4 The Calculus of Variations

4.1 The variational principle

In this section, we want to consider the problem of constructing a function which minimizes a given *functional*. (A functional is a map from functions to $\mathbb{R}$.)

Example 15. *Let $\Omega \subset\subset \mathbb{R}^n$ be a domain with smooth boundary. Then we consider the class*

$$\mathcal{F} = \{f \in C^2(\Omega) \cap C^0(\bar{\Omega}) : f = g \text{ on } \partial\Omega\}$$

for a given C^2 function g on $\partial\Omega$. Consider the graph of f

$$\{(x, f(x)) \in \bar{\Omega} \times \mathbb{R}\}.$$

By pulling back the Euclidean metric on $\mathbb{R}^{n+1}$, we can consider the n-volume of the graph. We have computed above

$$\mathrm{Vol}(f) = \int_\Omega \sqrt{1 + |\nabla f|^2}\, dx_n.$$

Then we want to consider the following question: Is there an $f \in \mathcal{F}$ which minimizes $\mathrm{Vol}(f)$ over all of $\mathcal{F}$?

If it exists, f must satisfy

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \mathrm{Vol}(f + \epsilon h) = 0$$

for every h so that $f + \epsilon h \in \mathcal{F}$. We compute and integrate by parts to find a differential equation f must satisfy. First of all, $f + \epsilon h \in \mathcal{F}$ if and only if

$h \in C^2(\Omega) \cap C^0(\bar{\Omega})$ and $h = 0$ on $\partial\Omega$.

$$\begin{aligned}
0 &= \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \mathrm{Vol}(f + \epsilon h) \\
&= \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \int_\Omega \sqrt{1 + |\nabla f + \epsilon \nabla h|^2}\, dx_n \\
&= \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \int_\Omega \sqrt{1 + |\nabla f|^2 + 2\epsilon\, df \cdot \nabla h + \epsilon^2 |\nabla h|^2}\, dx_n \\
&= \left.\int_\Omega \frac{2\nabla f \cdot \nabla h + 2\epsilon |\nabla h|^2}{2\sqrt{1 + |\nabla f + \epsilon \nabla h|^2}}\, dx_n \right|_{\epsilon=0} \\
&= \int_\Omega \frac{\nabla f \cdot \nabla h}{\sqrt{1 + |\nabla f|^2}}\, dx_n \\
&= -\int_\Omega h \nabla \cdot \left(\frac{\nabla f}{\sqrt{1 + |\nabla f|^2}} \right) dx_n + \int_{\partial\Omega} h \left(\frac{\nabla f}{\sqrt{1 + |\nabla f|^2}} \right) \cdot \mathbf{n}\, dV \\
&= -\int_\Omega h \nabla \cdot \left(\frac{\nabla f}{\sqrt{1 + |\nabla f|^2}} \right) dx_n.
\end{aligned}$$

This last integral must be equal to zero for every $h \in C^0(\bar{\Omega})$ which vanishes on $\partial\Omega$. We claim this forces

$$g = \nabla \cdot \left(\frac{\nabla f}{\sqrt{1 + |\nabla f|^2}} \right) = 0$$

on Ω.

To prove the claim, note that since f is C^2, g is continuous on Ω. We prove the claim by contradiction. If g is nonzero at any point $x \in \Omega$, assume without loss of generality that $g(x) > 0$. Then by continuity, $g > 0$ in a small ball $\mathcal{B}$ centered at x. Now it is easy to find a smooth bump function h whose support is contained in $\mathcal{B}$. In this case

$$\int_\Omega hg\, dx_n = \int_{\mathcal{B}} hg\, dx_n > 0,$$

which provides the contradiction.

Thus any function f which minimizes the functional Vol satisfies the Euler-Lagrange *equation of the functional*

$$\nabla \cdot \left(\frac{\nabla f}{\sqrt{1 + |\nabla f|^2}} \right) = 0.$$

This equation is known as the minimal surface equation*.*

So a solution to our problem satisfies the minimal surface equation, and the boundary condition $f = g$ on $\partial\Omega$. This sort of boundary condition of specifying the value of a solution f is called a Dirichlet boundary condition*. The problem of finding a solution to the equation with this boundary condition is a* Dirichlet boundary value problem*. Note that the Dirichlet boundary condition is essential in making sure the variational function h vanishes on the boundary, and thus there are no boundary terms when we integrate by parts. There is another useful type of boundary condition, the* Neumann boundary condition*, in which the normal derivative $\nabla f \cdot \mathbf{n} = 0$. Notice that this also makes the integral over $\partial\Omega$ vanish in the integration by parts.*

In the previous example, we computed the Euler-Lagrange equation for Vol. There may be solutions to the Euler-Lagrange equation which are not minimizers of Vol, since we have only checked the first-derivative test. A solution to the Euler-Lagrange equation may correspond to a local maximum, a saddle point or a local but non-global minimum. We'll see below specific techniques for finding a global minimizer, which we apply in another geometric problem.

The Euler-Lagrange equations come from the first variational formula that a minimizer must satisfy: Given a family f_ϵ with $f = f_0$, then if f minimizes a functional P,

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon=0} P(f_\epsilon) = 0.$$

This is the formula of the *first variation*, which comes from the first derivative test in calculus. We may also use the second derivative test. A minimizer f as above must satisfy the *second variation* formula

$$\left.\frac{d^2}{d\epsilon^2}\right|_{\epsilon=0} P(f_\epsilon) \geq 0.$$

Homework Problem 38. *Consider a variational problem for C^2 functions $y = y(x)$ from a domain $[a, b]$ and fixed endpoints $y(a) = y_0$, $y(b) = y_1$. Assume the function is of the form*

$$J(y) = \int_a^b F(y, y')dx,$$

for F a smooth function of 2 variables.

(a) *Compute the general Euler-Lagrange equation for J.*

(b) *Multiply the Euler-Lagrange equation by y' to show that any solution to the Euler-Lagrange equation must satisfy*

$$\frac{dG}{dx} = 0$$

for a function G depending on F, y and their derivatives.

(c) *A graph $y = y(x)$ of a C^1 positive function determines a surface of revolution around the x-axis with surface area*

$$A(y) = 2\pi \int_a^b y\sqrt{1 + (y')^2}\, dx.$$

Compute the Euler-Lagrange equation for A (assume y is C^2) and compute its general solution. (The graph of this solution is called a catenary.*)*

4.2 Geodesics

Given a C^1 path $\gamma: I \to X$ for $I = [\alpha, \beta]$ an interval and $X \subset \mathbb{R}^N$ a manifold with Riemannian metric g induced from the Euclidean metric on $\mathbb{R}^N$, the *length* of the path $\gamma(I)$ is given by

$$L(\gamma) = \int_\alpha^\beta |\dot\gamma|_g\, dt = \int_\alpha^\beta \sqrt{g(\dot\gamma, \dot\gamma)}\, dt = \int_\alpha^\beta \sqrt{g_{ij}(\gamma(t))\dot\gamma^i(t)\dot\gamma^j(t)}\, dt.$$

(In the last formulation, note the use of local coordinates. So the last formulation is strictly only true when $\gamma(I)$ is contained in a single coordinate chart.) $L(\gamma)$ is called the *length functional* which take paths γ to $\mathbb{R}$.

Proposition 49. *The length of a path is independent of the parametrization. In other words, if $\tilde{\gamma}(\tau) = \gamma(t(\tau))$ for $t = t(\tau)$ a C^1 diffeomorphism onto I, then $L(\tilde{\gamma}) = L(\gamma)$.*

Proof. Let $t = t(\tau)$ with $t(\tilde{\alpha}) = \alpha$, $t(\tilde{\beta}) = \beta$. Assume that $\tilde{\alpha} < \tilde{\beta}$ and since t is a diffeomorphism, then $dt/d\tau > 0$. Then compute

$$\begin{aligned} L(\tilde{\gamma}) &= \int_{\tilde{\alpha}}^{\tilde{\beta}} \sqrt{g\left(\frac{d\tilde{\gamma}}{d\tau}, \frac{d\tilde{\gamma}}{d\tau}\right)}\, d\tau \\ &= \int_{\tilde{\alpha}}^{\tilde{\beta}} \sqrt{g\left(\frac{d\gamma}{dt}\frac{dt}{d\tau}, \frac{d\gamma}{dt}\frac{dt}{d\tau}\right)}\, d\tau \\ &= \int_{\tilde{\alpha}}^{\tilde{\beta}} \sqrt{g\left(\frac{d\gamma}{dt}, \frac{d\gamma}{dt}\right)}\, \frac{dt}{d\tau}\, d\tau \\ &= \int_{\alpha}^{\beta} \sqrt{g\left(\frac{d\gamma}{dt}, \frac{d\gamma}{dt}\right)}\, dt \\ &= L(\gamma). \end{aligned}$$

The case when $dt/d\tau < 0$ and $\tilde{\alpha} > \tilde{\beta}$ is similar. □

So this definition corresponds to the usual definition of the arc length of a parametric curve. In particular, it is invariant under change of parametrization. This particular feature turns out to cause trouble analytically. In the following sections, we'll seek to find paths minimizing arc length by constructing a sequence of paths approaching a length-minimizing one. The fact that a potentially minimizing path has many different parametrizations will make the analysis more difficult, since it will be difficult to find a sequence of paths which approaches a particular minimizing path among all the possible parametrizations. Another analytic objection to the length functional is that it is the L^1 norm of the length of the tangent vector $\dot{\gamma}$. L^2 norms tend to behave better, since we can use the structure of Hilbert spaces.

Assume for convenience that the interval $I = [0, 1]$. This can always be achieved by using a linear map to take a given I to $[0, 1]$.

Thus we introduce a related functional, the *energy* of a C^1 path $\gamma\colon [0, 1] \to X$. Define

$$E(\gamma) = \int_0^1 |\dot{\gamma}|_g^2\, dt.$$

The energy is related to the length by the following proposition.

Proposition 50. *For a given homotopy class $\mathcal{C}$ of curves $\gamma\colon [0,1] \to X$, a C^1 curve γ is the minimizes E in $\mathcal{C}$ if and only if it minimizes L among C^1 curves in $\mathcal{C}$ and the speed $|\dot\gamma(t)|_g$ is constant.*

Before we start the proof, we recall a little about homotopy classes.

Two continuous curves $\gamma_i\colon [0,1] \to X$ $i = 0, 1$ are *homotopic* if $\gamma_i(0) = p$, $\gamma_i(1) = q$ for $i = 0, 1$, and if there is a continuous function (called a *homotopy*) $G\colon [0,1] \times [0,1] \to X$ so that $G(0,t) = \gamma_0(t)$, $G(1,t) = \gamma_1(t)$ for all $t \in [0,1]$, and $G(s,0) = p$ and $G(s,1) = q$ for all $s \in [0,1]$. (More generally, if Y and X are both metric spaces, then two continuous maps $f_0, f_1 : Y \to X$ are said to be homotopic if there is a continuous map $F : [0,1] \times Y \to X$ with $F(0,y) = f_0(y)$, $F(1,y) = f_1(y)$ for all $y \in Y$. In the present case, the space $Y = [0,1]$ and we impose the extra conditions that the values at the endpoints $t = 0, 1$ are fixed at p, q respectively as well.)

Since we are measuring length and energy, we are only interested in curves γ_i which are C^1, while we allow the homotopy G to be only continuous.

Proposition 51. *The condition of two paths being homotopic is an equivalence relation, and thus we may consider* homotopy classes *of paths.)*

Proof. We need to show the property is reflexive, symmetric, and transitive. If $\gamma\colon [0,1] \to X$ is a continuous path, then it is homotopic to itself via the homotopy $G(s,t) = \gamma(t)$ for $s \in [0,1]$. This shows the reflexive property.

If γ_0 is homotopic to γ_1 via the homotopy G, then we see γ_1 is homotopic to γ_0 via the homotopy $\tilde G(s,t) = G(1-s,t)$. This shows the symmetric property.

Finally, to show the transitive property, if γ_0 is homotopic to γ_1 via a homotopy G and γ_1 is homotopic to γ_2 via a homotopy F, then we construct a homotopy from γ_0 to γ_2 by the formula

$$H(s,t) = \begin{cases} G(2s,t) & \text{for} \quad s \in [0,1/2] \\ F(2s-1,t) & \text{for} \quad s \in [1/2,1] \end{cases}$$

Note this definition is well-defined, since for $H(1/2,t) = \gamma_1(t)$ for either definition above. This observation also shows that H is continuous. It is straightforward to show H is a homotopy. □

A C^1 diffeomorphism $t = t(\tau)$ of $[0,1]$ is called *orientation preserving* if $dt/d\tau > 0$. Another fact about homotopy we'll presently use is the following

Lemma 52. *If $\tilde{\gamma}(\tau) = \gamma(t(\tau))$ for $t = t(\tau)$ an orientation-preserving diffeomorphism of $[0,1]$, then $\tilde{\gamma}$ and γ are homotopic.*

Proof. For $s, \tau \in [0,1]$, define $\psi(s,\tau) = s\tau + (1-s)t(\tau)$. Then we will show that $G(s,\tau) = \gamma(\psi(s,\tau))$ is the required homotopy. First of all, since $t(\tau)$ is an orientation-preserving diffeomorphism, we see $t(0) = 0$, $t(1) = 1$. Now check that for $s, \tau \in [0,1]$, $\psi(s,\tau) \in [0,1]$: because $0 \leq \tau \leq 1$ and $0 \leq t(\tau) \leq 1$, then

$$0 = s(0) + (1-s)0 \leq s\tau + (1-s)t(\tau) \leq s(1) + (1-s)(1) = 1.$$

This shows the homotopy G is well-defined. It is obvious for $\tau \in [0,1]$ that $G(0,\tau) = \gamma_0(\tau)$ and $G(1,\tau) = \gamma_1(\tau)$. Also compute for $s \in [0,1]$, $G(s,0) = \gamma(0)$ and $G(s,1) = \gamma(1)$. □

Also, note the following

Lemma 53. *For any C^1 path γ, $E(\gamma) \geq L(\gamma)^2$ and they are equal if and only if $|\dot{\gamma}(t)|_g$ is constant.*

Proof. Apply Hölder's inequality

$$L(\gamma) = \int_0^1 |\dot{\gamma}(t)|_g \, dt \leq \left(\int_0^1 1^2 dt \right)^{\frac{1}{2}} \left(\int_0^1 |\dot{\gamma}(t)|_g^2 \, dt \right)^{\frac{1}{2}} = \sqrt{E(\gamma)}$$

with equality if and only if 1 is proportional to $|\dot{\gamma}(t)|_g$, which is the same as $|\dot{\gamma}(t)|_g$ being constant. □

Proof of Proposition 50. Let $\gamma \in \mathcal{C}$ satisfy $E(\gamma) \leq E(\gamma')$ for all $\gamma' \in \mathcal{C}$. Given γ, let γ_c be the constant speed reparametrization of γ (this exists by Problem 39 below). Then we have by Proposition 49 and Lemma 53

$$L(\gamma_c)^2 = L(\gamma)^2 \leq E(\gamma) \leq E(\gamma_c) = L(\gamma_c)^2.$$

Thus all the inequalities in the above equation must be equalities and $L(\gamma)^2 = E(\gamma)$. Then Lemma 53 implies γ must have constant speed. So we've shown so far that if γ minimizes E, then γ has constant speed.

Let γ minimize E. For each C^1 curve $\gamma' \in \mathcal{C}$, let γ'_c be a constant speed reparametrization. Then since γ has constant speed, Lemma 53 and Proposition 49 show

$$L(\gamma)^2 = E(\gamma) \leq E(\gamma'_c) = L(\gamma'_c)^2 = L(\gamma')^2.$$

So we've shown that if γ minimizes E in $\mathcal{C}$, then γ minimizes L in $\mathcal{C}$.

We leave the converse statement as Problem 40 below. □

Homework Problem 39. *Let $\gamma : [0,1] \to X$, $\gamma = \gamma(t)$ be a path into a Riemannian manifold X. Assume $|\dot\gamma(t)|_g \neq 0$ for all $t \in [0,1]$. Show that there is a reparametrization $t(\tau)$ so that $t(0) = 0$, $t(1) = 1$, $dt/d\tau > 0$, and $\left|\frac{d\gamma}{d\tau}\right|_g$ is constant.*

Hint: Show the constant must be equal to $L(\gamma)$. Then show the condition is an ODE in $\tau = \tau(t)$. (Note that if $dt/d\tau > 0$, then $t(\tau)$ is strictly increasing and thus has an inverse on $[0,1]$.)

Homework Problem 40. *For a given homotopy class $\mathcal{C}$ of curves $\gamma : [0,1] \to X$, assume γ has constant speed $|\dot\gamma(t)|_g$ and γ minimizes L among C^1 curves in $\mathcal{C}$. Then γ minimizes E among C^1 curves in $\mathcal{C}$.*

Now we compute the first variation of the energy functional. Let γ be a smooth curve from $[0,1]$ to X so that $\gamma(0) = p$, $\gamma(1) = q$. $X \subset \mathbb{R}^N$ has the Riemannian metric pulled back from $\mathbb{R}^N$. Assume γ minimizes E in a homotopy class $\mathcal{C}$, and that γ is C^2. Then for each smooth family $\gamma_\epsilon(t)$, we have

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon=0} E(\gamma_\epsilon) = 0.$$

Consider a variation of the following special form. Near a point in $\gamma([0,1])$, pick local coordinates $x\colon \mathcal{O} \to \mathcal{U} \subset \mathbb{R}^n$. Then there is a small time interval $I = \gamma^{-1}(\mathcal{O}) \subset [0,1]$. Assume for simplicity that I doesn't contain either endpoint 0 or 1. In terms of the local coordinates x, $x(\gamma(t)) = \gamma(t) \in \mathcal{U} \subset \mathbb{R}^n$, for $t \in I$. Then let $h\colon \mathbb{R} \to \mathbb{R}^n$ be a smooth function so that $\operatorname{supp}(h) \subset\subset I$. For ϵ near 0,

$$\gamma_\epsilon(t) = \gamma(t) + \epsilon h(t) \subset \mathcal{U}$$

for $t \in I$. We define γ_ϵ outside of $\mathcal{O}$ to be simply γ. Apply the first variational formula

$$\begin{aligned}
\left.\frac{d}{d\epsilon}\right|_{\epsilon=0} E(\gamma_\epsilon) &= \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \int_0^1 g(\dot\gamma_\epsilon(t), \dot\gamma_\epsilon(t))\, dt \\
&= \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \int_I g_{ij}(\gamma(t) + \epsilon h(t))[\dot\gamma^i(t) + \epsilon \dot h^i(t)][\dot\gamma^j(t) + \epsilon \dot h^j(t)]\, dt \\
&= \int_I \left[\frac{\partial g_{ij}}{\partial x^k}(\gamma(t)) h^k(t)\right] \dot\gamma^i(t)\, \dot\gamma^j(t)\, dt \\
&\quad + \int_I g_{ij}(\gamma(t))\, \dot h^i(t)\, \dot\gamma^j(t)\, dt \\
&\quad + \int_I g_{ij}(\gamma(t))\, \dot\gamma^i(t)\, \dot h^j(t)\, dt
\end{aligned}$$

Now we integrate by parts in the last two integrals. Note that since h has compact support, all the boundary terms involving h vanish. Compute

$$\begin{aligned}\int_I g_{ij}(\gamma(t))\,\dot{h}^i(t)\,\dot{\gamma}^j(t)\,dt &= -\int_I \left[\frac{\partial g_{ij}}{\partial x^k}(\gamma(t))\dot{\gamma}^k(t)\right] h^i(t)\,\dot{\gamma}^j(t)\,dt \\ &\quad - \int_I g_{ij}(\gamma(t))\,h^i(t)\,\ddot{\gamma}^j(t)\,dt.\end{aligned}$$

We may plug this in to find for a minimizer

$$\begin{aligned} 0 &= \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} E(\gamma_\epsilon) \\ &= \int_I \left\{\frac{\partial g_{ij}}{\partial x^k} h^k\dot{\gamma}^i\dot{\gamma}^j - \frac{\partial g_{ij}}{\partial x^k}\dot{\gamma}^k h^i\dot{\gamma}^j - g_{ij}h^i\ddot{\gamma}^j - \frac{\partial g_{ij}}{\partial x^k}\dot{\gamma}^k\dot{\gamma}^i h^j - g_{ij}\ddot{\gamma}^i h^j\right\} dt \\ &= \int_I h^k \left\{\frac{\partial g_{ij}}{\partial x^k}\dot{\gamma}^i\dot{\gamma}^j - \frac{\partial g_{kj}}{\partial x^i}\dot{\gamma}^i\dot{\gamma}^j - g_{kj}\ddot{\gamma}^j - \frac{\partial g_{ik}}{\partial x^j}\dot{\gamma}^j\dot{\gamma}^i - g_{jk}\ddot{\gamma}^j\right\} dt.\end{aligned}$$

Since this is true for each h with compact support in I, then we must have for each $k = 1, \dots, n$, and for all t in the open interval I,

$$0 = \frac{\partial g_{ij}}{\partial x^k}\dot{\gamma}^i\dot{\gamma}^j - \frac{\partial g_{kj}}{\partial x^i}\dot{\gamma}^i\dot{\gamma}^j - g_{kj}\ddot{\gamma}^j - \frac{\partial g_{ik}}{\partial x^j}\dot{\gamma}^j\dot{\gamma}^i - g_{jk}\ddot{\gamma}^j.$$

Since $g_{kj} = g_{jk}$, we have

$$\begin{aligned} 0 &= g_{jk}\ddot{\gamma}^j + \frac{1}{2}\left(-\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{kj}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j}\right)\dot{\gamma}^i\dot{\gamma}^j, \\ 0 &= \ddot{\gamma}^\ell + \frac{1}{2}g^{k\ell}\left(\frac{\partial g_{kj}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right)\dot{\gamma}^i\dot{\gamma}^j \\ &= \ddot{\gamma}^\ell + \Gamma^\ell_{ij}\dot{\gamma}^i\dot{\gamma}^j, \\ \Gamma^\ell_{ij} &= \frac{1}{2}g^{k\ell}\left(\frac{\partial g_{kj}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right).\end{aligned}$$

Γ^ℓ_{ij} are called the *Christoffel symbols* of the metric g_{ij}, and

$$\ddot{\gamma}^\ell + \Gamma^\ell_{ij}\dot{\gamma}^i\dot{\gamma}^j = 0 \tag{26}$$

is called the *geodesic equation* for the metric g. Note

$$\Gamma^\ell_{ij} = \Gamma^\ell_{ji}.$$

Any curve satisfying this second-order system is called a *geodesic* on the Riemannian manifold X.

Remark. Our definition of geodesic requires a specific parametrization to solve the equation (the constant speed parametrization). Many other authors define a geodesic to be a curve which satisfies the first variational equation of arc-length. These geodesics are the same as our geodesics as subsets of the Riemannian manifold, but the parametrization is not required to be constant speed.

Note that this analysis does not work at the endpoints 0 and 1. There, we simply have the conditions $\gamma(0) = p$ and $\gamma(1) = q$ to remain in the class $\mathcal{C}$. This is essentially a Dirichlet boundary condition on the problem.

Homework Problem 41. *Compute the Euler-Lagrange equations for the length functional $L(\gamma)$. Show that any $\gamma : [0,1] \to X$ which minimizes L must satisfy*

$$\ddot{\gamma}^\ell(t) + \Gamma^\ell_{ij}(\gamma(t))\dot{\gamma}^i(t)\dot{\gamma}^j(t) = c(t)\dot{\gamma}^\ell(t)$$

for $t \in (0,1)$ and $c(t)$ a real-valued function of t.

Homework Problem 42. *Let (X, g) be an n-dimensional smooth compact Riemannian manifold. By Nash's Theorem, we may assume that $g = i^*\delta$ the pull-back of the Euclidean metric δ on $\mathbb{R}^N$ for some embedding $i \colon X \to \mathbb{R}^N$. If $(p, v) \in TX$ (i.e. $p \in X$ and $v \in T_pX$), show that the solution to the geodesic equation (26) on X with initial conditions $\gamma(0) = p$ and $\dot{\gamma}(0) = v$ exists for all time.*

Hints:

- *(a) Show that if $\gamma(t)$ solves the geodesic equation (26), then the speed $|\dot{\gamma}(t)|_g$ is constant in t.*
- *(b) Reduce the problem to the case the initial speed $|v|_{g(p)} = 1$.*
- *(c) The* unit tangent bundle *UTX is defined by*

$$UTX = \{(p, v) \in TX : |v|_{g(p)} = 1\}.$$

  Show UTX is compact as long as X is compact.

- *(d) Mimic the proof of Theorem 15 to complete the proof.*

Example 16. *Euclidean space is $\mathbb{R}^n$ with the standard* Euclidean metric *$\delta = \delta_{ij}\, dx^i\, dx^j$. In this case, all the Christoffel symbols Γ^k_{ij} vanish, since each term involves differentiating the components of the metric tensor, all of which* *are constant. Therefore, the geodesic system is simply* $\ddot{\gamma}^k = 0$*. Solutions to this ODE are simply linear functions of* t*, and so geodesics are of the form* $\gamma = tv + w$ *for* $v, w \in \mathbb{R}^n$*. So geodesics on Euclidean space are straight lines traversed at constant speed.*

Example 17. *For hyperbolic space, recall the metric* $g_{ij} = (x^n)^{-2}\delta_{ij}$ *on* $\{x \in \mathbb{R}^n : x^n > 0\}$*. Compute the Christoffel symbols:*

$$\begin{aligned} g^{ij} &= (x^n)^2\delta^{ij}, \\ g_{ij,k} &= -2(x^n)^{-3}\delta_{ij}\delta^n_k, \\ \Gamma^k_{ij} &= \tfrac{1}{2}(x^n)^2\delta^{k\ell}(g_{i\ell,j} + g_{\ell j,i} - g_{ij,\ell}) \\ &= \tfrac{1}{2}(x^n)^2\delta^{k\ell}[-2(x^n)^{-3}](\delta_{i\ell}\delta^n_j + \delta_{\ell j}\delta^n_i - \delta_{ij}\delta^n_\ell) \\ &= -(x^n)^{-1}(\delta^k_i\delta^n_j + \delta^k_j\delta^n_i - \delta^{kn}\delta_{ij}). \end{aligned}$$

Now consider i, j, k *distinct integers in* $\{1, \ldots, n\}$*.*

$$\begin{aligned} \Gamma^k_{ij} &= 0, \\ \Gamma^i_{ik} = \Gamma^i_{ki} &= -(x^n)^{-1}\delta^n_k, \\ \Gamma^k_{ii} &= (x^n)^{-1}\delta^{kn}, \\ \Gamma^i_{ii} &= -(x^n)^{-1}\delta^n_i. \end{aligned}$$

First, we look for solutions in which $\dot{\gamma}^k = 0$ *for* $k = 1, \ldots, n-1$ *(so only* γ^n *varies in* t*). It is plausible to look for such solutions since the coefficients* g_{ij} *of the metric depend only on* x^n*.*

In this case, for $k < n$*, compute*

$$\begin{aligned} 0 = \ddot{\gamma}^k &= \Gamma^k_{ij}\dot{\gamma}^i\dot{\gamma}^j \\ &= \Gamma^k_{nn}\dot{\gamma}^n\dot{\gamma}^n \\ &= (x^n)^{-1}\delta^{kn}\dot{\gamma}^n\dot{\gamma}^n = 0. \end{aligned}$$

Thus if $\dot{\gamma}^1 = \cdots = \dot{\gamma}^{n-1} = 0$*, then the geodesic equations for* $\ddot{\gamma}^k$ *for* $k < n$ *are automatically solved.*

Now compute the geodesic equation for $\ddot{\gamma}^n$*:*

$$\begin{aligned} \ddot{\gamma}^n &= -\Gamma^n_{ij}\dot{\gamma}^i\dot{\gamma}^j \\ &= -\Gamma^n_{nn}\dot{\gamma}^n\dot{\gamma}^n \\ &= (x^n)^{-1}\dot{\gamma}^n\dot{\gamma}^n, \\ &= (\gamma^n)^{-1}\dot{\gamma}^n\dot{\gamma}^n. \end{aligned} \tag{27}$$

This is a second-order nonlinear equation in γ^n, and we do not have any general technique to solve such an equation. We can, however, make some educated guesses. In particular, note that

$$(\gamma^n \dot\gamma^n)^{\cdot} = \gamma^n \ddot\gamma^n + \dot\gamma^n \dot\gamma^n,$$

and that each of these terms is similar to those in the geodesic equation (27) above.

In particular, compute for a function f of γ^n

$$\begin{aligned} 0 &= (f(\gamma^n)\dot\gamma^n)^{\cdot} && (28) \\ &= f(\gamma^n)\ddot\gamma^n + f'(\gamma^n)\dot\gamma^n\dot\gamma^n, \\ 0 &= \ddot\gamma^n + \frac{f'(\gamma^n)}{f(\gamma^n)}\dot\gamma^n\dot\gamma^n. && (29) \end{aligned}$$

This last equation is the same as the geodesic equation (27) if

$$\frac{f'(\gamma^n)}{f(\gamma^n)} = -\frac{1}{\gamma^n},$$

and this is now a first-order separable equation for f. We may solve to find $f = (\gamma^n)^{-1}$ is a solution.

Now plug into (28) to find

$$\begin{aligned} 0 &= \left(\frac{\dot\gamma^n}{\gamma^n}\right)^{\cdot}, \\ C &= \frac{\dot\gamma^n}{\gamma^n} \\ &= (\log\gamma^n)^{\cdot}, \\ Ct + D &= \log\gamma^n, \\ \gamma^n &= Ae^{Ct} \end{aligned}$$

for A a positive constant (since in hyperbolic space, we have $x^n = \gamma^n > 0$) and C any real constant. Therefore,

$$\gamma^1 = \gamma^1_0, \quad \dots, \quad \gamma^{n-1} = \gamma^{n-1}_0, \quad \gamma^n = Ae^{Ct}$$

solves the geodesic system on hyperbolic space.

So far we have only found geodesics in the special case that $\dot\gamma^1 = \cdots = \dot\gamma^{n-1} = 0$. *To find all the geodesics on hyperbolic space, we introduce the notion of an isometry of a Riemannian manifold.*

Given a Riemannian manifold (X, g), *a diffeomorphism* $\Phi\colon X \to X$ *is an* isometry *if* $\Phi^* g = g$. *Isometries of* $\mathbb{H}^n$ *are well understood, and we introduce a specific type. For* $\alpha > 0$, *let*

$$\iota_\alpha\colon x \mapsto \alpha \frac{x}{|x|^2},$$

where $x \in \mathbb{H}^n \subset \mathbb{R}^n$ *and* $|x|^2 = (x^1)^2 + \cdots + (x^n)^2$ *comes from* $\mathbb{R}^n$. *It is easy to see that* ι_α *is a diffeomorphism of* $\mathbb{H}^n$. *To show that it is an isometry, let* $y = \iota_\alpha(x)$. *Then*

$$\iota_\alpha^* g = \iota_\alpha^* \left(\frac{\sum_{j=1}^n (dy^j)^2}{|y|^2} \right).$$

Dropping the pull back ι_α^* notation, we compute

$$
\begin{aligned}
y^j &= \alpha \frac{x^j}{|x|^2}, \\
dy^j &= \frac{\partial y^j}{\partial x^i}\, dx^i, \\
&= \alpha \sum_{i=1}^{n} \frac{|x|^2 \delta_i^j - 2x^i x^j}{|x|^4}\, dx^i, \\
(dy^j)^2 &= \alpha^2 \left(\sum_{i=1}^{n} \frac{|x|^2 \delta_i^j - 2x^i x^j}{|x|^4}\, dx^i \right)^2 \\
&= \alpha^2 \left(\sum_{i=1}^{n} \frac{|x|^2 \delta_i^j - 2x^i x^j}{|x|^4}\, dx^i \right) \left(\sum_{k=1}^{n} \frac{|x|^2 \delta_k^j - 2x^k x^j}{|x|^4}\, dx^k \right) \\
&= \frac{\alpha^2}{|x|^8} \sum_{i,k=1}^{n} \left[4x^i x^k (x^j)^2 - 2|x|^2 x^i x^j \delta_k^j - 2|x|^2 x^k x^j \delta_i^j + |x|^4 \delta_i^j \delta_k^j \right] dx^i\, dx^k \\
&= \frac{\alpha^2}{|x|^8} \left\{ 4(x^j)^2 \sum_{i,k=1}^{n} x^i x^k\, dx^i\, dx^k - 2|x|^2 x^j\, dx^j \sum_{i=1}^{n} x^i\, dx^i \right. \\
&\quad \left. - 2|x|^2 x^j\, dx^j \sum_{k=1}^{n} x^k\, dx^k + |x|^4 (dx^j)^2 \right\} \\
&= \frac{\alpha^2}{|x|^8} \left\{ 4(x^j)^2 \sum_{i,k=1}^{n} x^i x^k\, dx^i\, dx^k - 4|x|^2 x^j\, dx^j \sum_{i=1}^{n} x^i\, dx^i + |x|^4 (dx^j)^2 \right\},
\end{aligned}
$$

$$
\begin{aligned}
\sum_{j=1}^{n}(dy^j)^2 &= \frac{\alpha^2}{|x|^8}\Bigg\{4\left(\sum_{j=1}^{n}(x^j)^2\right)\left(\sum_{i,k=1}^{n} x^i x^k\, dx^i\, dx^k\right) \\
&\quad -4|x|^2\sum_{i,j=1}^{n} x^j x^i\, dx^i\, dx^j + |x|^4\sum_{j=1}^{n}(dx^j)^2\Bigg\} \\
&= \frac{\alpha^2}{|x|^8}\Bigg\{4|x|^2\sum_{i,k=1}^{n} x^i x^k\, dx^i\, dx^k - 4|x|^2\sum_{i,k=1}^{n} x^i x^k\, dx^i\, dx^k \\
&\quad + |x|^4\sum_{j=1}^{n}(dx^j)^2\Bigg\} \\
&= \frac{\alpha^2}{|x|^4}\sum_{j=1}^{n}(dx^j)^2, \\
\frac{\sum_{j=1}^{n}(dy^j)^2}{(y^n)^2} &= \frac{\frac{\alpha^2}{|x|^4}\sum_{j=1}^{n}(dx^j)^2}{\alpha^2\frac{(x^n)^2}{|x|^4}} \\
&= \frac{\sum_{j=1}^{n}(dx^j)^2}{(x^n)^2}.
\end{aligned}
$$

Therefore, $\iota_\alpha^* g = g$ and ι_α is an isometry.

Moreover, it is trivial to check that any translation $x \mapsto x + x_0$ is an isometry of $\mathbb{H}^n$ if the last component $x_0^n = 0$. Also, note that the composition of two isometries is again an isometry (indeed the set of isometries of a Riemannian manifold X forms a subgroup of the diffeomorphism group called the isometry group*).*

Proposition 55 below shows that for any geodesic $\psi\colon \mathbb{R} \to \mathbb{H}^n$, then $\iota_\alpha \circ \psi$ is also a geodesic. Recall we know so far that

$$\gamma = (\gamma_0^1, \dots, \gamma_0^{n-1}, Ae^{Ct})$$

are geodesics for $A > 0$, $C \in \mathbb{R}$. Compute for $\alpha > 0$,

$$\iota_\alpha \circ \gamma = \alpha\frac{\gamma}{|\gamma|^2} = \frac{\alpha(\gamma_0^1, \dots, \gamma_0^{n-1}, Ae^{Ct})}{(\gamma_0^1)^2 + \cdots + (\gamma_0^{n-1})^2 + A^2e^{2Ct}}.$$

The image $\iota_\alpha \circ \gamma(\mathbb{R})$ is then the half-circle in $\mathbb{R}^n$ which intersects $\{x^n = 0\}$ perpendicularly at

$$0 \quad \text{and} \quad \frac{(\gamma_0^1, \dots, \gamma_0^{n-1}, 0)}{(\gamma_0^1)^2 + \cdots + (\gamma_0^{n-1})^2}.$$

Then if we apply the isometry given by adding a constant x_0 *with* $x_0^n = 0$, *then every half-circle in* $\mathbb{H}^n$ *which intersects* $\{x^n = 0\}$ *perpendicularly at both endpoints is the image of a geodesic path in* $\mathbb{H}^n$.

All together, for constants

$$\gamma_0^1, \dots, \gamma_0^{n-1}, x_0^1, \dots, x_0^{n-1}, C \in \mathbb{R}, \;\; A, \alpha > 0,$$

the path for $t \in \mathbb{R}$

$$\psi(t) = \frac{\alpha(\gamma_0^1, \dots, \gamma_0^{n-1}, Ae^{Ct})}{(\gamma_0^1)^2 + \dots + (\gamma_0^{n-1})^2 + A^2 e^{2Ct}} + (x_0^1, \dots, x_0^{n-1}, 0) \qquad (30)$$

is a geodesic in $\mathbb{H}^n$, *and the image* $\psi(\mathbb{R})$ *is a ray or a half-circle in* $\mathbb{R}^n$ *perpendicular to* $\{x^n = 0\}$. *All such rays and semicircles are represented by such geodesic paths.*

We claim that we have found all the geodesics in $\mathbb{H}^n$. *The way to check this is to recognize that the geodesic system, as a second-order ODE system with smooth coefficients, has a unique solution for each initial value problem*

$$\ddot{\gamma}^k = -\Gamma_{ij}^k \dot{\gamma}^i \dot{\gamma}^j, \quad \gamma(0) = y_0, \quad \dot{\gamma}(0) = v_0.$$

Then if we can check that every initial condition $(y_0, v_0) \in T\mathbb{H}^n$ *occurs as* $(\psi(0), \dot{\psi}(0))$ *for a geodesic* $\psi(t)$ *in (30), uniqueness of the geodesic system will imply that we have found all the geodesics in* $\mathbb{H}^n$.

So we must check that every $(y_0, v_0) \in T\mathbb{H}^n = \mathbb{H}^n \times \mathbb{R}^n$ *can be represented by* $(\psi(0), \dot{\psi}(0))$ *for a* $\psi(t)$ *in (30). For a given point* $y_0 \in \mathbb{H}^n$, *and vector* $v_0 \in T_{y_0}\mathbb{H}^n = \mathbb{R}^n$, *consider first the case when*

$$v_0^1 = \dots = v_0^{n-1} = 0.$$

In this case, we can choose $A > 0$ *and* C *so that*

$$\psi(t) = (y_0^1, \dots, y_0^{n-1}, Ae^{Ct})$$

satisfies $\psi(0) = y_0$ *and* $\dot{\psi}(0) = v_0$. *Otherwise,* y_0 *and* v_0 *span a plane* $\mathcal{P}$ *in* $\mathbb{H}^n$. *Let* $\mathcal{L} = \mathcal{P} \cap \{x^n = 0\}$. *It is straightforward to check that there is a unique semicircle in the plane* $\mathcal{P}$ *which hits* $\mathcal{L}$ *perpendicularly, passes through* y_0 *and is tangent to* v_0 *at* y_0. *This is the image of some geodesic* $\psi(t)$ *in (30). Then we can adjust* C *and* A *to ensure that* $\psi(0) = y_0$ *and* $\dot{\psi}(0) = v_0$. *Therefore, every initial condition* (y_0, v_0) *is achieved by a geodesic on our list, and we have found all the geodesics in hyperbolic space.*

The following proposition was discussed in Example 17 above.

Proposition 54. *Consider a Riemannian manifold* (X, g)*. Given* $p \in X$*,* $v \in T_pX$*, there is an* $\epsilon > 0$ *and a unique geodesic* $\gamma : (-\epsilon, \epsilon) \to X$ *with* $\gamma(0) = p$*,* $\dot\gamma(0) = v$*.*

Remark. In general, the geodesic γ may not exist for all time, although we have seen that all the geodesics on hyperbolic space (Example 17) and on compact Riemannian manifolds (Problem 42) do exist for all time.

A map $\Phi : X \to Y$ for manifolds X and Y with Riemannian metrics g and h respectively is a *local isometry* if every point in X has a neighborhood $\mathcal{O}$ on which $\Phi \colon \mathcal{O} \to \Phi(\mathcal{O}) \subset Y$ is an isometry.

Proposition 55. *If* $\Phi \colon X \to Y$ *is a local isometry of Riemannian manifolds, then for every geodesic* $\psi \colon (-\epsilon, \epsilon) \to X$*,* $\Phi\circ\psi$ *is a geodesic on* Y*. Any geodesic on* $\Phi(X) \subset Y$ *is of this form.*

Proof. In local coordinates on X and Y, we can write the isometry as $y = y(x)$. Note this is the same form as a coordinate change, and the condition that the map is an isometry is simply that the metric pulls back as a $(0, 2)$ tensor when changing coordinates.

Therefore, the proof boils down the the following fact: for a local isometry, and for any C^2 path γ, the quantity

$$w^k = \ddot\gamma^k + \Gamma^k_{ij}\dot\gamma^i\dot\gamma^j$$

transforms like a tangent vector (i.e. a $(1, 0)$ tensor) under changes of coordinates. Therefore,

$$w^k \frac{\partial}{\partial x^k} = w^k \frac{\partial y^I}{\partial x^k}\frac{\partial}{\partial y^I}$$

and $w^k(x) = 0$ for $k = 1, \ldots, n$ is equivalent to $w^I(y) = 0$ for $I = 1, \ldots, n$. This is because $\frac{\partial y^I}{\partial x^k}$ is nonsingular for $y = y(x)$ a diffeomorphism.

In order to compute how w^k transforms, we use the following index convention. Indices $i, j, k, \ldots$ are with respect to the x variables, while indices $I, J, K, \ldots$ are with respect to the y variables. For example, g_{ij} is the metric in the x coordinates, while g_{IJ} is the metric in the y coordinates.

First of all, note

$$g_{IJ} = g_{ij}\,\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^J}, \qquad g^{IJ} = g^{ij}\,\frac{\partial y^I}{\partial x^i}\frac{\partial y^J}{\partial x^j}.$$

Compute

$$\begin{aligned}
g_{IJ,K} &= \frac{\partial g_{IJ}}{\partial y^K} \\
&= \frac{\partial}{\partial y^K}\left(g_{ij}\,\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^J}\right) \\
&= \frac{\partial g_{ij}}{\partial y^K}\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^J} + g_{ij}\,\frac{\partial^2 x^i}{\partial y^I \partial y^K}\frac{\partial x^j}{\partial y^J} + g_{ij}\,\frac{\partial x^i}{\partial y^I}\frac{\partial^2 x^j}{\partial y^J \partial y^K} \\
&= g_{ij,k}\,\frac{\partial x^k}{\partial y^K}\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^J} + g_{ij}\,\frac{\partial^2 x^i}{\partial y^I \partial y^K}\frac{\partial x^j}{\partial y^J} + g_{ij}\,\frac{\partial x^i}{\partial y^I}\frac{\partial^2 x^j}{\partial y^J \partial y^K}.
\end{aligned}$$

Then compute

$$\begin{aligned}
&g_{KJ,I} + g_{IK,J} - g_{IJ,K} \\
&\quad = g_{ij,k}\,\frac{\partial x^k}{\partial y^I}\frac{\partial x^i}{\partial y^K}\frac{\partial x^j}{\partial y^J} + g_{ij}\,\frac{\partial^2 x^i}{\partial y^I \partial y^K}\frac{\partial x^j}{\partial y^J} + g_{ij}\,\frac{\partial x^i}{\partial y^K}\frac{\partial^2 x^j}{\partial y^J \partial y^I} \\
&\qquad + g_{ij,k}\,\frac{\partial x^k}{\partial y^J}\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^K} + g_{ij}\,\frac{\partial^2 x^i}{\partial y^I \partial y^J}\frac{\partial x^j}{\partial y^K} + g_{ij}\,\frac{\partial x^i}{\partial y^I}\frac{\partial^2 x^j}{\partial y^J \partial y^K} \\
&\qquad - g_{ij,k}\,\frac{\partial x^k}{\partial y^K}\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^J} - g_{ij}\,\frac{\partial^2 x^i}{\partial y^I \partial y^K}\frac{\partial x^j}{\partial y^J} - g_{ij}\,\frac{\partial x^i}{\partial y^I}\frac{\partial^2 x^j}{\partial y^J \partial y^K} \\
&\quad = g_{ij,k}\,\frac{\partial x^k}{\partial y^I}\frac{\partial x^i}{\partial y^K}\frac{\partial x^j}{\partial y^J} + g_{ij,k}\,\frac{\partial x^k}{\partial y^J}\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^K} - g_{ij,k}\,\frac{\partial x^k}{\partial y^K}\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^J} \\
&\qquad + 2\,g_{ij}\,\frac{\partial^2 x^i}{\partial y^I \partial y^J}\frac{\partial x^j}{\partial y^K}.
\end{aligned}$$

Then the Christoffel symbols

$$\begin{aligned}
\Gamma^L_{IJ} &= \tfrac{1}{2}\, g^{KL}(g_{KJ,I} + g_{IK,J} - g_{IJ,K}) \\
&= \tfrac{1}{2}\, g^{m\ell}\, \frac{\partial y^K}{\partial x^m}\frac{\partial y^L}{\partial x^\ell}\left(g_{ij,k}\, \frac{\partial x^k}{\partial y^I}\frac{\partial x^i}{\partial y^K}\frac{\partial x^j}{\partial y^J} + g_{ij,k}\, \frac{\partial x^k}{\partial y^J}\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^K} \right. \\
&\quad \left. - g_{ij,k}\, \frac{\partial x^k}{\partial y^K}\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^J} + 2\, g_{ij}\, \frac{\partial^2 x^i}{\partial y^I \partial y^J}\frac{\partial x^j}{\partial y^K} \right) \\
&= \tfrac{1}{2}\, g^{m\ell}\, \frac{\partial y^L}{\partial x^\ell}\left(g_{mj,k}\, \frac{\partial x^k}{\partial y^I}\frac{\partial x^j}{\partial y^J} + g_{im,k}\, \frac{\partial x^k}{\partial y^J}\frac{\partial x^i}{\partial y^I} \right. \\
&\quad \left. - g_{ij,m}\, \frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^J} + 2\, g_{im}\, \frac{\partial^2 x^i}{\partial y^I \partial y^J} \right) \\
&= \Gamma^\ell_{ij}\, \frac{\partial y^L}{\partial x^\ell}\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^J} + \frac{\partial y^L}{\partial x^\ell}\frac{\partial^2 x^\ell}{\partial y^I \partial y^J}.
\end{aligned}$$

Note that the second term in the last formula shows that the Christoffel symbols do not transform as a tensor. In fact, this is fortunate, as the extra non-tensorial term will cancel out a similar term coming from the second derivative $\ddot{\gamma}^k$.

Note that

$$\begin{aligned}
\dot{\gamma}^I &= \frac{\partial y^I}{\partial x^i}\dot{\gamma}^i, \\
\ddot{\gamma}^L &= \frac{d}{dt}\left(\frac{\partial y^L}{\partial x^\ell}(\gamma)\, \dot{\gamma}^\ell \right) \\
&= \frac{\partial y^L}{\partial x^\ell}\, \ddot{\gamma}^\ell + \frac{\partial^2 y^L}{\partial x^\ell \partial x^j}\, \dot{\gamma}^j \dot{\gamma}^\ell.
\end{aligned}$$

Compute

$$\begin{aligned}
\Gamma^L_{IJ}\dot{\gamma}^I\dot{\gamma}^J &= \left(\Gamma^\ell_{ij}\, \frac{\partial y^L}{\partial x^\ell}\frac{\partial x^i}{\partial y^I}\frac{\partial x^j}{\partial y^J} + \frac{\partial y^L}{\partial x^\ell}\frac{\partial^2 x^\ell}{\partial y^I \partial y^J} \right) \dot{\gamma}^m \frac{\partial y^I}{\partial x^m}\dot{\gamma}^p \frac{\partial y^J}{\partial x^p} \\
&= \Gamma^\ell_{ij}\dot{\gamma}^i\dot{\gamma}^j \frac{\partial y^L}{\partial x^\ell} + \frac{\partial^2 x^k}{\partial y^I \partial y^J}\frac{\partial y^L}{\partial x^k}\frac{\partial y^I}{\partial x^j}\frac{\partial y^J}{\partial x^\ell}\dot{\gamma}^j\dot{\gamma}^\ell.
\end{aligned}$$

Therefore, $\ddot{\gamma}^L + \Gamma^L_{IJ}\dot{\gamma}^I\dot{\gamma}^J$ will transform like a tensor if we can show that the non-tensorial terms cancel: We need to show

$$\frac{\partial^2 y^L}{\partial x^\ell \partial x^j} + \frac{\partial^2 x^k}{\partial y^I \partial y^J}\frac{\partial y^L}{\partial x^k}\frac{\partial y^I}{\partial x^j}\frac{\partial y^J}{\partial x^\ell} = 0. \tag{31}$$

This equation follows from the formula for the first derivative of an inverse matrix. If $\dot{A}$ represents the first derivative of a matrix A (with respect to any parameter or variable), then

$$(A^{-1})^{\cdot} = -A^{-1}\dot{A}A^{-1}.$$

(Proof: Differentiate the equation $AA^{-1} = I$ to find $\dot{A}A^{-1} + A(A^{-1})^{\cdot} = 0$.) Then since $(\partial y^L/\partial x^\ell)$ is the inverse matrix of $(\partial x^\ell/\partial y^L)$,

$$\begin{aligned}
\frac{\partial^2 y^L}{\partial x^\ell \partial x^j} &= \frac{\partial}{\partial x^j}\left(\frac{\partial y^L}{\partial x^\ell}\right) \\
&= -\frac{\partial y^L}{\partial x^k}\left[\frac{\partial}{\partial x^j}\left(\frac{\partial x^k}{\partial y^J}\right)\right]\frac{\partial y^J}{\partial x^\ell} \\
&= -\frac{\partial y^L}{\partial x^k}\left[\frac{\partial y^I}{\partial x^j}\frac{\partial}{\partial y^I}\left(\frac{\partial x^k}{\partial y^J}\right)\right]\frac{\partial y^J}{\partial x^\ell}.
\end{aligned}$$

Upon plugging in, this proves formula (31) and the proposition. □

Remark. There is also a more geometric proof of the previous proposition. Recall that we derived the geodesic equation as the Euler-Lagrange equation of the energy functional. So any path which minimizes the energy satisfies the geodesic equation. It is easy to see that the energy of a path is invariant under an isometry; therefore, the notion of energy-minimizing path is invariant under isometries.

The problem is that there are geodesics which do not minimize the energy. (They may be saddle points of the energy functional.) This can be surmounted by restricting to small domains by using the following fact from Riemannian geometry: Every point in a Riemannian manifold has a neighborhood $\mathcal{O}$ so that all geodesic paths in $\mathcal{O}$ are energy-minimizing for endpoints in $\mathcal{O}$. (In Riemannian geometry books, this fact is usually stated in terms of the length functional instead; to translate to the present situation, recall that energy-minimizing paths are length-minimizing paths parametrized with constant speed.)

Homework Problem 43. *Given a smooth function on a Riemannian manifold, the Hessian of f is defined locally by the formula*

$$H(f)_{ij} = \frac{\partial^2 f}{\partial x^i \partial x^j} - \Gamma^k_{ij}\frac{\partial f}{\partial x^k}.$$

Show that the Hessian of f is a symmetric $(0,2)$ tensor.

Homework Problem 44. *Compute all the geodesics on* $\mathbb{S}^2$.

Hint: Use the expression for the metric in local coordinates (y^1, y^2) *from Example 13. Compute the Christoffel symbols. Analyze the case when* $y^2 = 0$ *and only* y^1 *varies. Solve the resulting second-order ODE for* $\gamma^1 = y^1$. *Then move these geodesics around via the isometry group of* $\mathbb{S}^2$.

(The isometry group of $\mathbb{S}^2$ *is given by the orthogonal group of* 3×3 *matrices*

$$O(3) = \{A : AA^\top = I\}.$$

Show that each such linear action is an isometry of $\mathbb{R}^3$ *which takes the unit sphere* $\mathbb{S}^2$ *to itself. For every line* $\mathcal{L}$ *though the origin in* $\mathbb{R}^3$, *show that rotating by an angle* θ *around the line* $\mathcal{L}$ *is a linear map in* $O(3)$. *Show that every initial condition* $(p, v) \in T\mathbb{S}^2$ *of the geodesic equation on* $\mathbb{S}^2$ *can be realized by the examples you computed above, when acted on by such a rotation in* $O(3)$.*)*

4.3 The direct method: An example

We have computed the Euler-Lagrange equations of the energy functional. Now we introduce an example of the direct method in the calculus of variations.

The direct method is this: Given a functional $E : \mathcal{C} \to \mathbb{R}$, if there is a lower bound $I = \inf_{\gamma \in \mathcal{C}} E(\gamma) > -\infty$, then there is a sequence of paths γ_i so that $E(\gamma_i) \to I$. The direct method is to show that there is a subsequence of $\{\gamma_i\}$ which converges to some γ, and to show that the limiting $\gamma \in \mathcal{C}$ and that $E(\gamma) = I$. Thus we have constructed a minimizer γ over the class $\mathcal{C}$ of the functional E. There are subtle points to deal with along the way. Typically, the class $\mathcal{C}$ is a closed subset of a Banach space, and in passing to the limit of a subsequence, the limit γ we construct may be in a weaker Banach space (for example, a sequence in C^1 may produce a limit only in C^0, which will be problematic if the functional involves any derivatives). A related issue is that in passing to the limit $\gamma_{i_j} \to \gamma$, we may not have $E(\gamma_{i_j}) \to E(\gamma)$. In particular, below we will have to deal with the situation in which we only know $\lim_{j\to\infty} E(\gamma_{i_j}) \geq E(\gamma)$—so that the functional is only *lower semi-continuous* under the limit. Thus we will typically need to spend time improving the regularity of the limit γ and showing some semi-continuity of the functional under the limiting subsequence.

The direct method of the calculus of variations is very useful in solving elliptic PDEs. The problem we approach involves geodesics, and thus the solution we produce be a solution to an ODE. This will allow us to proceed with much of the general picture of the calculus of variations while avoiding some of the more technical points. In particular, we will learn about distributions, weak derivatives, Hilbert spaces, and compact maps between Banach spaces in solving our problem.

Given a smooth manifold X, a *loop* is a continuous map from the circle $\mathbb{S}^1$ to X. Each such loop is equivalent to a continuous map $\gamma\colon \mathbb{R} \to X$ which is periodic in the sense that $\gamma(t+1) = \gamma(t)$ for all $t \in \mathbb{R}$. We will abuse notation by using the same γ for $\gamma : \mathbb{S}^1 \to X$ and the periodic $\gamma : \mathbb{R} \to X$. (This is because $\mathbb{S}^1$ is naturally the quotient $\mathbb{R}/\mathbb{Z}$, where $\mathbb{Z}$ acts on $\mathbb{R}$ by adding integers to real numbers.) Two loops $\gamma_0, \gamma_1 : \mathbb{S}^1 \to X$ are *freely homotopic* if there is a continuous homotopy

$$G\colon [0,1] \times \mathbb{S}^1 \to X, \qquad G(0,t) = \gamma_0(t), \quad G(1,t) = \gamma_1(t).$$

The condition of being freely homotopic is an equivalence relation, and thus each loop on a manifold X is a member of a *free homotopy class.*

Here is our problem:

Problem: Find a curve of least length in a free homotopy class of loops on a compact Riemannian manifold.

The problem may have no solution on a noncompact Riemannian manifold. There may be loops of arbitrarily small length in a given nontrivial free homotopy class, corresponding to a loops slipping off a narrowing end of the manifold.

Homotopy classes are objects defined by continuity, and the following result should come as no surprise.

Proposition 56. *For a smooth compact manifold $X \subset \mathbb{R}^N$, there is an $\epsilon > 0$ so that if two loops $\gamma_0, \gamma_1 : \mathbb{S}^1 \to X \subset \mathbb{R}^N$ satisfy*

$$\|\gamma_0 - \gamma_1\|_{C^0(\mathbb{S}^1,\mathbb{R}^N)} < \epsilon,$$

then γ_0 and γ_1 are homotopic as loops in X.

Proof. We apply the ϵ-Neighborhood Theorem (19): For $\epsilon > 0$, let X^ϵ be the open subset of $\mathbb{R}^N$ consisting of all points distance less than ϵ from X. There is a $\epsilon > 0$ small enough so that every point in X^ϵ has a unique closest point in X. Then the map $\pi : X^\epsilon \to X$ which sends a point in X^ϵ to its closest point in X is a smooth map of X^ϵ to X, and it fixes each point in $X \subset X^\epsilon$.

Let γ_0 and γ_1 be loops on X satisfying

$$\|\gamma_0 - \gamma_1\|_{C^0(\mathbb{S}^1,\mathbb{R}^N)} < \epsilon.$$

Then consider the homotopy in $\mathbb{R}^N$

$$\tilde{G}(s,t) = (1-s)\gamma_0(t) + s\gamma_1(t) \in \mathbb{R}^N.$$

For $s,t \in [0,1]$, the distance in $\mathbb{R}^N$

$$|\tilde{G}(s,t) - \gamma_0(t)| = s|\gamma_0(t) - \gamma_1(t)| < 1 \cdot \epsilon.$$

So $\tilde{G}(s,t) \in X^\epsilon$ for all $s,t \in [0,1]$, and we may define a homotopy in X by

$$G(s,t) = \pi(\tilde{G}(s,t)).$$

□

Remark. The homotopy $G(s,t)$ constructed is a smooth homotopy if γ_0 and γ_1 are smooth. Thus the same theorem works with smooth homotopy classes (as considered in Guillemin and Pollack).

Corollary 57. *If γ_i are a sequence of loops in a free homotopy class in $X \subset \mathbb{R}^N$, and*

$$\lim_{i\to\infty} \|\gamma_i - \gamma\|_{C^0(\mathbb{S}^1,\mathbb{R}^N)} = 0,$$

then the loop γ is in the same free homotopy class.

Proof. For the $\epsilon > 0$ of Proposition 56 above, there is a γ_i so that

$$\|\gamma_i - \gamma\|_{C^0(\mathbb{S}^1,\mathbb{R}^N)} < \epsilon.$$

Apply Proposition 56 to show γ and γ_i are in the same free homotopy class.

□

The ϵ-Neighborhood Theorem, together with the mollifier technique of approximation, allow us to prove an important foundational result in topology:

Theorem 20. *Let $f : \mathbb{R}^n \to Y$ be uniformly continuous, where $Y \subset \mathbb{R}^N$ is a compact submanifold without boundary. Then f is homotopic to a smooth map from $\mathbb{R}^n \to Y$.*

Proof. Since f is uniformly continuous, for all $\epsilon > 0$, there is a $\delta > 0$ so that if $|x - x'| < \delta$, then $|f(x) - f(x')| < \epsilon$. The ϵ-Neighborhood Theorem shows that there is an $\epsilon > 0$ so that the map $\pi \colon Y^\epsilon \to Y$ is well-defined and smooth. Let δ be the corresponding δ from the uniform continuity of f.

Let ρ be a smooth nonnegative bump function with support in the unit ball $B_1(0)$ in $\mathbb{R}^n$ so that $\int_{\mathbb{R}^n} \rho \, dx_n = 1$. Then for $\alpha > 0$, define $\rho_\alpha(x) = \alpha^{-n}\rho(x/\alpha)$. Note $\operatorname{supp} \rho_\alpha = B_\alpha(0)$. Define

$$f^\alpha(x) = \int_{\mathbb{R}^n} f(y)\rho_\alpha(x-y)\, dy_n = \int_{\{y:|x-y|\le\alpha\}} f(y)\rho_\alpha(x-y)\, dy_n.$$

(Note each f^α is $\mathbb{R}^N$-valued.) If $\alpha < \delta$, then $|f(y) - f(x)| < \epsilon$ for y in the domain of integration, and so

$$\begin{aligned} f^\alpha(x) &= \int_{\{y:|x-y|\le\alpha\}} f(y)\rho_\alpha(x-y)\, dy_n \\ &= \int_{\{y:|x-y|\le\alpha\}} [f(y)-f(x)]\rho_\alpha(x-y)\, dy_n \\ &\quad + \int_{\{y:|x-y|\le\alpha\}} f(x)\rho_\alpha(x-y)\, dy_n \\ &= \int_{\{y:|x-y|\le\alpha\}} [f(y)-f(x)]\rho_\alpha(x-y)\, dy_n + f(x) \end{aligned}$$

since

$$\int_{\{y:|x-y|\le\alpha\}} \rho_\alpha(x-y)\, dy_n = \int_{\mathbb{R}^n} \rho_\alpha(x-y)\, dy_n = \int_{\mathbb{R}^n} \rho_\alpha(z)\, dz_n = 1$$

for the substitution $z = x - y$. So

$$\begin{aligned} |f^\alpha(x) - f(x)| &= \left| \int_{\{y:|x-y|\le\alpha\}} [f(y)-f(x)]\rho_\alpha(x-y)\, dy_n \right| \qquad (32) \\ &\le \int_{\{y:|x-y|\le\alpha\}} |f(y)-f(x)|\rho_\alpha(x-y)\, dy_n \\ &< \epsilon \int_{\{y:|x-y|\le\alpha\}} \rho_\alpha(x-y)\, dy_n = \epsilon. \end{aligned}$$

Therefore if $\alpha \in (0,\delta)$, then $f^\alpha(x) \in Y^\epsilon$. Then we check that $\tilde{f}^\alpha(x) = \pi(f^\alpha(x))$ is the desired homotopy. In particular, as $\alpha \to 0$, $\tilde{f}^\alpha(x) \to f(x)$ uniformly by (32) (view ϵ as varying to zero instead of fixed for this interpretation). Since π and f^α are smooth, then $\tilde{f}^\alpha$ is smooth for small $\alpha > 0$. In particular, we have shown that

$$F(\alpha, x) = \begin{cases} \tilde{f}^\alpha(x) & \text{for } \alpha > 0 \text{ small} \\ f(x) & \text{for } \alpha = 0 \end{cases}$$

is the desired homotopy. □

Theorem 21. *Let $f: X \to Y$ be a continuous map between smooth manifolds. Then f is homotopic to a smooth map from $X \to Y$.*

Sketch of proof. We may assume $X \subset \mathbb{R}^M$ by Whitney's Embedding Theorem. Then there is a $\nu > 0$ so that $\pi_M : X^\nu \to X$ is well-defined and smooth. Define $g\colon \mathbb{R}^M \to \mathbb{R}^N$ by $g(p) = f(\pi_M(p))$ for $p \in X^\nu$ and $g(p) = 0$ for $p \notin X^\nu$. Note $g(p)$ is uniformly continuous on a neighborhood of X. Apply the mollifier argument as above to g and show that the homotopy constructed in the proof of Theorem 20, when restricted to $X \subset \mathbb{R}^M$, has the desired properties. □

The discussion above about energy and length still holds. Assuming the minimizer is smooth enough, then a constant-speed length-minimizing loop is the same as an energy-minimizing loop. Thus we may as well consider energy-minimizing loops, and we have the equivalent problem.

Problem: Find a curve of least energy in a free homotopy class of loops on a compact Riemannian manifold.

So far in our discussion, the formulation of length and energy depend on the loop γ being C^1 (so that the derivative $\dot{\gamma}$ is C^0 and thus can be integrated). If we look more closely, the energy is defined as the L^2 norm of $\dot{\gamma}$

$$E(\gamma) = \int_0^1 |\dot{\gamma}|_g^2 \, dt.$$

Therefore, we really do not need $\dot{\gamma}$ to be continuous, but only L^2. In terms of γ itself, we need to develop a theory of how to take a derivative which ends up not being continuous, but only L^2. For this purpose, we define derivatives in the sense of distributions, or *weak derivatives.*

4.4 Distributions

On $\mathbb{R}^n$, we consider each smooth function ϕ with compact support to be a *test function*. For any C^1 function f on $\mathbb{R}^n$ and test function ϕ, we have the following formula by integrating by parts:

$$\int_{\mathbb{R}^n} f_{,i}\phi \, dx_n = -\int_{\mathbb{R}^n} f\phi_{,i} \, dx_n. \tag{33}$$

For two locally L^1 functions f and h on $\mathbb{R}^n$, we say $f_{,i} = h$ *in the sense of distributions* if for all test functions ϕ,

$$\int_{\mathbb{R}^n} h\phi \, dx_n = -\int_{\mathbb{R}^n} f\phi_{,i} \, dx_n.$$

Let $\mathcal{D}(\mathbb{R}^n)$ be the vector space of all smooth functions with compact support in $\mathbb{R}^n$. A *distribution* on $\mathbb{R}^n$ is a linear map from $\mathcal{D}(\mathbb{R}^n) \to \mathbb{R}$. (We often allow $\mathbb{C}$-valued test functions and consider complex linear maps to $\mathbb{C}$; complex-valued functions are useful when doing Fourier analysis.) Recall a measurable function f is locally L^1 if over every compact subset K of the domain of f, $\int_K |f| < \infty$. Any locally L^1 function f on $\mathbb{R}^n$ gives a distribution by sending

$$f : \phi \mapsto f(\phi) = \int_{\mathbb{R}^n} f\phi \, dx_n.$$

Notice that there is a slight abuse of notation: $f(\phi)$ for ϕ a test function is not to be confused with $f(x)$ for $x \in \mathbb{R}^n$. Two locally L^1 functions f_1, f_2 are said to be equal in the sense of distributions if for every test function ϕ,

$$\int_{\mathbb{R}^n} f_1\phi \, dx_n = \int_{\mathbb{R}^n} f_2\phi \, dx_n \quad \Longleftrightarrow \quad \int_{\mathbb{R}^n} (f_1 - f_2)\phi \, dx_n = 0.$$

Remark. On $\mathbb{R}^N$, note that any locally L^p function for $p \geq 1$ is also locally L^1. This is because for $K \subset\subset \mathbb{R}^n$, $\frac{1}{p} + \frac{1}{q} = 1$, and f locally L^p, Hölder's inequality states

$$\int_K |f| \, dx_n \leq \left(\int_K 1 \, dx_n \right)^{\frac{1}{q}} \left(\int_K |f|^p \, dx_n \right)^{\frac{1}{p}} < \infty.$$

Example 18. *Any locally finite Borel measure $d\mu$ on $\mathbb{R}^n$ defines a distribution by sending*

$$\phi \mapsto \int_{\mathbb{R}^n} \phi \, d\mu$$

for any test function ϕ.

An important example of this is the inaptly named δ*-function, or unit point mass, at the origin. The* δ*-function is a measure on* $\mathbb{R}^n$ *so that for any subset* $\Omega \subset \mathbb{R}^n$,

$$\delta(\Omega) = \begin{cases} 1 & \textit{if} \quad 0 \in \Omega \\ 0 & \textit{if} \quad 0 \notin \Omega. \end{cases}$$

So the distribution defined by this measure is

$$\delta \colon \phi \mapsto \phi(0),$$

which is just evaluation of ϕ *at the origin. The following problem shows there is no locally* L^1 *function which is equal to the* δ*-function.*

Homework Problem 45. *Show that there is no* L^1 *function* f *on* $\mathbb{R}^n$ *so that*

$$\int_{\mathbb{R}^n} f\phi \, dx_n = \phi(0) \quad \textit{for all } \phi \in \mathcal{D}(\mathbb{R}^n).$$

Hint: Consider a smooth nonnegative function $\rho \colon \mathbb{R}^n \to \mathbb{R}$ *with support in* $B_1(0)$ *the unit ball centered at 0 and so that* $\int_{\mathbb{R}^n} \rho \, dx_n = 1$*. Use this* ρ *to define* $\rho_\epsilon(x) = \epsilon^{-n}\rho(x/\epsilon)$*. If there were such an* L^1 *function* f*, recall that if*

$$f^\epsilon(x) = \int_{\mathbb{R}^n} f(y)\rho_\epsilon(x-y)\, dy_n,$$

then $f^\epsilon \to f$ *in* L^1 *as* $\epsilon \to 0$.

(a) *Show that for all* $x \neq 0$ *that* $f^\epsilon(x) = 0$ *for* ϵ *small enough. (Follow the proof of Proposition 58.)*

(b) *Suppose a family of continuous functions* $f^\epsilon \to f$ *in* $L^1(\mathbb{R}^n)$ *as* $\epsilon \to 0^+$*, and let* $\mathcal{O} \subset \mathbb{R}^n$ *be a measurable subset on which* $f^\epsilon = 0$ *identically on* $\mathcal{O}$ *for all* ϵ *sufficiently small. Show that* $f = 0$ *almost everywhere on* $\mathcal{O}$*. (Split up the relevant integrals on* $\mathbb{R}^n$ *into integrals on* $\mathcal{O}$ *and* $\mathbb{R}^n \setminus \mathcal{O}$*.)*

(c) *Show our* $f = 0$ *almost everywhere on* $\mathbb{R}^n$.

(d) *Find a contradiction.*

We have just seen that distributions are more general than functions. In particular, it is possible to differentiate any distribution by mimicking formula (33). A distributional derivative of a function may no longer be a function, but it will be well-defined as a distribution. Given a distribution f defined by a map $f\colon \phi \mapsto f(\phi) \in \mathbb{R}$, the partial derivative $f_{,i}$ *in the sense of distributions* is defined to be the distribution

$$f_{,i}\colon \phi \mapsto -f(\phi_{,i}).$$

It is this innovation which allows us to define the derivatives of L^2 functions.

Remark. Note that the equation (33) motivating the distributional derivative is essentially the same as the integration by parts used to calculate the Euler-Lagrange equations for $\gamma + \epsilon h$. Thus if h is smooth with compact support, we can still integrate by parts even if γ is no longer regular enough for ordinary differentiation; we simply consider the derivatives to be taken in the sense of distributions.

Homework Problem 46. *Consider the Heaviside function*

$$h(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0. \end{cases}$$

Show that the derivative h' (taken in the sense of distributions) is the δ function on $\mathbb{R}$.

Homework Problem 47. *Consider for any test function $\phi \in \mathcal{D}(\mathbb{R})$,*

$$PV\left(\frac{1}{x}\right)(\phi) = \lim_{\epsilon \to 0^+} \left(\int_{-\infty}^{-\epsilon} \frac{1}{x}\,\phi(x)\,dx + \int_{\epsilon}^{\infty} \frac{1}{x}\,\phi(x)\,dx \right).$$

Part (a) *shows that $PV(\frac{1}{x})$ is a distribution. It is called the* principal value *of $\frac{1}{x}$.*

(a) Show $PV(\frac{1}{x})(\phi)$ converges for all smooth test functions ϕ. (Hint: The potential problem is clearly at $x = 0$. Use Taylor's Theorem to write $\phi = \phi(0) + O(x)$, where $O(x)$ represents a term so that $O(x)/x$ converges to a real limit as $x \to 0$.)

(b) Show that the first derivative in the sense of distributions of $PV(\frac{1}{x})$ is given in terms of $\phi \in \mathcal{D}(\mathbb{R})$ as

$$\lim_{\epsilon \to 0^+} \left[\int_{-\infty}^{-\epsilon} \left(-\frac{1}{x^2}\right) \phi(x)\,dx + \int_{\epsilon}^{\infty} \left(-\frac{1}{x^2}\right) \phi(x)\,dx + \frac{2}{\epsilon}\,\phi(0) \right].$$

One more thing is needed to complete the picture of distributions as generalizations of functions. Recall that every locally L^p function for $p \geq 1$ defines a distribution. The following proposition shows this map is injective.

Proposition 58. *If two locally L^1 functions f_1 and f_2 on $\mathbb{R}^n$ define the same distribution, then $f_1 = f_2$ almost everywhere.*

Proof. We first consider the case when f_1 and f_2 are both globally L^1 on $\mathbb{R}^n$. Then recall that we can use a mollifier to approximate each in L^1 by smooth functions. In particular, if ρ is a smooth nonnegative function with compact support so that $\int_{\mathbb{R}^n} \rho \, dx_n = 1$, then define

$$\rho_\epsilon(x) = \frac{1}{\epsilon^n} \rho\left(\frac{x}{\epsilon}\right), \quad f_i^\epsilon(x) = \int_{\mathbb{R}^n} \rho_\epsilon(x-y) f_i(y) \, dy_n, \quad i = 1, 2.$$

Then each f_i^ϵ is a smooth L^1 function on $\mathbb{R}^n$ and $f_i^\epsilon \to f_i$ in L^1 as $\epsilon \to 0$. Now for each fixed $x \in \mathbb{R}^n$, $\rho_\epsilon(x-y)$ is a smooth test function with compact support in y, and $f_i^\epsilon(x)$ is simply the evaluation of this test function by the distribution f_i. Since $f_1 = f_2$ in the sense of distributions, then $f_1^\epsilon(x) = f_2^\epsilon(x)$ for all $x \in \mathbb{R}^n$. So then

$$\|f_1 - f_2\|_{L^1} = \lim_{\epsilon \to 0} \|f_1^\epsilon - f_2^\epsilon\|_{L^1} = \lim_{\epsilon \to 0} 0 = 0.$$

Then $f_1 = f_2$ in L^1, which is equivalent to $f_1 = f_2$ almost everywhere.

If f_1 and f_2 are only locally L^1, consider a smooth function β_R with compact support which is identically equal to 1 on $B_R = \{|x| \leq R\}$. It is easy to check that the condition $f_1 = f_2$ in the sense of distributions implies $\beta_R f_1 = \beta_R f_2$ in the sense of distributions. Then since each f_i is locally L^1, each $\beta_R f_i$ is globally in L^1. We apply the argument of the previous paragraph; so $\beta_R f_1 = \beta_R f_2$ almost everywhere on $\mathbb{R}^n$. This implies that $f_1 = f_2$ almost everywhere on the ball B_R. Now let $R \to \infty$ to conclude that $f_1 = f_2$ almost everywhere on $\mathbb{R}^n$. $\square$

So far, we have discussed distributions on $\mathbb{R}^n$. On the circle $\mathbb{S}^1$, the definitions are similar, the main difference being that since $\mathbb{S}^1$ is compact, our test functions are simply all smooth functions on $\mathbb{S}^1$. In particular, we can think of test functions on $\mathbb{S}^1$ as smooth periodic functions on $\mathbb{R}$ with period 1. In this way, an L^1 function f on $\mathbb{S}^1$ acts on test functions by

$$f : \phi \to \int_0^1 f\phi \, dt.$$

One thing to check is that integration by parts still works. If f is C^1 on $\mathbb{S}^1$ and ϕ is smooth on $\mathbb{S}^1$, then

$$\begin{aligned}\int_{\mathbb{S}^1} \dot{f}\phi\,dt &= \int_0^1 \dot{f}\phi\,dt \\ &= -\int_0^1 f\dot{\phi}\,dt + (f\phi)\Big|_0^1 \\ &= -\int_{\mathbb{S}^1} f\dot{\phi}\,dt + f(1)\phi(1) - f(0)\phi(0) \\ &= -\int_{\mathbb{S}^1} f\dot{\phi}\,dt\end{aligned}$$

because $f(0) = f(1)$ and $\phi(0) = \phi(1)$ since f and ϕ are periodic. So we have the same basic formula as in (33), and we may define distributions and distributional derivatives in the same manner as above.

Now we return to our problem. We want to consider all loops $\gamma\colon \mathbb{S}^1 \to X \subset \mathbb{R}^N$ so that

$$E(\gamma) = \int_0^1 |\dot{\gamma}|_g^2\,dt = \|\dot{\gamma}\|^2_{L^2(\mathbb{S}^1,\mathbb{R}^N)} < \infty.$$

Therefore, we consider the *Sobolev space*

$$L^2_1(\mathbb{S}^1,\mathbb{R}^N) = \{\gamma\colon \mathbb{S}^1 \to \mathbb{R}^N : \|\gamma\|^2_{L^2_1} = \|\gamma\|^2_{L^2} + \|\dot{\gamma}\|^2_{L^2} < \infty\},$$

where the derivative $\dot{\gamma}$ is taken in the sense of distributions. Note that $\gamma \in L^2(\mathbb{S}^1,\mathbb{R}^N)$ implies that $\dot{\gamma}$, when defined in the sense of distributions, may be represented as a function (and an L^2 function at that).

We may consider each component $\gamma^1,\ldots,\gamma^N$ separately, and it should be clear that $\gamma_i \to \gamma$ in $L^2_1(\mathbb{S}^1,\mathbb{R}^N)$ if and only if each $\gamma^a_i \to \gamma^a$ in $L^2_1(\mathbb{S}^1,\mathbb{R})$ for each $a = 1,\ldots,N$. Thus we may work with each component of γ separately in $\mathbb{R}^N$. Below we will see that L^2_1 is a Hilbert space, but for now we are content to show that every function in $L^2_1(\mathbb{S}^1)$ is continuous. Recall that elements of $L^2_1(\mathbb{S}^1)$ are only equivalence classes of functions, two functions being equivalent if they agree almost everywhere.

Proposition 59. *Every element of $L^2_1(\mathbb{R})$ contains a continuous representative.*

Remark. This proposition is an important example of the Sobolev embedding theorem, which gives a means to embed Sobolev spaces $L^p_k(\mathbb{R}^n)$ into appropriate C^ℓ spaces $\ell = \ell(p,k,n)$. In particular, the present result depends strongly on the fact that the dimension of the domain $\mathbb{R}$ of the functions is one. (There are elements of $L^2_1(\mathbb{R}^2)$ which do not have continuous representatives.)

Proof. Let $f \in L^2_1(\mathbb{R})$. So $\int_{\mathbb{R}} |\dot f|^2\,dt = C^2 < \infty$. Then compute for $t_2 \geq t_1$

$$\begin{aligned} |f(t_2) - f(t_1)| &= \left| \int_{t_1}^{t_2} \dot f(t)\,dt \right| \\ &\leq \left(\int_{t_1}^{t_2} |\dot f(t)|^2\,dt \right)^{\frac12} \left(\int_{t_1}^{t_2} dt \right)^{\frac12} \\ &\leq C(t_2 - t_1)^{\frac12}. \end{aligned}$$

So this formula shows f is continuous, as long as we can justify using the Fundamental Theorem of Calculus

$$f(t_2) - f(t_1) = \int_{t_1}^{t_2} \dot f(t)\,dt.$$

We achieve this by defining $g(t) = \int_0^t \dot f(s)\,ds$. The previous argument implies that g is continuous. Now we argue that there is a constant K so that $f - g = K$ almost everywhere. This will show there is an continuous representative $g + K$ in the equivalence class of f.

First we show that $\dot g = \dot f$ in the sense of distributions. Consider a test function ϕ. Then

$$\begin{aligned} \dot g(\phi) &= -\int_{-\infty}^{\infty} g(t)\dot\phi(t)\,dt \\ &= -\int_{-\infty}^{\infty} \left(\int_0^t \dot f(s)\,ds \right) \dot\phi(t)\,dt \\ &= -\int_{R_1} \dot f(s)\dot\phi(t)\,ds\,dt + \int_{R_2} \dot f(s)\dot\phi(t)\,ds\,dt \end{aligned}$$

by Fubini's Theorem, for the regions in the plane

$$R_1 = \{(s,t) : s \geq 0, t \geq s\}, \qquad R_2 = -R_1.$$

Then again by Fubini, and since ϕ has compact support,

$$\begin{aligned}
\dot{g}(\phi) &= -\int_0^\infty \left(\int_s^\infty \dot{\phi}(t)\,dt\right)\dot{f}(s)\,ds + \int_{-\infty}^0 \left(\int_{-\infty}^s \dot{\phi}(t)\,dt\right)\dot{f}(s)\,ds \\
&= -\int_0^\infty (-\phi(s))\dot{f}(s)\,ds + \int_{-\infty}^0 \phi(s)\dot{f}(s)\,ds \\
&= \int_{-\infty}^\infty \phi(s)\dot{f}(s)\,ds \\
&= \dot{f}(\phi).
\end{aligned}$$

Therefore, $\dot{g} = \dot{f}$ in the sense of distributions.

The following proposition, applied to $f-g$, shows that there is a constant K so that $f = g + K$ in the sense of distributions. Then Proposition 58 above shows $f = g + K$ almost everywhere, and thus there is a continuous representative in the equivalence class of f. □

Proposition 60. *If a distribution h on $\mathbb{R}$ satisfies $\dot{h} = 0$ in the sense of distributions, then there is a constant K so that $h = K$ as distributions.*

Proof. Let ϕ be a test function with integral $\int_{\mathbb{R}} \phi\,dt = 1$. Let $K = h(\phi)$. Then for a test function ψ with $\int_{\mathbb{R}} \psi\,dt = L$, compute

$$h(\psi) = h(\psi - L\phi) + Lh(\phi) = h(\psi - L\phi) + LK.$$

But now

$$\int_{-\infty}^\infty (\psi - L\phi)\,dt = L - L\cdot 1 = 0,$$

and thus the function

$$\chi(t) = \int_{-\infty}^t [\psi(s) - L\phi(s)]ds \qquad (34)$$

is a smooth function with compact support—Proof: Let $\mathrm{supp}(\psi - \mathrm{L}\phi) \subset [\mathrm{T}, \mathrm{T}']$. It is clear that $\chi(t) = 0$ for $t < T$. For $t > T'$, note that $\chi'(t) = \psi(t) - L\phi(t) = 0$ and so χ is constant on (T', ∞). Then (34) shows that $\chi(t) \to 0$ as $t \to \infty$, and so $\chi = 0$ on (T', ∞).

Then since $\dot{\chi} = \psi - L\phi$,

$$h(\psi) = LK + h(\psi - L\phi) = LK + h(\dot{\chi}) = LK - \dot{h}(\chi) = LK$$

since $\dot{h} = 0$ in the sense of distributions. But then

$$h(\psi) = LK = K \int_{\mathbb{R}} \psi \, dt = \int_{\mathbb{R}} K\psi \, dt.$$

and $h = K$ as distributions. □

Homework Problem 48. *Prove Propositions 59 and 60 above for distributions on $\mathbb{S}^1$ instead of on $\mathbb{R}$. Here are the key steps:*

(a) Let $f : \mathbb{S}^1 \to \mathbb{R}$ be an L^2 function, and assume that the distributional derivative $\dot{f}$ is L^2 as well. Represent f and $\dot{f}$ as periodic functions from $\mathbb{R} \to \mathbb{R}$. For any $t \in \mathbb{R}$, define

$$g(t) = \int_0^t \dot{f}(s) \, ds.$$

Show that g is periodic and continuous (and so defines a continuous function on $\mathbb{S}^1$.) Note that the constant function 1 is a test function on $\mathbb{S}^1$.

(b) Show that $\dot{f} = \dot{g}$ in the sense of distributions. In other words, for every smooth periodic test function $\phi \in \mathcal{D}(\mathbb{S}^1)$, show that

$$\int_0^1 \dot{f}\phi \, dt = -\int_0^1 g\dot{\phi} \, dt.$$

(c) If h is a distribution on $\mathbb{S}^1$ which satisfies $\dot{h} = 0$ in the sense of distributions, show there is a constant K so that $h = K$ as distributions. In other words, show that for every periodic smooth $\psi \colon \mathbb{R} \to \mathbb{R}$,

$$h(\psi) = \int_0^1 K\psi \, dt.$$

Now since any L^2_1 map from $\mathbb{S}^1 \to X \subset \mathbb{R}^N$ is continuous, each one is in a free homotopy class of loops on X. With that in mind, we formulate our final version of the problem:

For $X \subset \mathbb{R}^N$ a smooth submanifold with Riemannian metric pulled back from the Euclidean metric on $\mathbb{R}^N$, define

$$L^2_1(\mathbb{S}^1, X) = \{\gamma \in L^2_1(\mathbb{S}^1, \mathbb{R}^N) : \gamma(\mathbb{S}^1) \subset X\}.$$

Here we assume that γ is continuous, as we may by Proposition 59 above.

Problem: Let $X \subset \mathbb{R}^N$ be a smooth compact manifold equipped with the Riemannian metric pulled back from the Euclidean metric on $\mathbb{R}^N$. Let $\mathcal{C}$ be the class of loops $\gamma\colon \mathbb{S}^1 \to X$ in a free homotopy class on X and in $L^2_1(\mathbb{S}^1, X)$. Find a loop of least energy in $\mathcal{C}$.

Proposition 61. *Let $\gamma \in L^2_1(\mathbb{S}^1, X)$ be energy minimizing in a free homotopy class on X for $X \subset \mathbb{R}^N$ a smooth manifold without boundary. Then γ solves the geodesic equation*

$$g_{k\ell}\left(\ddot{\gamma}^k + \Gamma^k_{ij}\dot{\gamma}^i\dot{\gamma}^j\right) = 0$$

for all $\ell = 1, \ldots, n$, in the sense of distributions.

Proof. First of all, note that we can choose γ to be continuous by Problem 48 above. Thus it makes sense that γ is in a free homotopy class. Since γ minimizes energy, then for each h smooth with compact support so that $\gamma(\operatorname{supp} h) \subset\subset$ a single coordinate chart in X, that

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon=0} E(\gamma + \epsilon h) = 0.$$

Compute the first variation as in the derivation of the Euler-Lagrange equations in Subsection 4.2 above:

$$\int_{\mathbb{S}^1} g_{ij,k}\, h^k\, \dot{\gamma}^i\, \dot{\gamma}^j\, dt + \int_{\mathbb{S}^1} g_{ij}\, \dot{h}^i\, \dot{\gamma}^j\, dt + \int_{\mathbb{S}^1} g_{ij}\, \dot{\gamma}^i\, \dot{h}^j\, dt = 0.$$

Since the components of h are smooth with compact support, they act as test functions, and we may then integrate by parts in the second and third integrals, in the sense of distributions, to conclude that

$$g_{k\ell}\left(\ddot{\gamma}^k + \Gamma^k_{ij}\dot{\gamma}^i\dot{\gamma}^j\right) = 0$$

in the sense of distributions. □

Remark. In the previous proposition, we cannot immediately remove the metric term $g_{k\ell} = g_{k\ell}(\gamma)$, because we only know that $g_{k\ell}(\gamma)$ is continuous in t (since γ is continuous in t by Proposition 59). In general, we cannot multiply a distribution by a continuous function—in this case, the inverse matrix $g^{\ell m}(\gamma)$—and get another distribution (this only works if the distribution is induced from a Borel measure). See the following homework problem.

Homework Problem 49. *Note that if $\lambda\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth function, and f is a locally L^1 function, then the product λf is also a locally L^1 function.*

(a) If $\lambda\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth function, and p is a distribution on $\mathbb{R}^n$, then show that it is possible to define the product λp in such a way that if p is induced from a locally L^1 function, then λp is induced from the usual product of two functions.

(b) Let δ be the δ-function on $\mathbb{R}$. Compute its first derivative $\dot{\delta}$ in the sense of distributions.

(c) Show that if $g\colon \mathbb{R} \to \mathbb{R}$ is a continuous function which is not differentiable at 0, then the formula for the product developed in part (a) above does not give a reliable answer for the product $g\dot{\delta}$ of the continuous function g and the distribution $\dot{\delta}$.

4.5 Hilbert spaces

Recall that a *Hilbert space* is a Banach space whose norm comes from a positive definite inner product. We now show that $L^2_1(\mathbb{S}^1, \mathbb{R})$ is a Hilbert space. Recall that $L^2_1(\mathbb{S}^1, \mathbb{R})$ consists of all L^2 functions on $\mathbb{S}^1$ whose derivative in the sense of distributions is also L^2. This suggests a natural inner product:

$$\langle f, h\rangle_{L^2_1} = \int_{\mathbb{S}^1} fh\, dt + \int_{\mathbb{S}^1} \dot{f}\dot{h}\, dt.$$

Then plug in $f = h$ to find

$$\|f\|^2_{L^2_1} = \int_{\mathbb{S}^1} |f|^2\, dt + \int_{\mathbb{S}^1} |\dot{f}|^2\, dt = \langle f, f\rangle_{L^2_1},$$

and so the norm on L^2_1 is induced by the inner product. Below in Corollary 67, we show that any positive definite inner product defines a norm.

Remark. $L^2_1(\mathbb{S}^1, \mathbb{R}^N)$ is also naturally a Hilbert space, with inner product given by

$$\langle f, h\rangle_{L^2_1} = \int_{\mathbb{S}^1} \langle f, h\rangle\, dt + \int_{\mathbb{S}^1} \langle \dot{f}, \dot{h}\rangle\, dt,$$

where $\langle\cdot,\cdot\rangle$ is the inner product on $\mathbb{R}^N$.

It is also useful to define complex Hilbert spaces, in which the inner product $\langle\cdot,\cdot\rangle$ is *Hermitian* and positive definite. A Hermitian inner product on a complex vector space V is a map from $V \times V \to \mathbb{C}$ which satisfies for $\lambda \in \mathbb{C}$ and $f, g, h \in V$,

$$\begin{aligned} \langle \lambda f + g, h \rangle &= \lambda \langle f, h \rangle + \langle g, h \rangle, \\ \langle f, \lambda g + h \rangle &= \bar{\lambda} \langle f, g \rangle + \langle f, h \rangle, \\ \langle f, g \rangle &= \overline{\langle g, f \rangle}. \end{aligned}$$

These three conditions are respectively that the inner product is *complex linear* in the first slot, *complex antilinear* in the second slot, and *skew-symmetric*. The first two conditions together are called *sesquilinear*.

Then $L^2_1(\mathbb{S}^1, \mathbb{C})$ is a complex Hilbert space with inner product

$$\langle f, g \rangle = \int_{\mathbb{S}^1} f \bar{g} \, dt + \int_{\mathbb{S}^1} \dot{f} \overline{\dot{g}} \, dt.$$

We can also define the Sobolev space $L^2_1(\mathbb{R}^n, \mathbb{R})$ by the inner product

$$\langle f, g \rangle = \int_{\mathbb{R}^n} f g \, dx_n + \sum_{i=1}^n \int_{\mathbb{R}^n} f_{,i} g_{,i} \, dx_n,$$

the derivatives taken in the sense of distributions. The elements of $L^2_1(\mathbb{R}^n, \mathbb{R})$ are then equivalence classes of functions in L^2 so that all the first partials in the sense of distributions are also in L^2.

We will work with $L^2_1(\mathbb{S}^1, \mathbb{R})$ instead of $L^2_1(\mathbb{S}^1, \mathbb{R}^N)$, since convergence in $L^2_1(\mathbb{S}^1, \mathbb{R}^N)$ is equivalent to each component converging in $L^2_1(\mathbb{S}^1, \mathbb{R})$. The proofs that follow will work with minor modifications for the spaces $L^2_1(\mathbb{S}^1, \mathbb{R}^N)$ and $L^2_1(\mathbb{S}^1, \mathbb{C})$.

We focus on $L^2_1(\mathbb{S}^1, \mathbb{R})$, which we refer to simply as L^2_1.

Proposition 62. *$L^2_1(\mathbb{S}^1, \mathbb{R})$ is a Hilbert space.*

Proof. We've exhibited an inner product on L^2_1, and it is easy to check that it is positive definite (if we consider elements to be equivalence classes of functions, two functions being equivalent if they agree almost everywhere). Thus the remaining thing to check is that the metric $L^2_1(\mathbb{S}^1, \mathbb{R})$ is complete (and so it is a Banach space).

First of all note that $f_n \to f$ in L^2_1 is equivalent to $f_n \to f$ in L^2 and $\dot{f}_n \to \dot{f}$ in L^2.

Let f_n be a Cauchy sequence in L^2_1. Then by the definition of the norm, it is clear that f_n and $\dot{f}_n$ are both Cauchy sequences in L^2. Then we have limits $f_n \to f$ and $\dot{f}_n \to g$ in L^2. In order to show that $f_n \to f$ in L^2_1, it suffices to show that $\dot{f} = g$ in the sense of distributions.

Let ϕ be a test function, and note that $f_n \to f$ in L^2 implies by Hölder's inequality that

$$|f_n(\phi) - f(\phi)| = \left| \int_{\mathbb{S}^1} (f_n - f)\phi \, dt \right| \leq \|f_n - f\|_{L^2} \|\phi\|_{L^2} \to 0$$

as $n \to \infty$. We use this fact for both $f_n \to f$ and $\dot{f}_n \to g$ to compute for a test function ϕ

$$\begin{aligned} g(\phi) = \int_{\mathbb{S}^1} g\phi \, dt &= \lim_{n\to\infty} \int_{\mathbb{S}^1} \dot{f}_n \phi \, dt \\ &= -\lim_{n\to\infty} \int_{\mathbb{S}^1} f_n \dot{\phi} \, dt \\ &= -\int_{\mathbb{S}^1} f\dot{\phi} \, dt \\ &= -f(\dot{\phi}) = \dot{f}(\phi). \end{aligned}$$

Therefore, $g = \dot{f}$ in the sense of distributions. □

Remark. Essentially the same proof shows that $L^2_1(\mathbb{R}^n, \mathbb{R}^m)$ is a Hilbert space.

A *Hilbert basis* of a Hilbert space is a generalization of the idea of an orthonormal basis. For a real Hilbert space H, a Hilbert basis is a collection of elements $\{e_\alpha\}_{\alpha\in A}$ which are orthonormal in that

$$\langle e_\alpha, e_\beta \rangle = \delta_{\alpha\beta}$$

and so that every element $v \in H$ can be written as

$$v = \sum_{\alpha\in A} v^\alpha e_\alpha$$

for $v^\alpha \in \mathbb{R}$. Here A is an index set, which may be finite, countably infinite, or uncountable (and of course the convergence of any infinite sum is controlled by the norm). A Hilbert space which has a countable (finite or infinite) Hilbert basis is called *separable*. The following is true:

Proposition 63. *Every Hilbert space has a Hilbert basis. In fact, every orthonormal set in a Hilbert space can be completed to a Hilbert basis.*

We omit the proof, which is similar to the proof of the corresponding fact for vector spaces (any linearly independent set can be completed to a basis). In particular, Zorn's Lemma is needed in the case of non-separable Hilbert spaces. But see Problem 54 below for a proof of this Proposition for separable Hilbert spaces, and for a discussion of how this special case is adequate for the proofs of the results in this section.

Theorem 22 (Pythagorean Theorem). *If $v, w \in H$ a Hilbert space, and $\langle v, w\rangle = 0$, then*

$$\|v\|^2 + \|w\|^2 = \|v + w\|^2.$$

Proof. Compute

$$\|v + w\|^2 = \langle v + w, v + w\rangle = \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle = \|v\|^2 + \|w\|^2.$$

□

Lemma 64 (Bessel's Inequality). *If $\{e_1, \ldots, e_n\}$ is a finite orthonormal set in H, then for all $y \in H$,*

$$\|y\|^2 \geq \sum_{i=1}^{n} |\langle y, e_i\rangle|^2.$$

Proof. Check that for $w = \sum_{i=1}^{n}\langle y, e_i\rangle e_i$, $\langle y - w, w\rangle = 0$. Then apply the Pythagorian Theorem to $y = (y - w) + w$, and note that $\|w\|^2 = \sum_{i=1}^{n} |\langle y, e_i\rangle|^2$. □

Corollary 65. *If e_i is a countable orthonormal set, then*

$$\|y\|^2 \geq \sum_{i=1}^{\infty} |\langle y, e_i\rangle|^2.$$

Proof. Use Bessel's Inequality and take limits of partial sums. □

Theorem 23. *Given a Hilbert space H with a Hilbert basis $\{e_\alpha\}_{\alpha\in A}$, for every element $v \in H$,*

$$v = \sum_{\alpha\in A}\langle v, e_\alpha\rangle e_\alpha, \tag{35}$$

$$\|v\|_H^2 = \langle v, v\rangle = \sum_{\alpha\in A} |\langle v, e_\alpha\rangle|^2, \tag{36}$$

where the (possibly uncountable) sums are defined by using Homework Problem 50 below. Moreover, if there are $v^\alpha \in \mathbb{R}$ so that $\sum_{\alpha\in A} |v^\alpha|^2 < \infty$, then $v = \sum_{\alpha\in A} v^\alpha e_\alpha$ converges to an element of H.

Remark. For each $v \in H$, only a countable number of the coefficients $v^\alpha = \langle v, e_\alpha\rangle$ are nonzero. This is due to the following fact:

Homework Problem 50. *Let A be an uncountable set, and for each $\alpha \in A$, let $x_\alpha \geq 0$.*

(a) If $A' \subset A$ is a finite set, let $S_{A'} = \sum_{\alpha\in A'} x_\alpha$. Show that if the set

$$\{S_{A'} : A' \subset A \text{ is a finite set}\}$$

is bounded, then $x_\alpha = 0$ for all but countably many $\alpha \in A$.

(b) Use part (a) *to define*

$$\sum_{\alpha\in A} x_\alpha = \sup\{S_{A'} : A' \subset A \text{ is a finite set}\}$$

as an element of $[0,\infty]$ for any $x_\alpha \geq 0$. In particular, if the sum is finite, show that

$$\sum_{\alpha\in A} x_\alpha = \sum_{\alpha\in\tilde{A}} x_\alpha,$$

where $\tilde{A} = \{\alpha \in A : x_\alpha > 0\}$ is countable. Show that if $\tilde{A}$ is infinite, the right-hand sum is the usual sum of a convergent countably infinite series (for any bijection between $\tilde{A}$ and the natural numbers).

Hint for (a)*: Each $x_\alpha > 0$ satisfies $x_\alpha \in [2^n, 2^{n+1})$ for some $n \in \mathbb{Z}$. Derive a contradiction if the number of positive x_α is uncountable.*

Remark. Note that

$$\sum_{\alpha\in A} x_\alpha = \int_A x\, dc$$

for dc the *counting measure* on A. If $A' \subset A$, then the counting measure $c(A') = |A'|$ the cardinality of A' (and so $c(A') = +\infty$ when A' is infinite).

Proof of Theorem 23. First assume that $v^\alpha \in \mathbb{R}$ and $\sum_{\alpha\in A} |v^\alpha|^2 < \infty$. Then Homework Problem 50 above shows that all but countably many of v^α are zero, and so we may write v as a countable sum $\sum_{i=1}^\infty v^i e_i$. Let $v_n = \sum_{i=1}^n v^i e_i$. Then for $n > m$

$$\|v_n - v_m\|^2 = \left\| \sum_{i=m+1}^n v^i e_i \right\|^2 = \sum_{i=m+1}^n |v^i|^2 \le \sum_{i=m+1}^\infty |v^i|^2.$$

Here, the second equality is by the Pythagorean Theorem. Since the series $\sum_{i=1}^\infty |v^i|^2$ converges, the tail of the series $\sum_{i=m+1}^\infty |v^i|^2$ must go to zero as $m \to \infty$, and thus $\{v_n\}$ is a Cauchy sequence in H. Since H is complete, v_n converges to the limit $v \in H$.

Now let $v \in H$ and $v^\alpha = \langle v, e_\alpha \rangle$. Then Bessel's Inequality shows that for all finite subsets $A' \subset A$, that

$$\sum_{\alpha\in A'} |v^\alpha|^2 \le \|v\|^2.$$

So for the collection $\{|v^\alpha|^2\}_{\alpha\in A}$, the set S of finite partial sums is bounded. So Homework Problem 50 shows that all but countably many $v^\alpha = 0$. Denumerate the countable number of nonzero terms as $v^1, v^2, \dots$, and the corresponding elements of the Hilbert basis as $e_1, e_2, \dots$.

Since the sequence $\sum_{i=1}^N |v^i|^2$ is bounded and increasing, it has a finite limit as $N \to \infty$. We have shown above that the series $\sum_{i=1}^\infty v^i e_i$ converges to a limit $v' \in H$. Compute

$$\langle v - v', e_i \rangle = \lim_{n\to\infty} \left\langle v - \sum_{j=1}^n v^j e_j, e_i \right\rangle = v^i - v^j \delta_j^i = 0.$$

And for any $e_\alpha \notin \{e_1, e_2, \dots\}$, compute

$$\langle v - v', e_\alpha \rangle = \lim_{n\to\infty} \left\langle v - \sum_{j=1}^n v^j e_j, e_\alpha \right\rangle = 0.$$

So for all e_α in the Hilbert basis,

$$\langle v, e_\alpha \rangle = \langle v', e_\alpha \rangle = \left\langle \sum_{i=1}^\infty v^i e_i, e_\alpha \right\rangle = v^\alpha.$$

Now the definition of Hilbert basis shows that there are $\tilde{v}^\alpha \in \mathbb{R}$ so that $\sum_{\alpha\in A} \tilde{v}^\alpha e_\alpha = v$. By the same analysis as above, all but countably many v^α are zero, and we may write $v = \sum_{i=1}^\infty \tilde{v}^i e_i$. Moreover, as in the previous paragraph,

$$v^\alpha = \langle v, e_\alpha\rangle = \left\langle \sum_{i=1}^\infty \tilde{v}^i e_i, e_\alpha \right\rangle = \tilde{v}^\alpha$$

and so (35) is proved.

To prove (36), note that (35) shows that $v = \lim_{n\to\infty} v_n$ in H, for $v_n = \sum_{i=1}^n \langle v, e_i\rangle e_i$. Since the norm is continuous, then

$$\|v\|^2 = \lim_{n\to\infty} \|v_n\|^2 = \lim_{n\to\infty} \sum_{i=1}^n |\langle v, e_i\rangle|^2 = \sum_{i=1}^\infty |\langle v, e_i\rangle|^2 = \sum_{\alpha\in A} |\langle v, e_\alpha\rangle|^2.$$

This concludes the proof of the theorem. □

Corollary 66. *If* $v = \sum_{i=1}^\infty v^i e_i$, $w = \sum_{i=1}^\infty w^i e_i$ *for* $\{e_i\}$ *a Hilbert basis of a separable Hilbert space, then*

$$\langle v, w\rangle = \sum_{i=1}^\infty v^i w^i.$$

Proof. Compute

$$\begin{aligned} \|v+w\|^2 &= \|v\|^2 + 2\langle v, w\rangle + \|w\|^2, \\ \langle v, w\rangle &= \tfrac{1}{2}[\|v+w\|^2 - \|v\|^2 - \|w\|^2] \\ &= \tfrac{1}{2}\sum_{i=1}^\infty [(v^i+w^i)^2 - (v^i)^2 - (w^i)^2] \\ &= \sum_{i=1}^\infty v^i w^i. \end{aligned}$$

□

Remark. The formula for a complex Hilbert space is

$$\langle v, w\rangle = \sum_{i=1}^\infty v^i \overline{w^i}.$$

Remark. Homework Problem 50 shows that this result still holds for non-separable Hilbert spaces, since the number of basis elements with nonzero coefficients for v and/or w is countable.

Here is another basic result in Hilbert spaces:

Homework Problem 51 (Cauchy-Schwartz Inequality). *If $v, w \in H$ a real Hilbert space, then $|\langle v, w\rangle| \leq \|v\|\|w\|$, and there is equality if and only if v and w are linearly dependent.*

Hint: Use calculus to compute the minimum value of $\|tv + w\|^2$ as a function of t, and note the minimum value must be nonnegative.

Remark. The Cauchy-Schwartz Inequality is also true for complex Hilbert spaces, but for the proof, note that the minimum value of $\|te^{i\theta}v + w\|^2$, for $t \in \mathbb{R}$ and θ so that $e^{i\theta}\langle v, w\rangle = |\langle v, w\rangle|$, is nonnegative.

Corollary 67. *Any positive definite inner product on a real vector space V produces a norm by the formula $\|v\|^2 = \langle v, v\rangle$.*

Proof. The main thing to check is the triangle inequality. Let $v, w \in V$ and note that

$$
\begin{aligned}
&\|v + w\| \leq \|v\| + \|w\| \\
&\qquad \iff \quad \|v + w\|^2 \leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \\
&\qquad \iff \quad \|v\|^2 + 2\langle v, w\rangle + \|w\|^2 \leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \\
&\qquad \iff \quad \langle v, w\rangle \leq \|v\|\|w\|.
\end{aligned}
$$

□

The main results we will use regarding Hilbert spaces involve another topology on the Hilbert space which is different from the topology defined by the metric. The usual metric convergence of sequences is called *strong convergence*. So a sequence $v_i \to v$ in H *strongly* if

$$\|v_i - v\|_H \to 0.$$

On the other hand, a sequence $v_i \in H$ is *weakly convergent* to a limit $v \in H$ if

$$\langle v_i, w\rangle \to \langle v, w\rangle \qquad \text{for all } w \in H.$$

If $v_i \to v$ strongly, then $v_i \to v$ weakly (Homework Problem 52 below), but the converse is not true in general, as the following example shows:

Example 19. *Let H be a Hilbert space with a countably infinite Hilbert basis $e_1, e_2, \dots$. Then $e_i \to 0$ weakly in H, but $\{e_i\}$ does not converge strongly.*

Proof. Let $w \in H$. Then since $\|w\|^2 = \sum_{i=1}^{\infty} |\langle w, e_i\rangle|^2 < \infty$, we must have each term $|\langle w, e_i\rangle|^2 \to 0$ as $i \to \infty$. This shows $e_i \to 0$ weakly in H as $i \to \infty$.

To show $\{e_i\}$ does not converge strongly, note that

$$\|e_i - e_j\|_H = \sqrt{2} \qquad \text{for } i \neq j$$

by the Pythagorean Theorem. Thus $\{e_i\}$ cannot be a Cauchy sequence in H, and thus cannot converge strongly. □

Homework Problem 52. *Show that if $v_i \to v$ converges strongly in a Hilbert space H, then $v_i \to v$ weakly in H.*

Hint: Use Cauchy-Schwartz.

Theorem 24. *Let $\{v_i\}$ be a sequence in a Hilbert space H satisfying $\|v_i\| \leq K$ for a uniform constant K. Then there is a weakly convergent subsequence to a limit v which satisfies $\|v\| \leq K$. In other words, the closed ball of radius K is compact in the weak topology on H.*

Proof. Let $\{e_\alpha\}_{\alpha \in A}$ be a Hilbert basis. Problem 50 shows that for each of $\{v_1, v_2, \dots\}$, only a countable subset $A_{v_1}, A_{v_2}, \dots \subset A$ have nonzero coefficients in the Hilbert decomposition. Then the union

$$\bigcup_{i=1}^{\infty} A_{v_i}$$

is also countable, and it represents all the basis elements with nonvanishing coefficients for all the v_i. Denumerate these elements as $e_1, e_2, \dots$, and write

$$v_i = \sum_{j=1}^{\infty} v_i^j e_j.$$

Since there is a constant K so that $\|v_i\| \leq K$, then Theorem 23 shows for each N

$$\sum_{j=1}^{N} |v_i^j|^2 \leq K^2. \tag{37}$$

Thus, since the interval $[-K, K] \subset \mathbb{R}$ is compact, there is a subsequence $\{{}_1v_i\}$ of $\{v_i\}$ so that

$$\lim_{i\to\infty} {}_1v_i^1 = v^1 \in [-K, K].$$

Now there is a subsequence $\{{}_2v_i\}$ of $\{{}_1v_i\}$ so that

$$\lim_{i\to\infty} {}_2v_i^1 = v^1, \quad \lim_{i\to\infty} {}_2v_i^2 = v^2, \quad |v^1|^2 + |v^2|^2 \le K^2.$$

This is because ${}_1v_i^2 \in [-K, K]$, which is compact, and the bound follows from (37). Recursively, we may define for each N a subsequence $\{{}_Nv_i\}$ and a real number v^N so that

$$\begin{aligned} &\{{}_Nv_i\} \text{ is a subsequence of } \{{}_{(N-1)}v_i\}, \\ &\lim_{i\to\infty} {}_Nv_i^j = v^j \text{ for } j = 1, \dots, N, \qquad (38) \\ &|v^1|^2 + \dots + |v^N|^2 \le K^2. \qquad (39) \end{aligned}$$

We use a diagonalization procedure to find a weakly convergent subsequence. $\{{}_iv_i\}$ is a subsequence of $\{v_i\}$, and we will show that it converges weakly to $v = \sum_{i=1}^\infty v^i e_i$ and $v \in H$. Note by construction that $\{{}_iv_i^j\} \to v^j$ as $i \to \infty$ for each $j = 1, 2, \dots$. This is because, for each j, $\{{}_iv_i\}_{i=j}^\infty$ is a subsequence of $\{{}_jv_i\}_{i=1}^\infty$ (after the j^{th} term at least) and by condition (38).

That $v \in H$ follows directly from (39) and Theorem 23. Now we show ${}_iv_i \to v$ weakly in H. Let $w \in H$, and let $\epsilon > 0$. Write

$$\begin{aligned} |\langle {}_iv_i, w\rangle - \langle v, w\rangle| &= |\langle {}_iv_i - v, w\rangle| \\ &\le \sum_{\alpha\in A} |({}_iv_i^\alpha - v^\alpha)w^\alpha| \\ &= \sum_{j=1}^\infty |({}_iv_i^j - v^j)w^j| \\ &\le \sum_{j=1}^n |({}_iv_i^j - v^j)w^j| + \sum_{j=n+1}^\infty |({}_iv_i^j - v^j)w^j| \end{aligned}$$

Here the third line follows from the second since ${}_iv_i^\alpha = v^\alpha = 0$ if $e_\alpha \notin \{e_1, e_2, \dots\}$.

Since $\|{}_iv_i\| \le K$ and $\|v\| \le K$, then $\|{}_iv_i - v\| \le 2K$ and Cauchy-Schwartz shows that

$$\begin{aligned}\sum_{j=n+1}^{\infty} |({}_iv_i^j - v^j)w^j| &\le \left(\sum_{j=n+1}^{\infty} |{}_iv_i^j - v^j|^2\right)^{\frac{1}{2}} \left(\sum_{j=n+1}^{\infty} |w^j|^2\right)^{\frac{1}{2}} \\ &\le \left(\sum_{j=1}^{\infty} |{}_iv_i^j - v^j|^2\right)^{\frac{1}{2}} \left(\sum_{j=n+1}^{\infty} |w^j|^2\right)^{\frac{1}{2}} \\ &\le 2K \left(\sum_{j=n+1}^{\infty} |w^j|^2\right)^{\frac{1}{2}}\end{aligned}$$

Since $w \in H$, $\sum_{j=1}^{\infty} |w^j|^2 \le \sum_{\alpha \in A} |w^\alpha|^2 = \|w\|^2$ converges, and there is an n so that

$$\left(\sum_{j=n+1}^{\infty} |w^j|^2\right)^{\frac{1}{2}} < \epsilon.$$

Now for $j = 1, 2, \dots, n$, each ${}_iv_i^j \to v^j$ as $i \to \infty$. So we may choose an I so that for all $i \ge I$, $|{}_iv_i^j - v_j| < \epsilon$. Therefore, for $i \ge I$,

$$\begin{aligned}|\langle {}_iv_i, w\rangle - \langle v, w\rangle| &\le \sum_{j=1}^{n} |({}_iv_i^j - v^j)w^j| + \sum_{j=n+1}^{\infty} |({}_iv_i^j - v^j)w^j| \\ &\le \epsilon(|w^1| + \dots + |w^n|) + 2K\epsilon\end{aligned}$$

Since n, K, and $|w^1| + \dots + |w^n|$ are independent of i (note that dependence on w is allowed for weak convergence), $\langle {}_iv_i, w\rangle \to \langle v, w\rangle$ as $i \to \infty$ and thus ${}_iv_i \to v$ weakly in H. □

Theorem 25. *Let $v_i \to v$ weakly in a Hilbert space. Then*

$$\|v\| \le \liminf_{i\to\infty} \|v_i\|.$$

In other words, the Hilbert space norm is lower semicontinuous under weak convergence.

Proof. The proof is to translate the current problem into Fatou's Lemma. Let $\{e_\alpha\}_{\alpha\in A}$ be a Hilbert basis of our Hilbert space H. Then put the *counting measure* c on the index set A. Let $f : A \to [0, \infty)$, $f : \alpha \mapsto f_\alpha$ be a nonnegative real-valued function on A. Then it is straightforward to check that

$$\int_A f\,dc = \sum_{\alpha\in A} f_\alpha,$$

and thus each sum may be thought of as an integral with respect to the counting measure.

In our case, if

$$v_i = \sum_{\alpha\in A} v_i^\alpha e_\alpha, \qquad v = \sum_{\alpha\in A} v^\alpha e_\alpha,$$

we may view v_i as a function from $A \to \mathbb{R}$ by $v_i : \alpha \to v_i^\alpha$. (The same holds for v.) Theorem 23 shows

$$\|v_i\|^2 = \sum_{\alpha\in A} |v_i^\alpha|^2, \qquad \|v\|^2 = \sum_{\alpha\in A} |v^\alpha|^2. \tag{40}$$

Now since $v_i \to v$ weakly, then

$$v_i^\alpha = \langle v_i, e_\alpha\rangle \to \langle v, e_\alpha\rangle = v^\alpha$$

as $i \to \infty$ for all α. Thus with respect to the counting measure on A, $v_i \to v$ everywhere on A. Thus each terms in the sums in (40) is nonnegative, and for each α, $|v_i^\alpha|^2 \to |v^\alpha|^2$, the limit $v_i \to v$ satisfies the hypotheses of Fatou's Lemma with respect to the counting measure, and so

$$\begin{aligned}
\liminf_{i\to\infty} \|v_i\|^2 &= \liminf_{i\to\infty} \sum_{\alpha\in A} |v_i^\alpha|^2 \\
&= \liminf_{i\to\infty} \int_A |v_i|^2\,d\mu \\
&\geq \int_A |v|^2\,d\mu = \sum_{\alpha\in A} |v^\alpha|^2 = \|v\|^2.
\end{aligned}$$

□

Note that the above proofs depend heavily on the existence of a Hilbert basis, Proposition 63, which we did not prove. The following problem outlines a standard procedure for getting around the proof of Proposition 63, by proving the existence of a Hilbert basis for any Hilbert space with a countable *spanning set*. A subset S of a Hilbert space H is said to be a spanning set if the (strong) closure of finite linear combinations of elements in S is equal to all of H. For example, in the proof of Theorem 24, we need only deal with the closure H' of the span of $\{v_1, v_2, \dots\}$. The existence of a Hilbert basis of H' is sufficient for the proof of Theorem 24.

Homework Problem 53. *Show that any strongly closed linear subspace of a Hilbert space H is again a Hilbert space (with the same inner product).*

We say a subset $\{v_\alpha\}_{\alpha \in A} \subset H$ is *linearly independent* (in the sense of Banach spaces) if any convergent sum

$$\sum_{\alpha \in A} b^\alpha v_\alpha = 0$$

implies $b^\alpha = 0$ for all $\alpha \in A$. Note in particular, the implication holds for any finite sum (and thus this notion of linearly independence in this Banach-space sense implies linear independence in the usual vector-space sense).

Homework Problem 54 (Gram-Schmidt Orthogonalization).

(a) Let H be a Hilbert space with a countable spanning set $\{v_1, v_2, \dots\}$ which is finite or countably infinite. Show that there is a subset of $\{v_1, v_2, \dots\}$ which is a linearly independent spanning set of H.

(b) Given a linearly independent spanning set $\{v_1, v_2, \dots\}$ on a Hilbert space H, define f_i and e_i recursively by

$$\begin{aligned} f_1 &= v_1, & e_1 &= \frac{f_1}{\|f_1\|}, \\ f_2 &= v_2 - \langle v_2, e_1 \rangle e_1, & e_2 &= \frac{f_2}{\|f_2\|}, \\ f_n &= v_n - \sum_{i=1}^{n-1} \langle v_n, e_i \rangle e_i, & e_n &= \frac{f_n}{\|f_n\|}. \end{aligned}$$

Show that this recursive definition can be carried out (in particular, show that $f_n \neq 0$). Then show that $\{e_1, e_2, \dots\}$ is a Hilbert basis for H. In other words, show that $\langle e_i, e_j \rangle = \delta_{ij}$ and that any v in H can be written as a convergent sum $v = \sum_{i=1}^{\infty} v^i e_i$.

The use of the previous problem isn't strictly necessary for our purposes, as $L^2_1(\mathbb{S}^1, \mathbb{R})$ is separable (though we won't prove that it is).

Recall that for every Banach space B, the dual space Banach space B^* is the space of all continuous linear functionals $\lambda\colon B \to \mathbb{R}$, with norm given by

$$\|\lambda\|_{B^*} = \sup_{x\in B\setminus\{0\}} \frac{|\lambda(x)|}{\|x\|_B}.$$

Also recall that for any $p \in (1, \infty)$, the dual Banach space of $L^p(\mathbb{R}^n)$ is $L^q(\mathbb{R}^n)$ for $p^{-1} + q^{-1} = 1$. Thus $L^2(\mathbb{R}^n)$ is dual to itself. This fact is true for all Hilbert spaces, as the following problem shows in the separable case.

Homework Problem 55. *Let H be a separable real Hilbert space. Show that the dual Banach space H^* is naturally equal to H. In particular, the inner product provides a map from $H \to H^*$ by*

$$x \mapsto \lambda_x = \langle \cdot, x\rangle.$$

Show that this map preserves the norm, is one-to-one and onto.

Hint: Use a Hilbert basis $\{e_i\}_{i=1}^\infty$. The most significant step is showing the map is onto. If λ corresponds to an $x \in H$, show that x must satisfy

$$x = \sum_{i=1}^\infty \lambda(e_i) e_i. \tag{41}$$

If $\lambda \in H^*$, then since λ is continuous, it is a bounded linear map (see Problem 56 below), and so $\|\lambda\|_{H^*}$ is finite. Then apply the definition of $\|\lambda\|_{H^*}$ to λ acting on the partial sums of (41). Show the series (41) converges in H.

A sequence v_i in a Banach space B converges to $v \in B$, in the *weak* topology* if for every $\lambda \in B^*$, $\lambda(v_i) \to \lambda(v)$. The previous problem shows that Theorem 24 is a special case of the following more general theorem about Banach spaces:

Theorem 26 (Banach-Alaoglu). *In a Banach space B, the unit ball $\{x \in B : \|x\|_B \le 1\}$ is compact in the weak* topology. In other words, if x_i is a sequence in the unit ball, then there is a subsequence x_{i_j} and a limit $x \in B$ so that for all $\lambda \in B^*$, $\lambda(x_{i_j}) \to \lambda(x)$ as $j \to \infty$.*

Example 20 (Fourier series). *In the following theorem, we compute perhaps the easiest nontrivial example of a Hilbert basis on an infinite-dimensional Hilbert space.* $L^2(\mathbb{S}^1, \mathbb{C})$ *is a complex Hilbert space with inner product given by*

$$\langle f, g\rangle = \int_{\mathbb{S}^1} f\bar{g}\, dt.$$

Theorem 27. *The complex exponential functions*

$$\{e^{2\pi ikt} : k \in \mathbb{Z}\}$$

form a Hilbert basis of $L^2(\mathbb{S}^1, \mathbb{C})$*.*

Proof. It is clear that each $e^{2\pi ikt} \in L^2(\mathbb{S}^1, \mathbb{C})$, and we compute

$$\begin{aligned} \langle e^{2\pi ikt}, e^{2\pi i\ell t}\rangle &= \int_{\mathbb{S}^1} e^{2\pi ikt}\overline{e^{2\pi i\ell t}}\, dt \\ &= \int_0^1 e^{2\pi i(k-\ell)t}\, dt \\ &= \begin{cases} \left.\dfrac{e^{2\pi i(k-\ell)t}}{2\pi i(k-\ell)}\right|_0^1 = 0 & \text{if } k \neq \ell \\ \displaystyle\int_0^1 dt \qquad\quad = 1 & \text{if } k = \ell. \end{cases} \end{aligned}$$

Therefore, $\{e^{2\pi ikt}\}_{k=-\infty}^{\infty}$ forms an orthonormal set in $L^2(\mathbb{S}^1, \mathbb{C})$.

We must show that every element $f \in L^2(\mathbb{S}^1, \mathbb{C})$ can be written as a *Fourier series*

$$f = \sum_{k=-\infty}^{\infty} \langle f, e^{2\pi ikt}\rangle e^{2\pi ikt},$$

with the convergence in the L^2 sense.

First, we address this problem for smooth functions $f \in C^\infty(\mathbb{S}^1, \mathbb{C})$. Recall that $C^\infty(\mathbb{S}^1, \mathbb{C})$ is dense in $L^2(\mathbb{S}^1, \mathbb{C})$ (which may be proved by mollifying L^2 functions).

Lemma 68. *If* $f \in C^\infty(\mathbb{S}^1, \mathbb{C})$*, then for every polynomial* $P = P(k)$*,*

$$\lim_{k\to\infty} P(k)\langle f, e^{2\pi ikt}\rangle = \lim_{k\to-\infty} P(k)\langle f, e^{2\pi ikt}\rangle = 0.$$

Proof. We use the following claim: For any L^2 function f, the Fourier coefficients $\langle f, e^{2\pi ikt}\rangle \to 0$ as $k \to \pm\infty$. This follows from Bessel's Inequality

$$\sum_{k=-\infty}^{\infty} |\langle f, e^{2\pi ikt}\rangle|^2 \leq \|f\|_{L^2}^2 < \infty.$$

If f is smooth, then $\dot{f}$ is also smooth (and thus is in L^2), and integration by parts gives us

$$\begin{aligned}
\langle \dot{f}, e^{2\pi ikt}\rangle &= \int_0^1 \dot{f} e^{-2\pi ikt}\, dt \\
&= -\int_0^1 f (e^{-2\pi ikt})^{\cdot}\, dt + f(t)e^{-2\pi ikt}\Big|_0^1 \\
&= 2\pi ik\langle f, e^{2\pi ikt}\rangle + 0.
\end{aligned}$$

Now by the claim, $\langle \dot{f}, e^{2\pi ikt}\rangle = 2\pi ik\langle f, e^{2\pi ikt}\rangle \to 0$ as $k \to \pm\infty$. Now we may apply induction to show that

$$\lim_{k\to\pm\infty} k^n\langle f, e^{2\pi ikt}\rangle = 0 \qquad \text{for each } n = 0, 1, 2, \ldots.$$

Thus any polynomial $P(k)$ times the Fourier coefficients also goes to zero as $k \to \pm\infty$. □

The previous lemma shows that for any smooth function $f \in C^\infty(\mathbb{S}^1, \mathbb{C})$, the Fourier series

$$g(t) = \sum_{k=-\infty}^{\infty} \langle f, e^{2\pi ikt}\rangle e^{2\pi ikt}$$

converges uniformly: This is because there is a constant $C > 0$ so that

$$|\langle f, e^{2\pi ikt}\rangle| \leq \frac{C}{1+k^2}$$

(why?), which shows that the C^0 norm of the Fourier series satisfies

$$\sum_{k=-\infty}^{\infty} \left\|\langle f, e^{2\pi ikt}\rangle e^{2\pi ikt}\right\|_{C^0} \leq \sum_{k=-\infty}^{\infty} \frac{C}{1+k^2} < \infty.$$

So the sup norm of the tails of the series

$$\sum_{k=-\infty}^{\infty} \langle f, e^{2\pi ikt} \rangle e^{2\pi ikt}$$

must go to zero, as they are bounded by the tails of an absolutely convergent series.

Therefore, uniform convergence implies that $g(t)$ is continuous (and thus is in L^2 as well—why?). (In fact, $g(t)$ is smooth—see Homework Problem 58 below.) If we let

$$h(t) = f(t) - g(t) = f(t) - \sum_{k=-\infty}^{\infty} \langle f, e^{2\pi ikt} \rangle e^{2\pi ikt},$$

then by the same techniques in the proof of Theorem 63 above, we see that

$$\langle h, e^{2\pi ikt} \rangle = 0 \qquad \text{for all } k \in \mathbb{Z}.$$

The following lemma shows that $h = 0$:

Lemma 69. *Given a function $h \in C^0(\mathbb{S}^1, \mathbb{C})$ all of whose Fourier coefficients $\langle h, e^{2\pi ikt} \rangle = 0$, then $h = 0$ identically.*

Proof. We prove by contradiction. If h is not identically zero, then there is a point $\tau \in \mathbb{S}^1$ at which $h(\tau) \neq 0$. Then we know that at least one of the following is true:

$$\operatorname{Re} h(\tau) > 0, \quad \operatorname{Re} h(\tau) < 0, \quad \operatorname{Im} h(\tau) > 0, \quad \operatorname{Im} h(\tau) < 0.$$

Assume that $\operatorname{Re} h(\tau) > 0$ (the other cases are similar), and let $\alpha(t) = \operatorname{Re} h(t)$. Since α is continuous, there is a $\delta > 0$ so that

$$\alpha(t) > \tfrac{1}{2}\alpha(\tau) > 0 \quad \text{if } t \in (\tau - \delta, \tau + \delta).$$

We will construct an approximate bump function to prove a contradiction. For n a positive integer, define

$$b_n(t) = [\tfrac{1}{2} + \tfrac{1}{2}\cos 2\pi(t - \tau)]^n = \left[\tfrac{1}{2} + \tfrac{1}{4}e^{-2\pi i\tau}e^{2\pi it} + \tfrac{1}{4}e^{2\pi i\tau}e^{-2\pi it}\right]^n.$$

It is obvious that $b_n(t)$ is real-valued, periodic with period 1 (and so defines a function on $\mathbb{S}^1$), and is equal to a finite Fourier series. Moreover, note that

$$\tfrac{1}{2} + \tfrac{1}{2}\cos 2\pi(t - \tau) \in [0, 1]$$

always, and is equal to 1 only if $t = \tau$ in $\mathbb{S}^1$. Thus the powers $b_n(t) \to 0$ as $n \to \infty$ away from $t = \tau$, while $b_n(\tau) \to 1$. This is the property that makes b_n similar to bump functions centered around $t = \tau$.

Now compute

$$\begin{aligned}
|\mathrm{Re}\,\langle h, b_n\rangle| &= \left|\mathrm{Re} \int_{\mathbb{S}^1} h(t)\overline{b_n(t)}\,dt\right| \\
&= \left|\int_{\mathbb{S}^1} \alpha(t)b_n(t)\,dt\right| \\
&= \left|\int_{\tau-\delta}^{\tau+\delta} \alpha(t)b_n(t)\,dt + \int_{\mathbb{S}^1\setminus[\tau-\delta,\tau+\delta]} \alpha(t)b_n(t)\,dt\right| \\
&\geq \left|\int_{\tau-\delta}^{\tau+\delta} \alpha(t)b_n(t)\,dt\right| - \left|\int_{\mathbb{S}^1\setminus[\tau-\delta,\tau+\delta]} \alpha(t)b_n(t)\,dt\right| \\
&> \int_{\tau-\frac{\delta}{2}}^{\tau+\frac{\delta}{2}} \alpha(t)b_n(t)\,dt - \left|\int_{\mathbb{S}^1\setminus[\tau-\delta,\tau+\delta]} \alpha(t)b_n(t)\,dt\right|.
\end{aligned}$$

(Note the last inequality follows since the integrand is positive.) Also, we have the following bounds:

$$\begin{aligned}
t \in [t - \tfrac{\delta}{2}, t + \tfrac{\delta}{2}] &\implies \alpha(t) > \tfrac{1}{2}\alpha(\tau) > 0, \quad b_n(t) \geq (\tfrac{1}{2} + \tfrac{1}{2}\cos\pi\delta)^n, \\
t \notin [t - \delta, t + \delta] &\implies |\alpha(t)| < C, \quad b_n(t) \leq (\tfrac{1}{2} + \tfrac{1}{2}\cos 2\pi\delta)^n.
\end{aligned}$$

for some constant C (since α is continuous). The bounds on b_n follow by examining the graph of the cosine function. The key point is that

$$\tfrac{1}{2} + \tfrac{1}{2}\cos\pi\delta > \tfrac{1}{2} + \tfrac{1}{2}\cos 2\pi\delta > 0. \tag{42}$$

Now compute

$$\begin{aligned}
|\mathrm{Re}\,\langle h, b_n\rangle| &> \int_{\tau-\frac{\delta}{2}}^{\tau+\frac{\delta}{2}} \alpha(t)b_n(t)\,dt - \left|\int_{\mathbb{S}^1\setminus[\tau-\delta,\tau+\delta]} \alpha(t)b_n(t)\,dt\right| \\
&\geq \delta\tfrac{1}{2}\alpha(\tau)(\tfrac{1}{2} + \tfrac{1}{2}\cos\pi\delta)^n - (1-2\delta)C(\tfrac{1}{2} + \tfrac{1}{2}\cos 2\pi\delta)^n.
\end{aligned}$$

Now (42) shows the ratio of the first term over the second goes to $+\infty$ as $n \to \infty$ and thus there is an n so that $|\mathrm{Re}\,\langle h, b_n\rangle| > 0$.

Now the contradiction is this: Since b_n is a finite Fourier series, $\langle h, b_n\rangle$ is a finite linear combination of Fourier coefficients $\langle h, e^{2\pi ikt}\rangle$, which we assume are all zero. Thus $\langle h, b_n\rangle = 0$, and we have a contradiction. □

Since h is the difference between the smooth f and its Fourier series, we have shown

Lemma 70. *Let* $f \in C^\infty(\mathbb{S}^1, \mathbb{C})$*. Then*

$$f(t) = \sum_{k=-\infty}^{\infty} \langle f, e^{2\pi i k t} \rangle \, e^{2\pi i k t},$$

and the series converges uniformly in t*.*

Uniform convergence on $\mathbb{S}^1$ implies L^2 convergence (since $\mathbb{S}^1$ has finite measure; why does this imply L^2 convergence?). Therefore, as in Theorem 23, we have

$$\|f\|_{L^2}^2 = \int_{\mathbb{S}^1} |f|^2 \, dt = \sum_{k=-\infty}^{\infty} |\langle f, e^{2\pi i k t} \rangle|^2,$$

for $f \in C^\infty(\mathbb{S}^1, \mathbb{C})$.

To complete the proof of Theorem 27, first define the Hilbert space $\ell^2 = L^2(\mathbb{Z}, \mathbb{C})$ for the counting measure on $\mathbb{Z}$. In other words, ℓ^2 is the set of all complex-valued integer-indexed sequences $\{v^k\}_{k\in\mathbb{Z}}$ so that $\sum_{k=-\infty}^{\infty} |v^k|^2 < \infty$. Then we have the operation $\mathcal{F}$ of defining Fourier series:

$$\mathcal{F} \colon L^2(\mathbb{S}^1, \mathbb{C}) \to \ell^2, \qquad \mathcal{F} \colon f \mapsto \hat{f}^k = \langle f, e^{2\pi i k t} \rangle.$$

Moreover, on the dense subset $C^\infty(\mathbb{S}^1, \mathbb{C}) \subset L^2(\mathbb{S}^1, \mathbb{C})$, $\mathcal{F}$ is an isometry. Bessel's Inequality and the fact that $\{e^{2\pi i k t}\}$ is an orthonormal set in $L^2(\mathbb{S}^1, \mathbb{C})$ shows that for all $f \in L^2(\mathbb{S}^1, \mathbb{C})$,

$$\|f\|_{L^2}^2 \geq \sum_{k=-\infty}^{\infty} |\langle f, e^{2\pi i k t} \rangle|^2 = \|\mathcal{F}(f)\|_{\ell^2}^2.$$

Therefore $\mathcal{F}$ is a *bounded linear map* from $L^2(\mathbb{S}^1, \mathbb{C})$ to ℓ^2. A linear map $\mathcal{L}$ from a Banach space $\mathcal{B}_1$ to another Banach space $\mathcal{B}_2$ is called bounded if there is a positive constant C so that for all $v \in \mathcal{B}_1$,

$$\|\mathcal{L}(v)\|_{\mathcal{B}_2} \leq C \|v\|_{\mathcal{B}_1}.$$

A linear map between Banach spaces is bounded if and only if it is continuous (see Problem 56 below). Therefore, $\mathcal{F}$ is continuous.

Also, define the linear map $\mathcal{G}\colon \ell^2 \to L^2(\mathbb{S}^1, \mathbb{C})$ by

$$\mathcal{G}(v) = \sum_{k=-\infty}^{\infty} v^k e^{2\pi ikt}.$$

The proof of Theorem 23 shows that $\mathcal{G}$ preserves the norms. In other words,

$$\|\mathcal{G}(v)\|_{L^2(\mathbb{S}^1,\mathbb{C})} = \|v\|_{\ell^2}.$$

Let $f \in L^2(\mathbb{S}^1, \mathbb{C})$. Since smooth functions are dense in L^2, there is a sequence $f_n \to f$ in L^2 for $f_n \in C^\infty(\mathbb{S}^1, \mathbb{C})$. Since $\mathcal{F}$ is continuous, then $\mathcal{F}(f_n) \to \mathcal{F}(f)$ in ℓ^2 as $n \to \infty$. In other words,

$$\begin{aligned} 0 &= \lim_{n\to\infty} \|\mathcal{F}(f_n) - \mathcal{F}(f)\|_{\ell^2}^2 \\ &= \lim_{n\to\infty} \|\hat{f}_n - \hat{f}\|_{\ell^2}^2 \\ &= \lim_{n\to\infty} \|\mathcal{G}(\hat{f}_n) - \mathcal{G}(\hat{f})\|_{L^2}^2 \end{aligned}$$

Now recall that

$$\mathcal{G}(\hat{f}_n) = \sum_{k=-\infty}^{\infty} \langle f_n, e^{2\pi ikt}\rangle e^{2\pi ikt} = f_n$$

since f_n is smooth. Therefore,

$$0 = \lim_{n\to\infty} \|\mathcal{G}(\hat{f}_n) - \mathcal{G}(\hat{f})\|_{L^2} = \lim_{n\to\infty} \|f_n - \mathcal{G}(\hat{f})\|_{L^2}.$$

So in L^2,

$$f_n \to \mathcal{G}(\hat{f}) = \sum_{k=-\infty}^{\infty} \hat{f}^k e^{2\pi ikt}.$$

Since we assumed $f_n \to f$ in L^2, this shows

$$f = \sum_{k=-\infty}^{\infty} \hat{f}^k e^{2\pi ikt}$$

in L^2, and since the sum converges in L^2, finite linear combinations of the orthonormal set $\{e^{2\pi ikt}\}$ are dense in $L^2(\mathbb{S}^1, \mathbb{C})$, and $\{e^{2\pi ikt}\}$ is a Hilbert basis of $L^2(\mathbb{S}^1, \mathbb{C})$. So Theorem 27 is proved. $\square$

Homework Problem 56. *Let $\mathcal{L}\colon \mathcal{B}_1 \to \mathcal{B}_2$ be a linear map between Banach spaces. Show that $\mathcal{L}$ is bounded if and only if $\mathcal{L}$ is continuous.*

Homework Problem 57. *Using the notation of the proof of Theorem 27 above, show that $\mathcal{F} : L^2(\mathbb{S}^1, \mathbb{C}) \to \ell^2$ is an isometry and that $\mathcal{F} \circ \mathcal{G}$ is the identity map.*

Homework Problem 58. *Let $f^k \in \mathbb{C}$ for all $k \in \mathbb{Z}$, and assume for all $n \geq 0$ that*

$$\lim_{k\to\infty} k^n f^k = \lim_{k\to-\infty} k^n f^k = 0.$$

Then the Fourier series

$$f(t) = \sum_{k=-\infty}^{\infty} f^k e^{2\pi i k t}$$

converges uniformly to a smooth function from $\mathbb{S}^1 \to \mathbb{C}$.

Hint: The key is being able to change the order of the derivative d/dt with the summation $\sum_{k=-\infty}^{\infty}$. Recall that the summation $\sum_{k=-\infty}^{\infty}$ can be interpreted as an integral over $\mathbb{Z}$ with respect to the counting measure $d\mu$. Thus for all $t \in \mathbb{S}^1$,

$$f(t) = \int_{\mathbb{Z}} f^k e^{2\pi i k t} d\mu(k).$$

To show that $f(t) \in C^1(\mathbb{S}^1, \mathbb{C})$, show that there is a constant $C > 0$ so that

$$|f^k| \leq \frac{C}{1 + |k|^3}.$$

Mimic the proof of Proposition 11: Show that the absolute value of the difference quotient

$$\frac{f^k e^{2\pi i k (t+h)} - f^k e^{2\pi i k t}}{h}$$

is uniformly $\leq \frac{C'(|k|+1)}{1+|k|^3}$ for a constant C'. (Apply the Mean Value Theorem to the real and imaginary parts of $e^{2\pi i k t}$ separately.) Show that the series

$$\sum_{k=-\infty}^{\infty} \frac{C'(|k|+1)}{1 + |k|^3}$$

converges by using the procedure in the proof of Lemma 70 above.

Use induction to show $f(t)$ is smooth.

4.6 Compact maps and the Ascoli-Arzelá Theorem

Recall that every element of $L^2_1(\mathbb{S}^1) = L^2_1(\mathbb{S}^1, \mathbb{R})$ has a continuous representative (Proposition 59). So there is a natural linear map $L^2_1(\mathbb{S}^1) \to C^0(\mathbb{S}^1)$. In this section, we show that this map is *compact*. A linear map between Banach spaces $\Lambda\colon \mathcal{B}_1 \to \mathcal{B}_2$ is called compact if the closure of the image of the unit ball in $\mathcal{B}_1$ is strongly compact in $\mathcal{B}_2$. In other words, if $v_i \in \mathcal{B}_1$ satisfy $\|v_i\|_{\mathcal{B}_1} \leq 1$, then $\{\Lambda(v_i)\}$ has a strongly convergent subsequence in $\mathcal{B}_2$: i.e. there is a subsequence $\{v_{i_j}\}$ and an element $w \in \mathcal{B}_2$ so that

$$\lim_{j\to\infty} \|\Lambda(v_{i_j}) - w\|_{\mathcal{B}_2} = 0.$$

The basic observation which allows us to conclude that the natural inclusion map $L^2_1(\mathbb{S}_1) \to C^0(\mathbb{S}^1)$ is compact comes from the proof of Proposition 59. If $f \in L^2_1(\mathbb{S}^1)$, then

$$\begin{aligned} |f(t_2) - f(t_1)| &= \left| \int_{t_1}^{t_2} \dot{f}(t)\, dt \right| \\ &\leq \left(\int_{t_1}^{t_2} |\dot{f}(t)|^2\, dt \right)^{\frac{1}{2}} \left(\int_{t_1}^{t_2} dt \right)^{\frac{1}{2}} \\ &\leq \left(\int_{\mathbb{S}^1} |\dot{f}(t)|^2\, dt \right)^{\frac{1}{2}} (t_2 - t_1)^{\frac{1}{2}} \\ &\leq \|f\|_{L^2_1} (t_2 - t_1)^{\frac{1}{2}} \end{aligned}$$

(Note that the first equality was justified in the proof of Proposition 59.) Therefore, f is continuous. But moreover, for every $\epsilon > 0$, we may choose

$$\delta = \left(\frac{\epsilon}{\|f\|_{L^2_1}} \right)^2$$

so that

$$|t_2 - t_1| < \delta \quad \Longrightarrow \quad |f(t_2) - f(t_1)| < \epsilon.$$

So the *modulus of continuity* δ does not depend on t, and depends only on the norm $\|f\|_{L^2_1}$, and on no other information about f.

A family of functions Ω of functions from a metric space X to a metric space Y is called *equicontinuous* at a point $x \in X$ if for all $\epsilon > 0$, there is a $\delta > 0$ so that

$$d_X(x, x') < \delta \quad \Longrightarrow \quad d_Y(f(x), f(x')) < \epsilon$$

for all $f \in \Omega$. The point is that δ does not depend on f. Such a family of functions Ω is called *equicontinuous* on X if it is equicontinuous at each point $x \in X$.

Note that if Ω is equicontinuous on X then each $f \in \Omega$ is continuous.

The computations above show

Lemma 71. *The unit ball in $L^2_1(\mathbb{S}^1)$ is equicontinuous on $\mathbb{S}^1$.*

Theorem 28 (Ascoli-Arzelá). *Let X be a compact metric space, and let Ω be an equicontinuous family of real-valued functions on X. Assume there is a uniform C so that $|f(x)| \le C$ for all $f \in \Omega$ and $x \in X$. Then each sequence $\{f_n\} \subset \Omega$ has a uniformly convergent subsequence.*

Proof. We'll prove the theorem with the help of a few lemmas.

Lemma 72. *Any compact metric space has a countable dense subset.*

Proof. Let X be the compact metric space. For $\epsilon = 1/n$, obviously

$$X = \bigcup_{x \in X} B_\epsilon(x), \qquad B_\epsilon(x) = \{y \in X : d_X(x,y) < \epsilon\}.$$

For each positive integer n, this open cover of X has a finite subcover consisting of balls of radius $1/n$ centered at points $x_{n,1}, \ldots, x_{n,m_n}$. The union

$$\bigcup_{n=1}^{\infty} \{x_{n,1}, \ldots, x_{n,m_n}\}$$

is a countable dense subset of X. □

Lemma 73. *Let $\mathcal{P}$ be a countable set, and let $f_n : \mathcal{P} \to \mathbb{R}$ be a sequence of functions. Assume there is a constant C so that $|f_n(p)| \le C$ for all $n = 1, 2, \ldots$ and all $p \in \mathcal{P}$. Then there is a subsequence of $\{f_n\}$ which converges everywhere on $\mathcal{P}$ to a function $f : \mathcal{P} \to \mathbb{R}$.*

Proof. See Problem 59 below. □

Lemma 74. *Let $\{f_n\}$ be an equicontinuous sequence of mappings from a compact metric space X to $\mathbb{R}$. If the sequence $\{f_n(x)\}$ converges for each x in a dense subset of X, then $\{f_n\}$ converges uniformly on X to a continuous limit function.*

Proof. First we show that $f_n(x)$ converges pointwise everywhere to a function $f(x)$. Let $y \in X$ and let $\epsilon > 0$. Then by equicontinuity, there is a $\delta > 0$ so that

$$d_X(x, y) < \delta \quad \Longrightarrow \quad |f_n(x) - f_n(y)| < \epsilon.$$

(Note δ is independent of n.) Since f_n converges on a dense subset of X, there is an $x \in B_\delta(y)$ for which $f_n(x)$ converges. Therefore, $\{f_n(x)\}$ is a Cauchy sequence in $\mathbb{R}$, and so there is an N so that

$$n, m \geq N \quad \Longrightarrow \quad |f_n(x) - f_m(x)| < \epsilon.$$

Therefore, for $n, m \geq N$,

$$|f_n(y) - f_m(y)| \leq |f_n(y) - f_n(x)| + |f_n(x) - f_m(x)| + |f_m(x) - f_m(y)| < 3\epsilon.$$

Therefore, $\{f_n(y)\}$ is a Cauchy sequence in the complete metric space $\mathbb{R}$, and so it converges to a limit which we call $f(y)$.

Let $y \in X$ and $\epsilon > 0$. Then equicontinuity shows that there is a $\delta > 0$ so that

$$x \in B_\delta(y) \quad \Longrightarrow \quad |f_n(x) - f_n(y)| < \epsilon \tag{43}$$

for all n. By letting $n \to \infty$, we also have

$$x \in B_\delta(y) \quad \Longrightarrow \quad |f(x) - f(y)| \leq \epsilon \tag{44}$$

These $B_\delta(y)$ form an open cover of X, and so there is a finite subcover

$$X = \bigcup_{i=1}^{k} B_{\delta_i}(y_i)$$

since X is compact. Choose N large enough so that

$$n \geq N \quad \Longrightarrow \quad |f_n(y_i) - f(y_i)| < \epsilon, \quad i = 1, \dots, k. \tag{45}$$

Then for $x \in X$, $x \in B_{\delta_i}(y_i)$ for some y_i, and so (43), (44) and (45) show

$$|f_n(x) - f(x)| \leq |f_n(x) - f_n(y_i)| + |f_n(y_i) - f(y_i)| + |f(y_i) - f(x)| < 3\epsilon.$$

Since the same N works for all $x \in X$, the convergence is uniform.

f, as the uniform limit of continuous functions, is continuous. □

This completes the proof of Theorem 28. □

Homework Problem 59. *Let $\mathcal{P}$ be a countable set, and let $f_n : \mathcal{P} \to \mathbb{R}$ be a sequence of functions. Assume that for each $p \in \mathcal{P}$, there is a constant $C = C_p$ so that $|f_n(p)| \le C$ for all $n = 1, 2, \ldots$. Show there is a subsequence of $\{f_n\}$ which converges everywhere on $\mathcal{P}$ to a function $f : \mathcal{P} \to \mathbb{R}$.*

Hint: Use a diagonalization argument.

An important version of the Ascoli-Arzelá Theorem is the following:

Theorem 29. *Let X be a metric space so that there is a countable number of open subsets $\mathcal{O}_i$ satisfying*

$$X = \bigcup_{i=1}^{\infty} \mathcal{O}_i, \qquad \mathcal{O}_i \subset\subset \mathcal{O}_{i+1}, \tag{46}$$

and let Ω be an equicontinuous set of real-valued functions on X. If for a sequence of functions $\{f_n\} \subset \Omega$, there is a uniform C so that $|f_n(x)| \le C$ for all n and all $x \in X$, then there is a subsequence of $\{f_n\}$ which converges pointwise to a function $f : X \to \mathbb{R}$, and the convergence is uniform on every compact subset of X.

Remark. Recall $A \subset\subset B$ for A a subspace of a topological space B means that the closure $\overline{A}$ relative to B is compact.

Remark. A sequence of functions converging uniformly on compact subsets of X is said to converge *normally* on X.

We relegate the proof of Theorem 29 to the following problem:

Homework Problem 60. *Prove Theorem 29.*

Hint: Consider X, $\mathcal{O}_i$ as in the previous theorem. Note we may apply Theorem 28 to each of the compact sets $\overline{\mathcal{O}_i}$. Use a diagonalization argument to find a uniformly convergent subsequence on each $\overline{\mathcal{O}_i}$. Show that every compact subset of X is contained in some $\mathcal{O}_i$.

Remark. For every smooth manifold X (which is Hausdorff and sigma-compact), there are a countable collection of open sets $\mathcal{O}_i$ satisfying condition (46). See the notes on "The Real Definition of a Smooth Manifold."

The Ascoli-Arzelá Theorem provides the following.

Proposition 75. *If $C > 0$ and $\{f_n\}$ is a sequence of functions in $L^2_1(\mathbb{S}^1, \mathbb{R})$ which satisfy $\|f_n\|_{L^2_1} \le C$, then there is a uniformly convergent subsequence.*

Proof. This follows from the Ascoli-Arzelá Theorem and Lemma 71 above, once we know in addition that there is a constant K so that $|f_n| \leq K$ pointwise. First of all, note that

$$|f_n(t_2) - f_n(t_1)| \leq \|f_n\|_{L^2_1} |t_2 - t_1|^{\frac{1}{2}} \leq C|t_2 - t_1|^{\frac{1}{2}}$$

shows that for every $t_2, t_1 \in \mathbb{S}^1$,

$$|f_n(t_2) - f_n(t_1)| \leq C$$

since we may choose $t_2, t_1 \in [0, 1)$. Since

$$\left(\int_0^1 |f_n|^2 \, dt \right)^{\frac{1}{2}} = \|f_n\|_{L^2} \leq \|f_n\|_{L^2_1} \leq C,$$

there must be a $t_1 \in \mathbb{S}^1$ so that $|f_n(t_1)| \leq C$. Then for any $t_2 \in \mathbb{S}^1$,

$$|f_n(t_2)| \leq |f_n(t_1)| + |f_n(t_2) - f_n(t_1)| \leq 2C.$$

Thus the hypotheses of the Ascoli-Arzelá Theorem are satisfied. □

Corollary 76. *The inclusion* $L^2_1(\mathbb{S}^1, \mathbb{R}) \hookrightarrow C^0(\mathbb{S}^1, \mathbb{R})$ *is compact.*

Proof. Take $C = 1$ in the above theorem. □

Corollary 77. *Let* $C > 0$ *and let* $X \subset \mathbb{R}^N$ *be a compact manifold, and let* $\gamma_n \in L^2_1(\mathbb{S}^1, X) \subset L^2_1(\mathbb{S}^1, \mathbb{R}^N)$ *satisfy* $E(\gamma_n) \leq C$*. Then there is a uniformly convergent subsequence of* $\{\gamma_n\}$*, and the limit is a continuous function* $\gamma : \mathbb{S}^1 \to X$*.*

Proof. Recall

$$\|\gamma_n\|^2_{L^2_1(\mathbb{S}^1,\mathbb{R}^N)} = \|\gamma_n\|^2_{L^2(\mathbb{S}^1,\mathbb{R}^N)} + \|\dot{\gamma}_n\|^2_{L^2(\mathbb{S}^1,\mathbb{R}^N)} = \|\gamma_n\|^2_{L^2(\mathbb{S}^1,\mathbb{R}^N)} + E(\gamma_n).$$

Since $\gamma_n(\mathbb{S}^1) \subset X$ and X is compact, there is a constant K so that $|\gamma_n(t)| \leq K$ for all n and t. Therefore,

$$\|\gamma_n\|^2_{L^2(\mathbb{S}^1,\mathbb{R}^N)} \leq \int_{\mathbb{S}^1} K^2 \, dt = K^2,$$

and moreover,

$$\|\gamma_n\|^2_{L^2_1(\mathbb{S}^1,\mathbb{R}^N)} \leq C + K^2$$

independently of n. So each component function γ_n^a for $a = 1, \ldots, N$ satisfies

$$\|\gamma_n^a\|_{L_1^2(\mathbb{S}^1,\mathbb{R})} \le \sqrt{C + K^2}.$$

Then Proposition 75 shows that there is a subsequence $\{{}_1\gamma_n\}$ of $\{\gamma_n\}$ so that the component ${}_1\gamma_n^1$ converges uniformly. Let $\{{}_2\gamma_n\}$ be a subsequence of $\{{}_1\gamma_n\}$ so that ${}_2\gamma_n^1$ and ${}_2\gamma_n^1$ converge uniformly. By induction, as in the proof of Theorem 24, there is a subsequence $\{{}_N\gamma_n\}$ of $\{\gamma_n\}$ so that ${}_N\gamma_n^a$ converges uniformly for $a = 1, \ldots, N$. Since this subsequence converges uniformly on each component in $\mathbb{R}^N$, ${}_N\gamma_n$ converges uniformly as $n \to \infty$ to a limit γ in $C^0(\mathbb{S}^1, \mathbb{R}^N)$.

Since X is closed in $\mathbb{R}^N$ and the subsequence converges pointwise, the limit $\gamma\colon \mathbb{S}^1 \to X$. □

It is also useful to define the Hölder norm for functions $f\colon \mathbb{S}^1 \to \mathbb{R}$

$$\|f\|_{C^{0,\frac{1}{2}}(\mathbb{S}^1)} = \|f\|_{C^0} + \sup_{t_1 \ne t_2} \frac{|f(t_1) - f(t_2)|}{d_{\mathbb{S}^1}(t_1, t_2)^{\frac{1}{2}}}.$$

(Here we define

$$d_{\mathbb{S}^1}(t_1, t_2) = \inf_{k \in \mathbb{Z}} |(t_1 + k) - t_2|.$$

This definition is necessary, since we identify the real numbers t and $t + k$ on the circle $\mathbb{S}^1$. For example, $d_{\mathbb{S}^1}(0, 0.9) = |1 - 0.9| = 0.1$.) It is easy to check that this defines a norm. Define the space $C^{0,\frac{1}{2}}(\mathbb{S}^1)$ to be all f from $\mathbb{S}^1 \to \mathbb{R}$ so that $\|f\|_{C^{0,\frac{1}{2}}(\mathbb{S}^1)} < \infty$.

$C^{0,\frac{1}{2}}(\mathbb{S}^1)$ is a Banach space (Proposition 78 below), and the calculations above show that there is a natural continuous inclusion map from $L_1^2(\mathbb{S}^1) \to C^{0,\frac{1}{2}}(\mathbb{S}^1)$. Moreover, the natural inclusion map from $C^{0,\frac{1}{2}}(\mathbb{S}^1) \to C^0(\mathbb{S}^1)$ is compact. Then Problem 63 below shows that composition inclusion from $L_1^2(\mathbb{S}^1) \to C^0(\mathbb{S}^1)$ is compact.

In general, for any metric space X, $\alpha \in (0, 1]$, we can define

$$\begin{aligned} C^{0,\alpha}(X) &= \{f\colon X \to \mathbb{R} : \|f\|_{C^{0,\alpha}} < \infty\}, \\ \|f\|_{C^{0,\alpha}} &= \sup_{x \in X} |f(x)| + \sup_{x \ne y \in X} \frac{|f(x) - f(y)|}{d_X(x, y)^\alpha}. \end{aligned}$$

These are called *Hölder spaces* and *Hölder norms* respectively.

Example 21. *This is the standard example for* $X = [-1, 1] \subset \mathbb{R}$*.* $f(x) = |x|^\alpha$ *is in* $C^{0,\alpha}(X)$*.*

Proof. It clearly suffices to bound the difference quotient

$$q(x, y) = \frac{||x|^\alpha - |y|^\alpha|}{|x - y|^\alpha}, \qquad x \neq y \in [-1, 1].$$

We will show that this is always ≤ 1. First, simplify to the case x and y have the same sign, since if they have opposite signs, $q(x, y) < q(-x, y)$. We may assume x and y have the same sign. By possibly interchanging $(x, y) \leftrightarrow (-x, -y)$ and switching x and y, we may assume $x > y \geq 0$. Then write

$$q(x, y) = \frac{x^\alpha - y^\alpha}{(x - y)^\alpha} = \frac{1 - \rho^\alpha}{(1 - \rho)^\alpha}, \qquad \rho = \frac{y}{x} \in [0, 1).$$

Then we compute

$$\frac{dq}{d\rho} = \frac{\alpha(1 - \rho^{\alpha - 1})}{(1 - \rho)^{\alpha + 1}} \leq 0.$$

Therefore, the max of $q(\rho)$ is achieved at $\rho = 0$, $q = 1$. ☐

We also say $f(x) = |x|^\alpha$ is *locally* $C^{0,\alpha}$ on $\mathbb{R}$, since the α Hölder norm of f is finite on any compact subset of $\mathbb{R}$.

In the case $\alpha = 1$, note that a function in $C^{0,1}$ is simply a C^0 function which is globally Lipschitz.

Homework Problem 61.

(a) Show that the inclusion $C^1(\mathbb{S}^1) \hookrightarrow C^0(\mathbb{S}^1)$ *is compact (Hint: use the Mean Value Theorem).*

(b) Show that every bounded sequence $f_n \in C^1(\mathbb{R})$ *(i.e., there is a uniform* C *so that* $\|f_n\|_{C^1} \leq C$ *for all* n*) has a subsequence which converges uniformly on compact subsets of* $\mathbb{R}$ *to a continuous limit* f*. Hint: It is easy to show that* $\mathbb{R}$ *satisfies condition (46).*

(c) Find an example of a bounded sequence of functions $f_n \in C^1(\mathbb{R})$ *which does not have a convergent subsequence in* $C^0(\mathbb{R})$*. Thus the inclusion* $C^1(\mathbb{R}) \hookrightarrow C^0(\mathbb{R})$ *is not compact. (Hint: How is this situation different from parts* (a) *and* (b)*? You must use the noncompactness of* $\mathbb{R}$*. Therefore, the interesting behavior of the* f_n *should be "moving off to infinity.")*

It is also useful to apply Hölder norms to the derivatives of a functions. In particular, on $\mathbb{R}^n$, we may define for k a positive integer, $\alpha \in (0, 1]$,

$$\|f\|_{C^{k,\alpha}} = \sum_{|\beta| \leq k} \|\partial_\beta f\|_{C^{0,\alpha}},$$

where, as in (3) above, we use the multi-index notation to denote all the partial derivatives of f of order $\leq k$.

Remark. It is not useful to define $C^{0,\alpha}$ for $\alpha > 1$, as the following problem shows.

Homework Problem 62. *Let $\alpha > 1$, and let $f : \mathbb{R} \to \mathbb{R}$. Assume that*

$$\sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha} = C < \infty.$$

Show that f is a constant function.

Hint: Use the definition of the derivative to show that $f'(x) = 0$ for all x.

Proposition 78. *Let X be a metric space and $\alpha \in (0, 1]$. Then $C^{0,\alpha}(X)$ is a Banach space.*

Proof. It is straightforward to show that $\| \cdot \|_{C^{0,\alpha}}$ is a norm. As always, we must check completeness carefully.

Let $\{f_n\}$ be a Cauchy sequence in $C^{0,\alpha}(X)$. We want to show that there is a limit $f \in C^{0,\alpha}$ and that $\|f_n - f\|_{C^{0,\alpha}} \to 0$ as $n \to \infty$.

First of all, it is obvious from the definition of the Hölder norm that $\{f_n\}$ is a Cauchy sequence in $C^0(X)$, and since C^0 is complete, there is a continuous limit function f, and $f_n \to f$ uniformly.

Now we show $f \in C^{0,\alpha}$. Let $\epsilon > 0$. Then there is an N so that

$$m, n \geq N \quad \Longrightarrow \quad \|f_m - f_n\|_{C^{0,\alpha}} < \epsilon. \tag{47}$$

Then for all $m \geq N$, $\|f_m\|_{C^{0,\alpha}} < \|f_N\|_{C^{0,\alpha}} + \epsilon \equiv C_\epsilon$. By the definition of the Hölder norm, for all $x, y \in X$,

$$|f_m(x) - f_m(y)| \leq C_\epsilon d_X(x, y)^\alpha.$$

Taking $m \to \infty$ shows that $f \in C^{0,\alpha}$. Now (47) also implies that for all $x, y \in X$,

$$|f_m(x) - f_n(x) - f_m(y) + f_n(y)| \leq \epsilon d_X(x, y)^\alpha,$$

and so again let $m \to \infty$ to show for all $x, y \in X$, and for all $n \geq N$,

$$|f(x) - f_n(x) - f(y) + f_n(y)| \leq \epsilon d_X(x,y)^\alpha.$$

Since we already know $f_n \to f$ in C^0, this is exactly the additional statement we need to show $f_n \to f$ in $C^{0,\alpha}$. □

Remark. If X is a smooth manifold, then it is possible (by using an atlas and a subordinate partition of unity) to define $C^{k,\alpha}(X)$. If X is compact, then $C^{k,\alpha}(X) \hookrightarrow C^k(X)$ is a compact inclusion.

Homework Problem 63. *Let $\Lambda\colon \mathcal{B}_1 \to \mathcal{B}_2$ and $\Phi\colon \mathcal{B}_2 \to \mathcal{B}_3$ be linear maps between Banach spaces.*

(a) Assume Λ is continuous and Φ is compact. Then $\Phi \circ \Lambda$ is compact.

(b) Assume Λ is compact and Φ is continuous. Then $\Phi \circ \Lambda$ is compact.

Homework Problem 64. *Let $\Lambda : \mathcal{B}_1 \to \mathcal{B}_2$ be a compact linear map of Banach spaces. Show Λ is continuous.*

Hint: It suffices to show Λ is bounded. For $B_1(0)$ the unit ball in $\mathcal{B}_1$, consider the image of the compact set

$$\overline{\Lambda B_1(0)} \subset\subset \mathcal{B}_2$$

under the norm map $\|\cdot\|_{\mathcal{B}_2}\colon \mathcal{B}_2 \to \mathbb{R}$.

Remark. The Hölder spaces $C^{k,\alpha}$, for $\alpha \in (0,1)$, and the Sobolev spaces L^p_k, for $p \in (1,\infty)$, play a very important role in the theory of partial differential equations. In particular, the behave much better than the more obvious spaces C^k. Our simple proofs that $L^2_1(\mathbb{S}^1)$ embeds continuously in $C^{0,\frac{1}{2}}(\mathbb{S}^1)$ and compactly in $C^0(\mathbb{S}^1)$ constitute some of the easiest cases of *Sobolev embedding theorem*. The Sobolev embedding theorem allow us to embed certain Sobolev spaces, in which derivatives are defined only in the sense of distributions, to Hölder and C^k spaces, in which we may take derivatives in the usual sense. These spaces are crucial to the regularity theory of solutions to PDEs.

4.7 Convergence

Now we have finally developed the tools needed to solve our problem. Recall

Problem: Let $X \subset \mathbb{R}^N$ be a smooth compact manifold equipped with the Riemannian metric pulled back from the Euclidean metric on $\mathbb{R}^N$. Let $\mathcal{C}$ be the class of loops $\gamma\colon \mathbb{S}^1 \to X$ in a free homotopy class on X and in $L^2_1(\mathbb{S}^1, X)$. Find a loop of least energy in $\mathcal{C}$.

Our strategy is as follows: Define

$$L = \inf_{\gamma \in \mathcal{C}} E(\gamma).$$

Since $E(\gamma) \geq 0$ always, $L \geq 0$. Now there is a sequence of $\gamma_i \in \mathcal{C}$ so that $E(\gamma_i) \to L$. We want to find a subsequence γ_{i_j} which converges to a limit $\gamma \in \mathcal{C}$ so that $E(\gamma) = L$. Moreover, we expect γ to be a geodesic—it should satisfy the geodesic equations not just in the sense of distributions, but also in the usual sense. Therefore, by the theory of ODEs, γ should be smooth.

First of all, we show the existence of a limit γ. Corollary 77 shows that there is a subsequence of γ_i which converges uniformly to a continuous $\gamma : \mathbb{S}^1 \to X$. (For simplicity, we just refer to this subsequence as γ_i again.) Since $\gamma_i \to \gamma$ uniformly, Corollary 57 shows that γ is in the same free homotopy class. Thus we have

Proposition 79. *There is a subsequence of γ_i which converges uniformly to a limit γ in the same free homotopy class.*

Proposition 80. *Let $X \subset \mathbb{R}^N$ be a compact manifold. If $\gamma_i\colon \mathbb{S}^1 \to X$ satisfy $E(\gamma_i) \to L$, then there is a constant K independent of i so that $\|\gamma_i\|_{L^2_1(\mathbb{R}^N)} \leq K$.*

Proof. Since X is compact, there is a uniform C so that $\|\gamma_i\|_{L^2(\mathbb{S}^1,\mathbb{R}^N)} \leq C$. Since $E(\gamma_i) \to L$, $\{E(\gamma_i)\}$ is a bounded sequence. Therefore,

$$\|\gamma_i\|^2_{L^2_1(\mathbb{S}^1,\mathbb{R}^N)} = E(\gamma_i) + \|\gamma_i\|^2_{L^2(\mathbb{S}^1,\mathbb{R}^N)}$$

is bounded independent of i. □

This proposition shows there is a further subsequence of γ_i which converges weakly to a $\tilde{\gamma} \in L^2_1(\mathbb{S}^1, \mathbb{R}^N)$ by Theorem 24. (Explanatory note: a further subsequence means that we take a subsequence not just of the original γ_i, but of the subsequence taken in the paragraph above Proposition 79.) We still refer to this further subsequence as γ_i. Then Theorem 25 shows that the Hilbert space norm

$$\|\tilde{\gamma}\|_{L^2_1(\mathbb{S}^1,\mathbb{R}^N)} \leq \liminf_{i\to\infty} \|\gamma_i\|_{L^2_1(\mathbb{S}^1,\mathbb{R}^N)}.$$

Note a potential problem: We have taken a subsequence of the original γ_i to converge uniformly to a continuous γ, and then we take a further subsequence to converge weakly in L^2_1 to $\tilde{\gamma}$ in L^2_1. We must show γ and $\tilde{\gamma}$ are the same. This will follow from the fact that they must be equal in the sense of distributions, and thus are equal almost everywhere (Proposition 58). Since both γ and $\tilde{\gamma}$ are continuous, they must be equal everywhere. In particular, we require

Proposition 81. *$\gamma = \tilde{\gamma}$ in the sense of distributions.*

Proof. It suffices to show each component $\gamma^a = \tilde{\gamma}^a$ in the sense of distributions for $a = 1, \ldots, N$.

For each $a = 1, \ldots, N$, $\gamma^a_i \to \gamma^a$ uniformly as $i \to \infty$. So if $\phi \in \mathcal{D}(\mathbb{S}^1)$ is a smooth test function, then

$$|\gamma^a_i(\phi) - \gamma^a(\phi)| = \left|\int_{\mathbb{S}^1} (\gamma^a_i - \gamma^a)\phi\, dt\right| \leq \|\phi\|_{L^1}\|\gamma^a_i - \gamma^a\|_{C^0},$$

which goes to 0 as $i \to \infty$ by uniform convergence. Therefore,

$$\gamma^a(\phi) = \lim_{i\to\infty} \gamma^a_i(\phi). \tag{48}$$

Also, $\gamma^a_i \to \tilde{\gamma}^a$ weakly in $L^2_1(\mathbb{S}^1)$. Let $\phi \in \mathcal{D}(\mathbb{S}^1) \subset L^2_1(\mathbb{S}^1)$ be a test function. Let $f_i = \gamma^a_i - \tilde{\gamma}^a$. Then $f_i \to 0$ weakly in L^2_1. Compute

$$\langle f_i, \phi\rangle_{L^2_1} = \int_{\mathbb{S}^1} (f_i\phi + \dot{f}_i\dot{\phi})\, dt = \int_{\mathbb{S}^1} (f_i\phi - f_i\ddot{\phi})\, dt = f_i(\phi - \ddot{\phi}),$$

the last term denoting f_i acting in the sense of distributions. Therefore, for all $\phi \in \mathcal{D}(\mathbb{S}^1)$,

$$\lim_{i\to\infty} f_i(\phi - \ddot{\phi}) = \lim_{i\to\infty} \langle f_i, \phi\rangle_{L^2_1} = 0.$$

By Proposition 82 below, for every $\psi \in \mathcal{D}(\mathbb{S}^1)$, there is a $\phi \in \mathcal{D}(\mathbb{S}^1)$ so that $\phi - \ddot{\phi} = \psi$. Therefore, for all $\psi \in \mathcal{D}(\mathbb{S}^1)$,

$$\lim_{i\to\infty} f_i(\psi) = 0 \quad\Longleftrightarrow\quad \lim_{i\to\infty} \gamma^a_i(\psi) = \tilde{\gamma}^a(\psi).$$

Therefore, by (48) above, $\gamma = \tilde{\gamma}$ in the sense of distributions. ∎

Proposition 82. *For every* $\psi \in \mathcal{D}(\mathbb{S}^1)$*, there is a* $\phi \in \mathcal{D}(\mathbb{S}^1)$ *so that* $\psi = \phi - \ddot{\phi}$.

Proof. Recall $\mathcal{D}(\mathbb{S}^1) = C^\infty(\mathbb{S}^1, \mathbb{R})$. Moreover, Lemma 70 and Problem 58 show that

$$C^\infty(\mathbb{S}^1, \mathbb{C}) = \left\{ \sum_{k=-\infty}^{\infty} f^k e^{2\pi i k t} : \lim_{k\to\pm\infty} f^k |k|^n = 0 \text{ for } n = 1, 2, \dots \right\}. \quad (49)$$

The convergence of each such series is uniform, and the sum commutes with the derivative d/dt.

Therefore, if

$$\phi = \sum_{k=-\infty}^{\infty} \hat{\phi}^k e^{2\pi i k t} \in C^\infty(\mathbb{S}^1, \mathbb{C}),$$

then

$$\begin{aligned} \ddot{\phi} &= \sum_{k=-\infty}^{\infty} (-4\pi^2 k^2) \hat{\phi}^k e^{2\pi i k t}, \\ \phi - \ddot{\phi} &= \sum_{k=-\infty}^{\infty} (1 + 4\pi^2 k^2) \hat{\phi}^k e^{2\pi i k t}. \end{aligned}$$

So if

$$\psi = \sum_{k=-\infty}^{\infty} \hat{\psi}^k e^{2\pi i k t} \in C^\infty(\mathbb{S}^1, \mathbb{C}),$$

then we may let

$$\phi = \sum_{k=-\infty}^{\infty} \frac{\hat{\psi}^k}{1 + 4\pi^2 k^2} e^{2\pi i k t},$$

so that $\phi - \ddot{\phi} = \psi$.

We must prove that $\phi \in C^\infty(\mathbb{S}^1, \mathbb{C})$. Let n be a positive integer. Then

$$\lim_{k\to\pm\infty} \hat{\phi}^k |k|^n = \lim_{k\to\pm\infty} \frac{\hat{\psi}^k |k|^n}{1 + 4\pi^2 k^2} = 0.$$

because $|\hat{\psi}^k||k|^{n-2} \to 0$. So ϕ is smooth. (Note that we went from a $|k|^n$ limit to a $|k|^{n-2}$ limit. This is because the differential equation is of order two.)

We have considered $\mathbb{C}$-valued functions so for. It is easy to check that $\psi \in C^\infty(\mathbb{S}^1, \mathbb{R})$ implies $\phi \in C^\infty(\mathbb{S}^1, \mathbb{R})$. ∎

Remark. The previous proposition uses a standard technique for solving constant-coefficient differential equations on $\mathbb{S}^1$. The differential equation then breaks into an algebraic equation for each Fourier coefficient, each of which can be typically be solved.

This also works for functions on the n-torus $(\mathbb{S}^1)^n$. In this case, the Fourier series is summed over $\mathbb{Z}^n$, and we can solve constant-coefficient PDEs. Also, on $\mathbb{R}^n$, the Fourier transform turns constant-coefficient PDEs into algebraic equations of the Fourier transform variable.

Homework Problem 65. *$L^2_2(\mathbb{S}^1, \mathbb{C})$ is the complex Hilbert space defined by the inner product*

$$\langle f, g\rangle_{L^2_2} = \int_{\mathbb{S}^1} (f\bar{g} + \dot{f}\bar{\dot{g}} + \ddot{f}\bar{\ddot{g}})\, dt.$$

The elements of $L^2_2(\mathbb{S}^1, \mathbb{C})$ are all complex-valued functions f on $\mathbb{S}^1$ which are L^2 and whose first and second derivatives $\dot{f}$ and $\ddot{f}$ in the sense of distributions are also L^2 functions. (You may assume $L^2_2(\mathbb{S}^1, \mathbb{C})$ is a Hilbert space, as in Proposition 62.)

Show that if $f_n \to f$ converges weakly in $L^2_2(\mathbb{S}^1, \mathbb{C})$, then for all $\phi \in \mathcal{D}(\mathbb{S}^1)$, $f_n(\phi) \to f(\phi)$.

Hint: Mimic the proofs of Propositions 81 and 82.

To recap, so far we have a sequence of loops γ_i in $\mathcal{C}$ so that

$$\begin{aligned} \lim_{i\to\infty} E(\gamma_i) &= L = \inf_{\alpha\in\mathcal{C}} E(\alpha), \\ \lim_{i\to\infty} \gamma_i &= \gamma \text{ uniformly and weakly in } L^2_1(\mathbb{S}^1, \mathbb{R}^N). \end{aligned}$$

Moreover, $\gamma \in \mathcal{C}$ the same free homotopy class of L^2_1 loops containing the γ_i. Since $\gamma_i \to \gamma$ uniformly, we have

$$\|\gamma_i - \gamma\|^2_{L^2(\mathbb{S}^1,\mathbb{R}^N)} = \int_{\mathbb{S}^1} |\gamma_i - \gamma|^2\, dt \le \sup_t |\gamma_i - \gamma|^2 \to 0,$$

and so $\gamma_i \to \gamma$ in L^2.

Now Theorem 25 shows that

$$\begin{aligned}
\|\gamma\|^2_{L^2_1(\mathbb{S}^1,\mathbb{R}^N)} &\leq \liminf_{i\to\infty} \|\gamma_i\|^2_{L^2_1(\mathbb{S}^1,\mathbb{R}^N)} \\
&= \liminf_{i\to\infty} \left[E(\gamma_i) + \|\gamma_i\|^2_{L^2(\mathbb{S}^1,\mathbb{R}^N)}\right] \\
&= L + \|\gamma\|^2_{L^2(\mathbb{S}^1,\mathbb{R}^N)}, \\
E(\gamma) &= \|\gamma\|^2_{L^2_1(\mathbb{S}^1,\mathbb{R}^N)} - \|\gamma\|^2_{L^2(\mathbb{S}^1,\mathbb{R}^N)} \\
&\leq L.
\end{aligned}$$

Since L is the infimum of the energy of all loops in $\mathcal{C}$, and $\gamma \in \mathcal{C}$, then $E(\gamma) \geq L$ as well. So $E(\gamma) = L$. Thus we have proved

Theorem 30. *Let X be a compact Riemannian manifold without boundary. Then in each free homotopy class of loops, there is a $\gamma \in L^2_1(\mathbb{S}^1, X)$ which minimizes the energy.*

Corollary 83. *This minimizing γ satisfies the geodesic equations (in local coordinates on X)*

$$g_{k\ell}\left(\ddot{\gamma}^k + \Gamma^k_{ij}\dot{\gamma}^i\dot{\gamma}^j\right) = 0$$

in the sense of distributions.

Proof. See Proposition 61. □

Note in the proof of Theorem 30 above, we implicitly use the fact that the map from $L^2_1(\mathbb{S}^1) \to L^2(\mathbb{S}^1)$ is compact, by using the inclusions

$$L^2_1(\mathbb{S}^1) \hookrightarrow C^0(\mathbb{S}^1) \hookrightarrow L^2(\mathbb{S}^1),$$

the first of which is compact and the second of which is continuous. The following problem gives a direct proof.

Homework Problem 66. *Show directly that the inclusion $L^2_1(\mathbb{S}^1, \mathbb{C}) \hookrightarrow L^2(\mathbb{S}^1, \mathbb{C})$ is a compact linear map.*

Hints:

(a) Use the characterization of $L^2_1(\mathbb{S}^1, \mathbb{C})$ in terms of Fourier series from Proposition 87 below.

(b) If $\|f_i(t)\|_{L^2_1} \leq 1$, then use a diagonalization argument to produce a subsequence $\{f_{i_j}\}$ so that for each $k \in \mathbb{Z}$, the Fourier coefficients $\hat{f}^k_{i_j}$ converge to constants $g^k \in \mathbb{C}$ as $j \to \infty$.

(c) *For all* $\epsilon > 0$*, show that there is an* N *so that if* $|k| \geq N$*, then*

$$\sum_{|k|\geq N} |\hat{f}^k|^2 < \epsilon$$

for all f *such that* $\|f\|_{L^2_1} \leq 1$*.*

(d) *Conclude that the subsequence* $\{f_{i_j}\}$ *converges strongly to*

$$\sum_{k\in\mathbb{Z}} g^k e^{2\pi ikt}$$

in $L^2(\mathbb{S}^1, \mathbb{C})$*.*

Remark. The proof presented in the previous problem works for Sobolev spaces in higher dimensions (for functions on the n-dimensional torus $\mathbb{S}^1 \times \cdots \times \mathbb{S}^1$), whereas the use of the Sobolev embedding theorem for the compact inclusion $L^2_1(\mathbb{S}^1, \mathbb{C}) \hookrightarrow C^0(\mathbb{S}^1, \mathbb{C})$ is only available in dimension $n = 1$.

4.8 Regularity

Now we show that γ is smooth. First of all, note that Γ^k_{ij} is a smooth in each set of local coordinates x on X. Also, since $\gamma \in L^2_1(\mathbb{S}^1, \mathbb{R}^N)$, then we know that γ is continuous in $t \in \mathbb{S}^1$, and so $\Gamma^k_{ij}(\gamma)$ is continuous on $\mathbb{S}^1$.

Until now, we've been lax about distinguishing between $\gamma = (\gamma^1, \dots, \gamma^N) \in X \subset \mathbb{R}^N$ and γ in local coordinates. There is an important point in which we should make a distinction. Recall we are working on a coordinate chart $\phi\colon \mathcal{U} \to \mathcal{O} \subset X \subset \mathbb{R}^N$, where $\mathcal{U} \subset \mathbb{R}^n$. Our notation has been this: γ^a is the a^{th} coordinate of γ in $\mathbb{R}^N \supset X$, while γ^i has been shorthand for $(\phi^{-1} \circ \gamma)^i$ the i^{th} coordinate of $\phi^{-1} \circ \gamma$ in $\mathbb{R}^n \supset \mathcal{U}$.

In the previous subsections, we have dealt with the L^2_1 norm of γ in $\mathbb{R}^N$, while in local coordinates, we should deal with the L^2_1 norm of $\phi^{-1} \circ \gamma$ in $\mathcal{U} \subset \mathbb{R}^n$. Let $\phi^{-1}\colon \mathcal{O} \to \mathcal{U}$ be restriction of the smooth map

$$y = (y^1, \dots, y^n)\colon \mathcal{Q} \to \mathcal{U},$$

where $\mathcal{Q}$ is an open subset of $\mathbb{R}^N$ which contains $\mathcal{O} \subset X \subset \mathbb{R}^N$. (Recall we may do this by the definition of smooth maps from $\mathcal{O}$ to $\mathbb{R}^n$.) Let $x = (x^1, \dots, x^N)$ represent coordinates on $\mathbb{R}^N$. Compute for $k = 1, \dots, n$

$$\frac{\partial}{\partial t}(y \circ \gamma)^k = \frac{\partial y^k}{\partial x^a}\,\dot{\gamma}^a,$$

where a is summed from 1 to N.

Proposition 84. *Let $\phi\colon \mathcal{U} \to \mathcal{O}$ be a smooth coordinate parametrization of X. Let $I \subset \mathbb{R}$ be a compact interval, and let $K \subset \mathcal{O}$ be compact. Then there are positive constants $C_1, \dots, C_5$ so that*

$$C_1 \|\gamma\|_{L^2_1(I,\mathbb{R}^N)} + C_3 \geq \|\phi^{-1} \circ \gamma\|_{L^2_1(I,\mathbb{R}^n)} + C_4 \geq C_2 \|\gamma\|_{L^2_1(I,\mathbb{R}^N)} + C_5 \qquad (50)$$

for all γ so that $\gamma(I) \subset K$. (The point is that C_1, C_2, C_3, C_4, C_5 are independent of γ.)

Corollary 85. *$\|\gamma\|_{L^2_1(I,\mathbb{R}^N)}$ is bounded if and only if $\|\phi^{-1} \circ \gamma\|_{L^2_1(I,\mathbb{R}^n)}$ is bounded.*

Remark. A related, simpler notion is the following: Two norms $\|\cdot\|_{\mathcal{B}_1}$ and $\|\cdot\|_{\mathcal{B}_2}$ on a single linear space $\mathcal{B}$ are called *equivalent* if there are constants $C_1 > C_2 > 0$ so that for all $x \in \mathcal{B}$,

$$C_1 \|x\|_{\mathcal{B}_1} \geq \|x\|_{\mathcal{B}_2} \geq C_2 \|x\|_{\mathcal{B}_1}.$$

Remark. As long as we restrict to compact subsets of coordinate charts, the norms in $\mathbb{R}^N \supset X$ and in local coordinates on $\mathbb{R}^n$ are equivalent. The corollary holds for all the Banach function spaces we have discussed, not just for L^2_1. Also, a similar proposition holds for Banach spaces of functions from X to $\mathbb{R}$, not simply spaces of maps from $\mathbb{S}^1$ to X:

For $K \subset\subset \mathcal{O}$, the norms on $L^2_1(K)$ and $L^2_1(\phi^{-1}K)$ are equivalent under the map

$$L^2_1(K) \to L^2_1(\phi^{-1}K), \qquad f \mapsto f \circ \phi.$$

Proof of Proposition 84. We claim it suffices to prove the bound (50) separately for the L^2 norm of γ and for the L^2 norm of $\dot{\gamma}$. Proof: if $A = \|\gamma\|_{L^2}$ and $B = \|\dot{\gamma}\|_{L^2}$, then

$$\|\gamma\|_{L^2_1} = \sqrt{A^2 + B^2}.$$

Then it is easy to check that for $A, B \geq 0$,

$$\frac{1}{\sqrt{2}}(A + B) \leq \sqrt{A^2 + B^2} \leq A + B.$$

In other words, the norm on γ given by the sum of the L^2 norm of γ and the L^2 norm of $\dot{\gamma}$ is equivalent to the L^2_1 norm. It is straightforward to use this fact to prove the claim.

Since ϕ^{-1} is C^1 on K, it is locally Lipschitz and thus globally Lipschitz on K (see Proposition 17). So for C the Lipschitz constant and x_0 a point in K, for all $x \in K$,

$$\begin{aligned} |\phi^{-1}(x)| &\leq |\phi^{-1}(x_0)| + C|x - x_0| \\ &\leq C' + C|x|, \\ C' &= |\phi^{-1}(x_0)| + C|x_0|. \end{aligned}$$

Therefore, the Triangle Inequality gives

$$\begin{aligned} \|\phi^{-1}(\gamma)\|_{L^2(\mathbb{S}^1,\mathbb{R}^n)} &= \left(\int_{\mathbb{S}^1} |\phi^{-1}(\gamma(t))|^2 \, dt \right)^{\frac{1}{2}} \\ &\leq \left(\int_{\mathbb{S}^1} (C' + C|\gamma(t)|)^2 \, dt \right)^{\frac{1}{2}} \\ &\leq \left(\int_{\mathbb{S}^1} (C')^2 \, dt \right)^{\frac{1}{2}} + \left(\int_{\mathbb{S}^1} [C|\gamma(t)|]^2 \, dt \right)^{\frac{1}{2}} \\ &= C' + C\|\gamma\|_{L^2(\mathbb{S}^1,\mathbb{R}^N)} \end{aligned}$$

This is essentially one half of (50) for the L^2 norm of γ. The other half follows from the fact that ϕ is a C^1 function on the compact set $\phi^{-1}K$.

We still must address the L^2_1 norm of γ. Recall for $y = \phi^{-1}$ as above, that

$$(\phi^{-1} \circ \gamma)^{\cdot} = \frac{\partial y}{\partial x^a} \, \dot{\gamma}^a.$$

On the compact set K, since ϕ^{-1} is C^1, there is a constant C so that

$$\left| \frac{\partial y}{\partial x^a} \right| \leq C \quad \text{on } K,$$

and so on K

$$\left| (\phi^{-1} \circ \gamma)^{\cdot} \right| = \left| \frac{\partial y}{\partial x^a} \, \dot{\gamma}^a \right| \leq C \sum_{a=1}^{N} |\dot{\gamma}^a| \leq CN|\dot{\gamma}|.$$

Thus, as in the previous paragraph,

$$\|(\phi^{-1} \circ \gamma)^{\cdot}\|_{L^2(\mathbb{S}^1,\mathbb{R}^n)} \leq CN\|\dot{\gamma}\|_{L^2(\mathbb{S}^1,\mathbb{R}^N)}.$$

The opposite inequality can be obtained by considering ϕ as a C^1 map instead of ϕ^{-1}. □

Remark. In the previous proof, it sufficed to consider the L^2 norms of γ and $\dot\gamma$ separately. For higher derivatives, this is no longer adequate: Compute

$$(\phi^{-1}\circ\gamma)^{\cdot\cdot} = \frac{\partial y}{\partial x^a}\,\ddot\gamma^a + \frac{\partial^2 y}{\partial x^a \partial x^b}\,\dot\gamma^a\dot\gamma^b.$$

So first derivative terms of γ come into the calculations of the second derivatives of $\phi^{-1}\circ\gamma$.

Now we show that $\ddot\gamma^k \in L^1$ in local coordinates and in the sense of distributions.

The geodesic equation is written in terms of the coordinates on $\mathcal{U}\subset\mathbb{R}^n$, and for an open interval $I\subset\mathbb{S}^1$, $\gamma(I)\subset\mathcal{O}$. On any compact subinterval of I, there is a constant C so that the Christoffel symbols Γ^k_{ij} and the metric $g_{k\ell}$ satisfy $|g_{k\ell}(\gamma)\Gamma^k_{ij}(\gamma)|\le C$ (this is since γ is continuous on the compact interval I). Since $\gamma\in L^2_1$, each $\dot\gamma^i\in L^2$. Therefore, Hölder's inequality shows that

$$\int_I |g_{k\ell}(\gamma)\Gamma^k_{ij}(\gamma)\dot\gamma^i\dot\gamma^j|\,dt \le C\sum_{i,j=1}^n \left(\int_I |\dot\gamma^i|^2\,dt\right)^{\frac12}\left(\int_I |\dot\gamma^j|^2\,dt\right)^{\frac12} < \infty.$$

Thus $g_{k\ell}(\gamma)\Gamma^k_{ij}(\gamma)\dot\gamma^i\dot\gamma^j\in L^1(I)$ for each ℓ. Corollary 83 then shows that

$$g_{k\ell}(\gamma)\ddot\gamma^k \in L^1(I) \tag{51}$$

in the sense of distributions. Now since the inverse metric $g^{\ell m}(\gamma)$ is continuous in t, we may multiply by the L^1 quantity (51) to find $\ddot\gamma^k\in L^1(I)$ in the sense of distributions.

Now for t_0 in the interior of I, we may define for $t\in I$,

$$g(t) = \int_{t_0}^t \ddot\gamma^k(s)\,ds,$$

as in the proof of Proposition 59. It also follows from that proof that there is a constant K so that $g(t)=\dot\gamma^k(t)+K$ for almost all $t\in I$. Then Lemma 86 below shows that g is continuous. Therefore, there is a natural continuous representation for each $\dot\gamma^k$, and so γ is locally C^1.

Lemma 86. *Let $f\in L^1_{\mathrm{loc}}(\mathbb{R})$. Then*

$$g(t) = \int_{t_0}^t f(s)\,ds$$

is continuous.

Proof. Let $t \in \mathbb{R}$, and let $h > 0$ (the case $h < 0$ is similar). Compute

$$\begin{aligned} g(t+h) - g(t) &= \int_t^{t+h} f(s)\, ds \\ &= \int_{\mathbb{R}} \chi_{[t,t+h]}(s)\, f(s)\, ds \end{aligned}$$

for $\chi_{[t,t+h]}$ the characteristic function of the interval $[t, t+h]$. Then as $h \to 0$,

$$\chi_{[t,t+h]}(s)\, f(s) \to 0$$

almost everywhere on $\mathbb{R}$. For small h,

$$\left|\chi_{[t,t+h]}(s)\, f(s)\right| \leq \left|\chi_{[t-1,t+1]}(s) f(s)\right|,$$

and the right-hand function is integrable since f is locally L^1. Then the Dominated Convergence Theorem says that

$$g(t+h) - g(t) = \int_{\mathbb{R}} \chi_{[t,t+h]}(s)\, f(s)\, ds \to \int_{\mathbb{R}} 0\, ds = 0$$

as $h \to 0^+$. The case $h \to 0^-$ is similar. Thus $g(t+h) \to g(t)$ as $h \to 0$ and g is continuous at each $t \in I$. $\square$

So far we know that γ is C^1 on I. So $\Gamma^k_{ij}(\gamma)$ is C^1 and $\dot\gamma$ is continuous. Again using Proposition 83, we see that

$$g_{k\ell}(\gamma)\ddot\gamma^k = -g_{k\ell}(\gamma)\Gamma^k_{ij}(\gamma)\dot\gamma^i\dot\gamma^j$$

in the sense of distributions, and $g_{k\ell}(\gamma)\ddot\gamma^k$ is thus equal to a continuous function in the sense of distributions. Therefore, by the same argument as above, $g_{k\ell}(\gamma)\ddot\gamma^k$ is continuous, and so $\ddot\gamma^k$ is continuous also, and thus γ^k is C^2.

Once we know that γ is C^2, we see that

$$g_{k\ell}\left(\ddot\gamma^k + \Gamma^k_{ij}\dot\gamma^i\dot\gamma^j\right) = 0$$

implies that the usual geodesic equation

$$\ddot\gamma^k + \Gamma^k_{ij}\dot\gamma^i\dot\gamma^j = 0$$

holds in the sense of ordinary derivatives, and since Γ^k_{ij} is smooth, the usual regularity theory for ODEs, Theorem 9, applies, and the geodesic γ is smooth.

4.9 Sobolev spaces, distributions, and Fourier series

In this subsection, we provide some more background results about Sobolev spaces and distributions on $\mathbb{S}^1$.

First of all, we describe $\mathbb{C}$ valued distributions. A complex valued distribution is a $\mathbb{C}$-linear map from $C^\infty(\mathbb{S}^1, \mathbb{C})$ to $\mathbb{C}$.

Example 22. *For $k \in \mathbb{Z}$, the map*

$$\phi \mapsto \hat{\phi}^k = \int_{\mathbb{S}^1} \phi \, \overline{e^{2\pi i k t}} \, dt$$

is a distribution.

Proposition 87.

$$L^2_1(\mathbb{S}^1, \mathbb{C}) = \left\{ \sum_{k \in \mathbb{Z}} f^k e^{2\pi i k t} : \sum_{k \in \mathbb{Z}} |f^k|^2 (k^2 + 1) < \infty \right\}.$$

Moreover, the norm $\|f\|_{L^2_1}$ is equivalent to

$$\left(\sum_{k \in \mathbb{Z}} |\hat{f}^k|^2 (k^2 + 1) \right)^{\frac{1}{2}}.$$

Proof. First we show $\subset$. Let $f \in L^2_1(\mathbb{S}^1, \mathbb{C})$ and compute

$$\dot{f}(e^{-2\pi i k t}) = \int_{\mathbb{S}^1} \dot{f}(t) \, e^{-2\pi i k t} \, dt = -\int_{\mathbb{S}^1} f(t)(-2\pi i k) e^{-2\pi i k t} \, dt = 2\pi i k \hat{f}^k.$$

Since $\dot{f} \in L^2$,

$$\sum_{k \in \mathbb{Z}} 4\pi^2 k^2 |\hat{f}^k|^2 = \|\dot{f}\|^2_{L^2} < \infty. \quad \Longleftrightarrow \quad \sum_{k \in \mathbb{Z}} k^2 |\hat{f}^k|^2 < \infty.$$

Now since $f \in L^2$ also, then

$$\sum_{k \in \mathbb{Z}} |\hat{f}^k|^2 < \infty \quad \text{and} \quad \sum_{k \in \mathbb{Z}} |\hat{f}^k|^2 (k^2 + 1) < \infty.$$

This proves $\subset$.

To show $\supset$, note that

$$\sum_{k\in\mathbb{Z}} |f^k|^2(k^2+1) < \infty \quad \Longleftrightarrow \quad \sum_{k\in\mathbb{Z}} |f^k|^2 < \infty \quad \text{and} \quad \sum_{k\in\mathbb{Z}} k^2|f^k|^2 < \infty.$$

Therefore,

$$f = \sum_{k\in\mathbb{Z}} f^k e^{2\pi ikt} \in L^2,$$

and by the computations in the previous paragraph $\dot{f}^k \equiv \dot{f}(e^{-2\pi ikt}) = 2\pi ikf^k$. Consider a test function $\phi \in C^\infty(\mathbb{S}^1, \mathbb{C})$. Then compute

$$\begin{aligned}
\dot{f}(\overline{\phi}) &= -f(\overline{\dot{\phi}}) \\
&= -\int_{\mathbb{S}^1} f\,\overline{\dot{\phi}}\,dt \\
&= -\langle f, \dot{\phi}\rangle_{L^2} \\
&= -\sum_{k\in\mathbb{Z}} f^k\,\overline{2\pi ik\hat{\phi}^k} \\
&= -\sum_{k\in\mathbb{Z}} (-2\pi ik) f^k\,\overline{\hat{\phi}^k} \\
&= \sum_{k\in\mathbb{Z}} \dot{f}^k \overline{\hat{\phi}^k} \\
&= \left\langle \sum_{k\in\mathbb{Z}} \dot{f}^k e^{2\pi ikt}, \phi \right\rangle_{L^2}.
\end{aligned}$$

This shows that

$$\dot{f} = \sum_{k\in\mathbb{Z}} \dot{f}^k e^{2\pi ikt} = \sum_{k\in\mathbb{Z}} (2\pi ik) f^k e^{2\pi ikt}$$

in the sense of distributions. Therefore, both f and $\dot{f}$ are in L^2, and thus $f \in L^2_1(\mathbb{S}^1, \mathbb{C})$.

The statement about equivalence of the norms follows easily. □

Remark. Similar easy calculations show that

$$L^2_m(\mathbb{S}^1, \mathbb{C}) = \left\{ \sum_{k\in\mathbb{Z}} f^k e^{2\pi ikt} : \sum_{k\in\mathbb{Z}} |f^k|^2(k^2+1)^m < \infty \right\}$$

for every $m = 0, 1, 2, \ldots$. Our characterization of smooth functions in (49) above then shows that

$$C^\infty(\mathbb{S}^1, \mathbb{C}) = \bigcap_{m=0}^{\infty} L^2_m(\mathbb{S}^1, \mathbb{C})$$

Proof: it is straightforward to show that $L^2_m(\mathbb{S}^1, \mathbb{C})$ compactly embeds in $C^{m-1}(\mathbb{S}^1, \mathbb{C})$ for all $m \geq 1$.

The Fourier series isometry between $L^2(\mathbb{S}^1, \mathbb{C})$ and sequences $\ell^2 = L^2(\mathbb{Z}, \mathbb{C})$ also allows us to define even more Sobolev spaces.

For any $s \in \mathbb{R}$, define $L^2_s(\mathbb{S}^1, \mathbb{C})$ to be the set of distributions f which act on

$$\phi = \sum_{k \in \mathbb{Z}} \hat{\phi}^k e^{2\pi i k t}$$

by

$$f(\phi) = \sum_{k \in \mathbb{Z}} \hat{f}^k \hat{\phi}^k, \tag{52}$$

where $\hat{f}^k = f(\overline{e^{2\pi i k t}})$ and we assume that

$$\sum_{k \in \mathbb{Z}} |\hat{f}^k|^2 (1 + k^2)^s < \infty. \tag{53}$$

Homework Problem 67. *Show that if $\hat{f}^k$ is a sequence of complex numbers satisfying (53), then for any $\phi \in C^\infty(\mathbb{S}^1, \mathbb{C})$, the sum in (52) converges.*

Now we are able to put a topology on $C^\infty(\mathbb{S}^1, \mathbb{C})$. We only describe this topology in terms of convergence of sequences. We say $\phi_j \to \phi$ in $C^\infty(\mathbb{S}^1, \mathbb{C})$, if $\phi_j \to \phi$ in $L^2_m(\mathbb{S}^1, \mathbb{C})$ for all $m \geq 0$.

Homework Problem 68. *Show that $\phi_j \to \phi$ in $C^\infty(\mathbb{S}^1, \mathbb{C})$ if and only if $\phi_j \to \phi$ in $C^p(\mathbb{S}^1, \mathbb{C})$ for all $p \geq 0$.*

Hint: You may use the fact that $L^2_m(\mathbb{S}^1, \mathbb{C})$ embeds compactly into $C^{m-1}(\mathbb{S}^1, \mathbb{C})$ for each $m \geq 1$. Also show that $C^p(\mathbb{S}^1, \mathbb{C})$ embeds continuously into $L^2_p(\mathbb{S}^1, \mathbb{C})$ for all $p \geq 0$.

Now we finally give the correct definition of complex distributions on $\mathbb{S}^1$. A distribution on $\mathbb{S}^1$ is a continuous $\mathbb{C}$-linear map from $C^\infty(\mathbb{S}^1, \mathbb{C})$ to $\mathbb{C}$. Denote the space of complex distributions on $\mathbb{S}^1$ by $\mathcal{D}'(\mathbb{S}^1, \mathbb{C})$.

Proposition 88. $\mathcal{D}'(\mathbb{S}^1,\mathbb{C}) = \bigcup_{m\in\mathbb{Z}} L^2_m(\mathbb{S}^1,\mathbb{C})$, *and the image of* $\mathcal{D}'(\mathbb{S}^1,\mathbb{C})$ *under the Fourier transform is the set of all polynomially bounded complex sequences. In other words, it is the set of all sequences* $\{f^k\}$ *so that there are* $m\in\mathbb{Z}$, $C>0$ *so that* $|f^k| \le C(k^2+1)^{\frac{m}{2}}$ *for all* $k\in\mathbb{Z}$.

Proof. We prove the first equality, and leave the rest as an exercise.

To prove $\supset$, if f is in the union, then $f\in L^2_{-m}(\mathbb{S}^1,\mathbb{C})$ for some positive m. To show $f\in\mathcal{D}'(\mathbb{S}^1,\mathbb{C})$, consider a sequence of $\phi_j\to\phi$ in $C^\infty(\mathbb{S}^1,\mathbb{C})$. Then by definition, $\phi_j\to\phi$ in L^2_m. Then

$$\begin{aligned}
|f(\phi_j)-f(\phi)| &= |f(\phi_j-\phi)| \\
&\le \sum_{k\in\mathbb{Z}} |\hat{f}^k(\hat{\phi}^k_j-\hat{\phi}^k)| \\
&= \sum_{k\in\mathbb{Z}} \frac{|\hat{f}^k|}{(1+k^2)^{\frac{m}{2}}}\left[|\hat{\phi}^k_j-\hat{\phi}^k|(1+k^2)^{\frac{m}{2}}\right] \\
&\le \left(\sum_{k\in\mathbb{Z}} \frac{|\hat{f}^k|^2}{(1+k^2)^m}\right)^{\frac{1}{2}} \left(\sum_{k\in\mathbb{Z}} |\hat{\phi}^k_j-\hat{\phi}^k|^2(1+k^2)^m\right)^{\frac{1}{2}}.
\end{aligned}$$

The second term in the last line goes to zero by the remark after Proposition 87, while the first term is finite by the fact $f\in L^2_{-m}$. Therefore, $f(\phi_j)\to f(\phi)$ for every test function ϕ, and $f\in\mathcal{D}'(\mathbb{S}^1,\mathbb{C})$.

We prove $\subset$ by contradiction. If $f\in\mathcal{D}'(\mathbb{S}^1,\mathbb{C})$ is not in $L^2_m(\mathbb{S}^1,\mathbb{C})$ for every $m\in\mathbb{Z}$, then for all $m\in\mathbb{Z}$,

$$\sum_{k=-\infty}^{\infty} |\hat{f}^k|^2(1+k^2)^m = \infty.$$

This implies that

$$\sup_{k\in\mathbb{Z}} |\hat{f}^k|^2(1+k^2)^m = \infty \quad \text{for all } m\in\mathbb{Z}.$$

(Proof of the contrapositive:

$$|\hat{f}^k|^2(1+k^2)^m \le C \quad\Longrightarrow\quad \sum_{k\in\mathbb{Z}} |\hat{f}^k|^2(1+k^2)^{m-1} \le \sum_{k\in\mathbb{Z}} \frac{C}{1+k^2} < \infty.)$$

So for each j, there is a k_j so that

$$\frac{|\hat{f}^{k_j}|}{(1+k_j^2)^{\frac{j}{2}}} \geq 1.$$

We may assume $k_j \neq 0$.

Now we construct a sequence ϕ_j which converges to 0 in $C^\infty(\mathbb{S}^1, \mathbb{C})$, but for which $f(\phi_j) \not\to 0$. Define

$$\phi_j = \frac{\overline{\hat{f}^{k_j}}}{|\hat{f}^{k_j}|(1+k_j^2)^{\frac{j}{2}}} \, e^{2\pi i k_j t}.$$

Compute

$$\|\phi_j\|_{L_n^2}^2 \approx \frac{(1+k_j^2)^n}{(1+k_j^2)^j} = (1+k_j^2)^{n-j},$$

where $\approx$ denotes equivalence of norms. For each fixed n, since each $k_j^2 \geq 1$, then

$$\lim_{j\to\infty} \|\phi_j\|_{L_n^2}^2 = 0,$$

and so $\phi_j \to 0$ in $C^\infty(\mathbb{S}^1, \mathbb{C})$. On the other hand,

$$f(\phi_j) = \hat{f}^{k_j} \cdot \frac{\overline{\hat{f}^{k_j}}}{|\hat{f}^{k_j}|(1+k_j^2)^{\frac{j}{2}}} = \frac{|\hat{f}^{k_j}|}{(1+k_j^2)^{\frac{j}{2}}} \geq 1.$$

So $f(\phi_j) \not\to 0 = f(0) = f(\lim \phi_j)$, where $\phi_j \to 0$ in $C^\infty(\mathbb{S}^1, \mathbb{C})$. $\square$