3

Geometric Preliminaries

In this chapter we discuss certain geometric preliminaries required for studying the geodesic X-ray transform on a general compact Riemannian manifold (M, g) with boundary. We will discuss the concept of a compact non-trapping manifold with strictly convex boundary. We will also introduce the exit time function τ, the geodesic vector field X, the geodesic flow φ_t, the scattering relation α, and the vector fields $X_\perp$ and V. The chapter will conclude with a discussion of conjugate points and with the important notion of a simple manifold, including several equivalent definitions.

3.1 Non-trapping and Strict Convexity

Let (M, g) be a compact, connected, and oriented Riemannian manifold with smooth boundary ∂M and dimension $n \geq 2$. We will denote the inner product induced by the metric g on tangent vectors by $\langle v, w \rangle_g$ and the norm by $|v|_g$. The subscript g will often be omitted for brevity.

Geodesics travel at constant speed, so we fix the speed to be one. We pack positions and velocities together in what we call the *unit sphere bundle* SM. This consists of pairs (x, v), where $x \in M$ and $v \in T_x M$ with norm $|v|_g = 1$. Given $(x, v) \in SM$, let $\gamma_{x,v}$ denote the unique geodesic determined by (x, v) so that $\gamma_{x,v}(0) = x$ and $\dot{\gamma}_{x,v}(0) = v$. For any $(x, v) \in SM$, the geodesic $\gamma_{x,v}$ is defined on a maximal interval of existence that we denote by $[-\tau_-(x, v), \tau_+(x, v)]$ where $\tau_\pm(x, v) \in [0, \infty]$, so that

$$\gamma_{x,v} : [-\tau_-(x, v), \tau_+(x, v)] \to M$$

is a smooth curve that cannot be extended to any larger interval as a smooth curve in M.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Definition 3.1.1 We let

$$\tau(x,v) := \tau_+(x,v).$$

Thus $\tau(x,v)$ is the *exit time* when the geodesic $\gamma_{x,v}$ exits M.

Exercise 3.1.2 Give examples of compact manifolds (M,g) with boundary and points $(x,v) \in SM$ where the following holds:

(a) The first time when $\gamma_{x,v}$ hits ∂M is different from the exit time $\tau(x,v)$.
(b) $\tau(x,v)$ is not continuous on SM.
(c) $\tau_\pm(x,v) = \infty$.
(d) $\tau_-(x,v)$ is finite but $\tau_+(x,v) = \infty$.

If some geodesic has infinite length, one needs to be careful when studying the geodesic X-ray transform since the integral of a smooth function over such a geodesic may not be finite. For the most part of this book, we will be working on manifolds where this issue does not appear.

Definition 3.1.3 We say that (M,g) is *non-trapping* if $\tau(x,v) < \infty$ for all $(x,v) \in SM$. Equivalently, there are no geodesics in M with infinite length.

Example 3.1.4 Compact subdomains in $\mathbb{R}^n$ and in hyperbolic space are non-trapping, and so are the small spherical caps in Example 2.5.4. Large spherical caps, catenoid-type surfaces, and flat cylinders have trapped geodesics (see Examples 2.5.5–2.5.7).

Unit tangent vectors at the boundary of M constitute the boundary ∂SM of SM and will play a special role. Specifically,

$$\partial SM := \{(x,v) \in SM : x \in \partial M\}.$$

We will need to distinguish those tangent vectors pointing inside ('influx boundary') and those pointing outside ('outflux boundary'), so we define two subsets of ∂SM as

$$\partial_\pm SM := \{(x,v) \in \partial SM : \pm\langle v, \nu(x)\rangle_g \geq 0\},$$

where ν denotes the **inward** unit normal vector to the boundary (cf. Figure 3.1). The convention of using the inward unit normal instead of the outward unit normal will eliminate some minus signs in the volume form $d\mu$ in Section 3.6 and certain other places. We also denote

$$\partial_0 SM := \partial_+ SM \cap \partial_- SM.$$

Note that one has $\partial_0 SM = S(\partial M)$.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

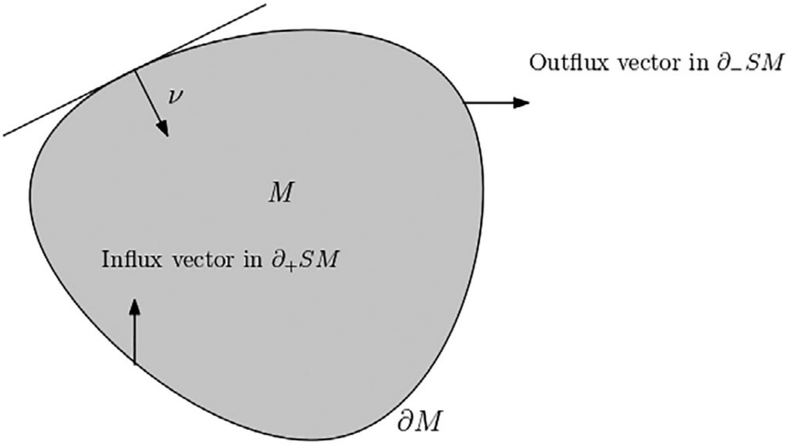

Figure 3.1 Influx and outflux boundaries.

Definition 3.1.5 The *geodesic X-ray transform* of a function $f \in C^\infty(M)$ on a compact non-trapping manifold (M, g) with smooth boundary is the function If defined by

$$If(x,v) = \int_0^{\tau(x,v)} f(\gamma_{x,v}(t))\, dt, \qquad (x,v) \in \partial_+ SM. \tag{3.1}$$

The idea is that if M is non-trapping, then any geodesic γ going through some point $(y,w) \in SM$ has an initial point $(x,v) = \gamma_{y,w}(-\tau_-(y,w))$. We must have $(x,v) \in \partial SM$, since if we had $(x,v) \in SM^{\mathrm{int}}$ then the geodesic could be extended further in both directions. Moreover, we must have $(x,v) \in \partial_+ SM$ since any geodesic starting at a point in $\partial SM \setminus \partial_+ SM$ could be extended further for small negative times.

The argument in the preceding paragraph shows that on non-trapping manifolds, there is a one-to-one correspondence between the set of unit speed geodesics and the set $\partial_+ SM$ of their initial points. Parametrizing geodesics by their initial points in $\partial_+ SM$ means that we are using the *fan-beam* parametrization of geodesics.

Remark 3.1.6 Note that the fan-beam parametrization is different from the *parallel-beam* parametrization that we used in Chapter 1, and also from the parametrization used in Section 2.4 for geodesics of a radial sound speed under the Herglotz condition based on their closest point to the origin.

Since f is smooth and the point $\gamma_{x,v}(t)$ depends smoothly on (x, v), the formula (3.1) shows that the regularity properties of If are decided by the regularity properties of the exit time function $\tau(x, v)$. If the boundary of M is not strictly convex, it can happen that τ is discontinuous. On the other hand, if ∂M is strictly convex then τ will be continuous and in fact smooth in most places, and the theory will be particularly clean.

For a precise definition of when the boundary ∂M is strictly convex, we will use the *second fundamental form* of ∂M that describes how ∂M sits inside M. Recall that the (scalar) second fundamental form is the bilinear form on $T\partial M$ given by

$$\Pi_x(v, w) := -\langle \nabla_v \nu, w \rangle_g,$$

where $x \in \partial M$ and $v, w \in T_x\partial M$. Here ∇ is the Levi-Civita connection of g, and on the right-hand side ν is extended arbitrarily as a smooth vector field in M (recall that $\nabla_X Y|_x$ only depends on $X|_x$ and the value of Y along any curve $\eta(t)$ with $\dot\eta(0) = X|_x$, so that $\Pi_x(v, w)$ does not depend on the choice of the extension of ν).

Definition 3.1.7 We shall say that ∂M is *strictly convex* if Π_x is positive definite for all $x \in \partial M$.

The combination of non-trapping with strict convexity of the boundary will produce several desirable properties. In fact, many results in this book will be stated either for compact non-trapping manifolds with strictly convex boundary, or for simple manifolds, which satisfy the additional condition that geodesics do not have conjugate points.

We already encountered the notion of strict convexity in Section 2.5, where this notion was related to the behaviour of tangential geodesics. We wish to show that a similar characterization exists in the general case. To do this, it is convenient to introduce the following notions.

Lemma 3.1.8 (Closed extension) *Let (M, g) be a compact manifold with smooth boundary. There is a closed (=compact without boundary) connected manifold (N, g) having the same dimension as M so that (M, g) is isometrically embedded in (N, g).*

Proof (Special case) The lemma has an easy proof in the special case where M is a subset of $\mathbb{R}^n$. In that case it is enough to consider some cube $N = [-R, R]^n$ with $M \subset N^{\mathrm{int}}$, and to extend g smoothly as a $2R$-periodic positive definite symmetric matrix function in N. Identifying the opposite sides of N, we see that (N, g) becomes a torus with (M, g) embedded in its interior. Then (N, g) is the required extension. $\square$

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Exercise 3.1.9 Prove Lemma 3.1.8 in general, by considering the double of the manifold M.

If (N, g) is a closed extension of (M, g), we continue to write $\gamma_{x,v}(t)$ for the geodesic in (N, g). One benefit of working with a closed extension is that now $\gamma_{x,v}(t)$ is well defined and smooth for all $t \in \mathbb{R}$.

Lemma 3.1.10 (Boundary defining function) *Let (M, g) be a compact manifold with smooth boundary, and let (N, g) be a closed extension. There is a function $\rho \in C^\infty(N)$, called a* boundary defining function, *so that $\rho(x) = d(x, \partial M)$ near ∂M in M, and*

$$\begin{aligned} M &= \{x \in N \,:\, \rho \geq 0\}, \\ \partial M &= \{x \in N \,:\, \rho = 0\}, \\ N \setminus M &= \{x \in N \,:\, \rho < 0\}. \end{aligned}$$

One has $\nabla\rho(x) = \nu(x)$ for all $x \in \partial M$.

Exercise 3.1.11 Prove Lemma 3.1.10.

The following result shows that the second fundamental form of ∂M is given by the Riemannian Hessian of ρ, defined in terms of the total covariant derivative ∇ by

$$\mathrm{Hess}(\rho) = \nabla^2\rho = \left(\partial_{x_j}\partial_{x_k}\rho - \Gamma^l_{jk}\partial_{x_l}\rho\right) dx^j \otimes dx^k.$$

Moreover, strict convexity of the boundary can indeed be characterized by the behaviour of tangential geodesics.

Lemma 3.1.12 (Strictly convex boundary) *If (M, g) is a compact manifold with smooth boundary and ρ is as in Lemma 3.1.10, then for any $(x, v) \in \partial_0 SM$, one has*

$$-\Pi_x(v, v) = \mathrm{Hess}_x(\rho)(v, v) = \frac{d^2}{dt^2}\rho(\gamma_{x,v}(t))\Big|_{t=0}.$$

Thus ∂M is strictly convex if and only if any geodesic in N starting from some point $(x, v) \in \partial_0 SM$ satisfies $\frac{d^2}{dt^2}\rho(\gamma_{x,v}(t))\big|_{t=0} < 0$. In particular, any geodesic tangent to ∂M stays outside M for small positive and negative times, and any maximal M-geodesic going from ∂M into M stays in M^{int} except for its end points.

The proof will follow from the next lemma, which will also be useful later.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Lemma 3.1.13 *Let ρ be as in Lemma 3.1.10, and consider the smooth function*

$$h\colon SN \times \mathbb{R} \to \mathbb{R}, \ \ h(x, v, t) = \rho(\gamma_{x,v}(t)).$$

If $(x, v) \in SN$ and if t_0 is such that $x_0 := \gamma_{x,v}(t_0) \in \partial M$, then one has

$$\begin{aligned} h(x, v, t_0) &= 0, \\ \frac{\partial h}{\partial t}(x, v, t_0) &= \langle \nu(x_0), \dot\gamma_{x,v}(t_0)\rangle, \\ \frac{\partial^2 h}{\partial t^2}(x, v, t_0) &= \langle \nabla_{\dot\gamma_{x,v}(t_0)} \nabla\rho, \dot\gamma_{x,v}(t_0)\rangle = \mathrm{Hess}_{x_0}(\rho)(\dot\gamma_{x,v}(t_0), \dot\gamma_{x,v}(t_0)). \end{aligned}$$

Proof Write $\gamma(t) = \gamma_{x,v}(t)$. Since $\rho|_{\partial M} = 0$ one has $h(x, v, t_0) = 0$. Moreover, using that $\nabla\rho|_{\partial M} = \nu$ we compute

$$\frac{\partial h}{\partial t}(x, v, t_0) = d\rho|_{x_0}(\dot\gamma(t_0)) = \langle \nu(x_0), \dot\gamma(t_0)\rangle.$$

Finally, one has

$$\begin{aligned} \frac{\partial^2 h}{\partial t^2}(x, v, t_0) &= \frac{d}{dt}(d\rho|_{\gamma(t)}(\dot\gamma(t)))\Big|_{t=t_0} = \frac{d}{dt}\langle \nabla\rho|_{\gamma(t)}, \dot\gamma(t)\rangle\Big|_{t=t_0} \\ &= \langle \nabla_{\dot\gamma(t)}\nabla\rho, \dot\gamma(t)\rangle + \langle \nabla\rho, \nabla_{\dot\gamma(t)}\dot\gamma(t)\rangle|_{t=t_0}. \end{aligned}$$

The last term is zero since γ is a geodesic (i.e. $\nabla_{\dot\gamma(t)}\dot\gamma(t) = 0$). The definition of the total covariant derivative ∇ gives that $\langle \nabla_{\dot\gamma(t)}\nabla\rho, \dot\gamma(t)\rangle|_{t=t_0} = \nabla^2\rho(\dot\gamma(t_0), \dot\gamma(t_0))$, which finishes the proof. □

Proof of Lemma 3.1.12 Let $(x, v) \in \partial_0 SM$ and write $\gamma(t) = \gamma_{x,v}(t)$ and $h(x, v, t) = \rho(\gamma(t))$. By Lemma 3.1.13 one has

$$\begin{aligned} h(x, v, 0) &= 0, \\ \frac{\partial h}{\partial t}(x, v, 0) &= 0, \\ \frac{\partial^2 h}{\partial t^2}(x, v, 0) &= \langle \nabla_v \nabla\rho, v\rangle = \mathrm{Hess}_x(\rho)(v, v). \end{aligned}$$

But $\nabla\rho|_{\partial M} = \nu$, which shows that $\langle \nabla_v\nabla\rho, v\rangle = -\Pi_x(v, v)$. This proves the required formula.

Now ∂M is strictly convex $\iff \Pi_x(v, v) > 0$ for all $(x, v) \in \partial_0 SM \iff \partial_t^2 h(x, v, 0) < 0$ for all $(x, v) \in \partial_0 SM$. By the Taylor formula,

$$\rho(\gamma(t)) = h(x, v, t) = -\frac{1}{2}\Pi_x(v, v)t^2 + O(t^3)$$

when $|t|$ is small. This shows that for small positive and negative times $\rho(\gamma(t)) < 0$, i.e. $\gamma_{x,v}(t)$ is in $N \setminus M$. □

3.2 Regularity of the Exit Time

We will now discuss in detail the regularity of the fundamental exit time function τ on a compact non-trapping manifold (M,g) with strictly convex boundary. Note that by definition, $\tau|_{\partial_- SM} = 0$.

Example 3.2.1 Let $M = \overline{\mathbb{D}}$ be the closed unit disk in the plane, and let $g = e$ be the Euclidean metric. Take $x = (0,-1)$ and let $v_\theta = (\cos\theta, \sin\theta)$. An easy geometric argument shows that

$$\tau(x, v_\theta) = \begin{cases} 2\sin\theta, & \theta \in [0,\pi], \\ 0, & \theta \in [-\pi, 0] \end{cases}.$$

Thus τ is continuous on ∂SM but fails to be continuously differentiable in tangential directions. However, the odd extension of $\tau|_{\partial_+ SM}$ with respect to $(x,v) \mapsto (x,-v)$,

$$\tilde{\tau}(x, v_\theta) := \begin{cases} 2\sin\theta, & \theta \in [0,\pi], \\ 2\sin\theta, & \theta \in [-\pi, 0], \end{cases}$$

is clearly smooth on ∂SM.

Exercise 3.2.2 Verify the claims in Example 3.2.1.

We will now show that the properties of the exit time function in Example 3.2.1 are valid in general.

Lemma 3.2.3 *Let (M,g) be a compact non-trapping manifold with strictly convex boundary. Then τ is continuous on SM and smooth on $SM \setminus \partial_0 SM$.*

Proof The proof that τ is continuous is left as an exercise. Let (N,g) be a closed extension of (M,g) and let ρ be a boundary defining function as in Lemma 3.1.10. Define $h\colon SN \times \mathbb{R} \to \mathbb{R}$, $h(x,v,t) := \rho(\gamma_{x,v}(t))$ as in Lemma 3.1.13. Then

$$\frac{\partial h}{\partial t}(x,v,t) = d\rho(\dot{\gamma}_{x,v}(t)) = \langle \nabla \rho(\gamma_{x,v}(t)), \dot{\gamma}_{x,v}(t)\rangle.$$

Assume that $(x,v) \in SM \setminus \partial_0 SM$, and set $y := \gamma_{x,v}(\tau(x,v)) \in \partial M$. Since y is the final point of the geodesic, one must have $\dot{\gamma}_{x,v}(\tau(x,v)) \in \partial_- SM$ (otherwise the geodesic could be extended further). By strict convexity, one must also have $\dot{\gamma}_{x,v}(\tau(x,v)) \notin \partial_0 SM$ (since otherwise $\tau(x,v) = 0$ and (x,v) would be in $\partial_0 SM$).

Thus $\dot{\gamma}_{x,v}(\tau(x,v)) \in \partial SM \setminus \partial_+ SM$, i.e. $\langle \dot{\gamma}_{x,v}(\tau(x,v)), \nu\rangle < 0$. Since $\nabla\rho$ agrees with ν on ∂M, we see that

$$\frac{\partial h}{\partial t}(x,v,\tau(x,v)) < 0.$$

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Since $h(x, v, \tau(x, v)) = 0$ and h is smooth, the implicit function theorem ensures that τ is smooth in $SM \setminus \partial_0 SM$. $\square$

The set $\partial_0 SM$, where geodesics are tangential to ∂M and τ is not smooth, is often called the *glancing region*. This terminology comes from the theory of boundary value problems for hyperbolic equations (Hörmander, 1983–1985, chapter 24).

Exercise 3.2.4 Show that τ is continuous in SM.

Exercise 3.2.5 Show that τ is indeed not smooth at the glancing region $\partial_0 SM$.

The next result shows that the odd extension of $\tau|_{\partial_+ SM}$ is smooth on ∂SM.

Lemma 3.2.6 (Odd extension of τ on ∂SM) *Let (M, g) be a compact non-trapping manifold with strictly convex boundary and define $\tilde{\tau} : \partial SM \to \mathbb{R}$ by*

$$\tilde{\tau}(x, v) := \begin{cases} \tau(x, v), & (x, v) \in \partial_+ SM, \\ -\tau(x, -v), & (x, v) \in \partial_- SM. \end{cases}$$

Then $\tilde{\tau} \in C^\infty(\partial SM)$; in particular, $\tau|_{\partial_+ SM} : \partial_+ SM \to \mathbb{R}$ is smooth.

Proof As before we let $h(x, v, t) := \rho(\gamma_{x,v}(t))$ for $(x, v) \in \partial SM$ and $t \in \mathbb{R}$. Note that by Lemma 3.1.13, with the choice $t_0 = 0$, one has

- $h(x, v, 0) = 0$;
- $\frac{\partial h}{\partial t}(x, v, 0) = \langle \nu(x), v \rangle$;
- $\frac{\partial^2 h}{\partial t^2}(x, v, 0) = \mathrm{Hess}_x(\rho)(v, v)$.

Hence the Taylor formula shows that for some smooth function $R(x, v, t)$, we can write

$$\begin{aligned} h(x, v, t) &= \langle \nu(x), v \rangle t + \frac{1}{2} \mathrm{Hess}_x(\rho)(v, v) t^2 + R(x, v, t) t^3 \\ &= t F(x, v, t), \end{aligned}$$

where F is the smooth function

$$F(x, v, t) := \langle \nu(x), v \rangle + \frac{1}{2} \mathrm{Hess}_x(\rho)(v, v) t + R(x, v, t) t^2.$$

Since $h(x, v, \tilde{\tau}(x, v)) = 0$, we have $\tilde{\tau} F(x, v, \tilde{\tau}) = 0$ and hence

$$F(x, v, \tilde{\tau}(x, v)) = 0. \tag{3.2}$$

Here we used that $\tilde{\tau}(x, v) = 0$ implies $\langle \nu(x), v \rangle = 0$ by strict convexity. Moreover,

$$\frac{\partial F}{\partial t}(x, v, 0) = \frac{1}{2} \mathrm{Hess}_x(\rho)(v, v).$$

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

But for $(x,v) \in \partial_0 SM$, $\mathrm{Hess}_x\rho(v,v) = -\Pi_x(v,v) < 0$ by strict convexity. Thus by the implicit function theorem, $\tilde{\tau}$ is smooth in a neighbourhood of $\partial_0 SM$. Since $\tilde{\tau}$ is smooth in $\partial SM \setminus \partial_0 SM$ by Lemma 3.2.3, the result follows. □

Remark 3.2.7 Note that we can define $\tilde{\tau}$ on all SM by setting $\tilde{\tau}(x,v) := \tau(x,v) - \tau(x,-v)$. The restriction of this function to ∂SM coincides with the definition of $\tilde{\tau}$ given by Lemma 3.2.6. It turns out that in fact $\tilde{\tau} \in C^\infty(SM)$. This stronger result is proved in Lemma 3.2.11.

Define

$$\mu(x,v) := \langle \nu(x), v\rangle, \qquad (x,v) \in \partial SM.$$

This expression appears in Santaló's formula, which is an important change of variables formula on SM (see Section 3.6). We record the following result for later purposes.

Lemma 3.2.8 *Let (M,g) be a compact non-trapping manifold with strictly convex boundary. The function $\mu/\tilde{\tau}$ extends to a smooth positive function on ∂SM whose value at $(x,v) \in \partial_0 SM$ is*

$$\frac{\Pi_x(v,v)}{2}.$$

Proof Using (3.2) we can write

$$\mu(x,v) = -\frac{1}{2}\mathrm{Hess}_x(\rho)(v,v)\tilde{\tau} - R(x,v,\tilde{\tau})\tilde{\tau}^2,$$

and hence for $(x,v) \in \partial SM \setminus \partial_0 SM$ near $\partial_0 SM$, we can write

$$\mu/\tilde{\tau} = -\frac{1}{2}\mathrm{Hess}_x(\rho)(v,v) - R(x,v,\tilde{\tau})\tilde{\tau}.$$

But the right-hand side of the last equation is a smooth function near $\partial_0 SM$ since R and $\tilde{\tau}$ are; its value at $(x,v) \in \partial_0 SM$ is $\Pi_x(v,v)/2$. Finally, observe that μ and $\tilde{\tau}$ are both positive for $(x,v) \in \partial_+ SM \setminus \partial_0 SM$ and both negative for $(x,v) \in \partial_- SM \setminus \partial_0 SM$. □

Even more precise regularity properties of the exit time function τ near $\partial_0 SM$ can be obtained from the next lemma. This will be the main tool when studying regularity properties of solutions to transport equations. The proof is motivated by the theory of Whitney folds, cf. (Hörmander, 1983–1985, Appendix C.4) and Section 5.2.

Lemma 3.2.9 *Let (M, g) be compact with smooth boundary, let $(x_0, v_0) \in \partial_0 SM$, and let ∂M be strictly convex near x_0. Assume that M is embedded in a compact manifold N without boundary. Then, near (x_0, v_0) in SM, one has*

$$\tau(x, v) = Q(\sqrt{a(x, v)}, x, v),$$
$$-\tau(x, -v) = Q(-\sqrt{a(x, v)}, x, v),$$

where Q is a smooth function near $(0, x_0, v_0)$ in $\mathbb{R} \times SN$, a is a smooth function near (x_0, v_0) in SN, and $a \geq 0$ in SM.

Proof This follows directly by applying Lemma 3.2.10 to $h(t, x, v) = \rho(\gamma_{x,v}(t))$ near $(0, x_0, v_0)$, where ρ is a boundary defining function for M as in Lemma 3.1.10. $\square$

Lemma 3.2.10 *Let $h(t, y)$ be smooth near $(0, y_0)$ in $\mathbb{R} \times \mathbb{R}^N$. If*

$$h(0, y_0) = 0, \qquad \partial_t h(0, y_0) = 0, \qquad \partial_t^2 h(0, y_0) < 0,$$

then one has

$$h(t, y) = 0 \text{ near } (0, y_0) \text{ when } h(0, y) \geq 0 \quad \Longleftrightarrow \quad t = Q(\pm\sqrt{a(y)}, y),$$

where Q is a smooth function near $(0, y_0)$ in $\mathbb{R} \times \mathbb{R}^N$, a is a smooth function near y_0 in $\mathbb{R}^N$, and $a(y) \geq 0$ when $h(0, y) \geq 0$. Moreover, $Q(\sqrt{a(y)}, y) \geq Q(-\sqrt{a(y)}, y)$ when $h(0, y) \geq 0$.

Proof We use the same argument as in Hörmander (1983–1985, Theorem C.4.2). Using that $\partial_t^2 h(0, y_0) < 0$, the implicit function theorem gives that

$$\partial_t h(t, y) = 0 \text{ near } (0, y_0) \quad \Longleftrightarrow \quad t = g(y),$$

where g is smooth near y_0 and $g(y_0) = 0$. Write

$$h_1(s, y) := h(s + g(y), y).$$

Then $\partial_s h_1(0, y) = 0$ and $\partial_s^2 h_1(0, y_0) < 0$. Thus by the Taylor formula we have

$$h_1(s, y) = h_1(0, y) - s^2 F(s, y),$$

where F is smooth near $(0, y_0)$ and $F(0, y_0) > 0$. We define

$$r(s, y) := sF(s, y)^{1/2},$$

and note that $r(0, y_0) = 0$, $\partial_s r(0, y_0) > 0$. Thus the map $(s, y) \mapsto (r(s, y), y)$ is a local diffeomorphism near $(0, y_0)$, and there is a smooth function S near $(0, y_0)$ so that

$$r(s, y) = \bar{r} \quad \Longleftrightarrow \quad s = S(\bar{r}, y).$$

Moreover, $\partial_r S(0, y_0) > 0$. Define the function

$$h_2(r, y) := h_1(0, y) - r^2.$$

Now

$$\begin{aligned} h(t, y) &= h_1(t - g(y), y) = h_1(0, y) - (t - g(y))^2 F(t - g(y), y) \\ &= h_2(r(t - g(y), y), y). \end{aligned}$$

Thus $h(t, y) = 0$ is equivalent with

$$r(t - g(y), y)^2 = h_1(0, y) = h(g(y), y). \tag{3.3}$$

We claim that

$$h(g(y), y) \geq 0 \text{ near } y_0 \text{ when } h(0, y) \geq 0. \tag{3.4}$$

If (3.4) holds, then we may solve (3.3) to obtain

$$\begin{aligned} &h(t, y) = 0 \text{ near } (0, y_0) \text{ when } h(0, y) \geq 0 \\ &\qquad\qquad \Longleftrightarrow \ r(t - g(y), y) = \pm\sqrt{h(g(y), y)}. \end{aligned}$$

The last condition is equivalent with

$$t - g(y) = S\left(\pm\sqrt{h(g(y), y)}, y\right).$$

This proves the lemma upon taking $Q(r, y) = g(y) + S(r, y)$ and $a(y) = h(g(y), y)$ (note that $r \mapsto Q(r, y)$ is increasing since $\partial_r S(0, y_0) > 0$). To prove (3.4), we use the Taylor formula

$$h(g(y) + s, y) = h(g(y), y) + \partial_t h(g(y), y)s + G(s, y)s^2$$

where $G(0, y_0) < 0$. Choosing $s = -g(y)$ and using that $\partial_t h(g(y), y) = 0$ shows that $h(g(y), y) \geq h(0, y)$ near $y = y_0$, and thus (3.4) indeed holds. □

Lemma 3.2.11 *Let (M, g) be a compact non-trapping manifold with strictly convex boundary. Then the functions*

$$\tilde{\tau}(x, v) := \tau(x, v) - \tau(x, -v) \quad \text{and} \quad T(x, v) := \tau(x, v)\tau(x, -v)$$

are smooth in SM.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Proof Given the properties of τ in Lemma 3.2.3 we just have to prove smoothness near a glancing point $(x_0, v_0) \in \partial_0 SM$. By Lemma 3.2.9 given $(x, v) \in SM$ near $(x_0, v_0) \in \partial_0 SM$, we have

$$\tilde{\tau}(x, v) = Q\left(\sqrt{a(x, v)}, x, v\right) + Q\left(-\sqrt{a(x, v)}, x, v\right).$$

Since we can write $Q(r, x, v) + Q(-r, x, v) = H(r^2, x, v)$, where H is smooth near $(0, x_0, v_0)$ (see Exercise 3.2.12), we deduce that

$$\tilde{\tau}(x, v) = H(a(x, v), x, v),$$

thus showing smoothness of $\tilde{\tau}$. The statement for T follows by taking products, rather than sums. $\square$

Exercise 3.2.12 If $f \in C^\infty(\mathbb{R})$ satisfies $f(t) = f(-t)$ for all $t \in \mathbb{R}$, show that there is $h \in C^\infty(\mathbb{R})$ with $f(t) = h(t^2)$ for all $t \in \mathbb{R}$.

Remark 3.2.13 Using Lemma 3.2.11, it is possible to write the functions Q and a from Lemma 3.2.9 in terms of $\tilde{\tau}$ and T. Indeed, since τ satisfies the quadratic equation

$$\tau(\tau - \tilde{\tau}) = T,$$

we have

$$\tau = \frac{\tilde{\tau} + \sqrt{\tilde{\tau}^2 + 4T}}{2},$$

with $\tilde{\tau}, T \in C^\infty(SM)$. Thus $Q(t, x, v) = (\tilde{\tau}(x, v) + t)/2$ and $a = \tilde{\tau}^2 + 4T$.

3.3 The Geodesic Flow and the Scattering Relation

Let (M, g) be a compact, connected, and oriented Riemannian manifold with boundary ∂M and dimension $n \geq 2$. By Lemma 3.1.8 we may assume that (M, g) is isometrically embedded into a closed manifold (N, g) of the same dimension.

The geodesics of (N, g) are defined for all times in $\mathbb{R}$. We pack them into what is called the *geodesic flow*. For each $t \in \mathbb{R}$ this is a diffeomorphism

$$\varphi_t \colon SN \to SN,$$

defined by

$$\varphi_t(x, v) := (\gamma_{x,v}(t), \dot{\gamma}_{x,v}(t)).$$

This is a *flow*, i.e. $\varphi_{t+s} = \varphi_t \circ \varphi_s$ for all $s, t \in \mathbb{R}$. The flow has an infinitesimal generator called the *geodesic vector field* and denoted by X. This is a smooth

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

section of TSN that can be regarded as the first-order differential operator $X\colon C^\infty(SN) \to C^\infty(SN)$ given by

$$(Xu)(x,v) := \frac{d}{dt}(u(\varphi_t(x,v)))\bigg|_{t=0}, \tag{3.5}$$

where $u \in C^\infty(SN)$. Observe that $X\colon C^\infty(SM) \to C^\infty(SM)$. The non-trapping property can be characterized using the operator X as follows:

Proposition 3.3.1 *Let (M,g) be a compact manifold with strictly convex boundary. The following are equivalent:*

(i) *(M,g) is non-trapping;*
(ii) *$X\colon C^\infty(SM) \to C^\infty(SM)$ is surjective;*
(iii) *there is $f \in C^\infty(SM)$ such that $Xf > 0$.*

Proof If (i) holds, let $f = -\tilde{\tau}$ where $\tilde{\tau}$ is smooth by Lemma 3.2.11. By Exercise 3.3.3 $Xf > 0$, thus (i) $\Longrightarrow$ (iii). Clearly (iii) $\Longrightarrow$ (i): if there is a geodesic in M with infinite length, since $Xf \geq c > 0$, integrating along it we would find $f(\varphi_t(x,v)) - f(x,v) \geq ct$ for all $t > 0$, which is absurd since f is bounded. The implication (ii) $\Longrightarrow$ (iii) is obvious, so it remains to prove that (i) $\Longrightarrow$ (ii).

Given $h \in C^\infty(SM)$, we need to find $u \in C^\infty(SM)$ with $Xu = h$. Consider (M,g) embedded in a closed manifold (N,g). Since strict convexity and $Xf > 0$ are open conditions, there is a slightly larger compact manifold M_1 with $M \subset M_1^{\text{int}} \subset N$ and such that ∂M_1 is strictly convex and (M_1,g) is non-trapping. Let τ_1 denote the exit time of M_1 and given $h \in C^\infty(SM)$, extend it smoothly to SM_1. For $(x,v) \in SM$, set

$$u(x,v) := -\int_0^{\tau_1(x,v)} h(\varphi_t(x,v))\,dt.$$

Since $\tau_1|_{SM}$ is smooth, $u \in C^\infty(SM)$. A calculation shows that $Xu = h$ and thus $X\colon C^\infty(SM) \to C^\infty(SM)$ is surjective. □

Remark 3.3.2 The assumption of ∂M being strictly convex is not necessary. See Duistermaat and Hörmander (1972, Theorem 6.4.1) for a proof of the same result for arbitrary vector fields.

Exercise 3.3.3 Let (M,g) be a compact non-trapping manifold with strictly convex boundary. Show that

$$X\tilde{\tau} = -2,$$

where $\tilde{\tau}$ is the function from Lemma 3.2.11.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Definition 3.3.4 Let (M, g) be a non-trapping manifold with strictly convex boundary. We define the *scattering relation* as the map $\alpha\colon \partial SM \to \partial SM$ given by

$$\alpha(x, v) := \varphi_{\tilde{\tau}(x,v)}(x, v).$$

Lemma 3.3.5 *Let (M, g) be a compact non-trapping manifold with strictly convex boundary. Then α is a diffeomorphism $\partial SM \to \partial SM$ whose fixed point set is $\partial_0 SM$. One has*

$$\alpha(\partial_\pm SM) = \partial_\mp SM,$$
$$\alpha \circ \alpha = \mathrm{Id}.$$

Proof By Lemma 3.2.6, the map α is smooth on ∂SM. By definition of $\tilde{\tau}$ we see that $\alpha\colon \partial_+ SM \to \partial_- SM$ and $\alpha\colon \partial_- SM \to \partial_+ SM$. One can check that $\tilde{\tau} \circ \alpha = -\tilde{\tau}$, which shows that $\alpha \circ \alpha = \mathrm{Id}$ and that α is a diffeomorphism whose fixed point set is $\partial_0 SM$. □

Exercise 3.3.6 Check that $\tilde{\tau} \circ \alpha = -\tilde{\tau}$.

3.4 Complex Structure

In this section we discuss the fact that on an oriented two-dimensional manifold M, a Riemannian metric g induces a complex structure and thus (M, g) becomes a *Riemann surface*. In fact, there is a one-to-one correspondence between conformal classes of Riemannian metrics and complex structures on M. In this way we can talk about holomorphic functions and harmonic conjugates in (M, g). We also discuss the important notion of isothermal coordinates (both local and global) on two-dimensional manifolds.

3.4.1 Generalities

We begin with some generalities.

Definition 3.4.1 (Complex manifold) An *N-dimensional complex manifold* is a $2N$-dimensional smooth (real) manifold with an open cover U_α and charts $\varphi_\alpha\colon U_\alpha \to \mathbb{C}^N$ such that $\varphi_\beta \circ \varphi_\alpha^{-1}$ is holomorphic $\varphi_\alpha(U_\alpha \cap U_\beta) \to \mathbb{C}^N$. The charts φ_α are called *complex* or *holomorphic coordinates*. The atlas $\{(U_\alpha, \varphi_\alpha)\}_\alpha$ is called a *complex atlas*. Two complex atlases are called *equivalent* if their union is a complex atlas. A *complex structure* is an equivalence class of complex atlases.

Definition 3.4.2 (Surface) A one-dimensional complex manifold is called a *surface* (or *Riemann surface*).

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

By Theorem 3.4.9, we will also use the term *surface* for any oriented two-dimensional (real) Riemannian manifold (M, g).

Definition 3.4.3 (Almost complex structure) If M is a differentiable manifold, an *almost complex structure* on M is a $(1,1)$ tensor field J such that the restriction $J_p \colon T_pM \to T_pM$ satisfies $J_p^2 = -\mathrm{Id}$ for any p in M. If g is a Riemannian metric on M, we say that J is *compatible* with g if $g(Jv, Jw) = g(v, w)$ for all $v, w \in T_pM$.

If M is a complex manifold, let $z = (z_1, \ldots, z_N)$ be a holomorphic chart $U_\alpha \to \mathbb{C}^N$, and write $z_j = x_j + iy_j$ with x_j and y_j real. There is a canonical almost complex structure J on M, defined for holomorphic charts by

$$J\left(\frac{\partial}{\partial x_j}\right) = \frac{\partial}{\partial y_j}, \qquad J\left(\frac{\partial}{\partial y_j}\right) = -\frac{\partial}{\partial x_j}.$$

Conversely, if M is a differentiable manifold equipped with an almost complex structure J (so it is necessarily even dimensional and orientable), then by the Newlander–Nirenberg theorem M has the structure of a complex manifold, having J as its canonical almost complex structure, if J satisfies an additional integrability condition.

Definition 3.4.4 (Holomorphic functions) If M is a complex manifold with complex charts $\varphi_\alpha \colon U_\alpha \to \mathbb{C}^N$, a C^1 function $f \colon M \to \mathbb{C}$ is called *holomorphic* (respectively *antiholomorphic*) if $f \circ \varphi_\alpha^{-1}$ is holomorphic (respectively antiholomorphic) from $\varphi_\alpha(U_\alpha) \subset \mathbb{C}^N$ to $\mathbb{C}$ for any α.

It is clear that all local properties of holomorphic functions in domains of $\mathbb{C}^N$ are valid also for holomorphic functions on complex manifolds.

3.4.2 Complex Structures in Two Dimensions

Let now (M, g) be a two-dimensional oriented (real) manifold with Riemannian metric g. In this case everything becomes very simple. In particular, the almost complex structures correspond to rotation by 90°.

Definition 3.4.5 (Rotation by 90°) For any $v \in T_xM$, let $v^\perp \in T_xM$ be the unique vector (the rotation of v by 90° counterclockwise) such that

$$|v^\perp|_g = |v|_g, \qquad \langle v, v^\perp \rangle = 0,$$

and $(v, v^\perp)$ is a positively oriented basis of T_xM when $v \neq 0$.

Exercise 3.4.6 Show that in local coordinates, if $g(x) = (g_{jk}(x))$, the vector $v^\perp$ is given by $v^\perp = g(x)^{-1/2}(-(g(x)^{1/2}v)_2, (g(x)^{1/2}v)_1)$, where $A^{1/2}$ is the square root of a positive definite matrix A.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Lemma 3.4.7 (Almost complex structures) *If (M, g) is an oriented two-dimensional manifold, then J is an almost complex structure compatible with g if and only if*

$$J(v) = \pm v^{\perp}, \qquad v \in TM.$$

Proof Let J be an almost complex structure compatible with g. Given $p \in M$ and $v \in T_pM$, the fact that J is compatible with g implies that $|Jv| = |v|$. Moreover, one has

$$\langle Jv, v\rangle = -\langle Jv, J^2 v\rangle = -\langle v, Jv\rangle,$$

which implies that $\langle Jv, v\rangle = 0$. Thus Jv is orthogonal to v and has the same length as v. Since T_pM is two dimensional, one must have $Jv = \pm v^{\perp}$. Conversely, $Jv = \pm v^{\perp}$ clearly satisfies $J^2 = -\mathrm{Id}$ and $\langle Jv, Jw\rangle = \langle v, w\rangle$. $\square$

We wish to find a complex structure on M associated with $J(v) = v^{\perp}$. The following fundamental result, proved by Gauss in 1822 in the real-analytic case, will yield complex coordinates that are compatible with J. We will prove later in Theorem 3.4.16 that if M is simply connected, then there exist *global* isothermal coordinates.

Theorem 3.4.8 (Isothermal coordinates) *Let (M, g) be an oriented two-dimensional manifold. Near any point of M there are positively oriented local coordinates $x = (x_1, x_2)$, called* isothermal coordinates*, so that the metric has the form*

$$g_{jk}(x) = e^{2\lambda(x)}\delta_{jk},$$

where λ is a smooth real-valued function.

Given the existence of isothermal coordinates, it is easy to show that any 2D Riemannian manifold has a complex structure. The proof uses the basic complex analysis fact that a smooth bijective map φ between open subsets of $\mathbb{R}^2$ is holomorphic if and only if it is conformal and orientation preserving. Recall that φ being conformal means that

$$\varphi^* h = ch$$

for some smooth positive function c where h is the Euclidean metric on $\mathbb{R}^2$.

Theorem 3.4.9 (Complex structure induced by g) *Let (M, g) be an oriented 2D manifold, and let (U_α) be an open cover of M so that there are isothermal coordinate charts $\varphi_\alpha \colon U_\alpha \to \mathbb{R}^2$. Then $\varphi_\beta^{-1} \circ \varphi_\alpha$ is holomorphic $\varphi_\alpha(U_\alpha \cap U_\beta) \to \mathbb{R}^2$ whenever $U_\alpha \cap U_\beta \neq \emptyset$. Thus the charts $(U_\alpha, \varphi_\alpha)$ induce a complex structure on M corresponding to $J(v) = v^{\perp}$. This complex structure* *is independent of the choice of the isothermal coordinate charts, and hence it is uniquely determined by g.*

Proof The fact that $g_{jk}(x) = e^{2\lambda(x)}\delta_{jk}$ in isothermal coordinates can be rewritten as

$$\left(\varphi_\alpha^{-1}\right)^* g = e^{2\lambda_\alpha} h,$$

where h is the Euclidean metric in $\mathbb{R}^2$. Suppose that $U_\alpha \cap U_\beta \neq \emptyset$ and let $\Phi = \varphi_\beta \circ \varphi_\alpha^{-1}$. Then Φ is a smooth map from an open set of $\mathbb{R}^2$ to $\mathbb{R}^2$, and one has

$$\Phi^* h = \left(\varphi_\alpha^{-1}\right)^* \varphi_\beta^* h = \left(\varphi_\alpha^{-1}\right)^* \left(e^{-2\varphi_\beta^*\lambda_\beta} g\right) = e^{2(\lambda_\alpha - \Phi^*\lambda_\beta)} h.$$

Since h is the Euclidean metric, the identity $\Phi^* h = ch$, where $c = e^{2(\lambda_\alpha - \Phi^*\lambda_\beta)}$ is a positive smooth function, means that Φ is a conformal bijective map between open sets in $\mathbb{R}^2$. Since isothermal coordinate charts are positively oriented, Φ is orientation preserving. Thus Φ must be holomorphic. This proves that any atlas consisting of isothermal coordinate charts is a complex atlas. It is also clear from this argument that if one uses different isothermal coordinate charts, then one obtains an equivalent atlas.

It remains to show that the almost complex structure J given by isothermal coordinates satisfies $J(v) = v^\perp$. But in isothermal coordinates $J(\partial_{x_1}) = \partial_{x_2} = (\partial_{x_1})^\perp$ and $J(\partial_{x_2}) = -\partial_{x_1} = (\partial_{x_2})^\perp$, so one must have $J(v) = v^\perp$. □

If (M, g) is a two-dimensional oriented Riemannian manifold, we will always use the complex structure induced by g on M. In fact the complex structure only depends on the conformal class

$$[g] = \{cg\,;\, c \in C^\infty(M) \text{ positive}\},$$

and conversely any complex structure on M arises from some conformal class.

Theorem 3.4.10 (Complex structures vs conformal classes) *Let M be an oriented two-dimensional manifold. There is a one-to-one correspondence between conformal classes of Riemannian metrics on M and complex structures on M.*

Proof Isothermal coordinates for a metric g are also isothermal for cg: if $(\varphi^{-1})^* g = e^{2\lambda} h$ with h the Euclidean metric, then $(\varphi^{-1})^*(cg) = e^{2\mu} h$ for $\mu = \lambda + \frac{1}{2}\log((\varphi^{-1})^* c)$. Thus the complex structure on M obtained in Theorem 3.4.9 is the same for g and cg.

Conversely, suppose that M is equipped with a complex structure. We wish to produce a metric g that induces this structure. Such a metric can be defined locally: if $p \in M$ and if (U, φ) is a complex coordinate chart near p, we can

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

define $g = \varphi^* h$ in U where h is the Euclidean metric in $\varphi(U) \subset \mathbb{R}^2$. More generally, if M is covered by complex coordinate charts $(U_\alpha, \varphi_\alpha)$ and if (χ_α) is a locally finite partition of unity subordinate to the cover (U_α), we can define

$$g = \sum \chi_\alpha \varphi_\alpha^* h.$$

Then g is a Riemannian metric on M. The complex coordinate charts $(U_\alpha, \varphi_\alpha)$ above are isothermal for g, since

$$(\varphi_\alpha^{-1})^* g = \sum_\beta \left((\varphi_\alpha^{-1})^* \chi_\beta\right)\left(\varphi_\beta \circ \varphi_\alpha^{-1}\right)^* h = \sum_\beta \left((\varphi_\alpha^{-1})^* \chi_\beta\right) c_{\alpha\beta} h = ch$$

for some positive smooth functions $c_{\alpha\beta}$ and c. Here we used that $\varphi_\beta \circ \varphi_\alpha^{-1}$ is holomorphic, hence conformal, and thus satisfies $(\varphi_\beta \circ \varphi_\alpha^{-1})^* h = c_{\alpha\beta} h$. This shows that the complex structure on M induced by g is the same as the original one. □

It remains to prove Theorem 3.4.8. It is convenient to consider rotations on T^*M instead of TM.

Definition 3.4.11 (Hodge star) For any $\xi \in T_x^*M$, let $\star\xi \in T_x^*M$ be the rotation of ξ by 90° counterclockwise, i.e.

$$\star\xi := \left((\xi^\sharp)^\perp\right)^\flat,$$

where $\sharp, \flat$ are the musical isomorphisms associated with g.

Clearly $\star\xi$ is the unique covector so that $|\star\xi|_g = |\xi|_g$, $\langle \xi, \star\, \xi \rangle = 0$, and $(\xi, \star\, \xi)$ is a positively oriented basis of T_x^*M when $\xi \neq 0$. The operator $\star$ is just the Hodge star operator specialized to 1-forms on a two-dimensional manifold. We can identify the almost complex structure $J(v) = v^\perp$ with the operator $\star$.

Proof of Theorem 3.4.8 Let $p \in M$. We wish to show that there are smooth functions u and v near p so that

$$|du|_g = |dv|_g > 0, \qquad \langle du, dv \rangle = 0 \qquad \text{near } p. \tag{3.6}$$

Since du and dv are linearly independent at p, the inverse function theorem shows that choosing $x_1 = u$, $x_2 = v$, and $\lambda = -\log |du|_g$ yields the required coordinate system near p.

The equations (3.6) state that du and dv should be orthogonal and have the same (positive) length. Since M is two dimensional, it follows that dv must be

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

the rotation of du by 90° (either clockwise or counterclockwise). Thus, given u with $du|_p \neq 0$, it would be enough to find v such that

$$dv = \star du, \tag{3.7}$$

where $\star$ is the Hodge star operator in Definition 3.4.11.

Now if the metric were Euclidean, the equations (3.7) would read

$$\partial_x u = \partial_y v, \qquad \partial_y u = -\partial_x v.$$

These are exactly the Cauchy–Riemann equations for an analytic function $f = u + iv$ in the complex plane. In particular, u and v would necessarily be harmonic. The same is true in the general case: by Exercise 3.4.14, on a two-dimensional oriented manifold one has

$$\Delta_g u = - \star d \star du.$$

Since $d^2 = 0$, it follows from (3.7) that u and v have to be harmonic.

We use Lemma 3.4.13 which shows that there is a harmonic function u near p with $du|_p \neq 0$. Then $\star du$ is a closed 1-form (since $d(\star du) = \star \Delta_g u = 0$), and the Poincaré lemma shows that in any small ball near p one can find a smooth function v satisfying (3.7). Since $du|_p \neq 0$, one has (3.6) in some neighbourhood of p which proves the theorem. □

We formulate part of the above proof as a lemma:

Lemma 3.4.12 (Harmonic conjugate) *Let (M, g) be a simply connected oriented 2-manifold. Given any $u \in C^\infty(M)$ satisfying $\Delta_g u = 0$ in M, there is $v \in C^\infty(M)$ satisfying*

$$dv = \star du \text{ in } M.$$

The function v, called a harmonic conjugate *of u, is harmonic and unique up to an additive constant. The function $f = u + iv$ is holomorphic in the complex structure induced by g. Conversely, the real and imaginary parts of any holomorphic function are harmonic.*

Lemma 3.4.13 *Let (M, g) be a Riemannian n-manifold and let $p \in M$. There is a harmonic function u near p with $du|_p \neq 0$.*

Proof We will work in normal coordinates at p. Writing out the local coordinate formula for Δ_g, it follows that

$$\Delta_g u = \Delta_e u + Qu, \qquad Qu = a^{jk}\partial_{jk} u + b^k \partial_k u,$$

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

where Δ_e is the Euclidean Laplacian and a^{jk}, b^k are smooth functions near 0. Since in normal coordinates one has $g_{jk}(0) = \delta_{jk}$ and $\partial_j g_{kl}(0) = 0$, it follows that

$$a^{jk}(0) = b^k(0) = 0.$$

We will look for u in the ball $B_r = B_r(0)$, where $r > 0$ is small, in the form

$$u(x) := x_1 + w(x).$$

The idea is that if r is small, then $\Delta_g x_1 \approx 0$ in B_r (since Δ_g is close to Δ_e and $\Delta_e x_1 = 0$), so there should be a solution of $\Delta_g u = 0$ close to x_1. We choose w as the solution of

$$\Delta_g w = -\Delta_g x_1 \text{ in } B_r, \qquad w|_{\partial B_r} = 0.$$

Clearly $\Delta_g u = 0$ in B_r. In order to estimate w, note that w solves

$$\Delta_e w = -Qu \text{ in } B_r, \qquad w|_{\partial B_r} = 0.$$

Writing $w_r(x) = w(rx)$ etc., we can rescale the previous equation to the unit ball:

$$\Delta_e w_r = -r^2 (Qu)_r \text{ in } B_1, \qquad w_r|_{\partial B_1} = 0.$$

For any $m \geq 0$, we may use elliptic regularity for the Dirichlet problem to get that

$$\|w_r\|_{H^{m+2}(B_1)} \lesssim r^2 \|(Qu)_r\|_{H^m(B_1)}$$

with the implied constant independent of r. Now $a^{jk}(0) = b^k(0) = 0$ and $u = x_1 + w$, so a short computation gives that

$$r^2 \|(Qu)_r\|_{H^m(B_1)} \lesssim r^3 + r\|w_r\|_{H^{m+2}(B_1)}.$$

If r is small enough, combining the last two equations gives

$$\|w_r\|_{H^{m+2}(B_1)} \lesssim r^3.$$

Choosing $m+2 > n/2+1$, the Sobolev embedding gives $\|\nabla w_r\|_{L^\infty(B_1)} \lesssim r^3$, which yields

$$\|\nabla w\|_{L^\infty(B_r)} \lesssim r^2.$$

If we choose r small enough, it follows that $du|_0 = dx_1|_0 + dw|_0 \neq 0$. $\square$

Exercise 3.4.14 Prove the formula $\Delta_g u = -\star d \star du$ used in the proof of Theorem 3.4.8.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

3.4.3 Global Isothermal Coordinates

We will now prove the existence of global isothermal coordinates on simply connected surfaces. This is part of the uniformization theorem for Riemann surfaces, and reduces to the following result. (Recall that $\mathbb{D}$ denotes the unit disk in $\mathbb{R}^2$.)

Theorem 3.4.15 (Riemann mapping theorem for surfaces) *Let (M, g) be a compact oriented simply connected 2-manifold with smooth boundary. There is a bijective holomorphic map*

$$\Phi\colon M^{\text{int}} \to \mathbb{D},$$

which extends smoothly as a diffeomorphism $M \to \overline{\mathbb{D}}$.

The result can be reformulated as follows:

Theorem 3.4.16 (Global isothermal coordinates) *If (M, g) is a compact oriented simply connected 2-manifold with smooth boundary, then there are global coordinates (x_1, x_2) in M so that in these coordinates*

$$g_{jk}(x) = e^{2\lambda(x)}\delta_{jk},$$

where λ is a smooth real-valued function.

Remark 3.4.17 By Proposition 3.7.22 any compact non-trapping manifold with strictly convex boundary is contractible. In particular, such manifolds are simply connected. Thus by Theorem 3.4.16 any compact non-trapping 2-manifold with strictly convex boundary is diffeomorphic to the unit disk and admits global isothermal coordinates.

There are several proofs of this theorem. Our proof, following Farkas and Kra (1992), will involve the Green function for the Laplacian in M and the fact that simply connected surfaces satisfy the monodromy theorem. To state this result, let Σ be a Riemann surface without boundary. If $\gamma\colon [0,1] \to \Sigma$ is a continuous curve and f_0 is analytic in a connected neighbourhood D_0 of $\gamma(0)$, we say that f_0 admits an analytic continuation along γ if for each $t \in [0,1]$ there is $\delta_t > 0$ and an analytic function f_t in a connected neighbourhood D_t of $\gamma(t)$, so that

$$f_s = f_t \text{ in } D_s \cap D_t \text{ whenever } s \in [0,1] \text{ and } |s - t| < \delta_t.$$

Theorem 3.4.18 (Monodromy theorem) *Let Σ be a simply connected Riemann surface without boundary. If f_0 is analytic near some $p \in \Sigma$ and admits an analytic continuation along any curve starting at p, then there is an analytic function f in Σ with $f = f_0$ near p.*

We first construct a candidate for the map Φ.

Lemma 3.4.19 *For any $p \in M^{\mathrm{int}}$, there is a holomorphic map*

$$\Phi\colon M^{\mathrm{int}} \to \mathbb{D},$$

which extends smoothly as a smooth map $M \to \overline{\mathbb{D}}$, so that p is a simple zero of Φ and there are no other zeros of Φ in M.

Proof Let z be a complex coordinate chart in a neighbourhood U of p so that $z(p) = 0$ and $g_{jk} = e^{2\lambda(x)}\delta_{jk}$ in these coordinates. Then locally near p the function $\Phi = z$ has the property that p is a simple zero and there are no other zeros. In order to obtain a global function in M with this property, we formally look for Φ in the form $\Phi = e^f$ where f is holomorphic in $M \setminus \{p\}$, near p one has $f = \log z + h$ where h is harmonic, and $\mathrm{Re}(f)|_{\partial M} = 0$. This argument is only formal since $\mathrm{Im}(\log z)$ is multivalued. To rectify this we instead construct the real part $u = \mathrm{Re}(f)$, which should be harmonic in $M \setminus \{p\}$, look like $\log|z| +$ harmonic near p, and vanish on ∂M. This means that u is just (a constant multiple of) the Green function for Δ_g in M.

To construct u precisely, note that $\Delta_g(\log|z|) = e^{-2\lambda}\Delta_e(\log|z|) = 0$ in $U \setminus \{p\}$, where Δ_e is the Laplacian in $\mathbb{R}^2$. Fix a cut-off function $\beta \in C_c^\infty(U)$ with $0 \leq \beta \leq 1$ and $\beta = 1$ near p. We define

$$u := \beta \log|z| + u_1,$$

where u_1 is the solution of the Dirichlet problem

$$\Delta_g u_1 = F \text{ in } M, \qquad u_1|_{\partial M} = 0,$$

and where F is the extension of $-\Delta_g(\beta \log|z|) \in C^\infty(M \setminus \{p\})$ by zero to p. Noting that $F \in C^\infty(M)$, elliptic regularity ensures that u_1 is a real-valued function in $C^\infty(M)$. Then we have the following desired properties:

$$u \text{ is harmonic in } M \setminus \{p\}, \quad u = \log|z| + u_1 \text{ near } p, \quad u|_{\partial M} = 0.$$

We want to prove that there is a holomorphic Φ in M^{int} with $|\Phi| = e^u$. First we show that such a function exists near p. In fact, since $\Delta_g u_1 = 0$ near p, by Lemma 3.4.12 there is a harmonic conjugate v_1 of u_1 in some small ball centred at p. The function

$$\Psi = z e^{u_1 + i v_1}$$

is holomorphic and satisfies $|\Psi| = e^u$ near p.

The above argument already proves the result if M is contained in a complex coordinate patch. In the general case, we wish to continue Ψ analytically to

M^{int}. If $\gamma\colon [0,1] \to M^{\text{int}}$ is any continuous curve with $\gamma(0) = p$, define the set

$$I := \{s \in [0,1]\colon \Psi \text{ admits an analytic continuation along } \gamma|_{[0,s]} \\ \text{so that } |f_t| = e^u \text{ for } t \in [0,s]\}.$$

Clearly $0 \in I$ and I is open. To show that I is closed, let $t_0 \in [0,1]$ be such that $[0,t_0) \subset I$. There is an analytic function $\tilde{\Psi}$ near $\gamma(t_0)$ with $|\tilde{\Psi}| = e^u$: if $\gamma(t_0) = p$ one can take $\tilde{\Psi} = \Psi$, and if $\gamma(t_0) \neq p$ one can take $\tilde{\Psi} = e^{u+iv}$ in a small ball $\tilde{U}$ centred at $\gamma(t_0)$ where v is a harmonic conjugate in $\tilde{U}$ of the smooth harmonic function u. Choose $\varepsilon > 0$ so that $\gamma([t_0-\varepsilon, t_0]) \subset \tilde{U}$. Since $t_0 - \varepsilon \in I$, Ψ admits an analytic continuation along $\gamma|_{[0,t_0-\varepsilon]}$. We continue this for $t \in [t_0-\varepsilon, t_0]$ by choosing $D_t = \tilde{U}$ and $f_t = \tilde{\Psi}$. It remains to show that $f_{t_0-\varepsilon} = \tilde{\Psi}$ near $\gamma(t_0-\varepsilon)$. But $|f_{t_0-\varepsilon}| = |\tilde{\Psi}| = e^u$ near $\gamma(t_0-\varepsilon)$, which means that the holomorphic function $f_{t_0-\varepsilon}/\tilde{\Psi}$ has modulus 1 near $\gamma(t_0-\varepsilon)$ (this is true also if $\gamma(t_0-\varepsilon) = p$, since both the numerator and denominator vanish simply at p). Thus $f_{t_0-\varepsilon}/\tilde{\Psi}$ is a constant $e^{i\theta} \in S^1$ near $\gamma(t_0-\varepsilon)$ (it must have vanishing derivative by the open mapping theorem). Replacing $\tilde{\Psi}$ by $e^{i\theta}\tilde{\Psi}$ above shows that Ψ admits an analytic continuation along $\gamma|_{[0,t_0]}$ so that $|f_t| = e^u$. Thus I is closed, and connectedness implies that $I = [0,1]$.

We have proved that Ψ admits an analytic continuation along any curve in M^{int}. By the monodromy theorem, there is an analytic function Φ in M^{int} extending Ψ, and one has $|\Phi| = e^u$ in M^{int}. In particular, Φ has a simple zero at p and no other zeros in M^{int}. Near any boundary point one has $\Phi = e^{u+iv}$ where the local harmonic conjugate v of u can be continued smoothly to ∂M, showing that Φ extends smoothly to M. Since $|\Phi||_{\partial M} = e^u|_{\partial M} = 1$, the maximum principle implies that Φ maps M to $\overline{\mathbb{D}}$. □

Remark 3.4.20 We sketch an alternative to the analytic continuation argument in the proof above, following Hubbard (2006). After constructing the Green function u, one could proceed by constructing a multivalued harmonic conjugate v for u in $M \setminus \{p\}$. The harmonic conjugate should formally satisfy $dv = \star du$ in $M \setminus \{p\}$. To solve the last equation, we fix $q \in M \setminus \{p\}$ and define

$$v(x) := \int_{\gamma_{q,x}} \star du, \qquad x \in M \setminus \{p\}, \tag{3.8}$$

where $\gamma_{q,x}$ is a smooth curve from q to x in $M \setminus \{p\}$. (Note that $M \setminus \{p\}$ is connected since M is.) Of course the value $v(x)$ depends on the choice of $\gamma_{q,x}$. If $\tilde{\gamma}_{q,x}$ is another such curve and if γ is the concatenation of $\gamma_{q,x}$ and the reverse of $\tilde{\gamma}_{q,x}$, then γ is a closed curve in $M \setminus \{p\}$.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

We now invoke the following topological fact: since M is simply connected and two dimensional, any closed curve γ in $M \setminus \{p\}$ is homologous to a small circle centred at p winding k times around p for some $k \in \mathbb{Z}$. Since $\star du$ is closed in $M \setminus \{p\}$ and $u = \log|z| +$ harmonic near p, an easy computation gives that

$$\int_\gamma \star du \in 2\pi\mathbb{Z}.$$

This shows that (3.8) defines $v(x)$ modulo $2\pi\mathbb{Z}$. It follows that e^{iv} is a well-defined smooth function in $M \setminus \{p\}$, and $\Phi = e^{u+iv}$ is holomorphic in $M \setminus \{p\}$. It is also bounded near p, and hence extends to the desired holomorphic function Φ near p.

Proof of Theorem 3.4.16 We shall show that the map from Lemma 3.4.19 gives the desired map Φ. First observe that by construction we have $\Phi(\partial M) \subset \partial\mathbb{D}$ and let γ denote a parametrization of ∂M. An application of the argument principle shows that $\Phi\colon M^{\mathrm{int}} \to \mathbb{D}$ is a bijection: indeed since Φ has a unique simple zero at p, the index of the curve $\Phi \circ \gamma$ around zero is one and thus there is a unique solution to $\Phi(z) = w$ for any $w \in \mathbb{D}$. A standard complex analysis argument gives that $\Phi\colon M^{\mathrm{int}} \to \mathbb{D}$ is a biholomorphism. It remains to show that the smooth extension $\Phi\colon M \to \overline{\mathbb{D}}$ is a diffeomorphism. We already know that the Jacobian determinant of Φ is non-zero for any $z \in M^{\mathrm{int}}$ and we claim that it is also non-zero for $z \in \partial M$. Since Φ is smooth on M, it satisfies the Cauchy–Riemann equations on M and thus it suffices to show that some directional derivative of Φ at $z \in \partial M$ is non-zero. But this is clearly the case since the harmonic function $\log|\Phi|$ attains its global maximum at every point of ∂M. It follows that the map $\Phi|_{\partial M}\colon \partial M \to \partial\mathbb{D}$ is a diffeomorphism since it has degree one. This gives that $\Phi\colon M \to \overline{\mathbb{D}}$ is a bijection with smooth inverse. □

3.5 The Unit Circle Bundle of a Surface

We consider now the unit sphere bundle SM when $\dim M = 2$. Many of the results in this section have natural counterparts in higher dimensions as discussed in Section 3.6, but when $\dim M = 2$ there is a special structure that simplifies many arguments.

3.5.1 The Vector Fields X, $X_\perp$, and V

When $\dim M = 2$ the manifold SM is three dimensional, and there is a very convenient frame of three vector fields on SM that will be used throughout this book. We will first consider this frame in the case of the Euclidean metric.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Example 3.5.1 (Frame of TSM in the Euclidean disk) Let $M = \overline{\mathbb{D}} \subset \mathbb{R}^2$ and let $g = e$ be the Euclidean metric. Then

$$SM = \{(x, v_\theta) \colon x \in M,\ \theta \in [0, 2\pi)\} = M \times S^1,$$

where $v_\theta = (\cos\theta, \sin\theta)$. We identify (x, v_θ) with (x, θ). The geodesic vector field acting on functions $u = u(x, \theta)$ on SM has the form

$$Xu(x,\theta) = \frac{d}{dt} u(x + t v_\theta, \theta)\Big|_{t=0} = v_\theta \cdot \nabla_x u(x,\theta).$$

Write $(v_\theta)_\perp = (\sin\theta, -\cos\theta)$ for the rotation of v_θ by 90° *clockwise*, and define another vector field

$$X_\perp u(x,\theta) = (v_\theta)_\perp \cdot \nabla_x u(x,\theta).$$

The vector fields X and $X_\perp$ encode all possible x-derivatives of a function on SM. We define a third vector field V by

$$Vu(x,\theta) = \partial_\theta u(x,\theta).$$

Now the vectors $\{X, X_\perp, V\}$ are linearly independent at each point of SM and thus give a frame on TSM. It is easy to compute the commutators of these vector fields:

$$[X, V] = X_\perp, \qquad [V, X_\perp] = X, \qquad [X, X_\perp] = 0.$$

Let now (M, g) be a two-dimensional oriented Riemannian manifold. We wish to define analogues of the vector fields $X_\perp$ and V in the example above.

Definition 3.5.2 (Rotation by 90° clockwise) For any $(x, v) \in SM$, we define

$$v_\perp := -v^\perp.$$

Definition 3.5.3 Define the vector field $X_\perp \colon C^\infty(SM) \to C^\infty(SM)$ by

$$X_\perp u(x,v) = \frac{d}{dt}(u(\psi_t(x,v)))\Big|_{t=0},$$

where $\psi_t(x,v) = (\gamma_{x,v_\perp}(t), W(t))$ and $W(t)$ is the parallel transport of v along the curve $\gamma_{x,v_\perp}(t)$.

Moreover, define the *vertical vector field* $V \colon C^\infty(SM) \to C^\infty(SM)$ by

$$Vu(x,v) = \frac{d}{dt} u(\rho_t(x,v))\Big|_{t=0},$$

where $\rho_t(x,v) = (x, e^{it} v)$ and $e^{it} v$ denotes the rotation of v by angle t counterclockwise in $(T_x M, g(x))$, i.e.

$$e^{it} v := (\cos t) v + (\sin t) v^\perp.$$

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Exercise 3.5.4 If the metric is Euclidean, show that $\psi_t(x, v) = (x + tv_\perp, v)$ and $e^{it} v_\theta = v_{\theta+t}$ and thus $X_\perp$ and V have the forms given in Example 3.5.1.

The next result gives all the commutators of the vector fields $X, X_\perp, V$. These are also called the *structure equations* (of the Lie algebra of smooth vector fields on SM).

Lemma 3.5.5 (Commutator formulas) *One has*

$$[X, V] = X_\perp,$$
$$[X_\perp, V] = -X,$$
$$[X, X_\perp] = -KV,$$

where K is the Gaussian curvature of (M, g).

One way to prove Lemma 3.5.5 is by local coordinate computations. For later purposes it will also be useful to have explicit forms of the three vector fields in local coordinates. Since M is two dimensional, it is particularly convenient to use the isothermal coordinates (x_1, x_2) introduced in Theorem 3.4.8. This induces special coordinates (x_1, x_2, θ) on SM, and the following local coordinate formulas are valid.

Lemma 3.5.6 (Special coordinates on SM) *Let (x_1, x_2, θ) be local coordinates on SM where (x_1, x_2) are isothermal coordinates on M and θ is the angle between a unit vector v and $\partial/\partial x_1$, i.e.*

$$v = e^{-\lambda}\left(\cos\theta \frac{\partial}{\partial x_1} + \sin\theta \frac{\partial}{\partial x_2}\right).$$

In these coordinates one has the formulas

$$X = e^{-\lambda}\left(\cos\theta \frac{\partial}{\partial x_1} + \sin\theta \frac{\partial}{\partial x_2} + \left(-\frac{\partial\lambda}{\partial x_1}\sin\theta + \frac{\partial\lambda}{\partial x_2}\cos\theta\right)\frac{\partial}{\partial\theta}\right),$$
$$X_\perp = -e^{-\lambda}\left(-\sin\theta \frac{\partial}{\partial x_1} + \cos\theta \frac{\partial}{\partial x_2} - \left(\frac{\partial\lambda}{\partial x_1}\cos\theta + \frac{\partial\lambda}{\partial x_2}\sin\theta\right)\frac{\partial}{\partial\theta}\right),$$
$$V = \frac{\partial}{\partial\theta}.$$

Remark 3.5.7 We will use the special coordinates (x_1, x_2, θ) on SM several times throughout this book. Note that (x_1, x_2, θ) are not isothermal coordinates on SM, since the Sasaki metric G introduced in Definition 3.5.10 is not even diagonal in these coordinates (one can check that $G(\partial_{x_1}, \partial_\theta) = -\partial_{x_2}\lambda$ and $G(\partial_{x_2}, \partial_\theta) = \partial_{x_1}\lambda$).

Exercise 3.5.8 Prove Lemma 3.5.6.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Exercise 3.5.9 Prove Lemma 3.5.5 by using Lemma 3.5.6 and the fact that the Gaussian curvature of a metric $g_{jk} = e^{2\lambda(x)}\delta_{jk}$ is $K = -\Delta_g \lambda = -e^{-2\lambda}(\partial_1^2\lambda + \partial_2^2\lambda)$.

3.5.2 Integration on SM

Above we introduced the fundamental vector fields $X, X_\perp, V$ on the unit sphere bundle of a two-dimensional manifold. These vector fields encode all possible derivatives of functions in SM. We will now discuss how to integrate functions on SM. We will consider the case $\dim M = 2$, but all the results in this subsection have natural counterparts in higher dimensions as discussed in Section 3.6.

Let (M, g) be a compact oriented Riemannian surface with smooth boundary. The manifold (M, g) has a volume form dV^2 induced by the Riemannian metric. In local coordinates,

$$dV^2 = |g(x)|^{1/2}\, dx_1 \wedge dx_2.$$

For any $x \in M$, the metric g induces a Riemannian metric (inner product) $g(x)$ on T_xM. The subset $S_xM = \{v \in T_xM : |v|_g = 1\}$ also becomes a Riemannian manifold. Denote by dS_x the volume form of $(S_xM, g(x))$. Defining a volume form requires a choice of orientation on S_xM, but we make the natural choice that S_xM is oriented according to the orientation of the surface.

Now the integral of a function $f \in C(SM)$ over SM is just

$$\int_M \int_{S_xM} f(x, v)\, dS_x(v)\, dV^2(x).$$

This integral induces a natural volume form (or measure) on SM called the *Liouville form*. We shall denote it by $d\Sigma^3$. At a point $(x, v) \in SM$ it can be written as

$$d\Sigma^3 = dV^2 \wedge dS_x.$$

In the special coordinates (x_1, x_2, θ) in Lemma 3.5.6, one has $dV^2 = e^{2\lambda(x)}\, dx_1 \wedge dx_2$ and $dS_x = d\theta$ (to see the latter, note that ∂_θ corresponds to $e^{-\lambda(x)}(-\sin\theta, \cos\theta)$ on TS_xM that has unit length). Thus

$$d\Sigma^3 = e^{2\lambda(x)}\, dx_1 \wedge dx_2 \wedge d\theta. \tag{3.9}$$

We will next show that $d\Sigma^3$ is actually the volume form of a canonical Riemannian metric on SM.

Definition 3.5.10 The *Sasaki metric* G on SM is the unique Riemannian metric on SM for which the vector fields $\{X, X_\perp, V\}$ are orthonormal at each point of SM.

Clearly, the Sasaki metric satisfies

$$G(aX + bX_\perp + cV, \tilde{a}X + \tilde{b}X_\perp + \tilde{c}V) = a\tilde{a} + b\tilde{b} + c\tilde{c}.$$

Defining the volume form dV_G of the Sasaki metric requires an orientation on SM. We already chose an orientation on S_xM, and then SM is oriented so that $(X, -X_\perp, V)$ is a positively oriented basis at each point of SM.

Lemma 3.5.11 $dV_G = d\Sigma^3$.

Proof The volume form dV_G is the unique 3-form on SM that satisfies $dV_G(X, -X_\perp, V) = 1$. On the other hand, a short computation using (3.9) and Lemma 3.5.6 shows that

$$d\Sigma^3(X, -X_\perp, V) = 1.$$

Thus it follows that $d\Sigma^3 = dV_G$. $\square$

Similarly as above, the integral of $h \in C(\partial SM)$ over ∂SM is

$$\int_{\partial M} \int_{S_x M} h(x,v)\, dS_x(v)\, dV^1(x),$$

where dV^1 is the volume form of $(\partial M, g)$. This integral induces a volume form on ∂SM given by

$$d\Sigma^2 := dV^1 \wedge dS_x.$$

The Sasaki metric on SM induces a metric G on ∂SM, and $d\Sigma^2$ coincides with the volume form of $(\partial SM, G)$. This follows as in Lemma 3.5.11 since $d\Sigma^2(w, \partial_\theta) = 1$ when w is a positively oriented unit vector in $T\partial M$.

The volume forms on SM and ∂SM induce L^2 inner products

$$(u, w)_{SM} = \int_{SM} u\bar{w}\, d\Sigma^3,$$

$$(h, r)_{\partial SM} = \int_{\partial SM} h\bar{r}\, d\Sigma^2.$$

We denote the corresponding L^2 spaces by $L^2(SM)$ and $L^2(\partial SM)$.

The next result establishes basic integration by parts formulas related to the vector fields X, $X_\perp$, and V. In particular, it shows that X, $X_\perp$, and V are formally skew-adjoint operators. Recall that ν is the *inward* unit normal of ∂M.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Proposition 3.5.12 (Integration by parts) *Let $u, w \in C^1(SM)$. Then*

$$\begin{aligned}(Xu, w)_{SM} &= -(u, Xw)_{SM} - (\langle v, \nu\rangle u, w)_{\partial SM},\\ (X_\perp u, w)_{SM} &= -(u, X_\perp w)_{SM} - (\langle v_\perp, \nu\rangle u, w)_{\partial SM},\\ (Vu, w)_{SM} &= -(u, Vw)_{SM}.\end{aligned}$$

Proof We only prove the first formula. Consider coordinates (x, θ) as in Lemma 3.5.6. Then

$$\begin{aligned}(Xu, w)_{SM} = \int_M \int_0^{2\pi} e^{\lambda}\Big(\cos\theta \frac{\partial u}{\partial x_1} + \sin\theta \frac{\partial u}{\partial x_2}\\ + \left(-\frac{\partial \lambda}{\partial x_1}\sin\theta + \frac{\partial \lambda}{\partial x_2}\cos\theta\right)\frac{\partial u}{\partial \theta}\Big)\bar{w}\, dx\, d\theta.\end{aligned}$$

Integrating by parts in x and θ, we see that the terms obtained when the x-derivatives hit e^{λ} and when the θ-derivative hits $\sin\theta$ and $\cos\theta$ add up to zero. The resulting expression is $-(u, Xw)_{SM} - (\langle v, \nu\rangle u, w)_{\partial SM}$ as required. □

Remark 3.5.13 Recall that if (N, g) is a compact manifold with boundary, if Y is a real vector field on N and $u, w \in C_c^\infty(N^{\mathrm{int}})$, one has

$$(Yu, w)_{L^2(N)} = -(u, Yw + \mathrm{div}_g(Y)w)_{L^2(N)},$$

where $\mathrm{div}_g(Y) = |g|^{-1/2}\partial_j(|g|^{1/2}Y^j)$ is the metric divergence. Moreover, the Lie derivative of the volume form dV_g satisfies

$$L_Y(dV_g) = \mathrm{div}_g(Y)\, dV_g.$$

Thus Proposition 3.5.12 implies that X, $X_\perp$, and V are divergence free with respect to the Sasaki metric, and they all preserve the volume form $d\Sigma^3$.

Next we state Santaló's formula, which is a fundamental change of variables formula on SM. The proof boils down to the fact that X is divergence free. Recall the notation $\mu(x, v) = \langle \nu(x), v\rangle$ for $(x, v) \in \partial SM$.

Proposition 3.5.14 (Santaló's formula) *Let (M, g) be a compact non-trapping surface with strictly convex boundary. Given $f \in C(SM)$ we have*

$$\int_{SM} f\, d\Sigma^3 = \int_{\partial_+ SM} \int_0^{\tau(x,v)} f(\varphi_t(x, v))\mu(x, v)\, dt\, d\Sigma^2.$$

Proof We give the proof for $f \in C_c^\infty(SM^{\text{int}})$ (the general case follows by approximation). For any $(x,v) \in SM$ define

$$u^f(x,v) := \int_0^{\tau(x,v)} f(\varphi_t(x,v))\,dt. \tag{3.10}$$

Since $\tau \in C(SM) \cap C^\infty(SM \setminus \partial_0 SM)$, clearly $u^f \in C(SM) \cap C^\infty(SM \setminus \partial_0 SM)$ and $u^f|_{\partial_- SM} = 0$. But if f has compact support in the interior of M, then u^f vanishes near tangential directions and thus u^f is in fact smooth. A simple computation shows that

$$Xu^f = -f. \tag{3.11}$$

We now apply Proposition 3.5.12 as follows:

$$\int_{SM} f\,d\Sigma^3 = -(Xu^f, 1)_{SM} = (\mu u^f, 1)_{\partial SM} = \int_{\partial SM} u^f(x,v)\mu(x,v)\,d\Sigma^2.$$

The result follows by inserting the formula (3.10) and using the fact that $u^f|_{\partial_- SM} = 0$. □

Exercise 3.5.15 Prove (3.11), and show that Santaló's formula holds for $f \in C(SM)$ (in fact for $f \in L^1(SM)$) using that it has been proved for $f \in C_c^\infty(SM^{\text{int}})$.

3.6 The Unit Sphere Bundle in Higher Dimensions

In this section we present some aspects of the geometry of the unit sphere bundle in arbitrary dimensions. We use this to describe how the strict convexity of ∂M reflects at level of the geodesic vector field and to give a proof of Santaló's formula in any dimension. We shall also use some of these preliminaries when discussing the various definitions of simple manifolds and in Section 5.2 to give an alternative proof for the main regularity result for transport equations.

Let (M,g) be a compact Riemannian manifold with unit sphere bundle $\pi : SM \to M$. For details of what follows, see, for example, Knieper (2002); Paternain (1999). It is well known that SM carries a canonical metric called the *Sasaki metric*. If we let $\mathcal{V}$ denote the vertical subbundle given by $\mathcal{V} = \ker d\pi$, then there is an orthogonal splitting with respect to the Sasaki metric:

$$TSM = \mathbb{R}X \oplus \mathcal{H} \oplus \mathcal{V}.$$

The subbundle $\mathcal{H}$ is called the horizontal subbundle. Elements in $\mathcal{H}(x,v)$ and $\mathcal{V}(x,v)$ are canonically identified with elements in the codimension one subspace $\{v\}^{\perp} \subset T_xM$. A vector in $\mathbb{R}X \oplus \mathcal{H}$ is canonically identified with the whole T_xM. In order to describe these identifications, we first introduce the *connection map* $\mathrm{K}\colon T_{(x,v)}SM \to T_xM$. Given $\xi \in T_{(x,v)}SM$, consider any curve $Z\colon (-\varepsilon,\varepsilon) \to SM$ such that $Z(0) = (x,v)$ and $\dot{Z}(0) = \xi$ and write $Z(t) = (\alpha(t), W(t))$. Then

$$\mathrm{K}\xi := D_t W|_{t=0},$$

where D stands for the covariant derivative of the vector field W along α given by the Levi-Civita connection. Using $d\pi$ and K, we set

$$\mathcal{V} := \ker d\pi, \quad \tilde{\mathcal{H}} := \ker \mathrm{K}.$$

It is straightforward to check that

$$d\pi|_{\tilde{\mathcal{H}}(x,v)}\colon \tilde{\mathcal{H}}(x,v) \to T_xM, \text{ and } \mathrm{K}|_{\mathcal{V}(x,v)}\colon \mathcal{V}(x,v) \to \{v\}^{\perp}$$

are linear isomorphisms and thus $\xi \in T_{(x,v)}SM$ may be written as

$$\xi = (\xi_H, \xi_V), \tag{3.12}$$

where $\xi_H = d\pi(\xi)$ and $\xi_V = \mathrm{K}\xi$. In this splitting, the geodesic vector field has a very simple form

$$X(x,v) = (v,0). \tag{3.13}$$

Using the splitting, one can also define the Sasaki metric G of SM as

$$\langle \xi, \eta \rangle_G := \langle \xi_H, \eta_H \rangle_g + \langle \xi_V, \eta_V \rangle_g. \tag{3.14}$$

Finally using the Sasaki metric, we decompose orthogonally $\tilde{\mathcal{H}} = \mathbb{R}X \oplus \mathcal{H}$ and we obtain the desired identifications of $\mathcal{H}(x,v)$ and $\mathcal{V}(x,v)$ with $\{v\}^{\perp}$. The canonical contact 1-form $\boldsymbol{\alpha}$ is uniquely defined by $\boldsymbol{\alpha}(X) = 1$ and $\ker \boldsymbol{\alpha} = \mathcal{H} \oplus \mathcal{V}$. Its differential $d\boldsymbol{\alpha}$ defines a symplectic form on $\mathcal{H} \oplus \mathcal{V}$, which can be shown to be

$$d\boldsymbol{\alpha}(\xi, \eta) = \langle \xi_V, \eta_H \rangle_g - \langle \xi_H, \eta_V \rangle_g. \tag{3.15}$$

The next lemma identifies the tangent spaces to ∂SM and $S\partial M = \partial_0 SM$ using this splitting.

Lemma 3.6.1

$$\begin{aligned} T_{(x,v)}\partial SM &= \{(\xi_H, \xi_V) : \ \xi_H \in T_x\partial M, \ \ \xi_V \in \{v\}^{\perp}\}; \\ T_{(x,v)}\partial_0 SM &= \{(\xi_H, \xi_V) : \ \xi_H \in T_x\partial M, \ \ \xi_V \in \{v\}^{\perp}, \\ &\qquad \langle \xi_V, \nu(x) \rangle = \Pi_x(v, \xi_H)\}. \end{aligned}$$

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Proof To prove the first statement consider a curve $Z\colon (-\varepsilon,\varepsilon) \to \partial SM$ with $Z(0) = (x,v)$ and $\xi = \dot{Z}(0)$. Then if we write $Z(t) = (\alpha(t), W(t))$ with $\alpha\colon (-\varepsilon,\varepsilon) \to \partial M$, we see that $\xi_H = d\pi(\xi) = \dot{\alpha}(0) \in T_x\partial M$. Differentiating $\langle W(t), W(t)\rangle = 1$ at $t = 0$ we get that $\langle \xi_V, v\rangle = 0$. The first statement follows by counting dimensions.

To prove the second statement we need to take a curve $Z\colon (-\varepsilon,\varepsilon) \to \partial_0 SM$ that gives the additional equation $\langle W(t), \nu(\alpha(t))\rangle = 0$. Differentiate this at $t = 0$, to get, using the definition of the connection map $\mathtt{K}$,

$$\langle \xi_V, \nu(x)\rangle + \langle v, \nabla_{\xi_H}\nu\rangle = 0.$$

This is equivalent to $\langle \xi_V, \nu(x)\rangle - \Pi_x(v, \xi_H) = 0$ and the result follows. $\square$

3.6.1 The Geodesic Vector Field and Strict Convexity

When does X fail to be transversal to ∂SM? Using Lemma 3.6.1 and (3.13) we see that this happens if and only if $(x,v) \in \partial_0 SM$. In addition, the characterization of $T_{(x,v)}\partial_0 SM$ tells us that X is always transversal to $\partial_0 SM$ under the assumption that the boundary ∂M is strictly convex.

We summarize this in the following lemma:

Lemma 3.6.2 *The geodesic vector field X is transversal to $\partial SM \setminus \partial_0 SM$. If ∂M is strictly convex, then X is transversal to $\partial_0 SM$. We always have $X(x,v) \in T_{(x,v)}\partial SM$ for $(x,v) \in \partial_0 SM$.*

The picture described by the lemma will be helpful later on when discussing regularity results for the transport equation and it may be visualized in Figure 3.2.

Exercise 3.6.3 Show that the horizontal vector $(\nu(x), 0)$ is a unit normal vector to ∂SM in the Sasaki metric. Moreover, show that the inner product of this vector with X is precisely the function μ introduced before Lemma 3.2.8.

3.6.2 Volume Forms and Santaló's Formula

Let (M,g) be a compact, connected, and oriented Riemannian manifold with smooth boundary, of dimension $n = \dim M \geq 2$. We wish to discuss integration of functions on SM and ∂SM. The manifold (M,g) has a volume form dV^n induced by the Riemannian metric. In local coordinates,

$$dV^n = |g(x)|^{1/2}\, dx_1 \wedge \cdots \wedge dx_n.$$

For any $x \in M$, the metric g induces a Riemannian metric (inner product) $g(x)$ on T_xM. The subset $S_xM = \{v \in T_xM \ :\ |v|_g = 1\}$ also becomes a Riemannian manifold. Denote by dS_x the volume form of $(S_xM, g(x))$.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

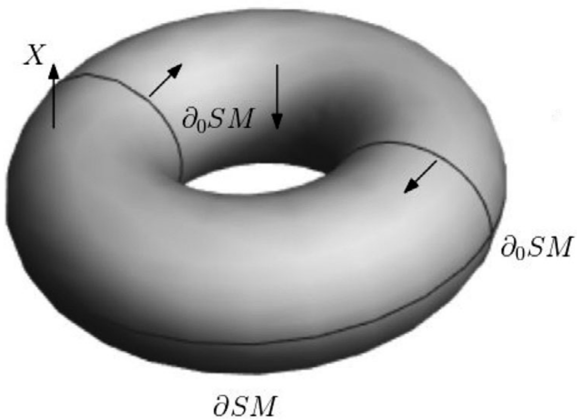

Figure 3.2 In the 2D case, ∂SM is a 2-torus (assuming M is a disk) and the glancing region $\partial_0 SM$ is given by two circles. The figure shows the geodesic vector field X being transversal to $\partial SM \setminus \partial_0 SM$ and at $\partial_0 SM$, X becomes tangent to ∂SM but remains transversal to $\partial_0 SM$ if ∂M is strictly convex.

Now the integral of a function $f \in C(SM)$ over SM is just

$$\int_M \int_{S_x M} f(x,v)\, dS_x(v)\, dV^n(x).$$

This integral induces a natural volume form (or measure) on SM called the *Liouville form*. We shall denote it by $d\Sigma^{2n-1}$. At a point $(x,v) \in SM$ it can be written as

$$d\Sigma^{2n-1} = dV^n \wedge dS_x.$$

This form can also be interpreted as the volume form of the Sasaki metric on SM or the volume form associated with the contact form of the geodesic flow. Liouville's theorem in classical mechanics asserts that the geodesic flow preserves $d\Sigma^{2n-1}$. In terms of the Lie derivative L_X this can be written as follows:

Lemma 3.6.4 $L_X(d\Sigma^{2n-1}) = 0$.

Similarly, the integral of $h \in C(\partial SM)$ over SM is

$$\int_{\partial M} \int_{S_x M} h(x,v)\, dS_x(v)\, dV^{n-1}(x),$$

where dV^{n-1} is the volume form of $(\partial M, g)$. This integral induces a volume form on ∂SM given by

$$d\Sigma^{2n-2} := dV^{n-1} \wedge dS_x,$$

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

where dV^{n-1} is the volume form of $(\partial M, g)$. This is just the volume form of the Sasaki metric restricted to ∂SM. Restricting $d\Sigma^{2n-2}$ to $\partial_\pm SM$ gives the natural volume form on these sets. The next lemma will be useful when proving Santaló's formula.

Lemma 3.6.5 *We have $j^* i_{\overline{\nu}} d\Sigma^{2n-1} = -d\Sigma^{2n-2}$, where $\overline{\nu} = (\nu, 0)$ is the horizontal lift of the unit normal ν and $j\colon \partial SM \to SM$ is the inclusion map. Moreover, $j^* i_X d\Sigma^{2n-1} = -\mu\, d\Sigma^{2n-2}$.*

Proof Consider a positively oriented orthonormal basis $(\xi_1, \dots, \xi_{2n-2})$ of $T_{(x,v)}\partial SM$. Since $\overline{\nu}$ is the inward unit normal in the Sasaki metric, by definition of boundary orientation, we have

$$d\Sigma^{2n-1}(\overline{\nu}, \xi_1, \dots, \xi_{2n-2}) = -1,$$

which gives the first claim. Writing $X = (X - \mu\overline{\nu}) + \mu\overline{\nu}$ and noting that $X - \mu\overline{\nu}$ is tangent to ∂SM, the second claim follows. □

The volume forms on SM and ∂SM induce L^2 inner products

$$(u, w)_{L^2(SM)} = \int_{SM} u\bar{w}\, d\Sigma^{2n-1},$$
$$(h, r)_{L^2(\partial SM)} = \int_{\partial SM} h\bar{r}\, d\Sigma^{2n-2}.$$

One has corresponding L^2 spaces $L^2(SM)$ and $L^2(\partial SM)$, with norms induced by the inner products.

Next we state and prove Santaló's formula. Recall that $\mu(x, v) = \langle \nu(x), v\rangle$ for $(x, v) \in \partial SM$.

Proposition 3.6.6 (Santaló's formula) *Let (M, g) be a compact non-trapping manifold with strictly convex boundary. Given $f \in C(SM)$ we have*

$$\int_{SM} f\, d\Sigma^{2n-1} = \int_{\partial_+ SM} d\mu(x, v) \int_0^{\tau(x,v)} f(\varphi_t(x, v))\, dt,$$

where $d\mu = \mu\, d\Sigma^{2n-2}$.

The proof will be very similar to the proof in two dimensions that we have already seen. We shall need the following lemma, which is an easy consequence of Stokes' theorem (its proof is left as an exercise).

Lemma 3.6.7 *Let N be a compact manifold with boundary, Θ a volume form, Y a vector field, and $u \in C^\infty(N)$. Then*

$$\int_N Y(u)\Theta = -\int_N u L_Y \Theta + \int_{\partial N} j^*(u i_Y \Theta),$$

where $j\colon \partial N \to N$ is the inclusion map.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Proof of Proposition 3.6.6 Recall that $\tau \in C(SM)$. Given $f \in C_c^\infty(SM)$, define for $(x,v) \in SM$,

$$u^f(x,v) := \int_0^{\tau(x,v)} f(\varphi_t(x,v))\,dt. \tag{3.16}$$

Clearly $u^f \in C(SM)$ and $u^f|_{\partial_- SM} = 0$. But if f has compact support in the interior of M, then u^f is in fact smooth. A simple computation shows that

$$Xu^f = -f. \tag{3.17}$$

We now apply Lemma 3.6.7 for the case $N = SM$, $Y = X$, and $u = u^f$. Since $L_X d\Sigma^{2n-1} = 0$ and $u^f|_{\partial_- SM} = 0$, we deduce

$$\int_{SM} f\,d\Sigma^{2n-1} = -\int_{\partial_+ SM} j^*\Big(u^f i_X d\Sigma^{2n-1}\Big).$$

The proposition now follows from the fact that $j^* i_X d\Sigma^{2n-1} = -\mu\, d\Sigma^{2n-2}$ (Lemma 3.6.5) and Exercise 3.5.15. $\square$

The next proposition shows that there is a natural positive smooth density that is preserved by the scattering relation. It also shows that the scattering relation is an orientation reversing diffeomorphism.

Proposition 3.6.8 *Let (M, g) be a non-trapping manifold with strictly convex boundary. Then*

$$\alpha^*\Big(\mu\,d\Sigma^{2n-2}\Big) = \mu\,d\Sigma^{2n-2}.$$

Moreover

$$\alpha^*\left(\frac{\mu}{\tilde{\tau}}\,d\Sigma^{2n-2}\right) = -\frac{\mu}{\tilde{\tau}}\,d\Sigma^{2n-2}.$$

Proof Recall that $\alpha(x,v) = \varphi_{\tilde{\tau}(x,v)}(x,v)$, thus using the chain rule we obtain for $\xi \in T_{(x,v)}\partial SM$:

$$d\alpha|_{(x,v)}(\xi) = d\tilde{\tau}(\xi)X(\alpha(x,v)) + d\varphi_{\tilde{\tau}(x,v)}(\xi). \tag{3.18}$$

Let us compute $\alpha^* j^* i_X d\Sigma^{2n-1} = (j\alpha)^* i_X d\Sigma^{2n-1}$. For this, take a basis $\{\xi_1, \ldots, \xi_{2n-2}\}$ of $T_{(x,v)}\partial SM$ and write

$$\begin{aligned}
&(j\alpha)^* i_X d\Sigma^{2n-1}(\xi_1, \ldots, \xi_{2n-2}) \\
&\quad = d\Sigma^{2n-1}(X(\alpha(x,v)), d\alpha|_{(x,v)}(\xi_1), \ldots, d\alpha|_{(x,v)}(\xi_{2n-2})) \\
&\quad = d\Sigma^{2n-1}(X(\alpha(x,v)), d\varphi_{\tilde{\tau}(x,v)}(\xi_1), \ldots, d\varphi_{\tilde{\tau}(x,v)}(\xi_{2n-2})) \\
&\quad = d\Sigma^{2n-1}(X(x,v), \xi_1, \ldots, \xi_{2n-2}),
\end{aligned}$$

where in the third line we used (3.18) and in the fourth we used that the geodesic flow preserves $d\Sigma^{2n-1}$. Thus

$$\alpha^* j^* i_X d\Sigma^{2n-1} = j^* i_X d\Sigma^{2n-1},$$

and the first identity in the proposition follows from Lemma 3.6.5. The second identity follows from $\tilde{\tau} \circ \alpha = -\tilde{\tau}$ and Lemma 3.2.8. □

3.7 Conjugate Points and Morse Theory

In this section we review basic properties of conjugate points (see e.g. Lee (1997); Jost (2017)). The following two facts will be important for later applications:

- Absence of conjugate points implies positivity of the index form. This will imply the positivity of certain terms in the Pestov identity used in the proof of injectivity of the geodesic X-ray transform on simple manifolds.
- Absence of conjugate points implies that the exponential map is a global diffeomorphism onto a simple manifold. This gives an analogue of polar coordinates, which can be used to prove that the normal operator of the geodesic X-ray transform is an elliptic pseudodifferential operator.

We will also state some related facts coming from Morse theory.

3.7.1 Conjugate Points and Jacobi Fields

Let (M,g) be a Riemannian manifold, and let $\gamma\colon [a,b] \to M$ be a geodesic segment. A family of curves $(\gamma_s)_{s\in(-\varepsilon,\varepsilon)}$ depending smoothly on s is called a *variation of γ through geodesics* if each $\gamma_s\colon [a,b] \to M$ is a geodesic (not necessarily unit speed) and if $\gamma_0 = \gamma$. We say that the variation γ_s *fixes the end points* if $\gamma_s(a) = \gamma(a)$ and $\gamma_s(b) = \gamma(b)$ for $s \in (-\varepsilon,\varepsilon)$.

Intuitively, conjugate points are related to situations where a family of geodesics starting at a fixed point converges to another point after finite time. The following is a basic example of this behaviour.

Example 3.7.1 (Family of geodesics joining the south and north pole) Let S^n, $n \geq 2$ be the sphere and consider the geodesic segment

$$\gamma\colon [-\pi/2,\pi/2] \to S^n, \ \ \gamma(t) = (\cos t)e_1 + (\sin t)e_{n+1}.$$

Define

$$\gamma_s\colon [-\pi/2,\pi/2] \to S^n, \ \ \gamma_s(t) = (\cos t)((\cos s)e_1 + (\sin s)e_2) + (\sin t)e_{n+1}.$$

Then (γ_s) is a variation of γ through geodesics that fixes the end points $-e_{n+1}$ (south pole) and e_{n+1} (north pole).

Any smooth variation (γ_s) of γ has a variation field $\partial_s \gamma_s(t)|_{s=0}$, which is a smooth vector field along γ. If (γ_s) is a variation through geodesics, then each $\gamma_s(t)$ satisfies the geodesic equation. Consequently the variation field $\partial_s \gamma_s(t)|_{t=0}$ satisfies the linearized geodesic equation, also known as the *Jacobi equation*. Below we write $D_t = \nabla_{\dot{\gamma}(t)}$ for the covariant derivative along $\gamma(t)$ and use the curvature operator

$$R_\gamma J := R(J, \dot{\gamma})\dot{\gamma},$$

where $R(X,Y)Z$ is the Riemann curvature tensor of (M,g).

Lemma 3.7.2 (Jacobi equation) *Let $\gamma\colon [a,b] \to M$ be a geodesic segment, and let (γ_s) be a variation of γ through geodesics. Then the variation field $J(t) = \partial_s \gamma_s(t)|_{s=0}$ satisfies the Jacobi equation*

$$D_t^2 J(t) + R_\gamma J(t) = 0, \qquad t \in [a,b].$$

Conversely, if $J(t)$ is a smooth vector field along γ satisfying the Jacobi equation, then there is a variation (γ_s) of γ through geodesics so that $\partial_s \gamma_s(t)|_{t=0} = J(t)$.

Proof Write $\Gamma(s,t) = \gamma_s(t)$, so that $\Gamma\colon (-\varepsilon,\varepsilon) \times [a,b] \to M$ is smooth. Then $J(t) = \partial_s \Gamma(0,t)$, and we wish to compute $D_t^2 J(t)$. Write $D_s = \nabla_{\partial_s \gamma_s}$. Since ∇ is torsion free, one has

$$D_t \partial_s \gamma_s(t) = D_s \partial_t \gamma_s(t).$$

Moreover, the definition of the Riemann curvature tensor gives that

$$D_t D_s W - D_s D_t W = R(\partial_t \gamma_s, \partial_s \gamma_s) W.$$

These facts imply that

$$\begin{aligned} D_t^2 J(t) &= D_t D_t \partial_s \gamma_s(t)|_{s=0} = D_t D_s \partial_t \gamma_s(t)|_{s=0} \\ &= D_s D_t \partial_t \gamma_s(t))|_{s=0} + R(\dot{\gamma}(t), J(t))\dot{\gamma}(t). \end{aligned}$$

One has $D_t \partial_t \gamma_s(t) = 0$ since each γ_s is a geodesic. Thus $J(t)$ satisfies the Jacobi equation.

For the converse, if $J(t)$ solves the Jacobi equation it is enough to consider a variation

$$\gamma_s(t) = \exp_{\eta(s)}(tW(s)) = \gamma_{\eta(s),W(s)}(t),$$

where η is a smooth curve with $\eta(0) = \gamma(a)$, and $W(s)$ is a smooth vector field along η with $W(0) = \dot{\gamma}(a)$. Then (γ_s) is a variation of γ through geodesics, and its variation field $Y(t) = \partial_s \gamma_s(t)|_{s=0}$ satisfies $Y(0) = \dot{\eta}(0)$ and

$$D_t Y(0) = D_s \partial_t \gamma_s(t)|_{s=t=0} = D_s W(0).$$

Now if we choose η and W so that $\dot{\eta}(0) = J(0)$ and $D_s W(0) = D_t J(0)$, then both $J(t)$ and $Y(t)$ satisfy the Jacobi equation with the same initial conditions. Uniqueness for linear ODEs shows that $Y \equiv J$. □

Definition 3.7.3 (Jacobi field) A smooth vector field along γ that solves the Jacobi equation is called a *Jacobi field*.

If a geodesic $\gamma : [a,b] \to M$ admits a variation through geodesics that fixes the end points, then by Lemma 3.7.2 it also admits a Jacobi field vanishing at the end points. This leads to the definition of conjugate points.

Definition 3.7.4 (Conjugate points) Let $\gamma : [a,b] \to M$ be a geodesic segment. We say that the points $\gamma(a)$ and $\gamma(b)$ are *conjugate along* γ if there is a nontrivial Jacobi field $J : [a,b] \to TM$ along γ satisfying $J(a) = J(b) = 0$.

Remark 3.7.5 If $\gamma(a)$ and $\gamma(b)$ are conjugate along γ, it follows from Lemma 3.7.2 (by choosing $\eta(s) \equiv \gamma(a)$ in the proof) that there is a variation (γ_s) of γ through geodesics that fixes the initial point $\gamma(a)$ and almost fixes the end point $\gamma(b)$ in the sense that $\partial_s \gamma_s(b)|_{s=0} = 0$.

The next lemma contains some basic properties of Jacobi fields. We say that a Jacobi field is *normal* (respectively *tangential*) if $J(t) \perp \dot{\gamma}(t)$ (respectively $J(t) \parallel \dot{\gamma}(t)$) for all t.

Lemma 3.7.6 *Let $\gamma : [a,b] \to M$ be a geodesic segment. Given any $v, w \in T_{\gamma(a)}M$, there is a unique Jacobi field with*

$$J(a) = v, \qquad D_t J(a) = w.$$

The space of Jacobi fields along γ is a $2n$-dimensional subspace of the set of smooth vector fields along γ. The space of normal Jacobi fields is $(2n-2)$-dimensional, and the space of tangential Jacobi fields is $\mathrm{span}\{\dot{\gamma}(t), t\dot{\gamma}(t)\}$ and hence 2-dimensional. The following conditions are equivalent:

(a) J is normal.
(b) $J(t_0)$ and $D_t J(t_0)$ are orthogonal to $\dot{\gamma}(t_0)$ at some t_0.
(c) $J(t_1) \perp \dot{\gamma}(t_1)$ and $J(t_2) \perp \dot{\gamma}(t_2)$ for some $t_1 \neq t_2$.

Proof The first claim follows from existence and uniqueness for linear ODEs. The map $(v,w) \mapsto J$ is linear and bijective, showing that the space of Jacobi fields is $2n$-dimensional. The geodesic equation $D_t\dot\gamma(t) = 0$ together with the antisymmetry of the curvature tensor imply that

$$\partial_t^2\langle J,\dot\gamma\rangle = \left\langle D_t^2 J,\dot\gamma\right\rangle = \left\langle D_t^2 J + R(J,\dot\gamma)\dot\gamma,\dot\gamma\right\rangle.$$

Thus for any Jacobi field, $\langle J,\dot\gamma\rangle = ct + d$ for some $c,d \in \mathbb{R}$, and taking the t-derivative gives that $\langle D_t J,\dot\gamma\rangle = c$. It follows that (a), (b), and (c) are equivalent. By part (b) one sees that the space of normal Jacobi fields is $(2n-2)$-dimensional, and it is easy to check that $\dot\gamma(t)$ and $t\dot\gamma(t)$ are linearly independent tangential Jacobi fields. □

The tangential Jacobi fields are not very interesting (they correspond to the variations $\gamma_s(t) = \gamma(t+s)$ and $\gamma_s(t) = \gamma(e^s t)$, which are just reparametrizations of $\gamma(t)$). Thus we will focus on normal Jacobi fields.

3.7.2 Jacobi Fields in Dimension Two

If $\dim M = 2$ there is a very simple description of Jacobi fields in terms of solutions of the ODE $\ddot y(t) + K(\gamma(t))y(t) = 0$, where K is the Gaussian curvature. Recall that $v^\perp$ is the rotation of v by 90° counterclockwise.

Lemma 3.7.7 (Jacobi fields in two dimensions) *Let (M,g) be two dimensional and $\gamma\colon [a,b] \to M$ a unit speed geodesic segment. The set of normal Jacobi fields along γ is spanned by $\alpha(t)\dot\gamma(t)^\perp$ and $\beta(t)\dot\gamma(t)^\perp$, where $\alpha,\beta \in C^\infty([a,b])$ satisfy the equations*

$$\begin{aligned}
&\ddot\alpha(t) + K(\gamma(t))\alpha(t) = 0, \qquad && \alpha(a) = 1,\ \ \dot\alpha(a) = 0,\\
&\ddot\beta(t) + K(\gamma(t))\beta(t) = 0, \qquad && \beta(a) = 0,\ \ \dot\beta(a) = 1.
\end{aligned}$$

Proof We first observe that $\dot\gamma(t)^\perp$ is parallel, i.e.

$$D_t\big(\dot\gamma(t)^\perp\big) = 0. \tag{3.19}$$

In fact, since $D_t\dot\gamma(t) = 0$ we have

$$\begin{aligned}
\left\langle D_t\dot\gamma^\perp,\dot\gamma\right\rangle &= \partial_t\big(\langle\dot\gamma^\perp,\dot\gamma\rangle\big) = \partial_t(0) = 0,\\
\left\langle D_t\dot\gamma^\perp,\dot\gamma^\perp\right\rangle &= \frac{1}{2}\partial_t\big(\langle\dot\gamma^\perp,\dot\gamma^\perp\rangle\big) = \frac{1}{2}\partial_t(1) = 0.
\end{aligned}$$

This proves (3.19).

When $\dim M = 2$ the Jacobi equation reduces to

$$D_t^2 J(t) + K(\gamma(t))J(t) = 0.$$

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

If $\alpha(t)$ and $\beta(t)$ satisfy the given equations, it follows from (3.19) that $\alpha(t)\dot{\gamma}(t)^{\perp}$ and $\beta(t)\dot{\gamma}(t)^{\perp}$ solve the Jacobi equation. Since they are linearly independent and normal, they span the space of normal Jacobi fields along γ. □

We can also present an alternative derivation of the Jacobi equation based on the structure equations given in Lemma 3.5.5 and the geodesic flow φ_t acting on SM.

Let (M, g) be an arbitrary Riemannian surface that we assume oriented for simplicity. Fix a point $(x, v) \in SM$. We adopt the following notation: let $X_\perp(t) = X_\perp(\varphi_t(x, v))$ and $X_\perp = X_\perp(0) = X_\perp(x, v)$, and similarly for $X(t)$, $V(t)$ etc. Let $\xi \in T_{(x,v)}SM$. We can write

$$\xi = aX - yX_\perp + zV$$

for some constants $a, y, z \in \mathbb{R}$. Moreover, there exist smooth functions $a(t), y(t), z(t)$ satisfying

$$d\varphi_t(\xi) = a(t)X(t) - y(t)X_\perp(t) + z(t)V(t), \tag{3.20}$$

subject to the initial conditions $a(0) = a$, $y(0) = y$ and $z(0) = z$.

Proposition 3.7.8 *The functions $a(t)$, $y(t)$, and $z(t)$ satisfy the equations*

$$\begin{aligned} \dot{a} &= 0, \\ \dot{y} - z &= 0, \\ \dot{z} + Ky &= 0. \end{aligned}$$

Proof We begin by applying $d\varphi_{-t}$ to both sides of (3.20) to obtain

$$\xi = a(t)d\varphi_{-t}(X(t)) - y(t)d\varphi_{-t}(X_\perp(t)) + z(t)d\varphi_{-t}(V(t)).$$

Differentiating both sides with respect to t and recalling the Lie derivative formula $L_X Y(\varphi_t) = \frac{d}{dt}(d\varphi_{-t}(Y(\varphi_t)))$, we obtain

$$\begin{aligned} 0 &= \frac{d}{dt}(\xi) \\ &= \dot{a}(t)d\varphi_{-t}(X(t)) + a(t)d\varphi_{-t}([X, X](t)) - \dot{y}d\varphi_{-t}(X_\perp(t)) \\ &\quad - y(t)d\varphi_{-t}([X, X_\perp]) + \dot{z}d\varphi_{-t}(V(t)) + z(t)d\varphi_{-t}([X, V](t)), \end{aligned}$$

and then applying Lemma 3.5.5 and grouping like terms, we obtain

$$0 = d\varphi_{-t}[\dot{a}(t)X(t) + (z(t) - \dot{y}(t))X_\perp(t) + (\dot{z} + K(t)y(t))V(t)].$$

Since $d\varphi_{-t}$ is an isomorphism and $\{X(t), X_\perp(t), V(t)\}$ is a basis of each tangent space $T_{\varphi_t(x,v)}SM$ the coefficients of $X(t)$, $X_\perp(t)$, and $V(t)$ must vanish for all t, and this is precisely what we wanted to show. □

The proposition implies, in particular, that $d\varphi_t$ leaves the 2-plane bundle spanned by $\{X_\perp, V\}$ invariant. Moreover, if $\xi = -yX_\perp + zV$, then

$$d\varphi_t(\xi) = -y(t)X_\perp(t) + \dot{y}(t)V(t),$$

where $y(t)$ is uniquely determined by the Jacobi equation $\ddot{y} + Ky = 0$ with initial conditions $y(0) = y$ and $\dot{y}(0) = z$. We see that $d\pi d\varphi_t(\xi) = -y(t)d\pi(X_\perp(t)) = y(t)\dot{\gamma}^\perp(t)$ is the normal Jacobi field J with initial conditions $J(0) = y\dot{\gamma}^\perp(0)$, $\dot{J}(0) = z\dot{\gamma}^\perp(0)$.

Thus Jacobi fields and their covariant derivatives describe how the differential of the geodesic flow evolves. The same is true in higher dimensions. Using the splitting described in Section 3.6, we may write for $\xi \in T_{(x,v)}SM$,

$$d\varphi_t(\xi) = (J_\xi(t), D_t J_\xi(t)), \tag{3.21}$$

where J_ξ is the unique Jacobi field with initial conditions $J_\xi(0) = d\pi(\xi)$ and $D_t J_\xi(0) = \mathtt{K}\xi$, where $\mathtt{K}$ is the connection map.

Exercise 3.7.9 Prove (3.21).

3.7.3 Exponential Map

We discuss the exponential map on a compact manifold with boundary and evaluate its derivative in terms of Jacobi fields.

Proposition 3.7.10 (Exponential map) *Let (M, g) be a compact non-trapping manifold with strictly convex boundary. For any $x \in M$ define*

$$D_x := \{tv \in T_xM : v \in S_xM \text{ and } t \in [0, \tau(x, v)]\}. \tag{3.22}$$

The exponential map

$$\exp_x : D_x \to M, \quad \exp_x(tv) = \gamma_{x,v}(t)$$

is smooth. For any $tv \in D_x$ and $w \in T_xM$, one has

$$(d\exp_x)|_{tv}(tw) = J(t),$$

where J is the Jacobi field along $\gamma_{x,v}$ with $J(0) = 0$ and $D_t J(0) = w$.

Proof The assumption on (M, g) guarantees that any point of D_x is the limit of some sequence in $(D_x)^{\mathrm{int}}$. Thus it is enough to verify the claims for any smooth extension of $\exp_x$ to some larger manifold containing D_x (the values of $d\exp_x$ on ∂D_x do not depend on the choice of the extension). Let (N, g) be a closed extension of (M, g). Then geodesics on N are well defined for all time and the exponential map of N, $\exp_x^N : T_xN \to N$, is smooth. It follows that $\exp_x = \exp_x^N |_{D_x}$ is also smooth.

Given $tv \in D_x$ and $w \in T_xM$, consider the smooth curve $\eta(s) = tv + stw$ on T_xN. By the definition of the derivative one has

$$\big(d\exp_x^N\big)|_{tv}(tw) = \frac{d}{ds}\exp_x^N(\eta(s))\Big|_{s=0}.$$

Consider $\gamma_s(r) = \exp_x^N(r(v+sw)) = \gamma_{x,v+sw}(r)$. Then $\gamma_s(r)$ is a variation of $\gamma_{x,v}(r)$ through geodesics in N, hence $J(r) = \partial_s\gamma_s(r)|_{s=0}$ is a Jacobi field along $\gamma_{x,v}$ with $J(0) = 0$ and $D_rJ(0) = D_s(v+sw)|_{s=0} = w$. It follows that $(d\exp_x^N)|_{tv}(tw) = \partial_s\gamma_s(t)|_{s=0} = J(t)$. □

Corollary 3.7.11 *Given $tv \in D_x$, the derivative $d\exp_x|_{tv}$ is invertible if and only if $\gamma_{x,v}(t)$ is not conjugate to x along $\gamma_{x,v}$.*

We will also need the Gauss lemma.

Proposition 3.7.12 (Gauss lemma) *Let $x \in M$ and $tv \in D_x$. For any $w \in T_xM$ one has*

$$\langle d\exp_x|_{tv}(v), d\exp_x|_{tv}(w)\rangle = \langle v, w\rangle.$$

In particular, $d\exp_x|_{tv}(w) \perp \dot\gamma_{x,v}(t)$ if and only if $v \perp w$.

Proof Note first that $d\exp_x|_{tv}(v) = \dot\gamma_{x,v}(t)$, and by Proposition 3.7.10 one has $d\exp_x|_{tv}(tw) = J_w(t)$ where $J_w(t)$ is the Jacobi field along $\gamma_{x,v}$ with $J_w(0) = 0$ and $D_tJ_w(0) = w$. Define

$$f(t) := \langle d\exp_x|_{tv}(v), d\exp_x|_{tv}(tw)\rangle = \langle\dot\gamma_{x,v}(t), J_w(t)\rangle.$$

Since $D_t\dot\gamma_{x,v}(t) = 0$, taking derivatives and using the Jacobi equation gives that

$$f''(t) = \big\langle\dot\gamma_{x,v}(t), D_t^2J_w(t)\big\rangle = -\langle\dot\gamma_{x,v}(t), R_\gamma J_w(t)\rangle.$$

The symmetries of the curvature tensor imply that the last quantity is zero. Thus $f(t)$ is an affine function, and

$$\begin{aligned}\langle d\exp_x|_{tv}(v), d\exp_x|_{tv}(tw)\rangle &= f(0) + f'(0)t = t\langle\dot\gamma_{x,v}(0), D_tJ_w(0)\rangle\\ &= t\langle v, w\rangle.\end{aligned}$$

This proves the result for $t > 0$, and the case $t = 0$ follows since $d\exp_x|_0 = \mathrm{id}$. □

The following result shows that among curves that are exponential images of curves in the domain of $\exp_x$, the radial geodesics always minimize length.

Proposition 3.7.13 (Minimizing curves in domain of $\exp_x$) *Let $x \in M$ and $w \in D_x$, let $\eta_0\colon [0,1] \to D_x$ be the curve $\eta_0(t) = tw$, and let $\eta\colon [0,1] \to D_x$ be any smooth curve with $\eta(0) = 0$ and $\eta(1) = w$. Then*

$$\int_0^1 |(\exp_x \circ \eta_0)'(t)|\, dt \le \int_0^1 |(\exp_x \circ \eta)'(t)|\, dt$$

with equality if and only if η is a reparametrization of η_0.

Proof We may assume that $w \neq 0$ and $\eta(t) \neq 0$ for $0 < t \le 1$ (if not, let t_0 be the last time with $\eta(t_0) = 0$ and replace η by $\eta|_{[t_0,1]}$ rescaled to the interval $[0,1]$). We write $\eta(t) = r(t)\omega(t)$ where $r(t) = |\eta(t)|$ and $|\omega(t)| = 1$. Then for $t > 0$ one has

$$\dot\eta(t) = \dot r(t)\omega(t) + r(t)\dot\omega(t).$$

The condition $|\omega(t)| = 1$ implies $\langle \omega(t), \dot\omega(t)\rangle = 0$. Using the Gauss lemma, we obtain that

$$\langle d\exp_x|_{\eta(t)}(\omega(t)), d\exp_x|_{\eta(t)}(\dot\omega(t))\rangle = 0,$$
$$|d\exp_x|_{\eta(t)}(\omega(t))| = |d\exp_x|_{\eta(t)}(\dot\omega(t))| = 1.$$

Combining these facts gives that

$$|(\exp_x \circ \eta)'(t)|^2 = |d\exp_x|_{\eta(t)}(\dot\eta(t))|^2 \ge \dot r(t)^2.$$

Thus the lengths satisfy

$$\begin{aligned}\int_0^1 |(\exp_x \circ \eta)'(t)|\, dt &\ge \int_0^1 |\dot r(t)| \ge r(1) - r(0) = |w| \\ &= \int_0^1 |(\exp_x \circ \eta_0)'(t)|\, dt.\end{aligned}$$

Equality holds if and only if $\dot\omega(t) = 0$ and $\dot r(t) \ge 0$, which corresponds to the case where η is a reparametrization of η_0. $\square$

3.7.4 Index Form

Next we consider a bilinear form on γ related to the Jacobi equation.

Definition 3.7.14 (Index form) Let $\gamma\colon [a,b] \to M$ be a geodesic segment, and let $H^1(\gamma)$ be the Sobolev space of vector fields along γ equipped with the norm

$$\|Y\|_{H^1(\gamma)} = \left(\int_a^b (|Y(t)|^2 + |D_t Y(t)|^2)\, dt\right)^{1/2}.$$

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Define $H^1_0(\gamma) = \{Y \in H^1(\gamma)\,;\, Y(a) = Y(b) = 0\}$. The *index form* of γ is the bilinear form

$$\mathtt{I}_\gamma(Y,Z) = \int_a^b (\langle D_t Y, D_t Z\rangle - \langle R_\gamma Y, Z\rangle)\,dt,$$

defined for $Y, Z \in H^1_0(\gamma)$.

The index form $\mathtt{I}_\gamma$ is the bilinear form associated with the elliptic operator $-D_t^2 - R_\gamma$ acting on $H^1_0(\gamma)$ (i.e. with vanishing Dirichlet boundary values). It arises as the second variation of the length or energy functionals. Namely, if $\gamma_s\colon [a,b] \to M$ is a variation of a unit speed geodesic γ through geodesics that fixes the end points, then

$$\frac{d^2}{ds^2} \int_a^b |\dot{\gamma}_s(t)|_g\,dt\Big|_{s=0} = \mathtt{I}_\gamma(Y,Y), \tag{3.23}$$

where $Y(t)$ is the component of $\partial_s \gamma_s(t)|_{s=0}$ normal to $\dot{\gamma}(t)$. Thus if γ minimizes length between its end points among the curves γ_s, then necessarily $\mathtt{I}_\gamma(Y,Y) \geq 0$.

The main result for our purposes is that absence of conjugate points guarantees that $\mathtt{I}_\gamma$ is positive definite.

Proposition 3.7.15 (Positivity of index form) *Let $\gamma\colon [a,b] \to M$ be a geodesic segment, and consider the index form $\mathtt{I}_\gamma$ on $H^1_0(\gamma)$. Then*

$$\mathtt{I}_\gamma > 0 \textit{ if and only if there is no pair of conjugate points on } \gamma.$$

There are many possible proofs of the above proposition. We will give one based on PDE (or in this case ODE) type ideas.

Proof For $r \in (a,b]$, let L_r be the elliptic operator $-D_t^2 - R_\gamma$ acting on $H^1_0(\gamma|_{[a,r]})$. Then L_r has a countable set of Dirichlet eigenvalues $\lambda_1(r) \leq \lambda_2(r) \leq \cdots$ with corresponding $L^2([a,r])$-normalized eigenfunctions $Y_j(\,\cdot\,;r)$ satisfying

$$\big(-D_t^2 - R_\gamma\big)Y_j(\,\cdot\,;r) = \lambda_j(r)Y_j(\,\cdot\,;r) \text{ on } (a,r), \quad Y_j(a;r) = Y_j(r;r) = 0.$$

We will be interested in the smallest eigenvalue $\lambda_1(r)$, also given by the Rayleigh quotient

$$\lambda_1(r) = \min_{Y \in H^1_0(\gamma|_{[a,r]})\setminus\{0\}} \frac{\mathtt{I}_\gamma(Y,Y)}{\|Y\|^2_{L^2(\gamma)}}.$$

Clearly $\mathtt{I}_\gamma > 0$ if and only if $\lambda_1(b) > 0$.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

We claim the following facts:

(1) $\lambda_1(r) > 0$ for r close to a.
(2) $\lambda_1(r)$ is Lipschitz continuous and decreasing on $(a,b]$.
(3) If $\lambda_1(r_0) = 0$, then $\gamma(a)$ and $\gamma(r_0)$ are conjugate.

The result now follows: if there are no conjugate points, then $\lambda_1(r)$ is never zero and hence $\lambda_1(r)$ is positive on $(a,b]$, showing that $\mathtt{I}_\gamma$ is positive definite. Conversely, if there is a pair of conjugate points then there is a nontrivial Jacobi field J vanishing at some a' and b'. Extending it by zero to $[a,b]$ gives a nontrivial vector field J in $H_0^1(\gamma)$, and integrating by parts shows that $\mathtt{I}_\gamma(J,J) = 0$. Thus $\mathtt{I}_\gamma$ is not positive definite.

Claim (1) above follows from a Poincaré inequality: if $Y \in H_0^1(\gamma|_{[a,a+\varepsilon]})$, then

$$\begin{aligned}\int_a^{a+\varepsilon} |Y|^2\,dt &= \int_a^{a+\varepsilon} \partial_t(t-a)|Y|^2\,dt = -2\int_a^{a+\varepsilon} (t-a)\langle D_tY, Y\rangle\,dt \\ &\le 2\varepsilon\|D_tY\|\|Y\| \le 2\varepsilon^2\|D_tY\|^2 + \frac{1}{2}\|Y\|^2.\end{aligned}$$

Absorbing the last term on the right to the left gives $\|D_tY\| \ge \frac{1}{2\varepsilon}\|Y\|$. If ε is chosen small enough, we get that $\mathtt{I}_\gamma(Y,Y) \ge c\|Y\|_{L^2}^2$ for some $c > 0$ whenever $Y \in H_0^1(\gamma|_{[a,a+\varepsilon]})$.

Claim (2) is standard: rescaling the interval $[a,r]$ to $[a,b]$, we see that $\lambda_1(r)$ is related to the smallest eigenvalue of a second order self-adjoint elliptic operator on $H_0^1(\gamma)$ whose coefficients depend smoothly on r. Hence $\lambda_1(r)$ is Lipschitz continuous and decreasing (both facts can be checked directly from the Rayleigh quotient). Claim (3) is immediate from the definition of conjugate points and elliptic regularity. □

Exercise 3.7.16 Prove claim (2) in the proof of Proposition 3.7.15.

The proof of Proposition 3.7.15 combined with the second variation formula (3.23) also gives the following result.

Proposition 3.7.17 (Geodesics do not minimize past conjugate points) *If $\gamma\colon [a,b] \to M$ is a geodesic segment having an interior point conjugate to $\gamma(a)$, then there is $X \in H_0^1(\gamma)$ with $\mathtt{I}_\gamma(X,X) < 0$ and γ is not length minimizing.*

The kernel of $\mathtt{I}_\gamma$ is the set of Jacobi fields vanishing at the end points,

$$\mathcal{J}(\gamma) = \{J \in H_0^1(\gamma)\colon D_t^2 J + R_\gamma J = 0,\ J(a) = J(b) = 0\}.$$

By elliptic regularity any $J \in \mathcal{J}(\gamma)$ is C^∞, and hence one can use H_0^1 vector fields J in the definition of conjugate points.

We will next state the Morse index theorem (cf. Jost (2017)) involving the two indices

$$\operatorname{Ind}(\gamma) = \dim V(\gamma),$$
$$\operatorname{Ind}_0(\gamma) = \dim V_0(\gamma),$$

where $V(\gamma)$ (respectively $V_0(\gamma)$) is a subspace of $H_0^1(\gamma)$ with maximal dimension so that the index form $\mathtt{I}_\gamma$ is negative definite (respectively negative semidefinite).

Theorem 3.7.18 (Morse index theorem) *Let $\gamma\colon [a,b] \to M$ be a geodesic segment. Then there are at most finitely many times $a < t_1 < \cdots < t_N \leq b$ so that $\gamma(t_j)$ is conjugate to $\gamma(a)$ along γ. The indices* $\operatorname{Ind}(\gamma)$ *and* $\operatorname{Ind}_0(\gamma)$ *are finite, and they satisfy*

$$\operatorname{Ind}(\gamma) = \sum_{t_j \in (a,b)} \dim \mathcal{J}\big(\gamma|_{[a,t_j]}\big),$$
$$\operatorname{Ind}_0(\gamma) = \sum_{t_j \in (a,b]} \dim \mathcal{J}\big(\gamma|_{[a,t_j]}\big).$$

3.7.5 Morse Theory Facts

The classical Morse theory of the energy functional on loop spaces provides several relevant results. These results are pretty standard on complete manifolds without boundary or closed manifolds. Given a compact manifold (M,g) with strictly convex boundary, throughout this subsection, we will assume that (N,g) is a no return extension with the following properties.

Lemma 3.7.19 (No return extension) *Let (M,g) be a compact manifold with strictly convex boundary. There is a complete manifold (N,g) of the same dimension as M so that (M,g) is isometrically embedded in (N,g) and geodesics leaving M never return to M. Moreover $N \setminus M$ can be taken as to be diffeomorphic to* $(0,\infty) \times \partial M$*, so that M is a deformation retract of N.*

Exercise 3.7.20 Prove that this extension exists (for a proof see Bohr (2021, Lemma 7.1)).

Proposition 3.7.21 *Let (M,g) be a compact manifold with strictly convex boundary. Then given any two points $x,y \in M$, any N-geodesic joining x and y is completely contained in M. Moreover, there is a minimizing geodesic in M connecting x to y.*

Proof If $\gamma\colon [0,1] \to N$ is a geodesic with $\gamma(0) = x$ and $\gamma(1) = y$, then $\gamma([0,1]) \subset M$ since otherwise some $\gamma(t_0)$ would be outside M and then also $\gamma(1) = y$ would be outside M, which is impossible. Moreover, since (N, g) is complete, the Hopf–Rinow theorem ensures that there is a minimizing geodesic in N connecting x and y and by the above argument this geodesic stays in M. □

Proposition 3.7.22 *Let (M, g) be a compact non-trapping manifold with strictly convex boundary. Then M is contractible.*

Proof Since M is a deformation retract of N, it follows that M is contractible if and only if N is. A classical result in Serre (1951, Proposition 13), proved using Morse theory, asserts that if $x, y \in N$ are distinct and if N is not contractible, there are infinitely many geodesics connecting x to y. Let now x be fixed and consider the map $f : T_x N \to N$, $f(w) = \exp_x(w)$. Sard's theorem applied to f shows that almost every $y \in N$ is a regular value. In particular, such points y are not conjugate to x. Moreover, given $T > 0$ there are only finitely many $w \in T_x M$ with $f(w) = y$ and $|w| \le T$. This shows that there are geodesics connecting x to y with arbitrarily large length.

Since N is a no return extension, if we pick x and y in M, then M itself admits geodesics of arbitrarily large length connecting x to y thus violating the non-trapping property. It follows that M is contractible. □

Remark 3.7.23 The proposition also follows from another well-known fact in Riemannian geometry: a compact connected and non-contractible Riemannian manifold with strictly convex boundary must have a closed geodesic in its interior (Thorbergsson, 1978, Theorem 4.2). This is also proved with Morse theory, but using the space of free loops.

Proposition 3.7.24 *Let (M, g) be a compact Riemannian manifold without conjugate points and with strictly convex boundary. Let γ be a geodesic with end points $x, y \in M$. If α is any other smooth curve in M connecting x to y that is homotopic to γ with a homotopy fixing the end points, then the length of α is larger than the length of γ. Moreover, there is a unique geodesic connecting x to y in a given homotopy class and this geodesic must be minimizing.*

Proof We follow Guillarmou and Mazzucchelli (2018, Lemma 2.2) where this very same proposition is proved. We let $\Omega(x, y)$ denote the Hilbert manifold of absolutely continuous curves $c : [0, 1] \to N$ with $c(0) = x$, $c(1) = y$ and finite energy

$$E(c) := \frac{1}{2} \int_0^1 |\dot{c}|^2 \, dt.$$

It is well known that $E : \Omega(x, y) \to \mathbb{R}$ is C^2 (Mazzucchelli, 2012, Proposition 3.4.3) and satisfies the Palais–Smale condition. The critical points of E are precisely the geodesics connecting x to y. Moreover, since there are no conjugate points, the Morse index theorem 3.7.18 guarantees that the Hessian of E at a critical point is positive definite (recall that N is a no return extension, so it suffices to assume that M has no conjugate points). Thus all critical points of E are local minimizers of E and are isolated. We now argue with E restricted to the connected component of $\Omega(x, y)$ containing γ, which we denote by $\Omega_{[\gamma]}(x, y)$. This coincides with the set of paths connecting x to y and homotopic to γ. We claim that γ is the unique minimizer of $E|_{\Omega_{[\gamma]}(x,y)}$. Indeed a mountain pass argument shows that if there is another local minimizer, then there is a geodesic $\sigma \in \Omega_{[\gamma]}(x, y)$ that is not a local minimum of $E|_{\Omega_{[\gamma]}(x,y)}$ (cf. Struwe (1996, Theorem 10.3) and Hofer (1985)). Again by the Morse index theorem, σ must contain conjugate points, and since it must be entirely contained in M, we get a contradiction. □

3.8 Simple Manifolds

In this section we introduce the notion of *simple manifold* and we prove several equivalent definitions. We start with the following:

Definition 3.8.1 Let (M, g) be a compact connected manifold with smooth boundary. The manifold is said to be *simple* if

- (M, g) is non-trapping,
- the boundary is strictly convex, and
- there are no conjugate points.

Our main goal will be to establish the following theorem.

Theorem 3.8.2 *Let (M, g) be a compact connected manifold with strictly convex boundary. The following are equivalent:*

(i) *M is simple;*
(ii) *M is simply connected and has no conjugate points;*
(iii) *for each $x \in M$, the exponential map $\exp_x$ is a diffeomorphism onto its image;*
(iv) *given two points there is a unique geodesic connecting them depending smoothly on the end points;*
(v) *consider (M, g) isometrically embedded in a complete manifold (N, g). Then M has a neighbourhood U in N such that any two points in U are joined by a unique geodesic;*
(vi) *the boundary distance function $d_g|_{\partial M \times \partial M}$ is smooth away from the diagonal.*

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

Remark 3.8.3 Conditions closely related to simplicity appear in Michel (1981/82); Muhometov (1977), and the term 'simple manifold' goes back at least to Sharafutdinov (1994). There may be other variations of the definition of simple manifold in the literature not listed above, but as far as we can see, they all follow easily from one of the statements above. An example is to say that a compact manifold (M, g) is simple if ∂M is strictly convex, every geodesic segment in M is minimizing and there are no conjugate points. Indeed, if every geodesic segment in M is minimizing, then (M, g) is non-trapping since all geodesic segments in M have length bounded by the diameter of M. We could also say that (M, g) is simple if ∂M is strictly convex, every two points are connected by a unique geodesic and there are no conjugate points.

We shall break down the proof of Theorem 3.8.2 into several propositions. The first is:

Proposition 3.8.4 *Let (M, g) be a simple manifold. Given $x, y \in M$, there is a unique geodesic connecting x to y and this geodesic is minimizing.*

Proof Since ∂M is strictly convex, Proposition 3.7.21 ensures that there is a minimizing geodesic connecting x to y. Since M is non-trapping, it must be simply connected by Proposition 3.7.22. Thus Proposition 3.7.24 implies that there is only one geodesic connecting x to y and this geodesic must be minimizing. □

Proposition 3.8.5 *Let (M, g) be simple. Given $x \in M$, let $D_x \subset T_x M$ be the domain of the exponential map given in* (3.22). *Then*

$$\exp_x : D_x \to M$$

is a diffeomorphism. In particular, M is diffeomorphic to a closed ball.

Proof The previous proposition asserts that if M is simple, then

$$\exp_x : D_x \to M$$

is a bijection. Since there are no conjugate points, Corollary 3.7.11 gives that $\exp_x$ is a local diffeomorphism at any $tv \in D_x$. Hence $\exp_x : D_x \to M$ is a diffeomorphism. This implies, in particular, that M is diffeomorphic to a closed ball in Euclidean space: if x is in the interior of M, then D_x is a closed star-shaped domain around zero with smooth boundary and hence diffeomorphic to a closed ball. □

Proposition 3.8.6 *Let (M, g) be a compact manifold with strictly convex boundary. The following are equivalent:*

(i) (M, g) *is simple;*
(ii) M *is simply connected and has no conjugate points.*

Any of these two properties implies:

- *Given two points in* M*, there is a unique geodesic connecting them and this geodesic is minimizing.*

Proof (i) $\Longrightarrow$ (ii): If M is simple, then it has no conjugate points by definition. It is simply connected due to Proposition 3.7.22.

(ii) $\Longrightarrow$ (i): Suppose M has strictly convex boundary, is simply connected and has no conjugate points. Proposition 3.7.24 implies that between two points in M there is a unique geodesic and this geodesic must be minimizing. It follows that all geodesics have length less than or equal to the diameter of M, hence the manifold is non-trapping and (M, g) is simple. $\square$

Proposition 3.8.7 *Let* (M, g) *be simple manifold. Any sufficiently small neighbourhood* U *of* M *in* N *whose boundary is* C^2*-close to that of* M *has the property that* $\overline{U}$ *is simple.*

Proof Clearly any sufficiently small neighbourhood U with ∂U C^2-close to ∂M has the property that its closure $\overline{U}$ has strictly convex boundary and is simply connected. To see that the property of having no conjugate points persists when we go to $\overline{U}$, let ρ be a boundary distance function for ∂M and let $U_r := \rho^{-1}[-r, \infty)$ with $r \geq 0$. If we cannot find a neighbourhood for M without conjugate points, there is a sequence $r_n \to 0$ and points $(x_n, v_n), (y_n, w_n) \in SU_{r_n}$ such that $\varphi_{t_n}(x_n, v_n) = (y_n, w_n)$, $d\varphi_{t_n}(\mathcal{V}(x_n, v_n)) \cap \mathcal{V}(y_n, w_n) \neq \{0\}$ with $t_n > 0$ and $\varphi_t(x_n, v_n) \in SU_{r_n}$ for all $t \in [0, t_n]$ (conjugate point condition, see (3.21)). By compactness we may assume that (x_n, v_n) converges to $(x, v) \in SM$ and (y_n, w_n) converges to $(y, w) \in SM$.

If the sequence t_n is bounded, by passing to a subsequence we deduce that there is $t_0 > 0$ such that $d\varphi_{t_0}(\mathcal{V}(x, v)) \cap \mathcal{V}(y, w) \neq \{0\}$ and thus M has conjugate points (the sequence t_n is bounded away from zero). Indeed, we have unit vectors (in the Sasaki metric) $\xi_n \in \mathcal{V}(x_n, v_n)$ such that

$$d\pi \circ d\varphi_{t_n}(\xi_n) = 0,$$

and passing to subsequences if necessary we find a unit norm $\xi \in \mathcal{V}(x, v)$ for which

$$d\pi \circ d\varphi_{t_0}(\xi) = 0.$$

If t_n is unbounded, we may assume by passing to a subsequence that $t_n \to \infty$. Since we are assuming that M is non-trapping there is $T > 0$ such that every

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

geodesic in M has length $\leq T$. Since $t_n \to \infty$, there is n_0 such that for all $n \geq n_0$, $\varphi_t(x_n, v_n) \in SU_{r_n}$ for all $t \in [0, T+1]$. Thus $\varphi_t(x, v) \in SM$ for all $t \in [0, T+1]$ and we have produced a geodesic in M with length $T+1$ which is a contradiction. $\square$

Exercise 3.8.8 Use the continuity of the cut time function $t_c \colon SN \to (0, \infty)$ (cf. Sakai (1996, chapter III, Proposition 4.1)) to give an alternative proof of Proposition 3.8.7 (take the extension N to be closed): if geodesics on M have no conjugate points and between two points there is only one, then cut points do **not** occur in M (again cf. (Sakai, 1996, chapter III, Proposition 4.1)), i.e. for all $(x, v) \in SM$, $\tau(x, v) < t_c(x, v)$. This means that one can go a bit further along any geodesic and by a uniform amount.

Exercise* 3.8.9 Construct an example of a compact surface with strictly convex boundary such that any two points are joined by a unique geodesic, but the surface is not simple. Such an example must have conjugate points between points at the boundary.

3.8.1 Proof of Theorem 3.8.2 Except for Item (vi)

The equivalence between (i) and (ii) is the content of Proposition 3.8.6. Proposition 3.8.5 gives that (i) implies (iii). To prove that (iii) implies (i), note that if $\exp_x$ is a diffeomorphism for each x, then every geodesic is minimizing by Proposition 3.7.13 and hence there are no geodesics with infinite length, thus M is non-trapping. We also know that the differential of $\exp_x$ is a linear isomorphism and hence there are no conjugate points (cf. Corollary 3.7.11). The equivalence between (iii) and (iv) follows right away if we note that $\gamma_{x, v(x,y)}(1) = \exp_x(v(x, y)) = y$, where $v(x, y)$ is defined uniquely if $\exp_x$ is a bijection. Smooth dependence of the geodesic on end points is precisely the statement that the map $(x, y) \mapsto v(x, y)$ is smooth. Let us complete the proof by showing that (i) $\iff$ (v). Proposition 3.8.7 gives that (i) $\implies$ (v). If we assume (v) we see right away that M is non-trapping and also that it is free of conjugate points (including boundary points) since U is a neighbourhood. $\square$

3.8.2 The Hessian of the Distance Function

The main purpose of this subsection is to complete the proof of Theorem 3.8.2 by establishing the equivalence of simplicity with item (vi) in the theorem. This result will not be subsequently used in the text.

Let (N, g) be a complete Riemannian manifold, fix $p \in N$ and let $f(x) := d(p, x)$. It is well known that f is smooth away from $\{p\} \cup \mathrm{Cut}_p$, where Cut_p

denotes the *cut locus* of p. It is also well known that the cut locus is a closed set of measure zero. Consider the open set $N_0 := N \setminus (\{p\} \cup \mathrm{Cut}_p)$ and define

$$\mathcal{I}_p := \{tv : \ t \in (0, t_c(v)), \ v \in S_pN\},$$

where t_c is the cut time function. Then

$$\exp_p : \mathcal{I}_p \to N_0$$

is a diffeomorphism; for a proof of these facts see Sakai (1996, chapter III, Lemma 4.4). The gradient of f on the full measure open set N_0 defines a vector field W that has unit norm and hence gives a smooth section $W : N_0 \to SN_0$. The vector field W has the property of being *geodesible*, i.e. its orbits are geodesics of g, or in other words $\nabla_W W = 0$, where ∇ is the Levi-Civita connection of g.

Exercise 3.8.10 Prove that $\nabla_W W = 0$.

For each $x \in N_0$, the Hessian of f at x, denoted by $\mathrm{Hess}_x(f)$, defines a bilinear form on T_xN. We shall consider its associated quadratic form for $v \in T_xN$ with unit norm, and we write this as $\mathrm{Hess}_x(f)(v, v)$. Moreover,

$$\mathrm{Hess}_x(f)(v, v) = \left.\frac{d^2}{dt^2}\right|_{t=0} f(\gamma_{x,v}(t)) = (X^2 f)(x, v),$$

where X is the geodesic vector field. In terms of the vector field W, we see right away that

$$\mathrm{Hess}_x(f)(v, v) = \langle \nabla_v W, v \rangle. \tag{3.24}$$

Exercise 3.8.11 Using that W is a gradient, show that $\langle \nabla_v W, w \rangle = \langle \nabla_w W, v \rangle$ for any $v, w \in T_xN$. In other words, the linear map $T_xN \ni v \mapsto \nabla_v W \in T_xN$ is symmetric.

In fact, since W has unit norm, $\langle \nabla_v W, W \rangle = 0$ for any $v \in T_xN$, and thus $\beta_x(v) := \nabla_v W$ defines a symmetric linear map $\beta_x : W(x)^\perp \to W(x)^\perp$.

Given $x \in N_0$ we define a subspace $E \subset T_{(x,W(x))}SN_0$ by setting

$$E(x, W(x)) := d\varphi_{f(x)}(\mathcal{V}(p, w)), \tag{3.25}$$

where $(p, w) = \varphi_{-f(x)}(x, W(x))$. The subspace E is a Lagrangian subspace in the kernel of the canonical contact form of SN with respect to the symplectic form given by (3.15) (since $\mathcal{V}$ is Lagrangian and $d\varphi_t$ preserves the symplectic form). Moreover, in terms of the horizontal and vertical splitting we may describe E as

$$E(x, W(x)) = \{(v, \nabla_v W) : v \in W(x)^\perp\}. \tag{3.26}$$

In other words E is the graph of the symmetric linear map β_x. To check the equality in (3.26), we proceed as follows. Fix $w \in S_p$ and $t < t_c(w)$. Let $x = \pi\varphi_t(p,w)$ so that $\varphi_t(p,w) = (x, W(x))$. Consider a curve $z\colon (-\varepsilon,\varepsilon) \to S_pM$ with $z(0) = w$, so that $\xi := \dot{z}(0) \in \mathcal{V}(p,w)$. We let J_ξ denote the normal Jacobi field with initial conditions determined by ξ as explained when discussing (3.21). Now write

$$\varphi_t(p, z(s)) = (\pi\varphi_t(p,z(s)), W(\pi\varphi_t(p,z(s))),$$

and differentiate this at $s = 0$ to obtain in terms of the vertical and horizontal splitting that

$$d\varphi_t(\xi) = (J_\xi(t), \nabla_{J_\xi(t)}W).$$

This gives (3.26) right away.

We wish to use the following well-known fact. We only sketch the proof leaving the details as exercise.

Proposition 3.8.12 *Let (N,g) be a complete Riemannian manifold. Take $x \neq y \in N$. Then the distance function d_g is smooth in a neighbourhood of (x,y) if and only if x and y are connected by a unique geodesic that is minimizing and free of conjugate points.*

Sketch If the condition on geodesics holds, write $d(x,y) = |\exp_x^{-1}(y)|$ and smoothness of d follows. For the converse fix x and set $f(y) := d(x,y)$. Then if f is differentiable at y and there is a unit speed minimizing geodesic γ connecting x to y, then $\nabla f(y)$ is the velocity vector of γ at y. If we have more than one minimizing geodesic the gradient would take two different values at the same point; absurd. For the conjugate points we have to go to the second derivatives of d and see that if x and y are conjugate along the unique minimizing geodesic joining them, then the Hessian blows up. □

Exercise 3.8.13 Complete the proof of Proposition 3.8.12.

Now we come to the main result of this subsection that completes the proof of Theorem 3.8.2.

Proposition 3.8.14 *Let (M,g) be a compact manifold with strictly convex boundary. Then M is simple if and only if the boundary distance function $d_g|_{\partial M\times\partial M}$ is smooth away from the diagonal.*

Proof Let (M,g) be a compact manifold with strictly convex boundary. We consider (M,g) isometrically embedded in a no return extension (N,g) as in Lemma 3.7.19.

https://doi.org/10.1017/9781009039901.006 Published online by Cambridge University Press

If M is simple, by Proposition 3.8.12 we know that the distance function d_g of N is smooth in a neighbourhood of $(x,y) \in \partial M \times \partial M$ for $x \neq y$. Hence its restriction to $\partial M \times \partial M$ is obviously smooth away from the diagonal.

The converse is more involved as we cannot use Proposition 3.8.12 directly since we are only assuming that the *restriction* to $\partial M \times \partial M$ is smooth away from the diagonal.

Take $x, y \in \partial M$ with $x \neq y$. We know (by strict convexity, see Proposition 3.7.21) that there is a minimizing geodesic between x and y. We claim there is only one. Let $f(z) = d(x,z)$ for $z \in M$ and let $h := f|_{\partial M}$. We know that h is C^1 (away from x). Thus if $\gamma\colon [0,\ell] \to M$ is a unit speed length minimizing geodesic joining x and y, then $\nabla h(y)$ is the orthogonal projection of $\dot\gamma(\ell)$ onto $T_y\partial M$. Indeed f is always C^1 on the interior of γ and

$$\nabla h(y) = \text{projection}\left(\lim_{t\to\ell^-} \nabla f(\gamma(t))\right).$$

This shows that the minimizing geodesic between x and y is unique.

Let $\mathcal{O}_x$ be the open set in S_xM given by those unit vectors pointing strictly inside M and consider the map $F\colon \partial M \setminus \{x\} \to \mathcal{O}_x$, where $F(y)$ is the initial velocity vector of the (unique) minimizing geodesic from x to y. This map is continuous and injective and by topological considerations it must also be onto.

Exercise 3.8.15 Prove that F is surjective.

Thus every $v \in \mathcal{O}_x$ is the initial velocity of some minimizing geodesic hitting the boundary. In particular, this implies that any geodesic starting on the boundary and ending in the interior is minimizing and has no conjugate points.

The next step is to show that (M,g) is non-trapping. Indeed let $p \in M$ be an interior point. Consider the set of all geodesics that start at p and hit the boundary. The set of their initial directions is open and closed (from minimality and transversality to the boundary due to strict convexity), hence it must be all S_pM.

The final step in the proof is to show that there are no conjugate points on the boundary. For this we will use the previous discussion on the Hessian of the distance function.

Let $p \in \partial M$ and consider as above $f(x) = d(p,x)$. We have seen that the interior of M is contained in N_0. Take $y \in \partial M$ and suppose that p and y are conjugate. Consider a sequence of points y_n along the unique minimizing geodesic connecting p to y such that they are in the interior of M, but $y_n \to y$. Using (3.25) we see that $E(y_n, W(y_n))$ converges to a Lagrangian subspace at $(y, W(y))$ that intersects the vertical subspace non-trivially (note that W

is defined at y). This in turn implies that there is a sequence of unit vectors $v_n \in W(y_n)^\perp$, such that $v_n \to v \in W(y)^\perp$ for which $\langle \nabla_{v_n} W, v_n \rangle \to \infty$. Going back to (3.24) we see that $\mathrm{Hess}_{y_n}(f)$ blows up as $y_n \to y$.

We are assuming that $h = f|_{\partial M}$ is smooth away from p, so to derive a contradiction from the blow-up of the Hessian of f we need to observe that $\nabla_W W = 0$, and thus $\mathrm{Hess}_{y_n}(f)(W(y_n), w) = 0$ for any $w \in T_{y_n} M$. But $W(y)$ is transversal to ∂M and thus the blow-up of the Hessian of h at y also holds contradicting the fact that h must be C^2 near y. $\square$