows from (3.6) that

$$\frac{\partial}{\partial \bar{z}} \left\langle \frac{\partial f}{\partial z}, \frac{\partial f}{\partial z} \right\rangle = 2 \left\langle \frac{D}{\partial \bar{z}} \left(\frac{\partial f}{\partial z} \right), \frac{\partial f}{\partial z} \right\rangle = 0, \quad \text{so} \quad \left\langle \frac{\partial f}{\partial z}, \frac{\partial f}{\partial z} \right\rangle$$

is a holomorphic function which extends to a C^2 function on S^2 by (3.12). This function must be constant by the maximum modulus principle, and since it vanishes at ∞, the constant must be zero, establishing the assertion. Note that the argument simply shows that the Hopf differential (3.11) is zero, so one could prove the Proposition by simply citing the well-known fact from Riemann surface theory that S^2 has no nonzero holomorphic quadratic differentials.

We next consider the case where $\Sigma = T^2 = S^1 \times S^1$, a torus. We imagine that we have fixed a basis (α, β) for $H_1(T^2; \mathbb{Z})$, representing the two S^1 factors. The covering transformations for the universal cover $\pi : \mathbb{C} \to T^2$ are invertible holomorphic maps from $\mathbb{C}$ to itself, and it follows from the maximum modulus principle they must be translations. After performing a possible rotation of $\mathbb{C}$ and a uniform stretching, we can arrange that one of the translations be horizontal with unit displacement. The other displacement can be represented by a point $\omega = u + iv$ in the upper half-plane

$$\mathbb{H} = \{\omega = u + iv \in \mathbb{C} : v > 0\}.$$

Thus we can arrange that the conformal structures on T^2, for fixed a basis for $H_1(T^2; \mathbb{Z})$, are in one-to-one correspondence with points in the upper half-plane $\mathbb{H}$, the point $\omega = u + iv \in \mathbb{H}$ corresponding to the conformal class of the torus $\mathbb{C}/\Lambda$, where Λ is the lattice in $\mathbb{C}$ generated by ω and 1, with α corresponding to ω and β corresponding to 1. We say that $\mathbb{H}$ is the Teichmüller space of the torus and write $\mathcal{T}_1 = \mathbb{H}$.

A change of basis for the lattice Λ is represented by multiplication by a matrix in $SL(2, \mathbb{Z})$,

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \qquad \text{for} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}). \tag{3.13}$$

The corresponding $SL(2,\mathbb{Z})$-action on the generators ω_1 and ω_2 of an arbitrary integral lattice in the complex plane is

$$\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix} \quad \mapsto \quad \begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}$$

while the corresponding action on the coordinate $\omega = u + iv = \omega_1/\omega_2$ is

$$\omega \quad \mapsto \quad \frac{a\omega + b}{c\omega + d}. \tag{3.14}$$

Note that while the action of the mapping class group $SL(2,\mathbb{Z})$ on $H^2(T^2,\mathbb{Z})$ is faithful, the action on the upper half plane $\mathbb{H}$ has kernel consisting of $\pm I$. By applying a suitable element of $SL(2,\mathbb{Z})$, we can arrange that $\omega = u + iv$ lie in the *fundamental domain* of $\mathcal{T}_1$ for this action:

$$D = \{\omega = u + iv \in \mathbb{C} : |u| \le 1/2, v > 0, |\omega| \ge 1\}.$$

The moduli space $\mathcal{R}_1$ is obtained from this fundamental domain by identifying edges. It can be shown that $\mathcal{R}_1$ has a Riemann surface structure and is holomorphically equivalent to the complex line $\mathbb{C}$.

Thus we have two ways of looking at the Riemann surface structures on T^2, the Teichmüller space $\mathcal{T}_1$ when a choice of basis for $H_1(T^2;\mathbb{Z})$ is fixed, and the moduli space $\mathcal{R}_1$ when we ignore choice of basis. The discrete action of $SL(2,\mathbb{Z})$ on $\mathcal{T}_1$ determines a branched covering $\mathcal{T}_1 \to \mathcal{R}_1$.

If $T^2 = \mathbb{C}/\Lambda$, then any map $f : T^2 \to M$ can be lifted to a doubly periodic map $\tilde{f} : \mathbb{C} \to M$ with periods in the lattice Λ. The standard coordinate $z = x + iy$ on $\mathbb{C}$ descends to local coordinates on T^2, and the standard metric $dx^2 + dy^2$ on T^2 descends to a flat metric on T^2 within the conformal equivalence class corresponding to ω. We can rescale the metric to $(1/v)(dx^2 + dy^2)$ so that it has total area one. In terms of the coordinates (t^1, t^2) on $\mathbb{C}$ defined by

$$\begin{cases} x = t^1 + ut^2, \\ y = vt^2, \end{cases}$$

the components of this flat metric and its inverse are

$$(\eta_{ab}) = \frac{1}{v}\begin{pmatrix}1 & u\\ u & u^2+v^2\end{pmatrix}, \qquad (\eta^{ab}) = \frac{1}{v}\begin{pmatrix}u^2+v^2 & -u\\ -u & 1\end{pmatrix},$$

while the area element is given by $dA = dt^1 dt^2$. Hence if $f \in C^2(T^2, M)$,

$$|df|^2 = \sum_{a,b}\eta^{ab}\frac{\partial f}{\partial t^i}\cdot\frac{\partial f}{\partial t^j} = \frac{1}{v}\left[(u^2+v^2)\left|\frac{\partial f}{\partial t^1}\right|^2 - 2u\frac{\partial f}{\partial t^1}\cdot\frac{\partial f}{\partial t_2} + \left|\frac{\partial f}{\partial t_2}\right|^2\right].$$

or equivalently,

$$|df|^2 = v\left|\frac{\partial f}{\partial t^1}\right|^2 + \frac{1}{v}\left|\frac{\partial f}{\partial t^2} - u\frac{\partial f}{\partial t^1}\right|^2,$$

From this, it is apparent that the energy

$$E(f,\omega) = \frac{1}{2}\int_0^1 \int_0^1 \left[v \left| \frac{\partial f}{\partial t^1} \right|^2 + \frac{1}{v} \left| \frac{\partial f}{\partial t^2} - u \frac{\partial f}{\partial t^1} \right|^2 \right] dt^1 dt^2 \tag{3.15}$$

is a smooth function on $C^2(T^2, M) \times \mathcal{T}_1$.

In the following proof, it will be helpful to utilize the Euclidean coordinates $x^1 = \frac{x}{\sqrt{v}}$ and $x^2 = \frac{y}{\sqrt{v}}$, in terms of which the energy of f is expressed as

$$E(f,\omega) = \frac{1}{2}\int_\Delta \left[\left| \frac{\partial f}{\partial x^1} \right|^2 + \left| \frac{\partial f}{\partial x^2} \right|^2 \right] dx^1 dx^2,$$

where Δ is a fundamental domain for the lattice Λ.

Proposition 3.2.4. *If (f,ω) is a critical point for*

$$E : C^2(T^2, M) \times \mathcal{T}_1 \longrightarrow \mathbb{R},$$

then f is a weakly conformal harmonic map, and hence a parametrized minimal surface.

Differentiation with respect to the first variable shows that if (f,ω) is a critical point for E, f must be harmonic. So we need only consider the derivative with respect to the second variable. If $f : T^2 \to M$ is any map, harmonic or not, we can construct the quadratic differential

$$\Omega(f,\omega) = \left[\int_{T^2} \left\langle \frac{\partial f}{\partial z}, \frac{\partial f}{\partial z} \right\rangle dx^1 dx^2 \right] dz^2,$$

which specializes to the Hopf differential of f when f is harmonic. This differential is closely associated with a symmetric bilinear form

$$H(f,\omega) : T_0T^2 \times T_0T^2 \longrightarrow \mathbb{R}$$

that we define on the tangent space to the torus at an origin 0 by

$$H(f,\omega) = \sum h_{ab} dx^a dx^b, \quad \text{where} \quad h_{ab} = \int_{T^2} \left\langle \frac{\partial f}{\partial x^a}, \frac{\partial f}{\partial x^b} \right\rangle dx^1 dx^2, \tag{3.16}$$

which is easily seen to be positive definite. The trace of the quadratic form is $2E(f)$, so the eigenvalues must be of the form

$$E(f,\omega) \pm a, \qquad \text{where} \qquad 0 \le a \le E(f,\omega),$$

and a short calculation shows that $a = |\Omega(f,\omega)|$.

If we change the conformal structure from ω to $\tilde{\omega}$, we must change the Euclidean coordinates by a linear transformation defined by $A \in SL(2,\mathbb{R})$,

$$\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \mapsto \begin{pmatrix} \tilde{x}^1 \\ \tilde{x}^2 \end{pmatrix} = A \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}, \quad \text{and hence} \quad \begin{pmatrix} d\tilde{x}^1 \\ d\tilde{x}^2 \end{pmatrix} = A \begin{pmatrix} dx^1 \\ dx^2 \end{pmatrix}.$$

It follows that the matrices $H = (h_{ij})$ and $\tilde{H} = (\tilde{h}_{ij})$ corresponding to $H(f, \omega)$ and $H(f, \tilde{\omega})$ transform by to the rule,

$$H = A^T \tilde{H} A, \qquad \text{for } A \in SL(2, \mathbb{R}).$$

If $A \in SO(2)$ this reduces to the transformation for a rotation of Euclidean coordinates for a fixed conformal structure.

Suppose now that $H(f, \omega)$ is degenerate; thus one of its eigenvalues is zero. We can rotate coordinates so that the x^1 direction corresponds to the zero eigenvalue. Then

$$\int_{T^2} \left\langle \frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^1} \right\rangle dx^1 dx^2 = 0 \qquad \Rightarrow \qquad \left\langle \frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^1} \right\rangle \equiv 0,$$

and the map f must degenerate to a curve. If f is harmonic with respect to ω, this curve must be a closed geodesic. The energy of the closed geodesic depends upon its length and the choice of conformal structure ω. It is easily seen that no such parametrized geodesic can be a critical point for E.

On the other hand, if $H(f, \omega)$ is nondegenerate, there exists an element $A \in SL(2, \mathbb{R})$ such that $H = A^T H_0 A$, where H_0 is a scalar multiple of the identity matrix, and hence there is a conformal structure ω_0 such that

$$H(f, \omega) = E(f, \omega_0)\langle \cdot, \cdot \rangle.$$

Since H_0 commutes with A, $H = A^T H_0 A = A^T A H_0$. We can rotate the Euclidean coordinates chosen for ω so that $A^T A$ and H are diagonal, and then

$$A^T A = \begin{pmatrix} e^{\lambda} & 0 \\ 0 & e^{-\lambda} \end{pmatrix}, \qquad H = E(f, \omega_0) \begin{pmatrix} e^{\lambda} & 0 \\ 0 & e^{-\lambda} \end{pmatrix},$$

for some choice of λ. Then $E(f, \omega) = (\cosh \lambda) E(f, \omega_0)$, and we see that the only critical points for E are not only harmonic, but also weakly conformal.

As a byproduct of the proof, we see that there are two types of harmonic tori. Those for which $H(f, \omega)$ is degenerate are simply parametrizations of smooth closed geodesics, and these can never be conformal with respect to any conformal structure. Those for which $H(f, \omega)$ is nondegenerate are either conformal or their energy can be decreased by replacing the conformal structure ω with a new conformal structure ω_0.

Note that the differential dz descends from $\mathbb{C}$ to the torus $\mathbb{C}/\Lambda$ and hence holomorphic differentials on a torus must be of the form hdz^2, where h is a holomorphic function on the torus. By the maximum modulus theorem, h must be constant. Hence the only possible Hopf differentials for harmonic tori are cdz^2, where c is a complex constant. If the harmonic torus $f : T^2 \to M$ is not weakly conformal, its Hopf differential is never zero, so it cannot have branch points. However, it may have fold points as the following example shows.

Example 3.2.5. If f is an immersion, the fibers of the line bundles $\mathbf{L}$ and $\bar{\mathbf{L}}$ are linearly independent at every point. However, a harmonic map f can have

points p at which the fibers $\mathbf{L}$ and $\bar{\mathbf{L}}$ coincide. At such points the rank of f can be at most one.

Indeed, there is a degree zero harmonic map $f : T^2 \to S^2$ when S^2 is given the standard Riemannian metric of constant curvature one, which has "fold points" along two circles parallel to the equator. To see how this is constructed, we first note that the metric on $S^2 \subset \mathbb{R}^2$ with equation $x^2 + y^2 + z^2 = 1$ is expressed in spherical coordinates $z = \cos\phi$, θ being the standard angular coordinate in the (x, y)-plane, is

$$ds^2 = (\cos^2\phi)d\theta^2 + d\phi^2 = \operatorname{sech}^2 u(d\theta^2 + du^2),$$

where u and ϕ are related by the equation $\tanh(u/2) = \tan(\phi/2)$. In terms of the standard coordinates (t^1, t^2) on T^2 which correspond to the factorization $T^2 = S^1 \times S^1$, the mapping $f : T^2 \to S^2$ can expressed as

$$\theta(t^1, t^2) = t_2, \qquad \phi(t^1, t^2) = \phi(t_1),$$

where ϕ is a (nonconstant speed) parametrization of the geodesic $\theta = (\text{constant})$.

Note that the circle $\phi = (\text{constant})$ has curvature $\kappa = 1/\cos\phi$ and normal curvature $\kappa_n = 1$. From the equation $\kappa_g^2 + \kappa_n^2 = \kappa^2$, where κ_g is the geodesic curvature, implies that $\kappa_g = \pm\tan\phi$. Moreover, the curve is traversed with constant speed $\cos\phi$. Hence

$$0 = \frac{D}{\partial t^1}\left(\frac{\partial f}{\partial t^1}\right) + \frac{D}{\partial t^2}\left(\frac{\partial f}{\partial t^2}\right) = \frac{d^2\phi}{dt^2} + (\tan\phi)(\cos^2\phi) = \frac{d^2\phi}{dt^2} + \frac{1}{2}\sin(2\phi).$$

The differential equation we must solve to obtain a harmonic map (the pendulum equation except for constant factors) is equivalent to the first order system

$$\frac{d\phi}{dt} = \psi, \qquad \frac{d\psi}{dt} = -\frac{1}{2}\sin(2\phi).$$

Eliminating dt yields

$$\frac{d\psi}{d\phi} = -\frac{(1/2)\sin(2\phi)}{\psi} \quad \text{which integrates to} \quad \psi^2 - \frac{1}{2}\cos(2\phi) = (\text{constant}).$$

Various choices of constant yield harmonic maps for appropriate conformal structures on T^2.

Note that the antipodal map $A : S^2 \to S^2$ induces an orientation reversing map $A : T^2 \to T^2$ such that $f \circ A = A \circ f$. We can take the quotient in both domain and range, obtaining thereby a harmonic map from a Klein bottle into the real projective plane $\mathbb{R}P^2$, which has as its image a Möbius band.

3.3 Minimal surfaces of higher genus

Finally, we consider the case where Σ is a sphere with g handles and $g \geq 2$. Classifying the conformal structures on Σ is now somewhat more challenging.

We let $\mathrm{Met}^k(\Sigma_g)$ denote the space of L^2_k Riemannian metrics

$$\eta = \sum_{a,b=1}^{2} \eta_{ab} du^a du^b$$

on Σ, an open subset of a Hilbert space, and let $\mathrm{Met}^k_0(\Sigma_g)$ denote the submanifold of metrics of constant Gaussian curvature and total area one. It is not difficult to conclude from the Uniformization Theorem that given any $\eta \in \mathrm{Met}^k(\Sigma_g)$, there is a rescaling $\lambda^2\eta$, where $\lambda^2 > 0$, which lies in $\mathrm{Met}^k_0(\Sigma_g)$. Moreover, one can show that this rescaling is unique and we thus obtain a strong deformation retraction

$$\pi : \mathrm{Met}^k(\Sigma_g) \longrightarrow \mathrm{Met}^k_0(\Sigma_g). \tag{3.17}$$

In addition, we consider

$$\mathrm{Diff}^{k+1}_+(\Sigma_g) = \{\phi \in L^2_{k+1}(\Sigma, \Sigma) \text{ such that } \phi \text{ is an orientation-preserving diffeomorphism }\},$$

which is a group under composition, and its subgroup

$$\mathrm{Diff}^{k+1}_0(\Sigma_g) = \{\phi \in \mathrm{Diff}^{k+1}_+(\Sigma_g) : \ \phi \text{ is homotopic to the identity }\}.$$

Unfortunately, although $\mathrm{Diff}^{k+1}_+(\Sigma_g)$ and $\mathrm{Diff}^{k+1}_0(\Sigma_g)$ are smooth Hilbert manifolds, the group multiplications are not smooth, so these are not infinite-dimensional Lie groups. However, the maps

$$\psi : \mathrm{Diff}^{k+1}_+(\Sigma_g) \times \mathrm{Met}^k(\Sigma_g) \longrightarrow \mathrm{Met}^k(\Sigma_g), \quad \psi(\phi, \eta) = \phi^*\eta,$$

and its restriction

$$\psi : \mathrm{Diff}^{k+1}_+(\Sigma_g) \times \mathrm{Met}^k_0(\Sigma_g) \longrightarrow \mathrm{Met}^k_0(\Sigma_g),$$

are indeed smooth. Moreover, $\psi(\phi, \pi(\eta)) = \pi \circ \psi(\phi, \eta)$, where π is the projection (3.17) into constant curvature metrics. Of course, either of these maps can be further restricted to the subgroup $\mathrm{Diff}^{k+1}_0(\Sigma_g)$.

Lemma 3.3.1. *The continuous action of $\mathrm{Diff}^{k+1}_0(\Sigma_g)$ on $\mathrm{Met}^k_0(\Sigma_g)$ is free. Moreover, the image of the map*

$$\tilde{\psi} : \mathrm{Diff}^{k+1}_+(\Sigma_g) \times \mathrm{Met}^k_0(\Sigma_g) \longrightarrow \mathrm{Met}^k_0(\Sigma_g) \times \mathrm{Met}^k_0(\Sigma_g), \quad \tilde{\psi}(\phi, \eta) = (\eta, \psi(\phi, \eta)), \tag{3.18}$$

is closed.

Indeed, if the action of $\mathrm{Diff}^{k+1}_0(\Sigma_g)$ on $\mathrm{Met}^k_0(\Sigma_g)$ were not free, there would be two distinct isometries

$$\mathrm{id}, \phi : (\Sigma_g, \phi^*\eta) \longrightarrow (\Sigma_g, \eta),$$

both homotopic to the identity. Each would be harmonic, contradicting Hartman's Theorem 3.1.2 which asserts that the harmonic map between two compact surfaces of negative curvature in a given homotopy class is unique. We leave the second statement as an easy exercise for the reader.

The orbit spaces,

$$\mathcal{T}_g = \frac{\mathrm{Met}_0^k(\Sigma_g)}{\mathrm{Diff}_0^{k+1}(\Sigma_g)} \quad \text{and} \quad \mathcal{R}_g = \frac{\mathrm{Met}_0^k(\Sigma_g)}{\mathrm{Diff}_+^{k+1}(\Sigma_g)},$$

are called the *Teichmüller space* and the *Riemann moduli space* of conformal structures on Σ_g, respectively. The fact that the image of the map $\tilde{\psi}$ in (3.18) is closed is equivalent to the orbit space $\mathcal{T}_g$ being Haussdorf.

Theorem 3.3.2. *If $g \geq 2$, the Teichmüller space $\mathcal{T}_g$ is a manifold of dimension $6g-6$.*

This is essentially due to Earle and Eells [15], and a modern argument using Banach manifolds is presented in Fischer and Tromba [20]. Our argument will mostly follow Chapter 2 of [73]. Our goal is to construct local coordinates in $\mathcal{T}_g$ about a base metric $\eta \in \mathrm{Met}_0^k(\Sigma_g)$. Let $G = \mathrm{Diff}_0^{k+1}(\Sigma_g)$. According to a well-known result regarding group actions (namely Theorem 2.9.10 in Varadarajan [81]), it suffices to construct a submanifold $S \subset \mathrm{Met}_0^k(\Sigma_g)$ such that $\eta \in S$ and

1. $T_\eta \mathrm{Met}_0^k(\Sigma_g) \cong T_\eta S \oplus T_\eta(G \cdot \eta)$ and
2. any G-orbit intersects S in only one point.

Such a submanifold is called a *slice* for the action, and makes it possible to construct an open neighborhood U of η in S and a diffeomorphism from $U \times G$ to an open neighborhood of the orbit through η.

To construct this slice, we first note that if

$$X = \sum_{a=1}^{2} X_a \frac{\partial}{\partial u^a}$$

is a smooth vector field on Σ with local one-parameter subgroup $\{\phi_t : t \in \mathbb{R}\}$, it follows from a familiar calculation that

$$\frac{d}{dt}(\phi_t^* \eta)\Big|_{t=0} = L_X \eta = \cdots = \sum_{a,b=1}^{2} X_{a;b} du^a du^b,$$

where the $X_{a;b}$'s are the components of the covariant derivative of X with respect to the metric η. We let $\langle \cdot, \cdot \rangle$ denote the G-invariant L^2 inner product on $T_\eta \mathrm{Met}^k(\Sigma_g)$ such that

$$\langle \dot{\eta}, \dot{\eta} \rangle = \int_\Sigma \frac{1}{\lambda^2} \sum_{a,b=1}^{2} \dot{\eta}_{ab} \dot{\eta}_{ab} du^1 du^2 \quad \text{and let} \quad \|\dot{\eta}\| = \sqrt{\langle \dot{\eta}, \dot{\eta} \rangle}.$$

We suppose that we have chosen coordinates so that the components of our base metric η are $\eta_{ab} = \delta_{ab}\lambda^2$, for some positive function λ^2, and consider when a one-parameter family of metrics,

$$\eta_{ab}(t) = \eta_{ab} + t\dot{\eta}_{ab} \quad \text{for} \quad t \in (\epsilon, \epsilon),$$

is perpendicular with respect to $\langle \cdot, \cdot \rangle$ at $t = 0$ to the orbit through η. An integration by parts shows that

$$\langle L_X \eta, \dot{\eta} \rangle = \int_\Sigma \frac{1}{\lambda^2} \sum_{a,b=1}^{2} \dot{\eta}_{ab} X_{a;b} du^1 du^2 = -\int_\Sigma \frac{1}{\lambda^2} \sum_{a,b=1}^{2} \dot{\eta}_{ab;b} X_a du^1 du^2,$$

so that $\dot{\eta}$ is perpendicular to the G-orbit when $\sum \dot{\eta}_{ab;b} = 0$. This suggests a slice for the action of G on the larger space $\mathrm{Met}^k(\Sigma_g)$, namely

$$\tilde{S}_\eta = \left\{ \eta + \dot{\eta} : \ \dot{\eta} \in T_\eta \mathrm{Met}^k(\Sigma_g), \ \|\dot{\eta}\| < \epsilon \text{ and } \sum_{b=1}^{2} \dot{\eta}_{ab;b} = 0 \right\}.$$

To obtain a section for the action on the smaller space $\mathrm{Met}_0^k(\Sigma_g)$, we utilize the following lemma:

Lemma 3.3.3. *If*

$$\eta_{ab}(t) = \eta_{ab} + t\dot{\eta}_{ab} \quad \textit{for} \quad t \in (\epsilon, \epsilon)$$

is a one-parameter family of metrics of constant Gaussian curvature K and total area one, $\dot{\eta}_{ab;b} = 0$ and

$$\Delta_{\eta(0)}(Tr(\dot{\eta})) = -2K\, Tr(\dot{\eta}), \quad \textit{where} \quad Tr(\dot{\eta}) = \frac{1}{\lambda^2}(\dot{\eta}_{11} + \dot{\eta}_{22})$$

and $\Delta_{\eta(0)}$ is the Laplace operator for the base metric $\eta(0)$.

To prove this, we can use geodesic normal coordinates centered for η at a given point $p \in \Sigma$. In terms of such coordinates, the curvature tensor of $\eta(t)$ is given by

$$K \det(\eta_{ab}(t)) = R_{1212}(t) = \frac{1}{2}\left(2\eta_{12,12}(t) - \eta_{11,22}(t) - \eta_{22,11}(t)\right) + (\text{higher order terms}),$$

where the commas denote coordinate derivatives. Differentiating with respect to t and setting $t = 0$, we obtain

$$K\lambda^2(\dot{\eta}_{11} + \dot{\eta}_{22}) = \frac{1}{2}\left(2\dot{\eta}_{12;12} - \dot{\eta}_{11;22} - \dot{\eta}_{22;11}\right),$$

where we have replaced ordinary derivatives by covariant derivatives and have evaluated at p. Finally, we use the identity $\sum \dot{\eta}_{ab;b} = 0$ to simplify the right-hand side,

$$K\lambda^2(\dot{\eta}_{11} + \dot{\eta}_{22}) = -\frac{1}{2}\left(\sum_{a,b=1}^{2} \dot{\eta}_{aa;bb}\right),$$

which yields the statement of the lemma:

$$\frac{1}{\lambda^2}\sum_{a,b=1}^{2}\left(\frac{1}{\lambda^2}\dot{\eta}_{aa}\right)_{;bb} = -2K\frac{1}{\lambda^2}(\dot{\eta}_{11}+\dot{\eta}_{22})$$

Since K is negative, it follows from the Lemma and the maximum principle applied to the operator $\Delta_{\eta(0)}+2K$ that $\dot{\eta}_{11}+\dot{\eta}_{22}=0$. This, together with the identity $\sum \dot{\eta}_{ab;b}=0$ implies that if $w=u^1+iu^2$,

$$\frac{\partial}{\partial \bar{w}}(\dot{\eta}_{11}+i\dot{\eta}_{12})=0,$$

and hence

$$(\dot{\eta}_{11}+i\dot{\eta}_{12})dw^2$$

is a holomorphic quadratic differential. Thus we set

$$\begin{aligned} S_\eta = \{\eta+\dot{\eta}:\ \dot{\eta}\in T_\eta \mathrm{Met}^k(\Sigma_g),\ \|\dot{\eta}\|<\epsilon \text{ and} \\ (\dot{\eta}_{11}+i\dot{\eta}_{12})dw^2 \text{ is a holomorphic quadratic differential }\}, \end{aligned}$$

a convex open subset of an affine subspace of a Hilbert space. One of the implications of the Riemann-Roch Theorem from Riemann surface theory is that the space of holomorphic quadratic differentials on a Riemann surface of genus g has complex dimension $3g-3$ or real dimension $6g-6$, so the affine space in which S_η lies has dimension $6g-6$.

A straightforward (if slightly tedious) calculation shows that the identity map

$$\mathrm{id}:(\Sigma_g,\eta)\longrightarrow(\Sigma_g,\eta+\dot{\eta})$$

is harmonic with Hopf differential $(1/2)(\dot{\eta}_{11}+i\dot{\eta}_{12})dw^2$. If the G-orbit intersected S_η in more than one point, we would have homotopic harmonic maps from (Σ_g,η) to the same target with two different Hopf differentials, contradicting Theorem 3.1.2 on uniqueness for harmonic maps. The section S_η is only tangent to the submanifold $\mathrm{Met}_0^k(\Sigma_g)$, but since the projection π to the space $\mathrm{Met}_0^k(\Sigma_g)$ commutes with the G-action, $S=\pi(S_\eta)$ is a slice for the G-action on $\mathrm{Met}_0^k(\Sigma_g)$ of dimension $6g-6$, completing the proof of the theorem.

Remark 3.3.4. In contrast to the Teichmüller space $\mathcal{T}_g$, the Riemann moduli space $\mathcal{R}_g$ is not even a manifold in general, but an orbifold. It is the quotient of Teichmüller space by the properly discontinuous action of the mapping class group $\Gamma_g=\mathrm{Diff}_+(\Sigma_g)/\mathrm{Diff}_0(\Sigma_g)$.

Choose a base Riemann surface (Σ,η_0) of genus g and let $\mathcal{O}(\kappa^2)$ denote the space of holomorphic quadratic differentials on (Σ,η_0), a real vector space of dimension $6g-6$. We can define a map

$$\Psi:\mathcal{T}_g\to\mathcal{O}(\kappa^2)\quad\text{by}\quad\Psi(\Sigma,\eta)=\Omega(f),$$

where $\Omega(f)$ is the Hopf differential of the unique harmonic map f from (Σ, η_0) to (Σ, η).

Theorem 3.3.5. (Teichmüller) *If $g \geq 2$, the map $\Psi : \mathcal{T}_g \to \mathcal{O}(\kappa^2)$ is a diffeomorphism, and hence $\mathcal{T}_g$ is diffeomorphic to $\mathbb{R}^{6g-6}$.*

Teichmüller's original proof that $\mathcal{T}_g$ is homeomorphic to $\mathbb{R}^{6g-6}$ was based upon the theory of quasiconformal maps, and was later simplified by Bers, as presented in [37]. Later a proof was found using harmonic maps, and we will return to discuss it later.

Let $\eta_0 \in \mathrm{Met}_0^k(\Sigma_g)$ be a base Riemannian metric. If η is any element of $\mathcal{T}_g$, it follows from Theorems 3.1.1 and 3.1.2 that there is a unique harmonic map

$$s(\eta) : (\Sigma, \eta) \to (\Sigma, \eta_0),$$

which depends smoothly on η in terms of the above coordinates. Using a variant of the Bochner Lemma to be treated in the next section, Schoen and Yau were able to prove that $s(\eta)$ is a diffeomorphism ([73], Chapter 1, §8). Thus we can define a map

$$\mathrm{Met}_0^k(\Sigma_g) \to \mathrm{Met}_0^k(\Sigma_g) \quad \text{by} \quad \sigma(\eta) = [s(\eta)^{-1}]^*(\eta).$$

Since $\phi \circ s(\phi^*\eta) = s(\eta)$, the metric $s(\eta)$ defines a section $\sigma : \mathcal{T}_g \to \mathrm{Met}_0^k(\Sigma_g)$.

We can now consider the energy as a map of two variables $E : C^2(\Sigma, M) \times \mathrm{Met}_0^k(\Sigma_g) \to \mathbb{R}$ defined by

$$E(f, (\eta_{ab})) = \frac{1}{2} \int_\Sigma \sum_{a,b} \eta^{ab} \left\langle \frac{\partial f}{\partial u^a}, \frac{\partial f}{\partial u^b} \right\rangle \sqrt{\eta} du_1 du_2, \tag{3.19}$$

where (η^{ab}) is the matrix inverse of (η_{ab}) and $\eta = \det(\eta_{ab})$, the integrand being independent of choice of local coordinates. If $\phi \in \mathrm{Diff}_0^{k+1}(\Sigma_g)$, then

$$E(f \circ \phi, \phi^*\eta) = E(f, \eta),$$

and hence E descends to a map on the space of orbits

$$\frac{C^2(\Sigma, M) \times \mathrm{Met}_0^k(\Sigma_g)}{\mathrm{Diff}_0^{k+1}(\Sigma_g)}.$$

However, using the map σ described above, one can exhibit this quotient as simply a product $C^2(\Sigma, M) \times \mathcal{T}_g$.

Theorem 3.3.6. *If (f, ω) is a critical point for the map*

$$E : C^2(\Sigma, M) \times \mathcal{T}_g \longrightarrow \mathbb{R},$$

then f is harmonic and weakly conformal with respect to ω, and hence a parametrized minimal surface.

Differentiation with respect to f shows that f is harmonic. We need to calculate the derivative with respect to the metric, so we take a perturbation of a given metric with support in a given isothermal coordinate system (u^1, u^2) on Σ. Suppose that the perturbation is given by the formula

$$\eta_{ab}(t) = \eta_{ab} + t\dot{\eta}_{ab} = \eta_{ab} + t\psi(u^1, u^2)\rho_{ab}(u^1, u^2),$$

where $\psi(x_1, x_2)$ has compact support, the variation in the metric is trace-free ($\dot{\eta}_{11} + \dot{\eta}_{22} = 0$) and the initial metric is given by the formulae

$$\eta_{ab} = \delta_{ab}\lambda^2, \qquad \sqrt{\eta} = \lambda^2,$$

for some conformal factor λ^2. Then a direct calculation shows that

$$\frac{d}{dt}\eta_{ab}(t) = \dot{\eta}_{ab}, \qquad \frac{d}{dt}\eta(t)\bigg|_{t=0} = \lambda^2(\dot{\eta}_{11} + \dot{\eta}_{22}) = 0,$$

From this, we can easily calculate

$$\frac{d}{dt}\begin{pmatrix} \sqrt{\eta}\eta^{11} & \sqrt{\eta}\eta^{12} \\ \sqrt{\eta}\eta^{21} & \sqrt{\eta}\eta^{22} \end{pmatrix}\bigg|_{t=0} = \lambda^{-2}\begin{pmatrix} \dot{\eta}_{22} & -\dot{\eta}_{12} \\ -\dot{\eta}_{21} & \dot{\eta}_{11} \end{pmatrix}.$$

Thus we find that

$$\begin{aligned} \frac{d}{dt}E(f, \eta_{ab}(t))\bigg|_{t=0} &= \int_\Sigma \sum_{a,b} \frac{d}{dt}\sqrt{\eta}\eta^{ab}\bigg|_{t=0} \left\langle \frac{\partial f}{\partial u^a}, \frac{\partial f}{\partial u^b} \right\rangle du^1 du^2 \\ = -\int_\Sigma \left[\frac{\dot{\eta}_{11}}{\lambda^2}\left(\left\langle \frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^1} \right\rangle - \left\langle \frac{\partial f}{\partial u^2}, \frac{\partial f}{\partial u^2} \right\rangle \right) + \frac{2\dot{\eta}_{12}}{\lambda^2}\left\langle \frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^2} \right\rangle \right] du^1 du^2 \\ = -4\int_\Sigma \mathrm{Re}\left[(\dot{\eta}_{11} + i\dot{\eta}_{12})\left\langle \frac{\partial f}{\partial w}, \frac{\partial f}{\partial w} \right\rangle \right] \frac{1}{\lambda^2} du^1 du^2, \quad (3.20) \end{aligned}$$

where $w = u^1 + iu^2$. Since this is true for all trace-free variations in the metric, we conclude that

$$\left\langle \frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^1} \right\rangle - \left\langle \frac{\partial f}{\partial u^2}, \frac{\partial f}{\partial u^2} \right\rangle = 0 = \left\langle \frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^2} \right\rangle,$$

and hence the Hopf differential $\Omega(f)$ must vanish. This finishes the proof of the theorem.

Of course, it is also true that $E(f \circ \phi, \phi^*\eta) = E(f, \eta)$ for ϕ in the larger group of orientation-preserving diffeomorphisms $\mathrm{Diff}_+^{k+1}(\Sigma_g)$, but this group does not act freely on $C^2(\Sigma, M) \times \mathrm{Met}_0^k(\Sigma_g)$ and the quotient space

$$\mathcal{M}(\Sigma, M) = \frac{C^2(\Sigma, M) \times \mathrm{Met}_0^k(\Sigma_g)}{\mathrm{Diff}_+^{k+1}(\Sigma_g)}$$

is not a manifold, but only an orbifold in general. We have two ways of looking at parametrized minimal surfaces of genus g, as critical points for the energy

$$E : \mathcal{M}(\Sigma, M) \longrightarrow \mathbb{R}$$

or as Γ_g-orbits of critical points for

$$E : C^2(\Sigma, M) \times \mathcal{T}_g \longrightarrow \mathbb{R},$$

where $\Gamma_g = \mathrm{Diff}_+^{k+1}(\Sigma_g)/\mathrm{Diff}_0^{k+1}(\Sigma_g)$ is the mapping class group.

3.4 The Bochner Lemma

The key step behind the existence theory for harmonic is the Bochner Lemma which gives an estimate for the Laplacian of the energy density in terms of curvature. We describe the Bochner Lemma in this section, following the treatment in [45]. Let Σ and M be compact Riemannian manifolds with metrics (η_{ab}) and (g_{ij}) respectively, with M as usual being isometrically imbedded in Euclidean space $\mathbb{R}^N$. We suppose that

$$f : \Sigma \longrightarrow M \subseteq \mathbb{R}^N$$

is a smooth map and for given choice of coordinates $(u^1, \ldots, u^m)$ on Σ, we consider the vector-valued maps

$$f_a = \frac{\partial f}{\partial u^a} \in \mathbb{R}^N \quad \text{which have components} \quad f_a^i = \frac{\partial f^i}{\partial u^a},$$

with respect to coordinates $(x^1, \ldots, x^n)$ on M. Recall that the energy density is given by

$$e(f) = \frac{1}{2} \sum \eta^{ab} f_a \cdot f_b = \frac{1}{2} \sum g_{ij} \eta^{ab} f_a^i f_b^j.$$

Finally, we let

$$R^\Sigma = (\tilde{R}_{abcd}) \quad \text{and} \quad R^M = (R_{ijkl})$$

be the curvature tensors of Σ and M, and let $\mathrm{Ric}^\Sigma = (\tilde{R}_{ab})$ be the Ricci curvature of Σ. Recall that if we divide the Laplacian of the vector-valued map $f : \Sigma \to \mathbb{R}^N$ into tangential and normal components,

$$\Delta_\eta f = (\Delta_\eta f)^\top + (\Delta_\eta f)^\perp,$$

the condition that f be harmonic is just $(\Delta_\eta f)^\top = 0$. On the other hand, the normal component of the Laplacian is expressed in terms of the second fundamental form

$$\alpha(f)(p) : T_pM \times T_pM \longrightarrow N_pM$$

of M within $\mathbb{R}^N$ by

$$(\Delta_\eta f)^\perp(p) = \sum \eta^{ab} \alpha(f)(p)(f_a, f_b).$$

Theorem 3.4.1 (Bochner Lemma). *If $f : \Sigma \to M$ is a smooth map, then*

$$\begin{aligned} \Delta_\eta(e(f)) = |\nabla df|^2 + \langle d[(\Delta_\eta f)^\top], df \rangle \\ + \sum \eta^{ac} \eta^{bd} \tilde{R}_{ab} f_c \cdot f_d - \sum \eta^{ac} \eta^{bd} R_{ijkl} f_a^i f_b^j f_c^k f_d^l. \end{aligned} \quad (3.21)$$

where ∇ denotes covariant derivative with respect to the connection in $T^*\Sigma \otimes f^*TM$. Hence if f is harmonic,

$$\Delta_\eta(e(f)) = |\nabla df|^2 + \sum \eta^{ac}\eta^{bd}\tilde{R}_{ab}f_c \cdot f_d - \sum \eta^{ac}\eta^{bd}R_{ijkl}f_a^i f_b^j f_c^k f_d^l. \quad (3.22)$$

To prove this, we use normal coordinates centered a point $p \in \Sigma$ and similar coordinates at $f(p)$ in M. Then

$$\begin{aligned}\Delta_\eta(e(f)) &= \sum |f_{a;b}|^2 + \sum \langle f_a, f_{a;bb}\rangle \\ &= \sum |f_{a;b}|^2 + \sum \langle f_a, f_{b;ba}\rangle - \sum \langle f_a, \tilde{R}_{bcab}f_c\rangle \\ &= \sum |f_{a;b}|^2 + \sum \langle f_a, (\Delta_\eta f)_a\rangle + \sum \tilde{R}_{ab}\langle f_a, f_b\rangle. \quad (3.23)\end{aligned}$$

On the other hand,

$$\begin{aligned}\langle f_a, (\Delta_\eta(f))_a\rangle &= \langle f_a, [(\Delta_\eta f)^\top]_a\rangle + \langle f_a, [(\Delta_\eta f)^\perp]_a\rangle \\ &= \langle f_a, [(\Delta_\eta f)^\top]_a\rangle + \sum \langle f_a, (\alpha(f)(f_b, f_b)_a\rangle.\end{aligned}$$

Now we use the fact that

$$\langle f_a, \alpha(f)(f_b^i, f_b^i)\rangle = 0$$

to conclude that

$$\begin{aligned}\langle f_a, (\Delta_\eta(f))_a\rangle &= \langle f_a, [(\Delta_\eta f)^\top]_a\rangle - \sum \langle \Delta_\eta(f), (\alpha(f)(f_b, f_b)\rangle \\ &= \langle f_a, [(\Delta_\eta f)^\top]_a\rangle - \sum \langle (\alpha(f)(f_a, f_a), (\alpha(f)(f_b, f_b)\rangle. \quad (3.24)\end{aligned}$$

Finally, we note that

$$\sum |f_{a;b}|^2 = |\nabla df|^2 + \sum \langle \alpha(f)(f_a, f_b), \alpha(f)(f_a, f_b)\rangle. \quad (3.25)$$

Finally, we substitute (3.24) and (3.25) into (3.24) and obtain

$$\begin{aligned}\Delta_\eta(e(f)) &= |\nabla df|^2 + \langle d[(\Delta_\eta f)^\top], df\rangle + \sum \tilde{R}_{ab}f_a \cdot f_b \\ &\quad - \left[\sum \langle \alpha(f)(f_a, f_a), (\alpha(f)(f_b, f_b)\rangle - \sum \langle \alpha(f)(f_a, f_b), \alpha(f)(f_a, f_b)\rangle\right].\end{aligned}$$

This last equation implies (3.21) by the Gauss equation.

If Σ is a compact surface with Gaussian curvature $\tilde{K}$, then (3.22) simplifies to

$$\Delta_\eta(e(f)) = |\nabla df|^2 + \tilde{K}|df|^2 - \sum R_{ijkl}f_a^i f_b^j f_a^k f_b^l. \quad (3.26)$$

In terms of isothermal coordinates $(u, v) = (u^1, u^2)$ on Σ, we can write

$$\sum \eta_{ab}du^a du^b = \lambda^2(du^2 + dv^2), \quad a(f) = \frac{1}{\lambda^2}\left|\frac{\partial f}{\partial u} \wedge \frac{\partial f}{\partial v}\right|.$$

Then (3.26) becomes

$$\Delta_\eta(e(f)) \geq |\nabla df|^2 + 2\tilde{K}e(f) - 2K(\sigma)a(f)^2, \tag{3.27}$$

where $K(\sigma)$ is the sectional curvature of the two-plane generated by $\partial f/\partial u$ and $\partial f/\partial v$. It follows from inequality (3.10) that if $K_0 \geq 0$, then

$$K(\sigma) \leq K_0 \quad \Rightarrow \quad \Delta_\eta(e(f)) \geq |\nabla df|^2 + 2\tilde{K}e(f) - 2K_0 e(f)^2, \tag{3.28}$$

equality holding when f is conformal.

Corollary 3.4.2. *If M has nonpositive sectional curvatures, there are no harmonic maps from the two-sphere S^2 into M and any harmonic map from the torus T^2 into M must be totally geodesic. If M has negative sectional curvatures, there are no harmonic maps from S^2 or T^2 into M.*

Proof: If $\Sigma = S^2$, we can choose the Riemannian metric on Σ so that $\tilde{K} \equiv 1$. The continuous function $e(f)$ must assume a maximum at some point. At this point the Bochner inequality shows that $\Delta(e(f)) > 0$, a contradiction.

If $\Sigma = T^2$, we can choose the Riemannian metric on Σ so that $\tilde{K} \equiv 0$. In this case, the inequality must be an equality. Hence $\nabla df = 0$ which implies that

$$\frac{\partial f}{\partial u_1} \qquad \text{and} \qquad \frac{\partial f}{\partial u_2}$$

are parallel sections of the pullback of the bundle f^*TM. This implies of course that f is totally geodesic.

The case where M has negative sectional curvatures is proven in a similar fashion.

3.5 The α-energy

In order to apply critical point theory to a function on an infinite-dimensional manifold, we need the function to be continuous and satisfy condition C of Palais and Smale. For the usual energy, however, the latter condition would require that we complete the space $C^2(\Sigma, M)$ with respect to the L^2_1-topology, a topology which is too weak for the usual techniques of global analysis, since the Sobolev inequality barely fails to show that L^2_1 functions are continuous.

Thus Sacks and Uhlenbeck [68], [69] were led to consider perturbations of the energy. Among the simplest such perturbations are functions

$$F : C^\infty(\Sigma, M) \to \mathbb{R} \qquad \text{defined by} \qquad F(f) = \int_\Sigma \phi(|df|^2) dA,$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is a smooth function. By calculating the differential of F and integrating by parts (assuming that f is sufficiently differentiable), we obtain the Euler-Lagrange equation for critical points

$$\frac{D}{\partial u}\left(\phi'(|df|^2)\frac{\partial f}{\partial u}\right) + \frac{D}{\partial v}\left(\phi'(|df|^2)\frac{\partial f}{\partial v}\right) = 0, \tag{3.29}$$

whenever (u, v) is a set of isothermal coordinates on Σ.

Example 3.5.1. We could take $\phi(t) = t^\alpha$, where $\alpha > 1$, which has continuous derivatives up to order two when regarded as a function on $L_1^{2\alpha}(\Sigma, M)$, a Banach space which does lie within Sobolev range. Thus for a given Riemannian metric η on Σ, we could define

$$\hat{E}_{\alpha,\eta} : L_1^{2\alpha}(\Sigma, M) \to \mathbb{R} \quad \text{by} \quad \hat{E}_{\alpha,\eta}(f) = \frac{1}{2}\int_\Sigma |df|^{2\alpha} dA, \tag{3.30}$$

However, the critical points of $\hat{E}_{\alpha,\eta}$ are not necessarily smooth. Indeed, in the special case where $M = \mathbb{R}$, the Euler-Lagrange equation

$$\frac{\partial}{\partial u}\left(|df|^{2\alpha-2}\frac{\partial f}{\partial u}\right) + \frac{\partial}{\partial v}\left(|df|^{2\alpha-2}\frac{\partial f}{\partial v}\right) = 0$$

can be rewritten in terms of polar coordinates (r, θ) as

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r|df|^{2\alpha-2}\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial}{\partial \theta}\left(|df|^{2\alpha-2}\frac{\partial f}{\partial \theta}\right) = 0,$$

which is satisfied by $f(r, \theta) = r^{(2\alpha-2)/(2\alpha-1)}$. Although this function is not C^1, it can be checked that it does lie in $L_1^{2\alpha}$. Thus we see that critical points in $L_1^{2\alpha}$ for this choice of ϕ are not necessarily C^∞.

However, the slight modification in which $\phi(t) = (1 + t)^\alpha$ provides a function which has C^∞ critical points, as we will see later, and that is the choice that we will make.

Definition. Suppose that M is a smooth manifold with Riemannian metric g. Given a Riemannian metric η on Σ, the α-energy is the function $E_{\alpha,\eta} : L_1^{2\alpha}(\Sigma, M) \to \mathbb{R}$ defined by

$$E_{\alpha,\eta}(f) = \frac{1}{2}\int_\Sigma (1 + |df|^2)^\alpha dA, \tag{3.31}$$

for $\alpha > 1$, where dA is calculated with respect to η and $|df|$ is calculated with respect to η and g. For $\omega \in \mathcal{T}$, the (α, ω)-*energy* $E_{\alpha,\omega}$ is just the function $E_{\alpha,\eta}$, where η is chosen to be the unique constant curvature metric of total area one on Σ within the conformal class ω.

By the normalization we have chosen for the metric, the α-energy $E_{\alpha,\omega}$ approaches $E_\omega + (1/2)$ as $\alpha \to 1$.

Note that the (α, ω)-energy on M is a composition of the map ω_i, induced via Theorem 1.4.7 (the so-called ω-lemma) by an isometric imbedding $i : M \to \mathbb{R}^N$, with the (α, ω)-energy on $\mathbb{R}^N$. The latter map, $E_{\alpha,\omega} : L_1^{2\alpha}(\Sigma, \mathbb{R}^N) \to \mathbb{R}$, is clearly continuous, and it is also C^2 with derivatives

$$dE_{\alpha,\omega}(f)(V) = \alpha\int_\Sigma (1 + |df|^2)^{\alpha-1} df \cdot dV dA \tag{3.32}$$

and

$$d^2E_{\alpha,\omega}(f)(V_1,V_2) = \alpha\int_\Sigma (1+|df|^2)^{\alpha-1}dV_1\cdot dV_2 dA \\ + 2\alpha(\alpha-1)\int_\Sigma (1+|df|^2)^{\alpha-2}(df\cdot dV_1)(df\cdot dV_2)dA, \quad (3.33)$$

by the same argument we used in Example 1.2.3. Since the composition of C^2 maps is C^2 we see that the map $E_{\alpha,\omega}$ on $L_1^{2\alpha}(\Sigma,M)$ is also C^2.

We will see shortly that $E_{\alpha,\omega}$ satisfies condition C and its critical points are smooth. If we let $\omega\in\mathcal{T}$ vary and set $E_\alpha(f,\omega)=E_{\alpha,\omega}(f)$, we get a C^2 function

$$E_\alpha : L_1^{2\alpha}(\Sigma,M)\times\mathcal{T}_g\to\mathbb{R}.$$

Definition. A critical point $f\in L_1^{2\alpha}(\Sigma,M)$ for $E_{\alpha,\omega}$ is called an (α,ω)-*harmonic map*, or sometimes an α-harmonic map. If $(f,\omega)\in L_1^{2\alpha}(\Sigma,M)\times\mathcal{T}$ is a critical point for E_α, f is called a (parametrized) α-*minimal surface*.

In complete analogy with ω-harmonic maps, we can calculate the first variation formula for (α,ω)-harmonic maps. Indeed, suppose that Σ is a compact Riemann surface. By the same argument that led to (3.29), we obtain the formula

$$dE_{\alpha,\omega}(f)(V) = -\int_\Sigma\left[\left\langle\frac{D}{\partial u}\left(\gamma^2\frac{\partial f}{\partial u}\right),V\right\rangle+\left\langle\frac{D}{\partial v}\left(\gamma^2\frac{\partial f}{\partial v}\right),V\right\rangle\right]dudv,$$

where $\gamma^2=(1+|df|^2)^{\alpha-1}$. In the integrand, (u,v) can be any choice of conformal coordinates. This leads to the Euler-Lagrange equation

$$\frac{D}{\partial u}\left(\gamma^2\frac{\partial f}{\partial u}\right)+\frac{D}{\partial v}\left(\gamma^2\frac{\partial f}{\partial v}\right)=0. \quad (3.34)$$

This equation can also be written in terms of the standard Laplace operator acting on f,

$$\frac{D}{\partial u}\left(\frac{\partial f}{\partial u}\right)+\frac{D}{\partial v}\left(\frac{\partial f}{\partial v}\right) \\ = -(\alpha-1)\left[\frac{\partial}{\partial u}(\log(1+|df|^2))\frac{\partial f}{\partial u}+\frac{\partial}{\partial v}(\log(1+|df|^2))\frac{\partial f}{\partial v}\right]. \quad (3.35)$$

Our next goal is to show that the function $E_{\alpha,\omega}:L_1^{2\alpha}(\Sigma,M)\to\mathbb{R}$ satisfies condition C. Recall that condition C asserts that if $\{f_i\}$ is a sequence of points in $L_1^{2\alpha}(\Sigma,M)$ such that

$$E_{\alpha,\omega}(f_i)\le E_0 \qquad \text{where } E_0 \text{ is some constant, and} \qquad \|dE_{\alpha,\omega}(f_i)\|\to 0,$$

then $\{f_i\}$ possesses a subsequence which converges in $L_1^{2\alpha}(\Sigma,M)$ to a critical point for $E_{\alpha,\omega}$.

Theorem 3.5.2. *If $\alpha > 1$, the function $E_{\alpha,\omega} : L_1^{2\alpha}(\Sigma, M) \to \mathbb{R}$ satisfies condition C.*

The proof is very similar to the one given before for the action integral in the theory of smooth closed geodesics. If V is tangent to $L_1^{2\alpha}(\Sigma, M)$,

$$dE_{\alpha,\omega}(f)(V) = 2\alpha \int_\Sigma (1 + |df|^2)^{\alpha-1} \left[\left\langle \frac{DV}{\partial u}, \frac{\partial f}{\partial u} \right\rangle + \left\langle \frac{DV}{\partial v}, \frac{\partial f}{\partial v} \right\rangle \right] dudv,$$

or equivalently,

$$dE_{\alpha,\omega}(f)(V) = -2\alpha \int_\Sigma \left\langle \frac{D}{\partial u}\left(\gamma^2 \frac{\partial f}{\partial u}\right) + \frac{D}{\partial v}\left(\gamma^2 \frac{\partial f}{\partial v}\right), V \right\rangle dudv,$$

for $V \in T_f(L_1^{2\alpha}(\Sigma, M))$, where $\gamma^2 = (1 + |df|^2)^{\alpha-1}$ and D denotes covariant derivative. More generally, we can consider an element $V \in T_f(L_1^{2\alpha}(\Sigma, \mathbb{R}^N))$ and project it into the tangent space to $L_1^{2\alpha}(\Sigma, M)$. Starting with the orthogonal projection $P : i^*T\mathbb{R}^N \to TM$ into the tangent space, we use the ω-Lemma to construct a smooth map

$$\omega_P : L_1^{2\alpha}(\Sigma, i^*T\mathbb{R}^N) \longrightarrow L_1^{2\alpha}(\Sigma, TM).$$

Lemma 3.5.3. *If $V \in T_f(L_1^{2\alpha}(\Sigma, i^*T\mathbb{R}^N)$,*

$$dE_{\alpha,\omega}(f)(\omega_P(V)) = \alpha \int_\Sigma (1 + |df|^2)^{\alpha-1} \left[\frac{\partial V}{\partial u} \cdot \frac{\partial f}{\partial u} + \frac{\partial V}{\partial v} \cdot \frac{\partial f}{\partial v} \right.$$
$$\left. - V \cdot \left(\alpha \left(\frac{\partial f}{\partial u}, \frac{\partial f}{\partial u} \right) + \alpha \left(\frac{\partial f}{\partial v}, \frac{\partial f}{\partial v} \right) \right) \right] dudv,$$

where $\alpha : TM \times TM \to NM$ is the second fundamental form of M in $\mathbb{R}^N$.

In the proof, we can suppose that f and V are C^∞, since the C^∞ maps are dense in $L_1^{2\alpha}$. We write $V = \omega_P(V) + V^\perp$, where $\omega_P(V)$ and $V^\perp$ are the components of V which are tangent and normal to M. Then

$$\left(\frac{\partial}{\partial u}\left(\gamma^2 \frac{\partial f}{\partial u}\right) \right) \cdot V = \left(\frac{\partial}{\partial u}\left(\gamma^2 \frac{\partial f}{\partial u}\right) \right) \cdot \omega_P(V) + \left(\frac{\partial}{\partial u}\left(\gamma^2 \frac{\partial f}{\partial u}\right) \right) \cdot V^\perp$$
$$= \left(\frac{D}{\partial x}\left(\gamma^2 \frac{\partial f}{\partial x}\right) \right) \cdot \omega_P(V) + \gamma^2 \alpha \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial x} \right) \cdot V^\perp,$$

and similarly

$$\left(\frac{\partial}{\partial v}\left(\gamma^2 \frac{\partial f}{\partial v}\right) \right) \cdot V = \left(\frac{D}{\partial v}\left(\gamma^2 \frac{\partial f}{\partial v}\right) \right) \cdot \omega_P(V) + \gamma^2 \alpha \left(\frac{\partial f}{\partial v}, \frac{\partial f}{\partial v} \right) \cdot V^\perp,$$

where the dot denotes the dot product in the ambient Euclidean space $\mathbb{R}^N$. It

follows that

$$dE_{\alpha,\omega}(f)(\omega_P(V)) = -2\alpha \int_\Sigma \left[\left(\frac{\partial}{\partial u}\left(\gamma^2 \frac{\partial f}{\partial u}\right) + \frac{\partial}{\partial v}\left(\gamma^2 \frac{\partial f}{\partial v}\right)\right) \cdot V + \gamma^2 \left(\alpha \left(\frac{\partial f}{\partial u}, \frac{\partial f}{\partial u}\right) + \alpha \left(\frac{\partial f}{\partial v}, \frac{\partial f}{\partial v}\right)\right) \cdot V^\perp \right] dudu.$$

An integration by parts now yields the statement of the Lemma.

Lemma 3.5.4. *There is a constant C depending only on M such that*

$$\|\omega_P(V)\|_{L_1^{2\alpha}} \leq C(1 + E_{\alpha,\omega}(f))\|V\|_{L_1^{2\alpha}} \quad \text{for } V \in T_f(L_1^{2\alpha}(\Sigma, i^*T\mathbb{R}^N).$$

This is a straightforward consequence of applying the Leibniz rule to the equation

$$(\omega_P(V))(p) = P_{f(p)}V(p),$$

where $P_{f(p)}$ is the projection from $\mathbb{R}^N$ into the tangent space $T_{f(p)}M$, to obtain

$$\frac{\partial}{\partial u}(\omega_P(V))(p) = \left(\frac{\partial P \circ f}{\partial u}\right)(V(p)) + P_{f(p)}\left(\frac{\partial V}{\partial u}(p)\right),$$
$$\frac{\partial}{\partial v}(\omega_P(V))(p) = \left(\frac{P \circ \partial f}{\partial v}\right)(V(p)) + P_{f(p)}\left(\frac{\partial V}{\partial v}(p)\right).$$

Thus that after possibly replacing the usual $L_1^{2\alpha}$ norm with an equivalent norm, we can write

$$\|\omega_P(V)\|_{L_1^{2\alpha}} = \|\omega_P(V)\|_{L^{2\alpha}} + \|d\omega_P(V)\|_{L^{2\alpha}},$$

but

$$\|d\omega_P(V)\|_{L^{2\alpha}} = \left\| \left(\frac{\partial P \circ f}{\partial u}\right)(V(p))du + P_{f(p)}\left(\frac{\partial V}{\partial u}(p)\right)du \right\|_{L^{2\alpha}} + \left\| \left(\frac{\partial P \circ f}{\partial v}\right)(V(p))dv + P_{f(p)}\left(\frac{\partial V}{\partial v}(p)\right)dv \right\|_{L^{2\alpha}} \leq C_1 \sup\{\|(V(p))\| : p \in \Sigma\} \|df\|_{L^{2\alpha}} + C_2 \|DV\|_{L^{2\alpha}},$$

where C_1 and C_2 are positive constants. Thus we see that

$$\|\omega_P(V)\|_{L_1^{2\alpha}} \leq C_3 \left[\|V\|_{L_1^{2\alpha}} E_{\alpha,\omega}(f) + C_4 \|DV\|_{L^{2\alpha}} + C_5 \|V\|_{L^{2\alpha}} \right],$$

where C_3, C_4 and C_5 are yet other positive constants and we have used the fact that the C^0-norm is less than some constant times the $L_1^{2\alpha}$-norm. The last estimate yields the statement of the lemma.

Let us turn now to the proof of Theorem 3.5.2 itself. Let $\{f_i\}$ be a sequence from $L_1^{2\alpha}(\Sigma, M) \subset L_1^{2\alpha}(\Sigma, \mathbb{R}^N)$ such that $E_{\alpha,\omega}(f_i)$ is bounded and $\|dE_{\alpha,\omega}(f_i)\| \to$

0. By the Sobolev Lemma 1.4.4 each $f_i \in C^0(\Sigma, M)$. Moreover, since M is compact, $\{f_i\}$ is uniformly bounded, and it follows from the Hölder estimate in Lemma 1.4.4 that $\{f_i\}$ is equicontinuous. Therefore, the Arzela-Ascoli Theorem from real analysis implies that a subsequence of $\{f_i\}$ converges uniformly to a continuous map $f_\infty : \Sigma \to M$.

Since $E_{\alpha,\omega}(f_i)$ is bounded and f_i is bounded in C^0, it follows that f_i is bounded in $L_1^{2\alpha}(\Sigma, \mathbb{R}^N)$. It then follows from Lemma 3.5.3 that $\omega_P(f_i - f_j)$ is bounded in $L_1^{2\alpha}(\Sigma, \mathbb{R}^N)$. Since $\|dE_{\alpha,\omega}(f_i)\| \to 0$,

$$|dE_{\alpha,\omega}(f_i)(\omega_P(f_i - f_j)) - dE_{\alpha,\omega}(f_j)(\omega_P(f_i - f_j))| \to 0 \qquad \text{as} \qquad i, j \to \infty.$$

Using Lemma 3.5.2, we can rewrite this as

$$\begin{aligned}
\Bigg| 2\alpha \int_\Sigma (1 + |df_i|^2)^{\alpha-1} \Bigg[\frac{\partial f_i}{\partial u} \cdot \left(\frac{\partial f_i}{\partial u} - \frac{\partial f_j}{\partial u} \right) + \frac{\partial f_i}{\partial v} \cdot \left(\frac{\partial f_i}{\partial v} - \frac{\partial f_j}{\partial v} \right) \\
- \left(\alpha \left(\frac{\partial f_i}{\partial u}, \frac{\partial f_i}{\partial u} \right) + \alpha \left(\frac{\partial f_i}{\partial v}, \frac{\partial f_i}{\partial v} \right) \right) \cdot (f_i - f_j) \Bigg] du dv \\
-2\alpha \int_\Sigma (1 + |df_j|^2)^{\alpha-1} \Bigg[\frac{\partial f_j}{\partial u} \cdot \left(\frac{\partial f_i}{\partial u} - \frac{\partial f_j}{\partial u} \right) + \frac{\partial f_j}{\partial v} \cdot \left(\frac{\partial f_i}{\partial v} - \frac{\partial f_j}{\partial v} \right) \\
- \left(\alpha \left(\frac{\partial f_j}{\partial u}, \frac{\partial f_j}{\partial u} \right) + \alpha \left(\frac{\partial f_j}{\partial v}, \frac{\partial f_j}{\partial v} \right) \right) \cdot (f_i - f_j) \Bigg] du dv \Bigg| \to 0
\end{aligned}$$

as i and j approach infinity. Note that since energy is bounded,

$$\begin{aligned}
&\left| \int_\Sigma \alpha \left(\frac{\partial f_i}{\partial u}, \frac{\partial f_i}{\partial u} \right) \cdot (f_i - f_j) du dv \right| \leq (\text{constant}) \sup |f_i - f_j| \to 0 \\
&\quad \text{and} \quad \left| \int_\Sigma \alpha \left(\frac{\partial f_i}{\partial v}, \frac{\partial f_i}{\partial v} \right) \cdot (f_i - f_j) du dv \right| \leq (\text{constant}) \sup |f_i - f_j| \to 0
\end{aligned}$$

as i and j approach infinity, and hence

$$\begin{aligned}
\Bigg| \int_\Sigma (1 + |df_i|^2)^{\alpha-1} \langle df_i, df_i - df_j \rangle dA \\
- \int_\Sigma (1 + |df_j|^2)^{\alpha-1} \langle df_j, df_i - df_j \rangle dA \Bigg| \to 0
\end{aligned}$$

as i and j approach infinity.

To proceed further, we need:

Lemma 3.5.5. *If $\alpha > 1$, there is a constant $c \geq (1/16)$ such that*

$$(|v|^{2\alpha-2} v - |w|^{2\alpha-2} w) \cdot (v - w) \geq c|v - w|^{2\alpha}, \quad \text{for all } v, w \in \mathbb{R}^{2N+1}.$$

Proof: It suffices to establish this inequality when both v and w are nonzero. If we set $f(v) = |v|^{2\alpha}$, then f is C^2 on $\mathbb{R}^{2N+1} - \{\mathbf{0}\}$, and a direct calculation shows that

$$Df(v)(w) = 2\alpha|v|^{2\alpha-2}v \cdot w,$$
$$D^2f(v)(w,w) = 4\alpha(\alpha-1)|v|^{2\alpha-4}(v\cdot w)^2 + 2\alpha|v|^{2\alpha-2}w\cdot w \geq 2\alpha|v|^{2\alpha-2}w\cdot w.$$

Hence

$$\begin{aligned} Df(v)(v-w) - Df(w)(v-w) &= \int_0^1 D^2f(w+t(v-w))(v-w,v-w)dt \\ &\geq 2\alpha\int_0^1 |w+t(v-w)|^{2\alpha-2}|v-w|^2dt \geq c|v-w|^{2\alpha}, \end{aligned}$$

for some constant $c > 0$. To see that c can be chosen to be larger than $1/16$, note that by the triangle inequality, either

$$|w| \geq \frac{1}{2}|v-w| \quad \Rightarrow \quad |w+t(v-w)| \geq \frac{1}{2}|w| \geq \frac{1}{4}|v-w| \quad \text{for } t \in [0,1/4],$$

or

$$|v| \geq \frac{1}{2}|v-w| \quad \Rightarrow \quad |v+(1-t)(w-v)| \geq \frac{1}{2}|v| \geq \frac{1}{4}|v-w| \quad \text{for } t \in [3/4,1].$$

This proves the lemma.

To apply Lemma 3.5.5, we set

$$v = \left(1, \frac{\partial f_i}{\partial u}, \frac{\partial f_i}{\partial v}\right), \qquad w = \left(1, \frac{\partial f_j}{\partial u}, \frac{\partial f_j}{\partial v}\right).$$

Then Lemma 3.5.5 implies that

$$\int_\Sigma |df_i - df_j|^{2\alpha}dA \to 0 \qquad \text{as} \qquad i,j \to \infty.$$

Since $f_i \to f_\infty$ in C^0, it follows that $\{f_i\}$ is a Cauchy sequence in $L_1^{2\alpha}(\Sigma, \mathbb{R}^N)$.

By completeness of the Banach space $L_1^{2\alpha}(\Sigma, \mathbb{R}^N)$, there exists an element $\tilde{f}_\infty \in L_1^{2\alpha}(\Sigma, \mathbb{R}^N)$ such that $f_i \to \tilde{f}_\infty$ in $L_1^{2\alpha}(\Sigma, \mathbb{R}^N)$. Clearly, $\tilde{f}_\infty = f_\infty \in L_1^{2\alpha}(\Sigma, M)$, and Theorem 3.5.2 is proven.

Of course, there are versions of Theorem 3.5.2 which hold for perturbations of the α-energy. For example, if $\alpha > 1$ and $\psi \in L_k^2(\Sigma, \mathbb{R}^N)$ for $k \geq 2$, the perturbed function

$$E_{\alpha,\psi,\omega} : L_1^{2\alpha}(\Sigma, M) \to \mathbb{R}, \quad \text{defined by} \quad E_{\alpha,\psi,\omega}(f) = E_{\alpha,\omega}(f) + \int_\Sigma f\cdot\psi dA$$

satisfies condition C. In fact, even more generally, we could consider the function $E^\beta_{\alpha,\psi,\omega} : L_1^{2\alpha}(\Sigma, M) \to \mathbb{R}$ defined by

$$E^\beta_{\alpha,\psi,\omega}(f) = \frac{1}{2}\int_\Sigma (\beta^2 + |df|^2)^\alpha dA + \int_\Sigma f\cdot\psi dA. \tag{3.36}$$

Theorem 3.5.5. *If $\alpha > 1$ and $\beta \in [0,1]$, then the function $E^{\beta}_{\alpha,\psi,\omega}$ defined by (3.36) satisfies condition C.*

The proof is a straightforward modification of that given for Theorem 3.5.2. Note that the argument works even if $\beta = 0$, in which case the critical points are not necessarily smooth.

Thus we can use Corollary 1.11.3 to show that any component of $L^{2\alpha}_1(\Sigma, M)$ contains an element f which minimizes $E_{\alpha,\omega}$ or more generally $E^{\beta}_{\alpha,\psi,\omega}$. Moreover, we can use Theorem 1.11.2 to prove existence of minimax critical points for these functions corresponding to various algebraic topology constraints. We note that

$$E(f) \leq E^0_\alpha(f) \leq E^\beta_\alpha(f) - \frac{\beta^{2\alpha}}{2}, \tag{3.37}$$

the second of these inequalities following from the fact that

$$\psi(t) = (\beta^2 + t^2)^\alpha - (\beta^{2\alpha} + t^{2\alpha}), \quad \text{then} \quad \psi'(t) > 0, \quad \text{for } t > 0,$$

an immediate consequence of differentiating ψ with respect to t.

Of course, our goal is to use these existence results to obtain existence of ω-harmonic maps in the limit as the various perturbations are turned off. Thus, for example, we could let $\beta \to 0$ first, obtaining a function that has simpler behavior under rescaling, and then let $\alpha \to 1$. Condition C is only lost when taking the second limit.

3.6 Regularity of (α, ω)-harmonic maps

At some point, it becomes useful to know that the (α, ω)-harmonic maps constructed by means of Condition C are actually smooth maps. Indeed, if the ambient manifold M is C^∞, so is every (α, ω)-harmonic map into M. This is the content of the following theorem:

Theorem 3.6.1. *If $1 < \alpha < 3/2$, any critical point $f \in L^{2\alpha}_1(\Sigma, M)$ for $E_{\alpha,\omega}$ is smooth. Moreover, if $\psi \in L^2_k(\Sigma, \mathbb{R}^N)$, for some $k \geq 3$, and $\beta \in (0,1]$, then any critical point $f \in L^{2\alpha}_1(\Sigma, M)$ for $E^{\beta}_{\alpha,\psi,\omega}$ is L^2_{k+2}.*

The proof given by Sacks and Uhlenbeck [68] relies on results from Morrey [54], a classic book which contains proofs of many such regularity theorems. A complete proof of this result goes beyond the scope of these notes. However, we hope the following outline will be helpful in giving an idea as to what is involved. Some readers may wish to skip the following sketch, which makes use of the Hölder spaces described in §5.1 of [19].

Following [68], we divide the proof into three steps: First we show that the critical point f is in L^2_2, then in a Hölder space $C^{1,\beta}$, and finally, we use the Schauder theory [25] together with elliptic bootstrapping to prove that f is C^∞. We will deal only with $E_{\alpha,\omega}$, the modification necessary for $E^{\beta}_{\alpha,\psi,\omega}$ being relatively straightforward.

Step 1. We show that f lies in $L^2_2(\Sigma, M)$. To do this, we can use regularity results for variational problems presented in Evans [19], §8.3.

Regularity is a local condition, and we need only show that a critical point f is regular near a given point $p \in \Sigma$. Let U and V be small open neighborhoods of p with $\bar{V} \subset U$. Our strategy is to work in local coordinates $(x^1, \ldots, x^n)$ defined on a neighborhood of $F(U)$ in M and a local conformal parameter (u^1, u^2) on $U \subset \Sigma$. The coordinate system on M allows us to regard $f|U$ as an $\mathbb{R}^n$-valued map. We abbreviate the composition $x^i \circ f$ to x^i, and let $\eta = \sum \eta_{ab} du^a du^b$ and $g = \sum g_{ij} dx^i dx^j$ denote the Riemannian metrics on Σ and M respectively. Let

$$p^i_a = \frac{\partial x^i}{\partial u^a} \quad \text{so} \qquad |df|^2 = \sum_{a,b=1}^{2} \sum_{i,j=1}^{n} g_{ij}(x^1, \ldots x^n) \eta^{ab} p^i_a p^j_b.$$

Then

$$E_{\alpha,\omega}(f|U) = \frac{1}{2} \int_U \mathcal{L}(p, x, u) du^1 du^2,$$
$$\text{where} \quad \mathcal{L}(p, x, u) = (1 + |df|^2)^\alpha \sqrt{\det(\eta_{ab})}.$$

The fact that $f|V$ is (α, ω)-harmonic is expressed by the Euler-Lagrange condition

$$\int_U \left[\sum_{a,i} \frac{\partial \mathcal{L}}{\partial p^i_a} \frac{\partial}{\partial u^a} (\zeta^2 v^i) + \sum_i \frac{\partial \mathcal{L}}{\partial x^i} \zeta^2 v^i \right] du^1 du^2 = 0, \tag{3.38}$$

for every smooth test function $v = (v^1, \ldots, v^n)$, $\zeta : \Sigma \to [0, 1]$ being a suitable smooth cutoff which is one on V and zero outside U.

Note that the first derivatives of the Lagrangian $\mathcal{L}$ are given by

$$\frac{\partial \mathcal{L}}{\partial p^i_a} = \alpha (1 + |df|^2)^{\alpha - 1} \frac{\partial}{\partial p^i_a} (|df|^2),$$

where

$$\frac{\partial}{\partial p^i_a} (|df|^2) = \sum_{b=1}^{2} \sum_{j=1}^{n} g_{ij}(x^1, \ldots x^n) \eta^{ab} p^j_b \sqrt{\det(\eta_{ab})}.$$

Similarly, the second derivatives of $\mathcal{L}$ are given by

$$\frac{\partial^2 \mathcal{L}}{\partial p^i_a \partial p^j_b} = \alpha (1 + |df|^2)^{\alpha - 1} \frac{\partial^2}{\partial p^i_a \partial p^j_b} (|df|^2)$$
$$+ \alpha(\alpha - 1)(1 + |df|^2)^{\alpha - 2} \frac{\partial}{\partial p^i_a} (|df|^2) \frac{\partial}{\partial p^j_b} (|df|^2),$$

where

$$\frac{\partial^2}{\partial p^i_a \partial p^j_b} (|df|^2) = g_{ij}(x^1, \ldots x^n) \eta^{ab} \sqrt{\det(\eta_{ab})} \frac{\partial^2}{\partial p^i_a \partial x^k} (|df|^2).$$

Note that by scaling the coordinates appropriately, we can make $|\partial g_{ij}/\partial x^k|$ and $|\partial^2 g_{ij}/\partial x^k \partial x^l|$ less than ϵ for any given $\epsilon > 0$. As a consequence we find that the Euler-Lagrange operator for the (α, ω)-energy is *uniformly elliptic*:

$$\sum_{a,b=1}^{2} \sum_{i,j=1}^{n} \frac{\partial^2 \mathcal{L}}{\partial p_a^i \partial p_b^j} \xi_a^i \xi_b^j \geq \alpha \gamma^2 \sum_{a,b=1}^{2} \sum_{i,j=1}^{n} g_{ij}(x^1, \ldots x^n) \eta^{ab} \sqrt{\det(\eta_{ab})} \xi_a^i \xi_b^j,$$

where $\gamma^2 = (1 + |df|^2)^{\alpha - 1}$. It is exactly this estimate on uniform ellipticity that allows us to apply the difference quotient approach described in Evans [19] to show that $u|V$ is in L_2^2. Since Σ can be covered by finitely many neighborhoods V, it follows that $f \in L_2^2(\Sigma, M)$, finishing Step 1.

We remark that it is uniform ellipticity that fails for the function $E^{\beta}_{\alpha,\psi,\omega}$ when $\beta = 0$ and this explains why critical points of this function are not necessariy smooth.

Step 2. For the second step, we recall that equation (3.35) for α-harmonic maps can be written in the form

$$(\Delta f)^{\top} = -(\alpha - 1) d(\log(1 + |df|^2)) \cdot df,$$
$$\text{where} \quad (\Delta f)^{\top} = \frac{Df}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right) + \frac{Df}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right),$$

D denoting the Levi-Civita connection on M. Since we know by Step 1 that f is L_2^2, we are justified in differentiating on the right-hand side, thereby obtaining the following result after a short calculation:

$$\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} = \alpha(f)(df, df) - (\alpha - 1)\frac{B(d^2 f, df)}{1 + |df|^2} df,$$

where $\alpha(f)$ and B are bilinear maps, $\alpha(f)$ being the familiar second fundamental form of M in $\mathbb{R}^N$. We can put the last term on the left-hand side obtaining $L(f) = \alpha(f)(df, df)$, where

$$L(u) = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + (\alpha - 1)\frac{B(d^2 u, df)}{1 + |df|^2} df,$$

L being a second-order differential operator with coefficients in L^{∞}. Moreover, since $\alpha - 1 < 1/2$, the operator L is uniformly elliptic. The linear operator L defines a bounded linear map from L_2^4 to L^4, which can be regarded as a small perturbation of the scalar Laplace operator when α is close to one.

If we restrict L to a small disk D about a given point p and $u \in \mathrm{Ker}(L)$, each component of u will assume its maximum value on the boundary ∂D of D. Thus if we impose Dirichlet boundary conditions that u and f agree on ∂D, the map $L : L_2^4 \to L^4$ will be injective. On the other hand, $\Delta : L_2^4 \to L^4$ has a continuous inverse $G_0 : L^4 \to L_2^4$, and we can think of G_0 as an approximate inverse to L when $\alpha - 1$ is small. Given $h \in L^4$, then for appropriate choices of norms on L_2^4 and L^4, the map $u \mapsto G_0(Lu - h) + u$ is a contraction, and it

must therefore possess a unique fixed point. This implies that L is surjective, and the open mapping theorem implies that the inverse G to L is continuous.

Since the critical point f is continuous and $df \in L^2$, $\alpha(f)(df, df)$ is in L^4, and there is a unique $u \in L^4_2(D, M)$ satisfying the Dirichlet boundary conditions such that $L(u) = \alpha(f)(df, df)$. This u must of course be f and hence $f \in L^4_2$. We thus conclude that the restriction of f to a small neighborhood of any point lies in L^4_2. Since L^4_2 is contained in the Hölder space $C^{1,\beta}$ for suitable β, we conclude that $f \subset C^{1,\beta}(\Sigma, M)$.

Step 3. The final step is via the technique of "elliptic bootstrapping" to the differential equation

$$L(f) = \alpha(f)(df, df), \tag{3.39}$$

using the Schauder theory. The necessary Schauder estimates are described in [25], Chapter 8 for scalar operators, the general case being a standard extension within PDE theory. The result is that:

$$f \in C^{1,\beta} \;\Rightarrow\; \text{right side of (3.39)} \in C^{0,\beta} \;\Rightarrow\; f \in C^{2,\beta}$$
$$\Rightarrow\; \text{right side of (3.39)} \in C^{1,\beta} \;\Rightarrow\; f \in C^{3,\beta} \;\Rightarrow \cdots .$$

This shows that f is C^∞ and completes our sketch of the argument for the theorem.

3.7 Morse theory for the perturbed energy

Once we have condition C, we can apply Liusternik-Schnirelmann theory to the functions

$$E_{\alpha,\psi,\omega}(f) = \frac{1}{2}\int_\Sigma (1 + |df|^2)^\alpha dA + \int_\Sigma f \cdot \psi dA,$$

but we would like to establish Morse inequalities and define a Morse-Witten complex when the critical points are Morse nondegenerate. This requires an extension of Morse theory to certain functions on Banach manifolds, and such a theory was in fact developed by Uhlenbeck [80]. In this section, we present the results of that theory, specialized to the perturbations of the energy we have been studying.

Note that if f is a critical point for $E^\beta_{\alpha,\psi,\omega}$, then a calculation similar to that which yields (3.32) shows that

$$dE_{\alpha,\psi,\omega}(f)(V) = \alpha \int_\Sigma (1 + |df|^2)^{\alpha-1} \langle df, DV\rangle dA + \int_\Sigma \psi \cdot V,$$

for all $V \in T_f L^{2\alpha}_1(\Sigma, M)$, where DV denotes the covariant differential of V with respect to the Levi-Civita connection on M.

We would also like a formula for the second derivative such as that given in Proposition 2.4.1. In order to state this formula, we define a linear operator

$\mathcal{K} : T_f L_1^{2\alpha}(\Sigma, M) \to T_f L_1^{2\alpha}(\Sigma, M)$ in terms of the Riemann curvature R of M by the formula

$$\langle \mathcal{K}(V), W \rangle dA = \left\langle R\left(V, \frac{\partial f}{\partial u}\right)\frac{\partial f}{\partial u} + R\left(V, \frac{\partial f}{\partial v}\right)\frac{\partial f}{\partial v}, W \right\rangle du dv,$$

where (u, v) are isothermal coordinates on M.

Proposition 3.7.1. *If f is a critical point for $E_{\alpha,\psi,\omega}$, then*

$$\begin{aligned} d^2 E_{\alpha,\psi,\omega}(f)(V, W) = \alpha \int_\Sigma (1 + |df|^2)^{\alpha-1} [\langle DV, DW \rangle - \langle \mathcal{K}(V), W \rangle] dA \\ + 2\alpha(\alpha - 1) \int_\Sigma (1 + |df|^2)^{\alpha-2} \langle df, DV \rangle \langle df, DW \rangle dA \\ + \int_\Sigma \psi \cdot \alpha(V, W) dA, \quad (3.40) \end{aligned}$$

for all $V, W \in T_f L_1^{2\alpha}(\Sigma, M)$, where α is the second fundamental form of M in $\mathbb{R}^N$.

Although (3.40) may look complicated at first, note that it specializes to (3.40) in the case where the Riemannian manifold M is just Euclidean space and $\psi = 0$. Moreover, we can set $\alpha = 1$, and obtain the formula for the Hessian of the ordinary energy E_ω at a critical point:

Corollary 3.7.2. *If f is ω-harmonic, then*

$$d^2 E_\omega(f)(V, W) = \int_\Sigma [\langle DV, DW \rangle - \langle \mathcal{K}(V), W \rangle] dA, \quad (3.41)$$

for all $V, W \in T_f L_1^{2\alpha}(\Sigma, M)$.

The proof of Proposition 3.7.1 is quite similar to the proof of Proposition 2.4.1. For simplicity, we assume that $\psi = 0$. We consider a variation of f which has its support within a given coordinate chart $(U, (x, y))$ on Σ. Recall that such a variation is a smooth family of maps $t \mapsto f(t)$ in $L_k^2(\Sigma, M)$ with $f(0) = f$, and let

$$\alpha(x, y, t) = f(t)(x, y), \qquad V(x, y) = \frac{\partial \alpha}{\partial t}(x, y, 0) \in T_{f(x,y)} M.$$

Setting $\gamma(t)^2 = (1 + |df(t)|^2)^{\alpha-1}$, we obtain

$$\begin{aligned} d^2 E_{\alpha,\omega}(f)(V, V) &= \frac{d^2}{dt^2}(E_\omega(f(t))\Big|_{t=0} \\ &= \alpha \int_\Sigma \frac{\partial}{\partial t}\left[\gamma^2 \left\langle \frac{\partial \alpha}{\partial x}, \frac{D}{\partial t}\frac{\partial \alpha}{\partial x} \right\rangle + \gamma^2 \left\langle \frac{\partial \alpha}{\partial y}, \frac{D}{\partial t}\frac{\partial \alpha}{\partial y} \right\rangle\right] dx dy \Big|_{t=0} \end{aligned}$$

where D as usual denotes the covariant derivative of the Levi-Civita connection on the ambient Riemannian manifold M. We can rewrite this as

$$\begin{aligned} d^2E_{\alpha,\omega}(f)(V,V) = \alpha \int_\Sigma \gamma^2 \Bigg[&\left\langle \frac{D}{\partial t}\frac{\partial \alpha}{\partial x}, \frac{D}{\partial t}\frac{\partial \alpha}{\partial x} \right\rangle + \left\langle \frac{D}{\partial t}\frac{\partial \alpha}{\partial y}, \frac{D}{\partial t}\frac{\partial \alpha}{\partial y} \right\rangle \\ &+ \left\langle \frac{\partial \alpha}{\partial x}, \frac{D^2}{\partial t^2}\frac{\partial \alpha}{\partial x} \right\rangle + \left\langle \frac{\partial \alpha}{\partial y}, \frac{D^2}{\partial t^2}\frac{\partial \alpha}{\partial y} \right\rangle \Bigg] dxdy \Bigg|_{t=0} \\ +2\alpha(\alpha-1) \int_\Sigma \frac{d\gamma^2}{dt} &\left[\left\langle \frac{\partial \alpha}{\partial x}, \frac{D}{\partial t}\frac{\partial \alpha}{\partial x} \right\rangle + \left\langle \frac{\partial \alpha}{\partial y}, \frac{D}{\partial t}\frac{\partial \alpha}{\partial y} \right\rangle \right] dxdy \Bigg|_{t=0}. \end{aligned} \tag{3.42}$$

When we evaluate at $t = 0$, we find that

$$\frac{d\gamma^2}{dt} = \frac{\langle df, DV \rangle}{1 + |df^2}},$$

where V is the variation field, so we can rewrite (3.42) as

$$\begin{aligned} d^2E_{\alpha,\omega}(f)V,V) = \alpha \int_\Sigma \gamma^2 \Bigg[&\left\langle \frac{DV}{\partial x}, \frac{DV}{\partial x} \right\rangle + \left\langle \frac{DV}{\partial y}, \frac{DV}{\partial y} \right\rangle \\ &+ \left\langle \frac{\partial f}{\partial x}, \frac{D}{\partial t}\frac{DV}{\partial x} \right\rangle + \left\langle \frac{\partial f}{\partial y}, \frac{D}{\partial t}\frac{DV}{\partial y} \right\rangle \Bigg] dxdy \\ &+ 2\alpha(\alpha-1) \int_\Sigma (1 + |df|^2)^{\alpha-2} \langle df, DV \rangle^2 dA. \end{aligned}$$

Finally, applying the definition of the Riemann-Christoffel curvature tensor R, we obtain

$$\begin{aligned} d^2E_{\alpha,\omega}(f)(V,V) = \int_\Sigma \gamma^2 \Bigg[&\left\langle \frac{DV}{\partial x}, \frac{DV}{\partial x} \right\rangle + \left\langle \frac{DV}{\partial y}, \frac{DV}{\partial y} \right\rangle \\ &+ \left\langle \frac{\partial f}{\partial x}, R\left(V, \frac{\partial f}{\partial x}\right)V \right\rangle + \left\langle \frac{\partial f}{\partial y}, R\left(V, \frac{\partial f}{\partial y}\right)V \right\rangle \\ &+ \left\langle \frac{\partial f}{\partial x}, \frac{D}{\partial x}\frac{DV}{\partial t} \right\rangle + \left\langle \frac{\partial f}{\partial y}, \frac{D}{\partial y}\frac{DV}{\partial t} \right\rangle \Bigg] dxdy \\ &+ 2\alpha(\alpha-1) \int_\Sigma (1 + |df|^2)^{\alpha-2} \langle df, DV \rangle^2 dA. \end{aligned}$$

An integration by parts and use of the Euler-Lagrange equation eliminates the third term, thereby yielding (3.40) in the case where $\psi = 0$.

If f is a critical point for the function $E_{\alpha,\psi,\omega}$, we can define a second-order linear partial-differential operator

$$L : L_1^2(f^*TM) \to L_{-1}^2(f^*TM) \quad \text{by} \quad d^2E_{\alpha,\psi,\omega}(f)(V,W) = \int_\Sigma \langle L(V), W \rangle dA,$$

and we call L the Jacobi operator. The Jacobi operator L restricts to a continuous linear map $L : L_k^2(f^*TM) \to L_{k-2}^2(f^*TM)$ for all $k \geq 2$, and if

$\iota : L^2_{k-2}(f^*TM) \to L^2_k(f^*TM)$ is the inclusion and $\lambda \in \mathbb{C}$, we have an associated operator $L_\lambda = L - \lambda\iota$. The fundamental Theorem 2.4.2 on formally self-adjoint elliptic operators now applies to the Jacobi operator L, and hence L is a Fredholm operator, which satisfies the additional conditions:

1. for each $\lambda \in \mathbb{C}$ the eigenspace $W_\lambda = \mathrm{Ker}(L_\lambda)$ is finite-dimensional.
2. all the elements of W_λ are C^∞,
3. if W_λ is empty, then L_λ possesses a Green's operator inverse,
4. if W_λ is nonempty, that is, λ is an eigenvalue, then $\lambda \in \mathbb{R}$, and
5. the eigenvalues can be arranged in a sequence

$$\lambda_1 < \lambda_2 < \cdots < \lambda_i < \cdots \qquad \text{with } \lambda_i \to \infty,$$

and only finitely many of the eigenvalues are negative.

On the other hand, if f is a critical point for $E_{\alpha,\psi,\omega}$, we can also define an inner product

$$\langle\langle \cdot, \cdot \rangle\rangle : T_f L_1^{2\alpha}(\Sigma, M) \times T_f L_1^{2\alpha}(\Sigma, M) \longrightarrow \mathbb{R}$$

by the somewhat complicated formula,

$$\langle\langle V, W \rangle\rangle = \alpha \int_\Sigma (1 + |df|^2)^{\alpha-1} [\langle DV, DW \rangle + \langle V, W \rangle] dA \\ + 2\alpha(\alpha - 1) \int_\Sigma (1 + |df|^2)^{\alpha-2} \langle df, DV \rangle \langle df, DW \rangle dA,$$

which is useful because it so close to (3.41). Since f is a smooth function, this is actually equivalent to the simpler L^2_1 inner product $\langle\langle \cdot, \cdot \rangle\rangle_0$ defined by

$$\langle\langle V, W \rangle\rangle_0 = \int_\Sigma [\langle DV, DW \rangle + \langle V, W \rangle] dA.$$

We can then define a second-order linear partial-differential operator

$$J : L^2_1(f^*TM) \to L^2_{-1}(f^*TM) \quad \text{by} \quad \langle\langle V, W \rangle\rangle = \int_\Sigma \langle J(V), W \rangle dA.$$

Once again J restricts to a Fredholm operator $J : L^2_k(f^*TM) \to L^2_{k-2}(f^*TM)$ for all $k \geq 2$. However, since $\langle\langle \cdot, \cdot \rangle\rangle$ is positive-definite, the eigenvalues of J are all positive and J itself has a Green's operator inverse. It is here that the complicated formula for the L^2_1-inner product on $L^2_1(f^*TM)$ pays off, because $L - J$ is of zero order, and hence $L - J : L^2_k(f^*TM) \to L^2_{k-2}(f^*TM)$ is a compact operator, for all $k \geq 1$.

In parallel with the theory of smooth closed geodesics, we can define a continuous linear map

$$A : L_1^2(f^*TM) \to L_1^2(f^*TM) \quad \text{by} \quad \langle\langle AV, W\rangle\rangle = d^2E_{\alpha,\psi,\omega}(f)(V, W). \quad (3.43)$$

Then the operator

$$A = J^{-1} \circ L : L_1^2(f^*TM) \to L_1^2(f^*TM)$$

is Fredholm, and restricts to a Fredholm operator on each $L_k^2(f^*TM)$, for $k \geq 2$. By the L^p theory for elliptic operators, it also restricts to a Fredholm operator on each $L_k^p(f^*TM)$, for $p \geq 2$ and $k \geq 2$.

But if we let $p = 2\alpha$ and choose q so that $(1/p)+(1/q) = 1$, then we can also define the operator A directly on $L_1^p(f^*TM)$ by (E:explicitformulaA). Indeed, any $V \in L_1^p(f^*TM)$ defines a continuous linear functional

$$W \mapsto d^2E'_{\alpha,\omega}(f)(V, W), \quad \text{for} \quad W \in L_1^q(f^*TM),$$

and by the duality between L_1^p and L_1^q this linear functional defines an element $AV \in L_1^p(f^*TM)$. This defines a continuous linear map

$$A : T_f L_1^{2\alpha}(\Sigma, M) \longrightarrow T_f L_1^{2\alpha}(\Sigma, M).$$

such that

$$A - \mathrm{id} : T_f L_1^{2\alpha}(\Sigma, M) \longrightarrow T_f L_1^{2\alpha}(\Sigma, M)$$

is a compact operator, and thus A is a Fredholm operator as before.

Definition. We say that the critical point f for $E_{\alpha,\psi,\omega}$ is *Morse nondegenerate* if A is an isomorphism. The *Morse index* of f is the sum of the dimensions of the eigenspaces of L for negative eigenvalues. Finally, $E_{\alpha,\psi,\omega} : L_1^{2\alpha}(\Sigma, M) \to \mathbb{R}$ is a *Morse function* if all of its critical points are Morse nondegenerate.

As in the theory of smooth closed geodesics, we can choose ψ so that $E_{\alpha,\psi,\omega}$ is a Morse function:

Theorem 3.7.3. *For a residual set of $\psi \in L_k^2(\Sigma, \mathbb{R}^N)$, the function $E_{\alpha,\psi,\omega}$ is a Morse function.*

The argument is virtually identical to the argument we presented for Theorem 2.7.1.

As in the theory of smooth closed geodesics, we can set

$$M_0 = \{f \in L_1^{2\alpha}(\Sigma, M) : \ f \text{ is constant } \},$$

and then M_0 is a nondegenerate critical submanifold for $E_{\alpha,\omega}$ of Morse index zero. Just as as at the end of §2.7, we can perturb $E_{\alpha,\omega}$ to a function $E'_{\alpha,\omega} : L_1^{2\alpha}(\Sigma, M) \to \mathbb{R}$ such that

1. $E'_{\alpha,\omega} = E_{\alpha,\omega}$ on a neighborhood U of M_0,

2. $E'_{\alpha,\omega} = E_{\alpha,\psi,\omega}$ outside a larger neighborhood V of M_0, for some $\psi \in L^2_k(\Sigma, \mathbb{R}^N)$, and

3. all critical points of $E'_{\alpha,\omega}$ either lie within M_0 or are Morse nondegenerate and do not lie in V.

The Handle Addition Theorem 2.9.1 can now be modified so that it applies to the Morse function $E'_{\alpha,\omega}$, and hence the Morse inequalities of §2.10 can be extended to $E'_{\alpha,\omega}$. Let $\mathcal{M} = L^{2\alpha}_1(\Sigma, M)$ with its standard Finsler metric, and let

$$\mathcal{M}^a = \{f \in L^{2\alpha}_1(\Sigma, M) : E'_{\alpha,\omega}(f) \leq a\}.$$

Theorem 3.7.4. *If the interval $[a, b]$ contains a single critical value c for $E'_{\alpha,\omega}$, there is exactly one critical point p for $E'_{\alpha,\omega}$ such that $E'_{\alpha,\omega}(p) = c$ and this critical point is Morse nondegenerate of Morse index λ, then $\mathcal{M}^b$ is homotopy equivalent to $\mathcal{M}^a$ with a handle of index λ attached.*

Indeed, the proof presented in §2.9 was designed precisely so that it would apply to this case. The main difference is that now the model space E is a Banach space instead of a Hilbert space. and one must make some small modifications to the argument to account for this.

Moreover, the notion of gradient-like vector field presented in §2.9 was designed so that it would apply directly to the function $E'_{\alpha,\omega} : L^{2\alpha}_1(\Sigma, M) \to \mathbb{R}$. Therefore, we can define stable and unstable manifolds of nondegenerate critical points just as we did before.

Suppose that $E'_{\alpha,\omega}$ is a perturbation of $E_{\alpha,\omega}$ which agrees with $E_{\alpha,\omega}$ in a neighborhood of M_0 and satisfies the condition that all its critical points are either in M_0 or are Morse nondegenerate. If $\mathcal{X}$ is a gradient-like vector field for $E'_{\alpha,\omega}$, we let $C_k(E'_{\alpha,\omega}, \mathcal{X})$ denote the free $\mathbb{Z}$-module generated by the critical points $f_{k,1}, f_{k,2} \ldots$ for $E'_{\alpha,\omega}$ of index k and let ∂ be the $\mathbb{Z}$-module homomorphism

$$\partial : C_k(E'_{\alpha,\omega}, \mathcal{X}) \longrightarrow C_{k-1}(E'_{\alpha,\omega}, \mathcal{X})$$

defined by

$$\partial(f_{k,j}) = \sum_q a_{jq} f_{k-1,q}, \tag{3.44}$$

where $a_{jq} \in \mathbb{Z}$ is the oriented number of trajectories from $f_{k,j}$ to $f_{k-1,q}$, the orientation being determined as in §2.11.

Theorem 3.7.5. *The $\mathbb{Z}$-module homomorphisms thus defined satisfy the identity $\partial \circ \partial = 0$ and the resulting Morse-Witten complex*

$$\cdots \to C_{k+1}(E'_{\alpha,\omega}, \mathcal{X}) \to C_k(E'_{\alpha,\omega}, \mathcal{X}) \to C_{k-1}(E'_{\alpha,\omega}, \mathcal{X}) \to \cdots$$

calculates the homology of the pair $(C^0(\Sigma, M), M_0)$.

The proof is a straightforward modification of the argument presented in §2.11.

3.8 Local control of energy density

Our next goal is to investigate the limit of α-energy critical points as $\alpha \to 1$. The key technique here is the Bochner Lemma, which gives a uniform C^1 estimate on harmonic and α-harmonic maps whose total energy is small (less than some positive constant ϵ_0). This in turn is crucial in understanding a key feature of harmonic and α-harmonic maps, the phenomenon of bubbling. Throughout this section, we use the notation

$$e(f) = \frac{1}{2}|df|^2 = \frac{1}{2}\left(\left|\frac{\partial f}{\partial u}\right|^2 + \left|\frac{\partial f}{\partial v}\right|^2\right) dudv,$$

for isothermal parameters (u, v) on Σ.

Lemma 3.8.1. *Suppose that $f : D_1 \to M$ is a harmonic map, where D_1 is the unit disk in the complex plane, with the standard Euclidean metric ds^2 and M is a compact Riemannian manifold with sectional curvatures satisfying the inequality $K(\sigma) \leq 1$. Then for any $r_0 > 0$,*

$$\int_{D_1} e(f)dA < \pi(r_0^2 - 8r_0^4) \qquad \Rightarrow \qquad \max_{\sigma\in(0,1]} \sigma^2 \sup_{D_{1-\sigma}} e(f) < 4r_0^2. \tag{3.45}$$

Proof: Choose $\sigma_0 \in (0, 1]$ so that

$$\sigma_0^2 \sup_{D_{1-\sigma_0}} e(f) \geq \sigma^2 \sup_{D_{1-\sigma}} e(f), \qquad \text{for all } \sigma \in (0, 1],$$

and choose $p_0 \in \overline{D_{1-\sigma_0}}$ so that

$$e_0 = e(f)(p_0) = \sup\{e(f)(p) : p \in \overline{D_{1-\sigma_0}}\}.$$

If $\sigma_0^2 e_0 < 4r_0^2$, the lemma is proven. So it suffices to assume that $\sigma_0^2 e_0 \geq 4r_0^2$, and derive a contradiction. But then

$$\left(\frac{\sigma_0}{2}\right)^2 \sup_{D_{1-\sigma_0/2}} e(f) \leq \sigma_0^2 \sup_{D_{1-\sigma_0}} e(f) = \sigma_0^2 e_0 \quad \Rightarrow \quad \sup_{D_{1-\sigma_0/2}} e(f) \leq 4e_0.$$

Moreover, $\sigma_0^2 e_0 \geq 4r_0^2$ implies that

$$\frac{r_0}{\sqrt{e_0}} \leq \frac{\sigma_0}{2},$$

so we can define a new harmonic map $g : D_{r_0} \to M$ by

$$g(q) = f\left(p_0 + \frac{q}{\sqrt{e_0}}\right) \quad \text{such that} \quad e(g)(0) = 1 \quad \text{and} \quad e(g) \leq 4.$$

It now follows from (3.28), which follows from the Bochner Lemma, that

$$\Delta(e(g)) \geq -2(e(g))^2 \geq -32.$$

Thus if we define a smooth function $h : D_{r_0} \to \mathbb{R}$ in terms of the standard coordinates (u, v) by

$$h = e(g) - [1 - 16(u^2 + v^2)],$$

then we find that

$$\Delta h = \Delta(e(g)) + 32 \geq 0,$$

and hence by the maximum principle,

$$\begin{aligned} 0 = h(0) \leq \int_{D_{r_0}} h dA &\leq \int_{D_{r_0}} [e(g) - (1 - 16(u^2 + v^2))] dA \\ &\leq \int_{D_{r_0}} e(g) dA - \int_0^{2\pi} \int_0^{r_0} (1 - 16r^2) r dr d\theta, \end{aligned}$$

and an integration yields the inequality

$$\int_{D_1} e(f) dA > \pi(r_0^2 - 8r_0^4).$$

Thus we have proven the contrapositive of the assertion in the lemma.

Corollary 3.8.2. *Suppose that $f : D_1 \to M$ is a harmonic map. Then*

$$\int_{D_1} e(f) dA < \frac{\pi}{16} \quad \Rightarrow \quad \max_{\sigma \in (0,1]} \sigma^2 \sup_{D_{1-\sigma}} e(f) < \frac{8}{\pi} \int_{D_1} e(f) dA.$$

To prove this we note that if $r_0 \leq (1/4)$,

$$\begin{aligned} \int_{D_1} e(f) dA = \frac{\pi}{2}(r_0^2) \leq \pi(r_0^2 - 8r_0^4) \\ \Rightarrow \quad \max_{\sigma \in (0,1]} \sigma^2 \sup_{D_{1-\sigma}} e(f) < 4r_0^2 = \frac{8}{\pi} \int_{D_1} e(f) dA. \end{aligned}$$

Corollary 3.8.3. *Suppose that D_r is the disk of radius r in the complex plane and $f : D_r \to M$ is a harmonic map. Then*

$$\int_{D_r} e(f) dA < \frac{\pi}{16} \quad \Rightarrow \quad \max_{\sigma \in (0,r]} \sigma^2 \sup_{D_{r-\sigma}} e(f) < \frac{8}{\pi} \int_{D_r} e(f) dA.$$

To prove this, simply note that under rescaling from D_1 to D_r, the energy integral remains unchanged, while

$$|df| \mapsto \frac{1}{r}|df|, \quad e(f) \mapsto \frac{1}{r^2} e(f)$$

and

$$\max_{\sigma \in (0,1]} \sigma^2 \sup_{D_{1-\sigma}} e(f) \mapsto \max_{\sigma \in (0,1]} \left(\frac{\sigma}{r}\right)^2 \sup_{D_{r-r\sigma}} e(f) = \max_{\sigma \in (0,r]} \tau^2 \sup_{D_{r-\tau}} e(f).$$

Of course, we would like versions of the previous lemmas for the α-energy $E_{\alpha,\omega}$ when α is sufficiently close to one.

Lemma 3.8.4. *Suppose that M is a compact Riemannian manifold whose sectional curvatures satisfy the bound $K(\sigma) \leq 1$. There exists $\alpha_0 \in (1,\infty)$ and for each $\alpha \in [1,\alpha_0]$ a homogeneous second-order elliptic operator L_α such that such that the coefficients of L_α tend uniformly to those of Δ as $\alpha \to 1$, and*

$$L_\alpha(e(f)) \geq [2\tilde{K}e(f) - 2e(f)^2],$$

where $\tilde{K}$ is the Gaussian curvature of the Riemannian metric on Σ.

Proof: It follows from (3.21) and the discussion leading to (3.28) that

$$\Delta_\eta(e(f)) = |\nabla df|^2 + \langle d[(\Delta_\eta f)^\top], df\rangle + \tilde{2}Ke(f) - 2(e(f))^2.$$

Recall that equation (3.35) for α-harmonic maps can be written in the form

$$(\Delta_\eta f)^\top = -(\alpha-1)d(\log\gamma^2)\cdot df, \qquad \text{where} \qquad \gamma^2 = 1+|df|^2,$$

Since

$$d(\log\gamma^2) = \frac{d(|df|^2)}{1+|df|^2} = \frac{2\langle\nabla df, df\rangle}{1+|df|^2},$$

$$\nabla d(\log\gamma^2) = \frac{\nabla d(|df|^2)(1+|df|^2) + (d(|df|^2))^2}{(1+|df|^2)^2} = \frac{\nabla d(|df|^2)}{(1+|df|^2)} + \frac{4\langle\nabla df, df\rangle^2}{(1+|df|^2)^2},$$

we conclude that there is a homogeneous second-order differential operator Q (with uniformly bounded coefficients which depend on df) such that

$$|\langle\nabla(\Delta f), df\rangle| \leq (\alpha-1)Q(e(f)) + (\alpha-1)(\text{constant})\frac{\langle\nabla df, df\rangle^2}{(1+|df|^2)^2},$$

the constant being independent of f and α. The second term can be absorbed in $|\nabla df|^2$ when α is sufficiently close to one, while the first term can be absorbed in the elliptic operator L_α.

One can now use Lemma 3.8.4 in place of the standard Bochner Lemma to prove a version of Lemma 3.8.3 for α-harmonic maps.

Lemma 3.8.5. *Suppose that M is a compact Riemannian manifold whose sectional curvatures satisfy the bound $K(\sigma) \leq 1$ and that D_r is the disk of radius r in the complex plane, with the standard Euclidean metric ds^2. There exists an $\alpha_0 > 1$ and an $\epsilon_0 > 0$ with the following property: If $f : D_r \to M$ is an α-harmonic map,*

$$\int_{D_r} e(f)dA < \epsilon_0 \qquad \Rightarrow \qquad \max_{\sigma\in(0,r]} \sigma^2 \sup_{D_{r-\sigma}} e(f) < 3\int_{D_r} e(f)dA. \tag{3.46}$$

The proof is a straightforward modification of the proof of Lemmas 3.8.1 to 3.8.3, in which we use Lemma 3.8.4 for the elliptic operator L_α instead of the ordinary Bochner Lemma for the Laplace operator Δ_η.

We can next relax the condition that the Riemannian metric on Σ be flat and rescale the metric on M so that it satisfies the condition $K(\Sigma) \leq K_0$ for $K_0 > 0$:

Lemma 3.8.6. *Suppose that M is a compact Riemannian manifold whose sectional curvatures satisfy the bound $K(\sigma) \leq K_0$ where $K_0 > 0$ and that D_r is the disk of radius r in the complex plane, a Riemannian metric ds^2 which is related to the flat metric ds_0^2 by inequalities of the form*

$$\frac{1}{L^2} ds_0^2 \leq ds^2 \leq L^2 ds_0^2,$$

L being a constant. There exists an $\alpha_0 > 1$ and constants $\epsilon_0 > 0$ and C_L depending on L such that: If $f : D_r \to M$ is an α-harmonic map with $\alpha \in (1, \alpha_0]$,

$$\int_{D_r} e(f) dA < \frac{\epsilon_0}{\sqrt{K_0}} \qquad \Rightarrow \qquad \max_{\sigma \in (0,r]} \sigma^2 \sup_{D_{r-\sigma}} e(f) < C_L \int_{D_r} e(f) dA. \quad (3.47)$$

Proof: The estimates (3.46) holds for the energy density $e_0(f)$ with respect to the flat metric ds_0^2. Now simply use the inequalities

$$\frac{1}{L^2} e_0(f) \leq e(f) \leq L^2 e_0(f)$$

to obtain (3.47) for $K_0 = 1$. Now note that multiplying the metric by $\sqrt{K_0}$ also multiplies the energy density by the same factor.

Inequality (3.47) is typically used as follows: Let Σ be a compact Riemann surface and M a compact Riemannian manifold, $r_0 > 0$ a fixed radius. Then there are constants $\epsilon_0 > 0$, $C > 0$ and $\alpha \in [1, \alpha_0]$ which depends upon Σ, M and r_0 such that if D_r is a disk of radius $r \leq r_0$ in Σ and $f : D_r \to M$ is an α-harmonic map,

$$\int_{D_r} e(f) dA < \epsilon_0 \qquad \Rightarrow \qquad \sup_{D_{r/2}} e(f) < \frac{C}{r^2} \int_{D_r} e(f) dA. \quad (3.48)$$

In other words, if the energy on a ball of small radius is sufficiently small, it gives a bound on the energy density itself on a ball of half the radius.

Theorem 3.8.7. *Suppose that M is a compact Riemannian manifold and Σ is a closed connected Riemann surface. If $\{\omega_m\}$ is a sequence of conformal structures $\in \mathcal{T}$ such that $\omega_m \to \omega_\infty \in \mathcal{T}$ and $\{f_m : \Sigma \to M\}$ is a sequence of (α_m, ω_m)-harmonic maps such that $\alpha_m \to 1$ and $E(f_m) \leq E_0$ for some constant $E_0 > 0$, then there is a finite collection of points*

$$\{p_1, \ldots, p_l\} \subseteq \Sigma,$$

and a subsequence of $\{f_m\}$ (which we still denote by $\{f_m\}$), such that $\{f_m\}$ converges uniformly in C^k on compact subsets of $\Sigma - \{p_1, \ldots, p_l\}$, for any non-negative integer k, to an ω_∞-harmonic map

$$f_\infty : \Sigma - \{p_1, \ldots, p_l\} \longrightarrow M.$$

Proof: We begin by choosing $\epsilon_0 > 0$ so that when $r \leq r_0$, (3.48) implies that

$$\int_{D_r} e(f) dA < \epsilon_0 \qquad \Rightarrow \qquad \sup_{D_{r/2}} e(f) < \frac{C\epsilon_0}{r^2}. \tag{3.49}$$

Let $r_m = (1/2)^m$, for each positive integer m and let $\mathcal{O}_m$ be an open cover of Σ by disks $D_{r_m}(p_i)$, so that each each point of Σ is covered by at most h disks and the disks $D_{r_m/2}(p_i)$ still cover M. (We can choose h to be independent of m.) Then

$$\sum_i \int_{D_{r_m}(p_i)} e(f_m) \leq hE_0.$$

Hence there are at most hE_0/ϵ_0 of the disks in the cover $\mathcal{O}_m$ on which

$$\int_{D_{r_m}(x_i)} e(f_m) \geq \epsilon_0.$$

For each j we let $\{p_{1m}, \ldots, p_{lm}\}$ be the center points of these disks. After possibly passing to a subsequence, we can arrange that

$$p_{1m} \to p_1, \ldots, p_{lm} \to p_l.$$

Then the $\{f_m\}$'s are uniformly bounded and equicontinuous on compact subsets of $\Sigma - \{p_1, \ldots, p_l\}$. By Arzela's Theorem the $\{f_m\}$'s converge to a continuous function f_∞ on $\Sigma - \{p_1, \ldots, p_l\}$, and the convergence is uniform on compact subsets.

To finish the proof, we use the process of "elliptic bootstrapping" to show that $f_m \to f_\infty$ in C^k on $\Sigma - \{p_1, \ldots, p_l\}$, uniformly on compact subsets, for any $k \geq 0$. Indeed, if M is isometrically imbedded in E^N, the equation for which f must satisfy to be an α-harmonic map is

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = -\alpha(f)(df, df) - (\alpha - 1)\frac{B(d^2 f, df)}{1 + |df|^2} df, \tag{3.50}$$

where $\alpha(f)$ and B are bilinear maps, $\alpha(f)$ being the familiar second fundamental form. It follows from (3.49) that $\{df_m\}$ is bounded in L^p for all p on any disk within $\Sigma - \{p_1, \ldots p_l\}$. It then follows from multiplication theorems for Sobolev spaces that the right-hand side of (3.50) is bounded in L^p for all p on any such disk, and hence elliptic estimates give that f is bounded in L^p_2 for all p. But now one of the Sobolev imbedding theorems implies that a subsequence of $\{f_m\}$

converges in $C^{1,\beta}$ for some $\beta > 0$ on any compact subset of $\Sigma - \{p_1, \dots p_l\}$. Now we can apply Schauder estimates to see that a subsequence of $\{f_m\}$ converges in $C^{2,\beta}$, then in $C^{3,\beta}$, and so forth. Finally we obtain a subsequence of $\{f_m\}$ which converges uniformly in every C^k on compact subsets of $\Sigma - \{p_1, \dots, p_l\}$, and the theorem is proven.

Theorem 3.8.8. *Suppose that M is a compact Riemannian manifold with nonpositive sectional curvatures. If ω is a conformal structure on the compact Riemann surface Σ, there is an ω-energy minimizing harmonic map in any free homotopy class $[\Sigma, M]$.*

In this case, Lemma 3.8.6 implies that for some $\alpha_0 \in (1, \infty)$ and some $r_0 \geq 0$, whenever D_r is a disk of radius $r \leq r_0$ in Σ and $f : D_r \to M$ is an α-harmonic map with $\alpha \in (1, \alpha_0]$,

$$\sup_{D_{r/2}} e(f) < 4C_L \int_{D_r} e(f) dA. \tag{3.51}$$

Thus we do not need the complicated covering argument utilized in the proof of Theorem 3.8.7.

Instead, we simply take a sequence $\alpha_m \to 1$ and for each α_m choose an α_m-harmonic map $f_m : \Sigma \to M$ which lies in a given component of $[\Sigma, M]$ and minimizes the α_m-energy on that component. Then (3.51) implies that $e(f_m)$ is bounded on any ball $D_{r/2}$, which implies that $\{f_m\}$ converges uniformly on compact subsets. Elliptic bootstrapping then implies that the sequence $\{f_m\}$ converges uniformly in all C^k, proving the theorem.

Theorem 3.8.8 is just the Eells-Sampson Theorem 3.1.1 in the special case where Σ has dimension two. Similarly, we can prove Hartman's Theorem 3.1.2 in the special case where Σ is a surface:

Theorem 3.8.9. *Suppose that M is a compact Riemannian manifold with negative sectional curvatures. If (Σ, ω) is any Riemann surface, there is at most one ω-harmonic map in any free homotopy class $[\Sigma, M]$.*

Indeed, the corresponding statement for (α, ω)-harmonic maps follows from the fact that since

$$\left\langle R\left(V, \frac{\partial f}{\partial u}\right)\frac{\partial f}{\partial u} + R\left(V, \frac{\partial f}{\partial v}\right)\frac{\partial f}{\partial v}, V \right\rangle \leq 0,$$

any (α, ω)-harmonic map is stable by (3.40). If there were two (α, ω)-harmonic maps in the same component of $C^0(\Sigma, M)$, there would have to be an unstable critical point by the Morse inequalities, which cannot happen.

To prove the corresponding statement for the limiting case of ω-harmonic maps, we can use the implicit function theorem. For this, we define a smooth map

$$E^\star_\omega : L^2_k(\Sigma, M) \times [1, \alpha_0) \longrightarrow \mathbb{R} \quad \text{by} \quad E^\star_\omega(f, \alpha) = E_{\alpha,\omega}(f),$$

and observe that at $\alpha = 1$ we obtain the usual energy up to a constant, $E^\star_\omega(f,1) = E_\omega(f) + (1/2)$. We can differentiate to obtain the Euler-Lagrange map

$$F^*_\omega : L^2_k(\Sigma, M) \times [1, \alpha_0) \longrightarrow L^2_{k-2}(\Sigma, TM).$$

If f is an (α,ω)-harmonic map, we can compose the differential DF^*_ω of F^*_ω at (f,α) with the projection onto the vertical obtaining the corresponding Jacobi operator,

$$L = \pi_V \circ DF^*_\omega(f,\alpha) : T_f L^2_k(\Sigma, M) \oplus \mathbb{R} \longrightarrow T_f L^2_{k-2}(\Sigma, M).$$

Note that L is a Fredholm operator with Fredholm index one, and that f is a Morse nondegenerate critical point for $E_{\alpha,\omega}$ if and only if L is surjective.

Suppose now that $f : \Sigma \to M$ is an ω-conformal harmonic map, and note that f must be Morse nondegenerate by Corollary 3.7.2. It follows from the implicit function theorem that if U is a sufficiently small C^2 neighborhood of $(f,1)$ in $L^2_k(\Sigma, M) \times [1, \alpha_0)$, then

$$(F^*_\omega)^{-1}(\text{zero section of } T(\text{Map}(\Sigma, M)))$$

is a smooth submanifold of U of dimension one. This submanifold consists of critical points of $E_{\alpha,\omega}$ for α varying in some interval $[1, \alpha_0)$ for which the corresponding Jacobi operators L are surjective. All the points of this submanifold are Morse nondegenerate critical points for $E_{\alpha,\omega}$.

Thus two distinct ω-harmonic maps $f, g : \Sigma \to M$ in the same component of $C^0(\Sigma, M)$ would give rise to two distinct one-parameter families f_α, g_α of (α,ω)-harmonic maps in the same component of $C^0(\Sigma, M)$, contradicting the uniqueness of (α,ω)-harmonic maps. This finishes the proof of Theorem 3.8.9.

3.9 Bubbling

In §3.7 we have shown existence of many α-energy critical points, corresponding to numerous topological constraints, while in §3.8 we have shown that given a sequence $\{f_m : \Sigma \to M\}$ of α_m-harmonic maps with $\alpha_m > 1$ such that $\alpha_m \to 1$ and $E(f_m) \leq E_0$ for some constant $E_0 > 0$, there is a collection of points $\{p_1, \ldots, p_l\}$ and a subsequence of $\{f_m\}$ (which we still denote by $\{f_m\}$) which converges uniformly on compact subsets of $\Sigma - \{p_1, \ldots, p_l\}$, together with the first k derivatives to a map

$$f_\infty : \Sigma - \{p_1, \ldots, p_l\} \longrightarrow M, \tag{3.52}$$

which is harmonic on $\Sigma - \{p_1, \ldots, p_l\}$. The following theorem implies that f_∞ extends to a harmonic map defined on the entire Riemann surface Σ:

Removeable Singularity Theorem 3.9.1. *Let D be the unit disk in the complex plane $\mathbb{C}$. If*

$$f : D - \{0\} \longrightarrow M$$

is a harmonic map of finite energy, then f extends to a smooth harmonic map on the entire disk D.

We apologize to the reader for not giving a proof of this theorem in these notes. It is proven as Theorem 3.6 in [68].

Returning now to our sequence $\{f_m\}$ of α_m-harmonic maps with $\alpha_m \to 1$ and $E(f_m) \leq E_0$, we note that as measures, the sequence of Radon measures $\{e(f_m)dA\}$ converges weakly to

$$e(f_\infty)dA + \sum_{i=1^l} m_i \delta_{p_i} dA,$$

where m_i is a nonnegative constant and δ_{p_i} is the Dirac delta-function located at p_i. If m_i is zero, we can throw the point p_i away, since in this case Lemma 3.8.6 implies that after possibly passing to a subsequence $\{f_m\}$ will converge on a neighborhood or p_i. If $m_i > 0$ is nonzero, we call p_i a *bubble point* of the sequence and m_i measures the amount of energy lost at the bubble point.

We next investigate what happens at the bubble points $p_1, \ldots, p_l$. For each m, choose $q_{im} \in \overline{D_{r_m}(p_{im})}$ so that

$$e(f_m)(q_{im}) = \sup\{e(f_m)(q) : q \in \overline{D_{r_m}(p_{im})}\}.$$

Note that $q_{im} \to x_i$ as j approaches ∞, and let

$$b_{im} = \sqrt{e(f_m)(q_{im})}.$$

If no subsequence of the f_{im}'s converge near p_i then $b_{im} \to \infty$. Assume that the latter alternative holds, and define

$$\tilde{f}_m : D_{r_m b_{im}}(0) \to M \qquad \text{by} \qquad \tilde{f}_m(q) = f_m\left(q_{im} + \frac{p}{b_{im}}\right).$$

One readily verifies that $e(f_{im}) \leq 1$ and $e(f_{im})(0) = 1$. Therefore a subsequence of the f_m's converges to a nonconstant harmonic map

$$\tilde{f}_\infty : \mathbb{C} - \{q_1, \ldots q_m\} \longrightarrow M,$$

the convergence being uniform in C^k for all k on compact subsets of $\mathbb{C} - \{q_1, \ldots q_m\}$. At the finitely many points $\{q_1, \ldots q_m\}$, further bubbling can occur.

Recall that as a Riemann surface the complex plane is just the Riemann sphere S^2 minus a point. Therefore the removable singularity theorem implies that $\tilde{f}_\infty$ extends to a smooth harmonic map $\hat{f}_\infty : S^2 \to M$.

It is clearly of interest to understand what α-energy critical points look like near bubble points. This problem has been studied by Parker and Wolfson [64], [63] and Chen and Tian [12], among others.

Remark 3.9.2. Each harmonic two-sphere which bubbles off carries with it a certain minimal amount of energy. Indeed, a harmonic two-sphere $h : S^2 \to M$

is automatically conformal and hence a minimal surface, and from the Gauss equation, its Gaussian curvature $K : S^2 \to \mathbb{R}$ is bounded above by K_0, the maximum value of all sectional curvatures on M. Hence by the Gauss-Bonnet theorem,

$$4\pi = \int_{S^2} K dA \leq K_0(\text{Area of } h(S^2)) \qquad \Rightarrow \qquad E(h) \geq \frac{4\pi}{K_0}.$$

In particular, if the energy of a sequence of α_m-harmonic maps is bounded by E_0, the number of harmonic two-spheres that can bubble off in the limit is $\leq (4\pi E_0)/K_0$.

We can now prove a theorem due to Sacks and Uhlenbeck [68], which can be thought of as the analog within minimal surface theory of the Fet-Lusternik Theorem 2.3.2 on existence of smooth closed geodesics:

Theorem 3.9.3. *If M is a compact simply connected Riemannian manifold, then M contains at least one nonconstant minimal two-sphere.*

Proof: In many ways, the argument is very similar to that given in §2.3. There is a least integer $q \geq 2$ such that $H_q(\tilde{M}; \mathbb{Z}) \neq 0$, and it follows from the Hurewicz theorem that

$$\pi_i(M) = 0, \quad \text{for } 0 < i < q, \quad \text{and } \pi_q(M) \cong H_q(M, \mathbb{Z}) \neq 0.$$

Let $\mathcal{M}$ be the space of smooth maps from S^2 to M and let $\pi : \mathcal{M} \to M$ by $\pi(f) = f(0)$, the evaluation of f at the basepoint $0 \in S^2 = \mathbb{C} \cup \{\infty\}$. It is a well-known fact from homotopy theory that π is a fibration with fiber $\mathcal{M}_p = \pi^{-1}(p)$, the set of basepoint preserving maps from S^2 to M. Moreover,

$$\pi_k(\mathcal{M}_p) \cong \pi_{k+2}(M)$$

and the fibration π induces a long exact sequence

$$\cdots \to \pi_k(\mathcal{M}_p) \to \pi_k(\mathcal{M}) \to \pi_k(M) \to \pi_{k-1}(\mathcal{M}_p) \to \cdots .$$

We note, moreover, that $\pi_* : \pi_k(\mathcal{M}) \to \pi_k(M)$ possesses a right inverse

$$i_* : \pi_k(M) \to \pi_k(\mathcal{M}) \quad \text{induced by the inclusion} \quad i : M \to \mathcal{M}$$

which takes a point to the constant map at that point. Hence the long exact sequence splits and we conclude that

$$\pi_k(\mathcal{M}) \cong \pi_k(M) \oplus \pi_k(\mathcal{M}_p) \cong \pi_k(M) \oplus \pi_{k+2}(M).$$

Thus the homotopy groups of $\mathcal{M}$ are completely determined by the homotopy groups of M. In particular,

$$\pi_{q-2}(\mathcal{M}) \cong \pi_q(M) \neq 0.$$

Since M is simply connected, $q \geq 2$ and $\pi_{q-1}(\mathcal{M}) \cong \pi_q(M)$ is abelian. Moreover, we can identify $\pi_{q-2}(\mathcal{M})$ with $[S^{q-2}, \mathcal{M}]$, the space of free homotopy classes of maps from S^{q-2} into $\mathcal{M}$. Choose a nonzero element $\alpha \in [S^{q-2}, \mathcal{M}]$. Let

$$\mathcal{F} = \{g(S^{q-2}) \text{ such that } g : S^{q-2} \to \mathcal{M} \text{ is a continuous map in } [\alpha]\}.$$

Clearly, $\mathcal{F}$ is an ambient isotopy invariant family of sets. Hence Minimax$(E_\alpha, \mathcal{F})$ is a critical value for E_α.

Recall that M is isometrically imbedded in an ambient Euclidean space $\mathbb{R}^N$ and let $M(\delta)$ denote the open δ-neighborhood of M in $\mathbb{R}^N$ for $\delta > 0$. By the tubular neighborhood theorem, if δ is sufficiently small, M is a strong deformation retract of $M(\delta)$. Moreover, for $\epsilon > 0$ sufficiently small, any map $f : S^2 \to M$ of energy $< \epsilon$ can be contracted to its center of mass in $\mathbb{R}^N$ without leaving $M(\delta)$.

Suppose now that $g(S^{q-1}) \subset J^{-1}([0, \epsilon])$. Then g is homotopic to a smooth map

$$\tilde{g} : S^{q-1} \to M_0, \quad \text{where} \quad M_0 = \{\gamma \in \mathcal{M} : \ \gamma \text{ is constant } \}.$$

Hence Minimax$(E_\alpha, \mathcal{F}) \geq \epsilon$. We now take a decreasing sequence of real numbers $\{\alpha_m\}$ with $\alpha_m \to 1$. The Minimax Theorem 1.11.2 gives a corresponding sequence $\{f_m\}$ of critical points for E_{α_m} which have energy $\geq \epsilon > 0$. Either a subsequence converges to a nonconstant harmonic map or a bubble forms in the limit. In the latter case, the bubble provides a nonconstant harmonic map. In either case we get a nonconstant harmonic map $f_\infty : S^2 \to M$ which is automatically conformal, since there is only one conformal structure on S^2. The image is a nonconstant minimal two-sphere, which proves the theorem.

3.10 Existence of minimizing spheres

Let M be a compact manifold and suppose that $f : D_1 \to M$ is a map with $f(\partial D_1) = p$ which represents an element $[f] \in \pi_2(M, p)$. Suppose, moreover, that $\gamma : [0, 1] \to M$ is a smooth map with $\gamma(0) = p = \gamma(1)$ which represents an element $[\gamma] \in \pi_1(M, p)$. We can then define $\gamma \star f$ as follows: Let (r, θ) be polar coordinates on D_1 and set

$$g(r, \theta) = (\gamma \star f)(r, \theta) = \begin{cases} f(2r, \theta), & \text{for } 0 \leq r \leq 1/2, \\ \gamma(2r - 1), & \text{for } 1/2 \leq r \leq 1. \end{cases}$$

Then $[\gamma], [f] \mapsto [g] = [\gamma \star f]$ gives an action of $\pi_1(M, p)$ on $\pi_2(M, p)$, which makes $\pi_2(M, p)$ into a $\mathbb{Z}[\pi_1(M, p)]$-module, where $\mathbb{Z}[\pi_1(M, p)]$ is the group ring of $\pi_1(M, p)$. This action is discussed in more detail in Chapter 4 of [35].

Theorem 3.10.1 (Sacks and Uhlenbeck). *Suppose that M is a compact Riemannian manifold. Then a generating set for $\pi_2(M, p)$ as a $\mathbb{Z}[\pi_1(M, p)]$-module can be represented by minimal two-spheres, possibly with branch points.*

Note that the minimal two-spheres need not pass through the point p. If M is simply connected, it follows from the Hurewicz Theorem that $\pi_2(M,p) \cong H_2(M;\mathbb{Z})$, and hence we obtain:

Corollary 3.10.2. *If M is a compact simply connected Riemannian manifold, then a set of generators for $H_2(M;\mathbb{Z})$ can be represented by minimal two-spheres.*

Before proving Theorem 3.10.1, we make a few remarks regarding the symmetry of the functions $E, E_\alpha : C^2(S^2, M) \to \mathbb{R}$. We regard S^2 as the one-point compactification $\mathbb{C} \cup \{\infty\}$ of the complex plane $\mathbb{C}$ with the standard coordinate

$$z = x + iy = re^{i\theta} = e^{u+i\theta},$$

and note that E is invariant under all the linear fractional transformations: If $\phi : S^2 \to S^2$ is the diffeomorphism defined by

$$\phi(z) = \frac{az+b}{cz+d}, \quad \text{for} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{C}),$$

then $E(f \circ \phi) = E(f)$. Note that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \Rightarrow \quad f \circ \phi = f,$$

so the group of linear fractional transformations is actually

$$G = PSL(2,\mathbb{C}) = \frac{SL(2,\mathbb{C})}{\pm I}.$$

In contrast, when $\alpha > 1$, the maps $E^0_\alpha, E_\alpha : C^2(S^2, M) \to \mathbb{R}$ defined by

$$E^0_\alpha(f) = \frac{1}{2}\int_\Sigma |df|^{2\alpha} dA \quad \text{and} \quad E_\alpha(f) = \frac{1}{2}\int_\Sigma (1+|df|^2)^\alpha dA.$$

are invariant only under the smaller group of isometries $SO(3) \subseteq PSL(2,\mathbb{C})$. Indeed, critical points of E^0_α must be parametrized so that the "center of mass" is zero:

Lemma 3.10.3. *Let $\mathbf{X} : S^2 \to \mathbb{R}^3$ denote the standard inclusion, and let $\mathbf{0}$ denote the origin in $\mathbb{R}^3$. If $f \in C^2(S^2, M)$ is a critical point for E^0_α with $\alpha \in (1,\infty)$, then*

$$\int_{S^2} \mathbf{X}|df|^{2\alpha} dA = \mathbf{0}.$$

Moreover, there is a smooth function $\psi_\alpha : [0,\infty) \to \mathbb{R}$ such that $\psi_\alpha(0) = 0$, $\psi_\alpha(t), \psi'_\alpha(t) > 0$ for $t > 0$, and if f is a critical point for E_α, then

$$\int_{S^2} \mathbf{X}\psi_\alpha(|df|^2) dA = \mathbf{0}. \tag{3.53}$$

In the argument for this lemma, we utilize the standard metric on S^2 of constant curvature one, expressed in terms of our standard coordinates by

$$ds^2 = \frac{4}{(1+r^2)^2}(dr^2 + r^2 d\theta^2) = \frac{1}{\cosh^2 u}(du^2 + d\theta^2),$$

where $r = e^{-u}$. (This metric is related by scaling to the constant curvature metric of total area one.) For $t \in \mathbb{R}$, we define a linear fractional transformation

$$\phi_t : S^2 \to S^2 \quad \text{by} \quad u \circ \phi_t = u + t, \quad \theta \circ \phi_t = \theta,$$

so that $t \mapsto \phi_t$ is a one-parameter subgroup. The energy density is then given by the formula

$$e(f \circ \phi_t) = \frac{1}{2}\left(\left|\frac{\partial f}{\partial u}\right|^2 + \left|\frac{\partial f}{\partial \theta}\right|^2\right)^2 \cosh^2(u+t),$$

and it is straightforward to calculate its derivative at $t = 0$:

$$\begin{aligned}\frac{d}{dt} e(f \circ \phi_t)\bigg|_{t=0} &= \frac{1}{2}\,\frac{d}{dt}|d(f \circ \phi_t)|^2\bigg|_{t=0} \\ &= \left(\left|\frac{\partial f}{\partial u}\right|^2 + \left|\frac{\partial f}{\partial \theta}\right|^2\right) \sinh u \cosh u = |df|^2 \tanh u. \quad (3.54)\end{aligned}$$

Using this fact, we can perform a straightforward calculation to obtain

$$\frac{d}{dt} E^0_\alpha(f \circ \phi_t)\bigg|_{t=0} = (\alpha - 1)\int_{S^2} |df|^{2\alpha}(\tanh u) dA.$$

In terms of the standard Euclidean coordinates (x, y, z) on $\mathbb{R}^3$, S^2 is represented by the equation $x^2 + y^2 + z^2 = 1$, and a straightforward computation using stereographic projection from the north pole to the (x, y)-plane shows that

$$z = \frac{r^2 - 1}{r^2 + 1} = \frac{\sinh u}{\cosh u} = \tanh u,$$

so if f is a critical point for E^0_α,

$$0 = \frac{d}{dt} E^0_\alpha(f \circ \phi_t)\bigg|_{t=0} = (\alpha - 1)\int_{S^2} |df|^{2\alpha} z dA.$$

But we can take the z-axis to be any line passing through the origin, and hence

$$\int_{S^2} \mathbf{X}|df|^{2\alpha} dA = \mathbf{0},$$

for the metric of constant curvature one, and hence for the constant curvature metric of total area one.

More generally, we can use (3.54) to calculate the derivative of E_α to obtain

$$\begin{aligned}\frac{d}{dt}E_\alpha(f\circ\phi_t)\Big|_{t=0} &= \frac{1}{2}\,\frac{d}{dt}\int_{\S^2}(1+|d(f\circ\phi_t)|^2)^\alpha \mathrm{sech}^2(u+t)dud\theta\Big|_{t=0}\\ &= \alpha\int_{S^2}(1+|df|^2)^{\alpha-1}\tanh u|df|^2dA - \int_{S^2}(1+|df|^2)^\alpha\tanh u dA.\end{aligned}$$

Thus, if f is a critical point for E_α,

$$\begin{aligned}0 &= \int_{S^2}[\alpha(1+|df|^2)^{\alpha-1}|df|^2 - (1+|df|^2)^\alpha]zdA,\\ &= \int_{S^2}[\alpha(1+|df|^2)^{\alpha-1}|df|^2 - (1+|df|^2)^\alpha - 1]zdA,\end{aligned}$$

where we have used the fact that the average value of z on S^2 is zero. We obtain (3.53) by setting

$$\psi_\alpha(t) = \frac{\alpha(1+t)^{\alpha-1}t - (1+t)^\alpha - 1}{\alpha-1} = \int_0^t \alpha\frac{\tau}{(1+\tau)^{2-\alpha}}d\tau. \tag{3.55}$$

One easily verifies that this function has the desired properties, and has a smooth limit as $\alpha\to 1$, namely

$$\psi_1(t) = \int_0^t \frac{\tau}{(1+\tau)}d\tau. \tag{3.56}$$

Remark 3.10.4. It is sometimes convenient to demand that harmonic maps $f: S^2\to M$ satisfy the side condition

$$\int_{S^2}\mathbf{X}\psi_1(|df|^2)dA = \mathbf{0},$$

thereby reducing the symmetry group of the problem from $PSL(2,\mathbb{C})$ to a maximal compact subgroup $SO(3)$.

Lemma 3.10.5. *Suppose that M is a compact Riemannian manifold and $K_0 > 0$ is an upper bound for the sectional curvatures of M and $\alpha_0\in(1,\infty)$ is sufficiently close to one. Then there is an $\epsilon_0>0$ such that every nonconstant α-harmonic map $f: S^2\to M$ satisfies*

$$E_\alpha(f) - \frac{1}{2} \geq E(f) \geq \epsilon_0.$$

We give S^2 the constant curvature metric of total area one, which has Gaussian curvature $\tilde{K} = 2\pi$. Then it follows from Lemma 3.8.4 that there is an elliptic operator L_α such that

$$L_\alpha(e(f)) \geq 2(2\pi - e(f))e(f).$$

Moreover, it follows from Lemma‘3.8.5 that there is an $\epsilon_0 > 0$ such that

$$\alpha \in (1, \alpha_0] \quad \text{and} \quad E(f) < \epsilon_0 \quad \Rightarrow \quad e(f) < 2\pi.$$

This yields a contradiction to the maximum principle for L_α unless $e(f)$ is identically zero.

Lemma 3.10.6. *Suppose that $\{f_m\}$ is a sequence of nonconstant critical points for E_{α_m} with $\alpha_m \to 1$ such that f_m converges to a harmonic map $f_\infty : S^2 \to M$ on $S^2 - \{0\}$, uniformly on compact subset in C^k for all k. Then f_∞ cannot be a map to a point.*

Indeed, it follows from the previous lemma that $E_\alpha(f_m) \geq \epsilon_0$ and hence if f_∞ were constant a nontrivial bubble would form near the point $0 \in S^2$. Thus energy density would have to concentrate at the point 0 and this would contradict (3.53).

Proof of Theorem 3.10.1: The proof is an induction which makes use of a succession of variational problems. Let

$$\mathcal{F}_1 = \{f \in C^2(S^2, M) : \text{ the free homotopy class of } f \text{ is nontrivial } \},$$

and let

$$\mu_1 = \inf\{E(f) : f \in \mathcal{F}_1\}.$$

Suppose that $\alpha_m \to 1$ and let f_m be an element of $\mathcal{F}_1$ which achieves a minimum for E_{α_m} on $\mathcal{F}_1$. Either a subsequence of $\{f_m\}$ converges without bubbling to a minimal two-sphere $f_\infty : S^2 \to M$ which lies within $\mathcal{F}_1$, in which case we have a nonconstant minimal two-sphere $f \in \mathcal{F}_1$ such that $E(f_1) = \mu_1$, or at least one nonconstant minimal minimal two-sphere $h : S^2 \to M$ bubbles off at a point $p_1 \in S^2$. We need to show that the second possibility leads to a contradiction.

Recall that a subsequence of $\{f_m\}$ will converge to a harmonic map $f_\infty : S^2 \to M$, uniformly on compact subsets of $S^2 - \{p_1, \ldots, p_l\}$, where $p_1, \ldots, p_l$ are the bubble points. By Lemma 3.10.6, either f_∞ is nonconstant or there is at least one other bubble point. Let us suppose that f_∞ is nonconstant. (The other case can be treated in a similar fashion.) We will perform surgeries on small circles about p_1 to divide each α_m-harmonic map into a base map $\hat{f}_m : S^2 \to M$ and a bubble $\hat{h}_m : S^2 \to M$.

In more detail, let $D_{r_m}(p_1)$ be a disk of radius r_m chosen so that $r_m \to 0$ as $m \to \infty$,

$$\int_{D_{r_m}(p_1)} e(f_m) dA \geq \epsilon_0 \quad \text{and} \quad f_m(D_{r_m}(p_1) - D_{(1/3)r_m}(p_1)) \subseteq N_\delta(q)$$

where $q = f_\infty(p_1)$ and $N_\delta(q)$ is the domain of a geodesic coordinate system of radius δ in M. Then define a map $\hat{f}_m : S^2 \to M$ by

$$\hat{f}_m(p) = \begin{cases} f_m(p), & \text{for } p \in S^2 - D_{r_m}(p_1), \\ \exp_q(\eta(p)(\exp_q)^{-1} f_m(p)), & \text{for } p \in D_{r_m}(p_1), \end{cases}$$

where $\eta : S^2 \to [0,1]$ is a smooth function such that

$$\eta \equiv \begin{cases} 1 & \text{on } S^2 - D_{(2/3)r_m}(p_1), \\ 0 & \text{on } D_{(1/3)r_m}(p_1), \end{cases}$$

Let $R_m : S^2 \to S^2$ be the reflection through the circle $\partial D_{(1/2)r_m}(p_1)$, and define $\hat{h}_m : S^2 \to M$ by

$$\hat{h}_m(p) = \begin{cases} \exp_q(\eta \circ R_m(p)(\exp_q)^{-1} R_m \circ f_m(p)), & \text{for } p \in S^2 - D_{r_m}(p_1), \\ f_m(p), & \text{for } p \in D_{r_m}(p_1), \end{cases}$$

Thus $\hat{f}_m$ agrees with f_m outside $D_{r_m}(p_1)$ while $R_m \circ \hat{h}_m$ agrees with f_m inside $D_{r_m}(p_1)$.

We can clearly arrange that

$$E_{\alpha_m}(\hat{f}_m) + E_{\alpha_m}(\hat{h}_m) < E_{\alpha_m}(f_m) + \frac{\epsilon_0}{2}$$

if m is sufficiently large, where ϵ_0 is the constant appearing in Lemma 3.10.5. At least one of the maps $\hat{f}_m$ or $\hat{h}_m$ must be homotopically nontrivial with energy less than the infimum over $\mathcal{F}_1$, and this provides the desired contradiction.

Thus we obtain a homotopically nontrivial minimal two-sphere $f_1 : S^2 \to M$ and we let Γ_1 denote the $\mathbb{Z}[\pi_1(M,p)]$-submodule of $\pi_2(M,p)$ generated by $[f_1]$. If $\Gamma_1 \neq \pi_2(M,p)$, we let

$$\mathcal{F}_2 = \{f \in C^2(S^2, M) : \text{ the free homotopy class of } f \text{ is not in } \Gamma_1 \},$$

set $\mu_2 = \inf\{E(f) : f \in \mathcal{F}_2\}$, and proceed exactly as before. One thereby obtains a minimal two-sphere $f_2 : S^2 \to M$ which does not lie in Γ_1. we then let Γ_2 be the $\mathbb{Z}[\pi_1(M,p)]$-submodule of $\pi_2(M,p)$ generated by $[f_1]$ and $[f_2]$, and so forth. For the inductive step, we suppose we have already constructed a $\mathbb{Z}[\pi_1(M,p)]$-submodule Γ_{k-1} of $\pi_2(M,p)$ generated by minimal two-spheres $[f_1], \ldots [f_{k-1}]$. If $\Gamma_{k-1} \neq \pi_2(M,p)$, we let

$$\mathcal{F}_k = \{f \in C^2(S^2, M) : \text{ the free homotopy class of } f \text{ is not in } \Gamma_{k-1} \},$$

let $\mu_k = \inf\{E(f) : f \in \mathcal{F}_k\}$, and verify that there is a nonconstant minimal two-sphere $f_k \in \mathcal{F}_k$ such that $E(f_k) = \mu_k$. Theorem 3.10.1 follows by induction.

The above theorem can be applied to compact three-dimensional Riemannian manifolds. In this case, surfaces which minimize area are free of branch points by a theorem of Osserman and Gulliver (see [58] and [31]). In fact the minimal spheres which generate $\pi_2(M,p)$ are either imbedded or double coverings of projective planes:

Theorem 3.10.7 (Meeks and Yau). *Suppose that M is an oriented compact three-dimensional Riemannian manifold. Then there is a finite collection $\{f_1, \ldots, f_l\}$ of generators for $\pi_2(M,p)$ as a $\mathbb{Z}[\pi_1(M,p)]$-module, each of which*

is either an embedded minimal two-sphere or a doubly covered imbedded projective plane.

We refer the reader to [48] for the proof, which is based upon the tower construction of Papakyriakopoulos.

Thus one can divide a compact orientable three-dimensional manifold M along the homotopically nontrivial minimal two-spheres which separate M, obtaining a connected sum decomposition

$$M = M_0 \sharp M_1 \sharp \cdots \sharp M_k. \tag{3.57}$$

Suppose that an element $\pi_2(M_i)$ is nonzero for some i. Then there exists an imbedded two sphere $N \subset M_i$ and two points close to N but on opposite sides can be connected by an arc C within M_i. A small neighborhood of $N \cup C$ will have boundary diffeomorphic to a separating two-sphere. It then follows from the positive resolution to the Poincaré conjecture that M_i is diffeomorphic to $S^1 \times S^2$. Thus each summand in (3.57) either has vanishing π_2 or is $S^1 \times S^2$.

A compact orientable three-dimensional manifold is said to be *prime* if it cannot be expressed as a nontrivial connected sum. It also follows from the Poincaré conjecture that each factor in the connected sum decomposition (3.57) is prime, and the decomposition must be the prime decomposition for a compact oriented three-manifold M discovered by Kneser and Milnor. The remarkable fact is that the decomposition occurs along imbedded two-spheres which are minimal with respect to any preassigned Riemannian metric.

Example 3.10.8. Here is an explicit example illustrating how much gets lost of the critical point theory for α-harmonic maps as $\alpha \to 1$. Starting with a lens space $L(3,1)$ of constant curvature one, we consider the connected sum $M = L(3,1)\sharp L(3,1)\sharp L(3,1)$ along isolated embedded minimal two-spheres N_1 and N_2 of very small radius of curvature which minimize within their free homotopy classes. From van Kampen's theorem it follows that the fundamental group is generated by elements a, b and c, with the relations $a^3 = b^3 = c^3 = 1$. Explicit construction of the universal cover shows that $\pi_2(M;p)$ is generated as a $\pi_1(M)$-module by the two imbedded minimal spheres

$$f_1 : S^2 \to N_1, \quad f_2 : S^2 \to N_2.$$

We can construct an additional embedded two-sphere $N = N_1 \sharp N_2$ by connecting N_1 with N_2 by a very thin tube and a corresponding imbedding $f : S^2 \to N$ which represents the homotopy class $[f_1] + [f_2]$, which is not freely homotopic to any multiple of $[f_1]$ or $[f_2]$. If α_0 is sufficiently close to one, then for each $\alpha \in (1, \alpha_0]$ there is a minimizing α-minimal two-sphere f_α in the component of $\mathrm{Map}(S^2, M)$ representing this free homotopy class. As $\alpha \to 1$ a subsequence of these α-minimal two-spheres should approach a configuration consisting of N_1, N_2 and a minimal geodesic connecting N_1 and N_2 of some length bounded away from zero.

Although $\pi_2(M,p)$ is generated by two imbedded spheres as a $\mathbb{Z}[\pi_1(M,p)]$-module, it is interesting to note that as an abelian group, $\pi_2(M,p)$ has infinitely

many generators, as one sees by noting that it is isomorphic to the second homotopy group of its universal cover.

3.11 Existence of minimal tori

Suppose we want to investigate the topology of a compact oriented three-dimensional manifold. We start by using $\pi_2(M)$ to express M as a direct sum decomposition (3.57) in which each of the summands is prime. One can show that each prime summand is one of three types: it is diffeomorphic to $S^1 \times S^2$, it is finitely covered by S^3, or all of its homotopy groups are trivial except for $]pi_1 = \pi$, so it is a $K(\pi, 1)$.

The next step in exploring the topology is to try to divide up the $K(\pi, 1)$ summands along tori. An imbedded torus or Klein bottle in M is *incompressible* if the inclusion induces an injection on π_1. To construct incompressible tori, we look for subgroups of $\pi_1(M)$ which are isomorphic to $\mathbb{Z} \oplus \mathbb{Z}$ and are maximal among subgroups with this property. Indeed, according to Thurston's geometrization program (as described, for example, in [54]), any prime compact oriented three-manifold can be decomposed along incompressible imbedded tori and Klein bottles into manifolds which have locally homogeneous Riemannian structures. It is natural to ask whether the incompressilble tori and Klein bottles used in the torus decomposition can be taken to be minimal with respect to any Riemannian metric on M.

This question provides some motivation for the next existence theorem for minimal surfaces, due to Schoen and Yau [72] or Sacks and Uhlenbeck [69] with different proofs. But the theory also applies to minimal tori in Riemannian manifolds of arbitrary dimension.

For $k \in \{0\} \cup \mathbb{N}$, we let

$$\begin{aligned}\mathrm{Map}^{(k)}(T^2, M) = \{f \in C^2(T^2, M) :\ &\text{the image of} \\ &f_\sharp : \pi_1(T^2) \longrightarrow \pi_1(M) \text{ is an abelian group with } k \text{ generators } \}.\end{aligned}$$

For example, if $f_\sharp(\pi_1(T^2)) \cong \mathbb{Z} \oplus \mathbb{Z}_2$, then $f \in \mathrm{Map}^{(2)}(T^2, M)$. Note that the mapping class group $\Gamma = SL(2, \mathbb{Z})$ preserves $\mathrm{Map}^{(k)}(T^2, M)$, so E_α induces a map

$$E_\alpha : \mathcal{M}^{(k)}(T^2, M) \longrightarrow \mathbb{R}, \quad \text{where} \quad \mathcal{M}^{(k)}(T^2, M) = \frac{\mathrm{Map}^{(k)}(T^2, M) \times \mathcal{T}}{\Gamma}.$$

Moreover, if $f \in \mathrm{Map}^{(2)}(T^2, M)$,

$$f \circ \phi = f \quad \text{for some} \quad \phi \in \Gamma \quad \Rightarrow \quad \phi = \text{identity},$$

so the mapping class group $SL(2, \mathbb{Z})$ acts freely on $\mathrm{Map}^{(2)}(T^2, M) \times \mathcal{T}$. Thus if $\mathcal{M}^{(k)}_\alpha(T^2, M)$ denotes the completion with respect to the $L^{2\alpha}_1$ norm, then

$\mathcal{M}_\alpha^{(k)}(T^2, M)$ will be a smooth Banach manifold. Recall that the α-energy descends to a C^2 map on the quotient

$$E_\alpha : \mathcal{M}_\alpha^{(k)}(T^2, M) \longrightarrow \mathbb{R}.$$

We expect that sequences tending to a minimum for α-energy in $\mathcal{M}_\alpha^{(1)}(T^2, M)$ would degenerate to a closed geodesic.

Lemma 3.11.1. *The map $E_\alpha : \mathcal{M}_\alpha^{(2)}(T^2, M) \to \mathbb{R}$ satisfies Condition C.*

What Condition C asserts is that if $[f_i, \omega_i]$ is a sequence of points in $\mathcal{M}^{(2)}(\Sigma, M)$ on which E_α is bounded and for which $\|dE_\alpha([f_i, \omega_i])\| \to 0$, and if for each i, $(f_i, \omega_i) \in \mathrm{Map}^{(2)}(T^2, M) \times \mathcal{T}$ is a representative for $[f_i, \omega_i]$, then there are elements $\phi_i \in \Gamma$ such that a subsequence of $(f_i \circ \phi_i, \phi_i^* \omega_i)$ converges to a critical point for E_α on $\mathrm{Map}^{(2)}(T^2, M) \times \mathcal{T}$.

To prove this, we recall that for the torus, the Teichmüller space $\mathcal{T}$ is the upper half plane, and after a change of basis we can arrange that an element $\omega \in \mathcal{T}$ lies in the fundamental domain

$$D = \{u + iv \in \mathbb{C} : v > 0, -(1/2) \leq u \leq (1/2), u^2 + v^2 \geq 1\} \tag{3.58}$$

for the action of the mapping class group $\Gamma = SL(2, \mathbb{Z})$. The moduli space $\mathcal{R}$ is obtained from D by identifying points on the boundary. The complex torus corresponding to $\omega \in \mathcal{T}$ can be regarded as the quotient of $\mathbb{C}$ by the abelian subgroup generated by d and ωd, where d is any positive real number, or alternatively, this torus is obtained from a fundamental parallelogram spanned by d and ωd by identifying opposite sides. The fundamental parallelogram of area one can be regarded as the image of the unit square $\{(t^1, t^2) \in \mathbb{R}^2 : 0 \leq t^i \leq 1\}$ under the linear transformation

$$\begin{pmatrix} t^1 \\ t^2 \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{v}} \begin{pmatrix} 1 & u \\ 0 & v \end{pmatrix} \begin{pmatrix} t^1 \\ t^2 \end{pmatrix},$$

where $z = x + iy$ is the usual complex coordinate on $\mathbb{C}$. A straightforward calculation gives a formula for the usual energy

$$\begin{aligned} E(f, \omega) = \frac{1}{2} \int_P \left(\left| \frac{\partial f}{\partial x} \right|^2 + \left| \frac{\partial f}{\partial y} \right|^2 \right) dx dy \\ = \frac{1}{2} \int_P \left(v \left| \frac{\partial f}{\partial t^1} \right|^2 + \frac{1}{v} \left| \frac{\partial f}{\partial t^2} - u \frac{\partial f}{\partial t^1} \right|^2 \right) dt^1 dt^2, \end{aligned}$$

P denoting the image of the unit square. The only way that ω can approach the boundary of Teichmüller space while remaining in the fundamental domain D is for $v \to \infty$. The rank two condition implies that the maps $t^1 \mapsto f(t^1, b)$ must be homotopically nontrivial for each choice of $t^2 = b$, and hence the length in

M of $t^1 \mapsto f(t^1, b)$ is bounded below by a positive constant c. This implies that

$$E(f,\omega) \geq \frac{1}{2}\int_0^1 \int_0^1 v \left|\frac{\partial f}{\partial t^1}\right|^2 dt^1 dt^2$$
$$\geq \frac{v}{2}\int_0^1 (\text{length of } t^1 \mapsto f(t^1, b))^2 db \geq \frac{c^2 v}{2} \quad (3.59)$$

by the Cauchy-Schwarz inequality, and hence $E_\alpha(f,\omega)$ (which is $\geq E(f,\omega)$) must approach infinity.

Suppose now that $[f_i, \omega_i]$ is a sequence of points in $\mathcal{M}^{(2)}(T^2, M)$ on which E_α is bounded and for which $\|dE_\alpha([f_i,\omega_i])\| \to 0$, and for each i, $(f_i,\omega_i) \in \mathrm{Map}(T^2, M) \times \mathcal{T}$ is a representative for $[f_i,\omega_i]$. Then the projection $[\omega_i] \in \mathcal{R}$ of $\omega_i \in \mathcal{T}$ is bounded, and must therefore have a subsequence which converges to an element $[\omega_\infty] \in \mathcal{R}$. Hence there are elements $\phi_i \in \Gamma$ such that a subsequence of $\phi_i^*\omega_i$ converges to an element $\omega_\infty \in \mathcal{T}$. Then $E_{\alpha,\omega_\infty}(f_i \circ \phi_i)$ is bounded and $\|dE_{\alpha,\omega_\infty}(f_i \circ \phi_i)\| \to 0$, so by Condition C for E_{α,ω_∞}, a subsequence of $\{(f_i \circ \phi_i, \phi_i^*\omega_i)\}$ converges to a critical point for E_α on $\mathrm{Map}(T^2, M) \times \mathcal{T}$. This establishes Condition C for the function E_α.

Just as for a generic perturbation of $E_{\alpha,\omega}$, one can establish Morse inequalities for a generic perturbation E'_α of E_α. Thus we have a Morse-Witten complex for this function, and we can use it to investigate critical points of E in the limit as $\alpha \to 1$.

The following theorem was proven by Schoen and Yau [72] and Sacks and Uhlenbeck [69]:

Theorem 3.11.2. *Every component of $\mathcal{M}^{(2)}(\Sigma, M)$ contains an element $[f,\omega]$ which minimizes the function*

$$E : \mathcal{M}^{(2)}(T^2, M) \longrightarrow \mathbb{R}$$

If (f,ω) is a representative, then $f : T^2 \to M$ is conformal and harmonic with respect to ω, and hence a minimal surface.

Proof: Let C be a component of $\mathcal{M}^{(2)}(T^2, M)$. Choose a decreasing sequence $\alpha_m \to 1$ and for each α_m, a corresponding critical point

$$[f_m, \omega_m] \in \mathcal{M}^{(2)}(T^2, M) \quad \text{which minimizes } E_{\alpha_m} \text{ on } C.$$

We can assume that

$$E_{\alpha_m}([f_m,\omega_m]) \to \inf\{E(f,\omega) : [f,\omega] \in C\}. \quad (3.60)$$

Since $[\omega_m]$ must lie in a bounded region of the Riemann moduli space $\mathcal{R}_1$, we can arrange that $[\omega_m]$ converges after passing to a subsequence, and hence after choosing suitable representatives, arrange that ω_m converges to an element $\omega \in \mathcal{T}_1$. By Theorem 3.8.7, we can pass to a further subsequence so that either $\{f_m\}$ converges uniformly on compact subsets of T^2 to an ω-harmonic map

$f : T^2 \to M$ or nonconstant minimal two-spheres bubble off as $\alpha \to 1$. But in the latter case, we can perform a surgery just like we described in the proof of Theorem 3.10.1 and obtain a new parametrized torus $\hat{f}_m : T^2 \to M$ which lies in the same component C and has smaller energy

$$E_{\alpha_m}(\hat{f}_m, \omega_m) < E_{\alpha_m}(f_m, \omega_m) - \frac{\epsilon_0}{2},$$

contradicting (3.60). So no bubbling can occur, and we obtain a critical point (f, ω) for E which minimizes E within the component C.

Remark 3.11.3. The local regularity result of Osserman and Gulliver implies that the minimal tori found by Theorem 3.11.2 are immersed. Using a variant of the tower construction, it was proven by Freedman, Hass and Scott [24] that the tori and Klein bottles needed for the torus decomposition of a prime compact oriented three-manifold M can be taken to be minimal.

3.12 Higher genus minimal surfaces*

The existence theory for incompressible minimal tori presented in the previous section can be extended to surfaces of higher genus. Suppose that Σ is a compact oriented compact surface of genus $g \geq 2$. We set

$$\mathrm{Map}'(\Sigma, M) = \{ f \in C^2(\Sigma, M) : \ f_\sharp : \pi_1(\Sigma) \to \pi_1(M) \text{ is injective } \},$$

define $\mathcal{M}'(\Sigma, M)$ to be the quotient space

$$\mathcal{M}'(\Sigma, M) = \frac{\mathrm{Map}'(\Sigma, M) \times \mathcal{T}}{\Gamma},$$

where Γ is the mapping class group, and let $\mathcal{M}'_\alpha(\Sigma, M)$ be the completion of $\mathcal{M}'(\Sigma, M)$ with respect to the $L^{2\alpha}_1$ norm.

Lemma 3.12.1. *If Σ is a compact oriented surface of genus $g \geq 2$, then the map $E_\alpha : \mathcal{M}'_\alpha(\Sigma, M) \to \mathbb{R}$ satisfies Condition C.*

The proof makes use to two ingredients, a collar theorem of Halpern and Keen, and the structure of the Bers compactification of the moduli space $\mathcal{R}$ of conformal structures on Σ.

We suppose that Σ is given the hyperbolic metric of Gaussian curvature -1 corresponding to the conformal structure $\omega \in \mathcal{T}$. The collar theorem is a fundamental result of Riemann surface theory, and states that if γ is a closed geodesic of this metric of length $\leq k_1$, where k_1 is a positive constant, then there is a collar region $C \subseteq \Sigma$ of fixed area $k_2 > 0$ about γ. Indeed, we can arrange that γ lifts to the map

$$\tilde{\gamma} : [0, l] \to \mathbb{H}^2 \quad \text{defined by} \quad \tilde{\gamma}(t) = \exp\left(t + \frac{i\pi}{2}\right),$$

and the collar region is of the form $C = \pi(\tilde{C})$, where

$$\tilde{C} = \{re^{i\theta} \in \mathbb{H}^2 : 1 \leq r < e^l, \pi - \theta_0 < \theta < \theta_0\}, \quad \text{where} \quad \cot\theta_0 = \frac{k_2}{2l}. \quad (3.61)$$

Lemma 3.12.2. *Suppose that Σ has a closed geodesic γ whose length with respect to the hyperbolic metric corresponding to ω is $l \leq k_1$, and that C is the collar region about γ described above. If $f : \Sigma \to M$ is any smooth map, then the ω-energy of $f|C$ is at least k_4/l, for some positive constant k_4.*

We consider the lift $\tilde{f} : \tilde{C} \to M$, which has energy density

$$e(\tilde{f}) = \frac{1}{2}|df|^2 \geq \frac{1}{2}\frac{|(\tilde{f} \circ \tilde{\gamma}_\theta)'(t)|^2}{|\tilde{\gamma}_\theta'(t)|^2}, \quad \text{where} \quad \tilde{\gamma}_\theta(t) = \exp(t + i\theta).$$

Straightforward calculation shows that

$$|\tilde{\gamma}_\theta'(t)|^2 = \frac{1}{\sin^2\theta} \quad \text{and} \quad |(\tilde{f} \circ \tilde{\gamma}_\theta)'(t)|^2 = r^2\left|\frac{\partial\tilde{f}}{\partial r}\right|^2,$$

$$\text{so} \quad e(f) \geq \frac{r^2\sin^2\theta}{2}\left|\frac{\partial\tilde{f}}{\partial r}\right|^2.$$

Thus we find that

$$E(\tilde{f}|\tilde{C}) = \int_{\theta_0}^{\pi-\theta_0}\int_1^{e^l} e(\tilde{f})\frac{dr d\theta}{r\sin^2\theta}$$

$$\geq \int_{\theta_0}^{\pi-\theta_0}\int_1^{e^l}\frac{r}{2}\left|\frac{\partial\tilde{f}}{\partial r}\right|^2 dr d\theta = \frac{1}{2}\int_{\theta_0}^{\pi-\theta_0}\int_0^l\left|\frac{\partial\tilde{f}}{\partial t}\right|^2 dt d\theta,$$

where we have set $r = e^t$ in the last integral. But by the Cauchy-Schwarz inequality,

$$L(\tilde{f} \circ \tilde{\gamma}_\theta)^2 = \int_0^l\left|\frac{\partial\tilde{f}}{\partial t}\right| dt \leq l\int_0^l\left|\frac{\partial\tilde{f}}{\partial t}\right|^2 dt,$$

and since $\tilde{f} \circ \tilde{\gamma}_\theta$ is a closed curve in M which is not homotopic to a constant, $L(\gamma_\theta) \geq k_3$, for some positive constant, k_3. Hence

$$E(f, \omega) \geq E(\tilde{f}|\tilde{C}) \geq \frac{k_3^2(\pi - 2\theta_0)}{2l},$$

which yields the assertion of the lemma, since $\theta_0 \to 0$ as $l \to 0$.

To finish the proof of Lemma 3.12.1, we suppose that $\{[f_m, \omega_m]\}$ is a sequence of points in $\mathcal{M}'_\alpha(\Sigma, M)$ such that $E_\alpha([f_m, \omega_m])$ is bounded. We claim that the sequence $\{[\omega_m]\}$ must stay in a compact region of the moduli space $\mathcal{R}$. To

see this, we make use of the Bers compactification of the moduli space $\mathcal{R}$, as described in Appendix B of [37]. After passing to a subsequence, we can assume that $\{[\omega_m]\}$ converges to a point in the compactification. But points in the compactification are Riemann surfaces with nodes and and as one approaches a Riemann surface with nodes from inside the moduli space, some homotopically nontrivial loop must have its length go to zero, and then the energy will go to infinity by Lemma 3.12.2. Thus $\{[\omega_m]\}$ must remain bounded, and after passing to a subsequence we can assume that $[\omega_m] \to [\omega_\infty] \in \mathcal{R}$. Now we use the fact that E_{α,ω_∞} satisfies Condition C to conclude that $E_\alpha : \mathcal{M}'_\alpha(\Sigma, M) \to \mathbb{R}$ satisfies Condition C.

Thus we have a well-behaved Morse theory for suitable perturbations of the α-energy on $\mathcal{M}'_\alpha(\Sigma, M)$, just as in the case of the torus.

Theorem 3.12.3. *If Σ is a compact oriented Riemann surface of genus $g \geq 2$, then every component of the space $\mathcal{M}'(\Sigma, M)$ of incompressible maps from Σ to M contains an element $[f, \omega]$ which minimizes the function*

$$E : \mathcal{M}'(\Sigma, M) \longrightarrow \mathbb{R}$$

If (f, ω) is a representative, then $f : \Sigma \to M$ is conformal and harmonic with respect to ω, and hence a minimal surface.

To prove this, we use essentially the same argument as we used in the proof of Therem 3.11.2.

3.13 Complex form of second variation

In order to apply the apply the Morse theory of α-harmonic maps to derive geometric consequences, it is useful to be able to estimate the Morse index of a harmonic map. Recall that by Corollary 3.7.2, the second variation of energy at a harmonic map $f : \Sigma \to M$ is given by the formula

$$d^2 E_\omega(f)(V, W) = \int_\Sigma \langle L_f(V), W \rangle dA,$$

where L_f is the *Jacobi operator*, defined by

$$L_f(V) = -\frac{1}{\lambda^2}\left[\frac{D}{\partial x} \circ \frac{D}{\partial x} + \frac{D}{\partial y} \circ \frac{D}{\partial y} + \mathcal{K}(V)\right]. \tag{3.62}$$

Recall that we say that an element $V \in T_f\mathcal{M}$ is a *Jacobi field* along f if $L_f(V) = 0$.

Theorem 3.13.1. *The Jacobi operator can be written in the following complex form:*

$$L_f(V) = -\frac{4}{\lambda^2}\left[\frac{D}{\partial z} \circ \frac{D}{\partial \bar{z}} + R\left(\cdot, \frac{\partial f}{\partial z}\right)\frac{\partial f}{\partial \bar{z}}\right],$$

where the Riemann-Christoffel curvature tensor R has been extended to be complex linear.

Straightforward calculations show that on the one hand,

$$4\frac{D}{\partial z}\circ\frac{D}{\partial \bar{z}} = \frac{D}{\partial x}\circ\frac{D}{\partial x} + \frac{D}{\partial y}\circ\frac{D}{\partial y} + \sqrt{-1}\left(\frac{D}{\partial x}\circ\frac{D}{\partial y} - \frac{D}{\partial y}\circ\frac{D}{\partial \bar{x}}\right)$$
$$= \frac{D}{\partial x}\circ\frac{D}{\partial x} + \frac{D}{\partial y}\circ\frac{D}{\partial y} + \sqrt{-1}R\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right),$$

while on the other,

$$4R\left(\cdot, \frac{\partial f}{\partial z}\right)\frac{\partial f}{\partial \bar{z}} = R\left(\cdot, \frac{\partial f}{\partial x}\right)\frac{\partial f}{\partial x} + R\left(\cdot, \frac{\partial f}{\partial y}\right)\frac{\partial f}{\partial y}$$
$$+ \sqrt{-1}\left[R\left(\cdot, \frac{\partial f}{\partial x}\right)\frac{\partial f}{\partial y} - R\left(\cdot, \frac{\partial f}{\partial y}\right)\frac{\partial f}{\partial x}\right]$$
$$= R\left(\cdot, \frac{\partial f}{\partial x}\right)\frac{\partial f}{\partial x} + R\left(\cdot, \frac{\partial f}{\partial y}\right)\frac{\partial f}{\partial y} - \sqrt{-1}R\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right),$$

the last step following from the Bianchi symmetry. Substitution into (3.62) now yields the Theorem.

Corollary 3.13.2. *The second variation of ω-energy is given by the formula:*

$$d^2E_\omega(f)(V,\bar{V}) = 4\int_\Sigma \left[\left|\frac{DV}{\partial \bar{z}}\right|^2 - \left\langle \mathcal{R}\left(V\wedge\frac{\partial f}{\partial z}\right), \bar{V}\wedge\frac{\partial f}{\partial \bar{z}}\right\rangle\right] dxdy,$$

for sections V of the complex vector bundle E, where the bar indicates complex conjugate and we use the notation $\langle\mathcal{R}(x\wedge y), v\wedge w\rangle = \langle R(x,y)w, v\rangle$.

Proof: Simply substitute into (3.62) and integrate by parts.

Corollary 3.13.3. *A section V of the subbundle $\mathbf{L}$ of $f^*TM\otimes\mathbb{C}$, characterized by the fact that $\partial f/\partial z$ is a locally defined section of $\mathbf{L}$, is a Jacobi field if and only if it is holomorphic.*

Proof: If σ is a holomorphic section of $\mathbf{L}$, then $\sigma = g(\partial f/\partial z)$ where g is a holomorphic function. It follows that

$$\frac{D}{\partial z}\circ\frac{D}{\partial \bar{z}}(\sigma) = 0 \qquad \text{and} \qquad R\left(\sigma, \frac{\partial f}{\partial z}\right) = 0.$$

We leave the converse to the reader.

Remark 3.13.4. Note first that since L_f is a real operator, real and imaginary parts of Jacobi fields are also Jacobi fields.

Remark 3.13.5. The dimension of the space of holomorphic sections $\mathcal{O}(\mathbf{L})$ of the line bundle $\mathbf{L}$ can often be estimated by the Riemann-Roch theorem from the

theory of Riemann surfaces. (One can apply the usual Riemann-Roch theorem to $\underline{\mathbf{L}}$ since $\underline{\mathbf{L}}$ has a canonical holomorphic structure by the Koszul-Malgrange Theorem 3.1.3. Alternatively, one can apply the Atiyah-Singer Index Theorem directly to the operator $Z \mapsto (DZ/\partial\bar{z})dz$ and use the fact that this operator has the same symbol as the $\bar{\partial}$-operator, avoiding the proof of the Koszul-Malgrange Theorem.) First consider the case in which Σ has genus zero. In this case, one recalls that there is exactly one holomorphic line bundle $\mathbf{H}^k$ over $\Sigma = S^2$ (up to isomorphism) whose Chern class satisfies $\langle c_1\mathbf{H}^k), [\Sigma]\rangle = k$, and the space of holomorphic sections of this bundle has complex dimension $k+1$ if $k \geq 0$. The reason for the notation $\mathbf{H}^k$ is that the bundle of first Chern class k is the k-th tensor power of a bundle $\mathbf{H}$ called the hyperplane bundle over S^2. Thus $\mathbf{L} = \mathbf{H}^k$, where $k = c_1(\mathbf{L})[S^2]$.

According to (3.7), the line bundle $\mathbf{L}$ defined by the harmonic map f has first Chern class given by the formula

$$\langle c_1(\mathbf{L}), [\Sigma]\rangle = 2 + \sum\{\nu_p : \ p \text{ is a branch point of } f \ \}.$$

In particular, $\langle c_1(L), [\Sigma]\rangle$ is always positive and $\mathbf{L}$ always has a space of holomorphic section which has complex dimension at least three. A holomorphic section of $\mathbf{L}$ can be identified with a meromorphic section σ of the holomorphic tangent bundle $T\Sigma$ with the property that its divisor (σ) satisfies

$$(\sigma) + \sum\{\nu_p p : \ p \text{ is a branch point of } f \ \} \geq 0.$$

Counting up the possibilities, we see that there is a holomorphic section of $\mathbf{L}$ which has every possible principal part at every branch point of f.

We can do a similar analysis for line bundles over a torus T^2, except that this time the line bundles of a given Chern class form a torus of dimension two, and if $c_1(\mathbf{L})[T^2] \geq 1$, the dimension of holomorphic sections of $\mathbf{L}$ will miss by one the dimension that would be possible if every principal part were realized. Thus in the case of the torus, there is a section σ of $\mathbf{L}$ such that the corresponding section of the tangent bundle to T^2 has arbitrary principal part at all but one branch point.

3.14 Isotropic curvature

Just as the theory of geodesics uncovers relations between curvature and topology of Riemannian manifolds (through classical theorems with names such as Synge's theorem, Myers' theorem and the theorem of Hadamard and Cartan among others), one might hope to apply the theory of minimal surfaces to curvature and topology. Pursuing this idea, however, leads to a different notion of curvature. Just as studying the stability of geodesics leads to sectional curvature, we find stability theory for minimal surfaces leads to the notion of isotropic curvature.

Recall that in normal coordinates $(x^1, \ldots, x^n)$ centered at p on a Riemannian manifold M, the Riemannian metric can be expressed by a Taylor series

$$g_{ij} = \delta_{ij} - \frac{1}{3} \sum_{k,l} R_{ikjl}(p) x^k x^l + \text{ higher order terms},$$

where the R_{ikjl}'s are components of the Riemann-Christoffel curvature tensor. The curvature operator is defined in terms of these components to be the linear map

$$\mathcal{R} : \Lambda^2 T_p M \longrightarrow \Lambda^2 T_p M$$

such that

$$\mathcal{R}\left(\left. \frac{\partial}{\partial x^i} \right|_p \wedge \left. \frac{\partial}{\partial x^j} \right|_p \right) = \sum_{k,l} R_{ijkl}(p) \left. \frac{\partial}{\partial x^k} \right|_p \wedge \left. \frac{\partial}{\partial x^l} \right|_p .$$

If z and w are linearly independent elements of $T_p M \otimes \mathbb{C}$, the sectional curvature of the complex two-plane σ spanned by z and w is

$$\frac{\langle \mathcal{R}(z \wedge w), \bar{z} \wedge \bar{w} \rangle}{\langle z \wedge w, \bar{z} \wedge \bar{w} \rangle},$$

where the bar denotes complex conjugation. The complex two-plane σ is said to be *isotropic* if $\langle z, z \rangle = \langle w, w \rangle = \langle z, w \rangle = 0$.

Definition. The Riemannian manifold M is said to have *positive isotropic curvature* if $K(\sigma) > 0$, whenever σ is an isotropic complex two-plane.

To see why isotropic curvature is related to stability properties of minimal surfaces, we first recall the second variation formula for a harmonic map $f : S^2 \to M$ (Corollary 3.13.3):

$$d^2 E_\omega(f)(V, \bar{V}) = 4 \int_\Sigma \left[\left| \frac{DV}{\partial \bar{z}} \right|^2 - \left\langle \mathcal{R}\left(V \wedge \frac{\partial f}{\partial z} \right), \bar{V} \wedge \frac{\partial f}{\partial \bar{z}} \right\rangle \right] dxdy. \qquad (3.63)$$

We need one further ingredient:

Grothendieck Theorem 3.14.1. *Any holomorphic line bundle over the Riemann sphere $S^2 = \mathbb{C}P^1$ divides into a holomorphic direct sum of holomorphic line bundles.*

This theorem, proven in [30], allows us to write $\mathbf{E} = f^* TM \otimes \mathbb{C}$ as a direct sum of line bundles,

$$\mathbf{E} = \mathbf{L}_1 \oplus \mathbf{L}_2 \oplus \cdots \oplus \mathbf{L}_n, \qquad \text{where} \qquad c_1(\mathbf{L}_1)[S^2] \geq \cdots \geq c_1(\mathbf{L}_n)[S^2].$$

Since the Riemannian metric is invariant under the Levi-Civita connection, it extends to a holomorphic complex bilinear form

$$\langle \cdot, \cdot \rangle : \mathbf{E} \times \mathbf{E} \longrightarrow \mathbb{C},$$

and in particular, $\mathbf{E}$ is isomorphic to its dual. Thus the line bundles in the above sequence can be arranged so that $c_1(\mathbf{L}_i) = -c_1(\mathbf{L}_{n-i+1})$ for each i.

If V is a holomorphic section of one of the line bundles $\mathbf{L}_i$ inthe direct sum decomposition of E, then

$$\langle V, V\rangle : S^2 \to \mathbb{C} \qquad \text{and} \qquad \left\langle V, \frac{\partial f}{\partial z}\right\rangle : S^2 \to \mathbb{C}$$

are holomorphic, and hence

$$\langle V, V\rangle = (\text{constant}), \qquad \left\langle V, \frac{\partial f}{\partial z}\right\rangle = (\text{constant}).$$

In particular, if V is a holomorphic section of $\mathbf{L}_i$ where $\mathbf{L}_i$ has positive first Chern class, or more generally if V is any holomorphic section such that $\langle V, V\rangle = 0$, then

$$\langle V, V\rangle = \left\langle V, \frac{\partial f}{\partial z}\right\rangle = \left\langle \frac{\partial f}{\partial z}, \frac{\partial f}{\partial z}\right\rangle = 0,$$

and

$$V \quad \text{and} \quad \frac{\partial f}{\partial z} \quad \text{span an isotropic two-plane}$$

at every point where neither vanishes. If M has positive isotropic curvatures, it therefore follows from the index formula (3.63) that $d^2E_\omega(f)(V, \bar{V}) < 0$.

Let m be the number of line bundle summands L_i of E with positive first Chern number,

$$c_1(\mathbf{L}_1)[S^2] \geq \cdots \geq c_1(\mathbf{L}_m)[S^2] \geq 1, \qquad c_1(\mathbf{L}_{m+1})[S^2] \leq 0$$

and let $\mathbf{E}_0$ be the direct sum of all of the line bundle summands with zero Chern class; the dimension of the space $\mathcal{O}(\mathbf{E}_0)$ of holomorphic sections of $\mathbf{E}_0$ is $n-2m$. If $V_1, \ldots, V_m$ are nonzero holomorphic sections of $\mathbf{L}_1, \ldots, \mathbf{L}_m$ respectively, then $\langle V_i, \mathbf{E}_0\rangle = 0$ for $1 \leq i \leq m$. If $\{W_1, \ldots, W_l\}$ is a basis for $\mathcal{O}(\mathbf{E}_0)$, then $\langle W_i, W_j\rangle$ is constant, and therefore there is a maximal isotropic holomorphic subbundle $\mathbf{I}_0 \subset \mathbf{E}_0$) which has rank $\geq (1/2)$(rank of $\mathbf{E}_0$)-1).

If $\mathcal{V}$ is the space of holomorphic sections of $\mathbf{L}_1 \oplus \cdots \mathbf{L}_m \oplus I_0$, then $\dim \mathcal{V} \geq (1/2)(\dim M - 1)$ and

$$V \in \mathcal{V} \qquad \Rightarrow \qquad d^2E_\omega(f)(V, \bar{V}) < 0.$$

Thus we have proven a lemma which will be needed for the proof of the next theorem:

Lemma 3.14.2. *Suppose that M is a Riemannian manifold which has positive isotropic curvature. Then the Morse index of any harmonic two-sphere $f : S^2 \to M$ is at least $(1/2)(\dim M - 1)$.*

The following theorem [49] generalizes the earlier classical sphere theorem due to Berger, Klingenberg and Toponogov:

Sphere Theorem 3.14.3. *Suppose that M is a compact smooth simply connected Riemannian manifold of dimension at least four which has positive isotropic curvature. Then M is homeomorphic to a sphere.*

The idea behind the proof is to show that M is a homotopy sphere and apply the solutions to the generalized Poincaré conjecture in dimensions greater than four to conclude that M is homeomorphic to a sphere.

By definition, a compact connected manifold M of dimension n is a homotopy sphere if $\pi_q(M) = 0$, for all integers q such that $0 < q < n$. Thus if M is not a homotopy sphere there must be an integer q with $0 < q < n$ such that $\pi_q(M) \neq 0$, and we choose the smallest such integer. By the Hurewicz isomorphism theorem, q is also the smallest positive integer such that $H_q(M; \mathbb{Z}) \neq 0$. By Poincaré duality, we can assume that $q \leq n/2$.

Just as in §5.4, we let $\mathcal{M}$ be the space of smooth maps from S^2 to M and let $\pi : \mathcal{M} \to M$ be the evaluations map defined by $\pi(f) = f(0)$ where $0 \in S^2 = \mathbb{C} \cup \{\infty\}$ is the basepoint. Since the fibration π possesses a section, the long exact homotopy sequence of π splits, and if $\mathcal{M}_p$ denotes the subset of $\mathcal{M}$ consisting of maps such that $f(0) = p$, we conclude that

$$\pi_k(\mathcal{M}) \cong \pi_k(M) \oplus \pi_k(\mathcal{M}_p) \cong \pi_k(M) \oplus \pi_{k+2}(M),$$

which implies that

$$\pi_{q-2}(\mathcal{M}) \cong \pi_q(M) \neq 0.$$

We now apply Morse theory to a perturbation $E_{\alpha,\psi}$ of the energy E, where ψ is chosen so that all nonconstant critical points of $E_{\alpha,\psi}$ are Morse nondegenerate, and let $\alpha \to 1$ and $\psi \to 0$, in accordance with Theorem 2 from §3.5. We thereby obtain a sequence $\{f_m\}$ of critical points for E_{α_m,ψ_m}, each having Morse index no more than $q-2$. By the bubbling argument presented in §3.9, we find that a subsequence (still denoted by $\{f_m\}$) converges uniformly in every C^k on every compact subset of $S^2 - \{p_1, \cdots, p_l\}$, where $p_1, \ldots, p_l$ are a finite number of bubble points, to a harmonic map on $\Sigma - \{p_1, \cdots, p_l\}$, which can be extended to a smooth harmonic map $f_\infty : \Sigma \to M$ by the Sacks-Uhlenbeck removeable singularity theorem.

Suppose first that f_∞ is nonconstant. In that case, we claim that the Morse index of f_∞ is no larger than $q - 2$. For this, we need the following lemma:

Lemma 3.14.4. *Suppose that $f_m : \Sigma \to M$ is a sequence of α_m-harmonic maps in M which converge in C^k on compact subsets of $\Sigma - \{x_1, \cdots, x_l\}$ to a smooth harmonic map $f_\infty : \Sigma \to M$. Then*

$$\textit{Morse index of } f_\infty \leq \liminf \textit{Morse index of } f_m.$$

To prove the Lemma, we first recall the formula the second variation formula

(3.40) for α-harmonic two-spheres:

$$d^2E_\alpha(f)(V,W) = \alpha\int_\Sigma (1+|df|^2)^{\alpha-1}[\langle\nabla V,\nabla W\rangle - \langle\mathcal{K}(V),W\rangle]dA \\ + 2\alpha(\alpha-1)\int_\Sigma (1+|df|^2)^{\alpha-2}\langle df,\nabla V\rangle\langle df,\nabla W\rangle dA.$$

Note that the second term is dominated by the first and goes to zero when α is close to one, and the first term approaches $d^2E(f)(V,W)$ when the support of V and W does not contain any bubble points. We claim that if $d^2E(f_\infty)$ is negative definite on a fixed linear space $\mathcal{V}$ of dimension $q-2$, so is $d^2E_\alpha(f_m)$ when m is sufficiently large.

To see this, we make use of a cutoff function near the bubble points. First, define a map $\phi:\mathbb{R}\to\mathbb{R}$ by

$$\phi(r) = \begin{cases} 0, & \text{if } r\le\epsilon^2,\\ (\log(\epsilon^2)-\log r)/(\log\epsilon), & \text{if } \epsilon^2\le r\le\epsilon,\\ 1, & \text{if } \epsilon\le r,\end{cases}$$

so that

$$\frac{d\phi}{dr}(r) = \begin{cases} 0, & \text{if } r\le\epsilon^2,\\ (-1)/(r\log\epsilon), & \text{if } \epsilon^2\le r\le\epsilon,\\ 0, & \text{if } \epsilon\le r,\end{cases}$$

and

$$\int_0^{2\pi}\int_0^{\epsilon}\left(\frac{d\phi}{dr}(r)\right)^2 r\,dr\,d\theta = \int_{\epsilon^2}^{\epsilon}\frac{2\pi}{r(\log\epsilon)^2}dr = -\frac{2\pi}{\log\epsilon}.$$

Thus if we define $\psi_i : S^2\to\mathbb{R}$ so that it is one outside an ϵ-neightborhood of the bubble point x_i and in terms of polar coordinates (r_i,θ_i) about the bubble point x_i satisfies the condition $\psi_i = \phi\circ r_i$, then

$$\int_{S^2}|d\psi_i|^2dA \le \frac{-C}{\log\epsilon}, \qquad \text{where } C \text{ is a positive constant,}$$

and a similar estimate holds for $\psi=\psi_1\cdots\psi_l$, a cutoff function which vanishes at every bubble point.

If $\epsilon>0$ is chosen sufficiently small, then

$$V\in\mathcal{V} \quad\Rightarrow\quad d^2E(f)(\psi V,\psi V)<0 \quad\Rightarrow\quad d^2E_\alpha(f)(\psi V,\psi V)<0.$$

This shows that $d^2E_\alpha(f)$ is negative definite on a space of dimension $q-2$ and proves the Lemma.

It follows from Lemma 3.14.4 that if f_∞ is nonconstant, it must have index $\le q-2\le(1/2)(\dim M-2)$ which contradicts Lemma 3.14.2.

But if f_∞ is constant a nonconstant sphere must bubble off; that is there must be a family of conformal reparametrizations $g_m : S^2\to S^2$, such that

$f_m \circ g_m$ converges to a nonconstant harmonic sphere $\hat{f}_\infty$ in C^k on compact subsets of $S^2 - \{p_1, \cdots, p_l\}$, where $p_1, \ldots, p_l$ are a finite number of bubble points. We can choose $p_m \in S^2$ such that

$$\|df_m(p_m)\| = \sup\{\|df_p\| : p \in S^2\},$$

and after rotations we can arrange that all the p_m's are equal and in fact that $z(p_m) = 0$, where z is the standard coordinate on $S^2 - \{\infty\} = \mathbb{C}$. Let $r_m = \|df_m(p_m)\|$, and let $g_m : S^2 \to S^2$ be the conformal map expressed in terms of the standard coordinate as $h_m(z) = r_m z$. Finally, let $h_m = f_m \circ g_m$, a sequence of maps which converge to $\hat{f}_\infty$ on compact subsets of $S^2 - \{p_1, \cdots, p_l\}$.

We must now replace $d^2E_\alpha(f_m)$ by $d^2E_\alpha(f_m \circ g_m)$ in the above argument. Once again, one obtains a nonconstant two-sphere $\hat{f}_\infty$ in the limit. Moreover, a calculation similar to that for f_∞ leads to the conclusion that $\hat{f}_\infty$ has index $\leq q - 2 \leq (1/2)(\dim M - 2)$, which once again contradicts Lemma 3.14.2.

Thus M must be a homotopy sphere, and follows from positive resolutions of the generalized Poincaré conjecture ([51] and [23]) in dimensions ≥ 4 that M is homeomorphic to a sphere.

Remark 3.14.5. It can be shown that if the real sectional curvatures $K(\sigma)$ of a Riemannian manifold satisfy the inequalities

$$\frac{1}{4} < K(\sigma) \leq 1, \tag{3.64}$$

then the manifold has positive isotropic curvatures, but not conversely. The sphere theorem proven by Berger, Klingenberg and Toponogov made the hypothesis (3.64) on real sectional curvatures and is therefore weaker than the sphere theorem we have proven. The complex projective space with the standard Fubini-Study metric is simply connected, has real sectional curvatures lying in the range $[1/4, 1]$ and has nonnegative isotropic curvature, but is not homeomorphic to a sphere.

Remark 3.14.6. Conformally flat four-manifolds of positive scalar curvature automatically have positive isotropic curvature. It is possible to construct a conformally flat metric of positive scalar curvature on the connected sum of a finite number of $S^3 \times S^1$'s. Therefore the fundamental group of a compact manifold of positive isotropic curvature can be a free group of arbitrary rank. However, a recent article of Fraser [22] shows that the fundamental group of such a manifold cannot contain a free abelian group of rank two.

Bibliography

[1] R. Abraham, *Lectures of Smale on differential topology*, Columbia University, 1963.

[2] R. Abraham, J. Marsden and T. Ratiu, *Tensor analysis*, Second Edition, Addison-Wesley, 1988.

[3] M. Anderson, *Geometrization of 3-manifolds via the Ricci flow*, Notices Amer. Math. Soc. **51** (2004),184-193.

[4] V. Anosov, *On generic properties of closed geodesics*, Math. USSR Izvestiya **21** (1983), 1-29.

[5] M. Atiyah, N. Hitchin and I. Singer, *Self-duality in four-dimensional Riemannian geometry*, Proc. Roy. Soc. London Ser. A **362** (1978), 425-261.

[6] S. M. Bates, *Toward a precise smoothness hypothesis in Sard's theorem*, Proc. Amer. Math. Soc. **117** (1993), 279-283.

[7] L. Biliotti, M. A. Javaloyes and P. Piccione, *On the semi-Riemannian bumpy metric theorem*, to appear.

[8] R. Bott, *On the iteration of closed geodesics and the Sturm intersection theory*, Comm. Pure and Applied Math. **9** (1956), 171-206.

[9] R. Bott, *Lectures on Morse theory, old and new*, Bull. Amer. Math. Soc. **7** (1982), 331-358.

[10] R. Bott and L. Tu, *Differential forms in algebraic topology*, Springer Verlag, New York, 1982.

[11] H. Brezis, *The interplay between analysis and topology in some nonlinear PDE problems*, Bull. Amer. Math. Soc. **40** (2003), 179-201.

[12] J. Chen and G. Tian, *Compactification of moduli space of harmonic mapping*, Comm. Math. Helv. **74** (1999), 201-237.

[13] P. Deligne, P. Griffiths, J. Morgan and D. Sullivan, *Real homotopy theory of Kähler manifolds*, Inventiones mathematicae **29** (1975), 245-274.

[14] S. Donaldson and P. Kronheimer, *The geometry of four-manifolds*, Clarendon Press, Oxford, 1990.

[15] C. Earle and J. Eells, *A fiber bundle description of Teichmüller theory*, J. Differential Geometry **3** (1969), 19-43

[16] J. Eells, *A setting for global analysis*, Bull. Amer. Math. Soc. **72** (1966), 751-807.

[17] J. Eells and S. Salamon, *Twistorial construction of harmonic maps*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **12** (1986), 589-640.

[18] J. Eells and J. Sampson, *Harmonic maps of Riemannian manifolds*, Amer. J. Math. **86** (1964), 109-160.

[19] L. C. Evans, *Partial differential equations*, Amer. Math. Soc., Providence, RI, 1998.

[20] A. Fischer and A. Tromba, *On a purely "Riemannian" proof of the structure and dimension of the unramified moduli space of a compact Riemann surface*, Math. Ann. **267** (1984), 311-345.

[21] A. Fischer and A. Tromba, *A new proof that Teichmüller space is a cell*, Trans. Amer. Math. Soc. **303** (1987), 257-262.

[22] A. Fraser, *Fundamental groups of manifolds with positive isotropic curvature*, Ann. of Math. 158 (2003), 345-354.

[23] M. H. Freedman and F. Quinn, *Topology of* 4*-manifolds*, Princeton Univ. Press, Princeton NJ, 1990.

[24] M. Freedman, J. Hass and P. Scott, *Least area incompressible surfaces in* 3*-manifolds*, Inventiones math. **71** (1983), 609-642.

[25] D. Gilbarg and N. Trudinger, *Elliptic partial differential equations of second order*, second edition, Springer, New York, 1983.

[26] P. Griffiths and J. Morgan, *Rational homotopy theory and differential forms*, Birkhäuser, Boston, 1981.

[27] D. Gromoll and W. Meyer, *Periodic geodesics on compact Riemannian manifolds*, J. Differential Geometry **3** (1969), 493-510.

[28] D. Gromoll and W. Meyer, *On differentiable functions with isolated critical points*, Topology **8** (1969), 361-369.

[29] M. Gromov, *Homotopical effects of dilatation*, J. Differential Geometry **13** (1978), 303-310.

[30] A. Grothendieck, *Sur la classification des fibres holomorphes sur la sphere de Riemann*, Amer. J. Math. **79** (1957), 121-138.

[31] R. Gulliver, R. Osserman and H. Royden, *A theory of branched immersions of surfaces*, Amer. J. Math **95** (1973), 750-812.

[32] R. Hamilton, *The inverse function theorem of Nash and Moser*, Bull. Amer. Math. Soc. **7** (1982), 65-222.

[33] F. Hang and F. Lin, *Topology of Sobolev mappings II*, Acta Math. **191** (2003), 55-107.

[34] P. Hartman, *On homotopic harmonic maps*, Canadian Math. J. **19** (1967), 673-687.

[35] A. Hatcher, *Algebraic topology*, Cambridge Univ. Press, Cambridge, UK, 2002.

[36] M. Hirsch, *Differential topology*, Sixth Edition, Springer, New York, 1997.

[37] Y. Imayoshi and M. Taniguchi, *An introduction to Teichmüller spaces*, Springer, New York, 1992.

[38] W. Jäger and H. Kaul, *Uniqueness and stability of harmonic maps and their Jacobi fields*, Manuscripta Math. **28** (1979), 269-291.

[39] J. Jost, *Two-dimensional geometric variational problems*, John Wiley and Sons, New York, 1991.

[40] J. Jost, *Riemannian geometry and geometric analysis*, Fifth edition, Springer, New York, 2008.

[41] J. Jost and M. Struwe, *Morse-Conley theory for minimal surfaces of varying topological type*, Invent. math. **102** (1990), 465-499.

[42] W. Klingenberg, *Lectures on closed geodesics*, Springer, New York, 1978.

[43] S. Lang, *Differential and Riemannian manifolds*, Springer, New York, 1995.

[44] H. B. Lawson and M. L. Michelsohn, *Spin geometry*, Princeton Univ. Press, Princeton, NJ, 1989.

[45] F. Lin and C. Wang, *The analysis of harmonic maps and their heat flows*, World Scientific, Singapore, 2008.

[46] D. McDuff and D. Salamon, *J-holomorphic curves and symplectic topology*, AMS Colloquium Publications **52**, Amer. Math. Soc., Providence Rhode Island, 2004.

[47] Y. Matsumoto, *An introduction to Morse theory*, Translations of Mathematical Monographs **208**, Amer. Math. Soc., Providence Rhode Island, 2002.

[48] W. Meeks and S. T. Yau, *The topology of three-dimensional manifolds and embedding problems in minimal surface theory*, Annals of Math. **112** (1980), 441-484.

[49] M. Micallef and J. D. Moore, *Minimal two-spheres and the topology of manifolds with positive curvature on totally isotropic two-planes*, Annals of Math. **127** (1988), 199-227.

[50] J. Milnor, *Morse theory*, Annals of Math. Studies **51**, Princeton Univ. Press, Princeton, NJ, 1963.

[51] J. Milnor, *Lectures on the h-cobordism theorem*, Princeton Univ. Press, Princeton, NJ, 1965.

[52] J. D. Moore, *Bumpy Riemannian metrics and closed parametrized minimal surfaces in Riemannian manifolds*, revised version, arXiv preprint 1012.3906.

[53] J. D. Moore, *Nondegeneracy of coverings of minimal tori in Riemannian manifolds*, Pacific J. Math. **230** (2007), 147-166.

[54] J. Morgan, *Recent progress on the Poincaré conjecture and the classification of* 3*-manifolds*, Bull. Amer. Math. Soc. **42** (2004), 57-78.

[55] M. Morse, *Relations between the critical points of a real function of n independent variables*, Trans. Amer. Math Soc. **27** (1925), 345-396.

[56] M. Morse, *The calculus of variations in the large*, American Math. Soc. Colloquium Publications **18**, Ann Arbor, Mich., 1934.

[57] J. Nash, *The imbedding problem for Riemannian manifolds*, Annals of Math. **63** (1956), 20-63.

[58] R. Osserman, *A proof of the regularity everywhere of the classical solution to Plateau's problem*, Annals of Math. **91** (1970), 550-569.

[59] R. Palais, *Lusternik-Schnirelmann theory on Banach manifolds*, Topology **5** (1966), 115-132.

[60] R. Palais, *Foundations of global nonlinear analysis*, Benjamin, New York, 1968.

[61] R. Palais, *Critical point theory and the minimax principle*, Proc. Symposia in Pure Math. **15** (1970), 185-212.

[62] R. Palais and S. Smale, *A generalized Morse theory*, Bull. Amer. Math. Soc. **70** (1964), 165-171.

[63] T. Parker, *Bubble tree convergence for harmonic maps curvature flow*, J. Differential Geometry **44** (1996), 595-633.

[64] T. Parker and J. Wolfson, *Pseudo-holomorphic maps and bubble trees*, J. Geometric Analysis **3** (1993), 63-98.

[65] M. Reed and B. Simon, *Methods of mathematical physics I: Functional analysis*, Academic Press, New York, 1980.

[66] H. L. Royden, *Real analysis*, Third edition, Prentice-Hall, New York, 1988.

[67] W. Rudin, *Real and complex analysis*, Third Edition, McGraw-Hill, New York, 1986.

[68] J. Sacks and K. Uhlenbeck, *The existence of minimal immersions of* 2*-spheres*, Annals of Math. **113** (1981), 1-24.

[69] J. Sacks and K. Uhlenbeck, *Minimal immersions of closed Riemann surfaces*, Trans. Amer. Math. Soc. **271** (1982), 639-652.

[70] J. H. Sampson, *Some properties and applications of harmonic mappings*, Annales Scientifiques de l'École normale Supérieure **11** (1978), 211-228.

[71] R. Schoen and S. T. Yau, *On univalent harmonic maps between surfaces*, Inventiones math. **44** (1978), 265-278.

[72] R. Schoen and S. T. Yau, *Existence of incompressible minimal surfaces and the topology of three dimensional manifolds with non-negative scalar curvature*, Annals of Math. **110** (1979), 127-142.

[73] R. Schoen and S. T. Yau, *Lectures on harmonic maps*, International Press, Boston, 1997.

[74] M. Schwarz, *Morse homology*, Birkhäuser Verlag, Basel, 1993.

[75] J. P. Serre, *Homologie singulière des espaces fibrées*, Annals of Math. **54** (1951), 425-505.

[76] S. Smale, *Morse theory and a nonlinear generalization of the Dirichlet problem*, Annals of Math. **80** (1964), 382-396.

[77] S. Smale, *On the Morse index theorem*, J. Math. Mech. **14** (1965), 1049-1055; Corrigendum **16** (1967), 1069-1070.

[78] S. Smale, *An infinite-dimensional version of Sard's theorem*, Amer. J. Math. **87** (1966), 861-866.

[79] M. Taylor, *Partial differential equations: basic theory*, Springer, New York, 1996.

[80] K. Uhlenbeck, *Morse theory on Banach manifolds*, J. Functional Analysis **10** (1972), 430-445.

[81] V. S. Varadarajan, *Lie groups, Lie algebras and their representations*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[82] M. Vigué-Poirrier and D. Sullivan, *The homology theory of the closed geodesic problem*, J. Differential Geometry **11** (1976), 633-644.

[83] E. Witten, *Supersymmetry and Morse theory*, J. Differential Geometry **17** (1982), 661-692.