CHAPTER 6

Boundary control method

1. Brief introduction to the boundary control method

1.1. Wave equation and Gel'fand inverse problem. Let $\mathcal{N}$ be an n-dimensional complete connected Riemannian manifold with boundary $\partial\mathcal{N}$. We shall consider an IBVP (initial-boundary value problem) for the wave equation

$$\partial_t^2 u = \Delta_g u \quad \text{on} \quad \mathcal{N} \times (0, \infty),$$

where Δ_g is the Laplace-Beltrami operator. In local coordinates

$$\Delta_g = g^{-1/2}\partial_i(g^{ij}g^{1/2}\partial_j), \quad g = \det(g_{ij}).$$

We impose the initial condition

$$u\big|_{t=0} = \partial_t u\big|_{t=0} = 0,$$

and the boundary condition

$$\partial_\nu u\big|_{\partial\mathcal{N}\times(0,\infty)} = f \in C_0^\infty(\partial\mathcal{N} \times (0, \infty)).$$

Here ν is the outer unit normal to $\partial\mathcal{N}$. Let $u^f(x,t)$ be the solution to the above IBVP. We measure u^f on $\partial\mathcal{N} \times (0, \infty)$, and call

$$\Lambda^h : f \to u^f\big|_{\partial\mathcal{N}\times(0,\infty)} \tag{1.1}$$

a *hyperbolic Neumann-to-Dirichlet map.* The basic question we address is the following one.

Question Assume we know Λ^h. Can we determine $(\mathcal{N}, g)$, i.e. the manifold $\mathcal{N}$ and the metric g?

This is the *Gel'fand inverse problem* (stated in a slightly different form, [**37**]). Note that Λ^h is an operator defined on $\partial\mathcal{N} \times (0, \infty)$. Starting from the knowledge on $\partial\mathcal{N} \times (0, \infty)$, the first issue is the topology of $\mathcal{N}$, and the second issue is the Riemannian structure.

The answer to the above question is affirmative when $\mathcal{N}$ is compact, and also for non-compact $\mathcal{N}$ with some additional geometric assumption. To fix the idea, in this chapter, $\mathcal{N}$ means either any compact connected Riemannian manifold with boundary, or when dealing with the non-compact case, the manifold Ω^c discussed in Chap. 5, §4. However, the arguments given below also work for non-compact manifolds possesing the spectral representation as in the case of Ω^c. Note that in both cases $\partial\mathcal{N}$ is compact.

1.2. Spectral formulation. Let us begin with the compact manifold case. Consider the Neumann Laplacian H^N:

$$H^N u = -\Delta_g u, \quad u \in H^2(\mathcal{N}), \quad \partial_\nu u\big|_{\partial\mathcal{N}} = 0.$$

The spectrum of H^N consists of real numbers

$$0 = \lambda_1 < \lambda_2 \le \lambda_3 \le \cdots \to \infty.$$

Let φ_k be the associated eigenvectors

$$-\Delta_g \varphi_k = \lambda_k \varphi_k, \quad \partial_\nu \varphi_k\big|_{\partial\mathcal{N}} = 0.$$

Without loss of generality we can assume φ_k to be real-valued. The set $\{\varphi_k\}_{k=1}^\infty$ can be made to form an orthonormal basis in $L^2(\mathcal{N})$ and orthogonal basis in $H^1(\mathcal{N})$, where the inner products of $L^2(\mathcal{N})$ and $H^1(\mathcal{N})$ are defined by

$$(f,g)_{L^2(\mathcal{N})} = \int_{\mathcal{N}} f(x)\overline{g(x)} dV_g, \quad dV_g = g^{1/2} dx^1 \cdots dx^n,$$

$$(f,g)_{H^1(\mathcal{N})} = \int_{\mathcal{N}} g^{ij}\, \partial_i f\, \overline{\partial_j g}\, dV_g + (f,g)_{L^2}.$$

We call $\big\{\big(\lambda_k, \varphi_k\big|_{\partial\mathcal{N}}\big)\big\}_{k=1}^\infty$ the *boundary spectral data* (BSD). The original Gel'fand inverse problem is equivalent to:

Question Given BSD, can we determine $(\mathcal{N}, g)$?

The relation of BSD to the hyperbolic Neumann-to-Dirichlet map is represented by the following (formal) formula:

$$\left(\Lambda^h f\right)(x,t) = \int_{\partial\mathcal{N}} \int_{\mathbf{R}_+} G(x,y,t-s) f(y,s)\, dS_y\, ds.$$

$$G(x,y,t) = \sum_{k=1}^{\infty} \frac{\sin(\sqrt{\lambda_k} t)}{\sqrt{\lambda_k}} \varphi_k(x)\varphi_k(y)\big|_{\partial\mathcal{N}\times\partial\mathcal{N}}. \tag{1.2}$$

One can also deal with the Dirichlet Laplacian, i.e.

$$H^D u = -\Delta_g u, \quad u \in H^2(\mathcal{N}) \cap H_0^1(\mathcal{N}).$$

Let $0 < \mu_1 < \mu_2 \le \mu_3 \le \cdots \to \infty$ be the Dirichlet eigenvalues, and ψ_k the associated eigenvectors. Considering IBVP

$$\begin{cases} \partial_t^2 w = \Delta_g w, \\ w\big|_{\partial\mathcal{N}\times\mathbf{R}_+} = f \in C_0^\infty(\partial\mathcal{N}\times\mathbf{R}_+), \\ w\big|_{t=0} = \partial_t w\big|_{t=0} = 0, \end{cases}$$

we define the hyperbolic Dirichlet-to-Neumann map by

$$R^h f : f \to \partial_\nu w^f\big|_{\partial\mathcal{N}\times\mathbf{R}_+}.$$

The integral kernel of R^h is formally written as

$$R^h(x,y,t) = \sum_{k=1}^{\infty} \frac{\sin(\sqrt{\mu_k} t)}{\sqrt{\mu_k}} \partial_\nu\psi_k(x)\partial_\nu\psi_k(y)\big|_{\partial\mathcal{N}\times\partial\mathcal{N}}.$$

The method we are going to talk about is called the *Boundary Control* (BC) method, whose history goes back to the famous results by M. G. Krein, in the mid-fifties, on the 1−dimensional inverse scattering theory ([**83**], [**84**]). Compared with the fundamental methods by Gel'fand-Levitan and Marchenko, the method of Krein is distinguished by the systematic use of the finite propagation speed for the wave equation. However, the ideas based upon the domain of influence, etc. coming from this finite velocity are "disguised" in the work of Krein due to their formulation in the frequency domain (or the stationary equation), where they turn out to be conditions on analyticity of the corresponding Fourier transform of the solution. This principal hyperbolic nature of Krein's method was revealed by Blagovestchenskii who was working in the time-domain (or the time-dependnet equation) using the finite velocity of the wave propagation and ideas of controllability in the filled domain to derive a Volterra-type equation for unknown functions ([**18**]). These ideas have become crucial for the extension of the method to multidimensions pioneered by Belishev [**10**], see also [**77**]. One more important ingredient of the BC-method, namely, the possibility to evaluate the inner product of waves sent into $\mathcal{N}$ from $\partial\mathcal{N}$ also goes back to the 1-dimensional case to the work of Blagovestchenskii [**19**]. See [**12**] for the multidimensional case.

The BC method has the following features.

(1) BC method is hyperbolic.

Since the propagation speed of wave motion is finite, and singularities of waves are related with geodesics, this implies the close connection of BC method with geometry.

(2) BC method is not perturbative.

We do not assume that the given metric is close to some standard one. In this sense, the BC method does not have the character of perturbation theory.

1.3. Outline of the procedure. The crucial tool of the BC-method is the Kuratowski space of boundary distance functions $R(\mathcal{N})$ to be defined in §5, and the reconstruction of the manifold $\mathcal{N}$ is done by the following 3 steps :

- In §8, we show that BSP determines $R(\mathcal{N})$.
- In §5, we show that $R(\mathcal{N})$ is topologically isomorphic to $\mathcal{N}$.
- In §7, we show that $R(\mathcal{N})$ determines the Riemannian metric of $\mathcal{N}$.

This is an effective interplay of linear partial differential equations and geometry. The main ingredients of the 1st step are Blagovestchenskii's idenitity, which represents the solution of the initial boundary value problem (IBVP) of the wave equation by BSD, and Tataru's uniqueness theorem, which guarantees the conrollablity of IBVP. The 2nd step is of the character of general topology. The 3rd step is purely from differential geometry, in which the coordinate system of $\mathcal{N}$ is constructed by $R(\mathcal{N})$ and the metric tensor is computed. The analytic and geometric preliminaries are done in §2, §4, and in §5, §6, respectively.

2. Blagovestchenskii idenitity

Given a solution u^f of the wave equation

$$
\begin{cases}
\partial_t^2 u = \Delta_g u, \\
\partial_\nu u\big|_{\partial\mathcal{N}\times\mathbf{R}_+} = f, \\
u\big|_{t=0} = \partial_t u\Big|_{t=0} = 0,
\end{cases}
\tag{2.1}
$$

we expand it by eigenvectors to get

$$u^f(x,t) = \sum_k u_k^f(t)\varphi_k(x), \quad u_k^f(t) = \int_{\mathcal{N}} u^f(y,t)\varphi_k(y)dV_g.$$

Then we have

$$\begin{aligned}\frac{d^2}{dt^2}u_k^f(t) &= \int_{\mathcal{N}} \Delta_g u^f(y,t)\varphi_k(y)dV_g \\ &= \int_{\partial\mathcal{N}} \left[\partial_\nu u^f \varphi_k - u^f \partial_\nu \varphi_k\right] dS_g + \int_{\mathcal{N}} u^f \Delta_g \varphi_k dV_g \\ &= \int_{\partial\mathcal{N}} f(y,t)\varphi_k(y,t)dS_g - \lambda_k \int_{\mathcal{N}} u^f(y,t)\varphi_k(y)dV_g.\end{aligned}$$

We have thus derived

$$\frac{d^2}{dt^2}u_k^f(t) + \lambda_k u_k^f(t) = \int_{\partial\mathcal{N}} f(y,t)\varphi_k(y)dS_g,$$

and, due to the initial condition in IBVP,

$$u_k^f(0) = \frac{d}{dt}u_k^f(0) = 0.$$

Solving this differential equation, we obtain *Blagovestchenskii idenitity*

$$u_k^f(t) = \int_0^t ds \int_{\partial\mathcal{N}} dS_g \frac{\sin(\sqrt{\lambda_k}(t-s))}{\sqrt{\lambda_k}} f(y,s)\varphi_k(y). \tag{2.2}$$

This formula shows that $u_k^f(t)$ is represented by λ_k and $\varphi_k\big|_{\partial\mathcal{N}}$, i.e. BSD.

Lemma 2.1. *The following holds:*

$$(u^f(t), u^h(s)) = \sum_k u_k^f(t)\overline{u_k^h(s)}, \tag{2.3}$$

i.e. BSP determines the inner product $(u^f(t), u^h(s))_{L^2(\mathcal{N})}$, $\forall t, s \in \mathbf{R}$, $\forall f, h \in C_0^\infty(\partial\mathcal{N} \times \mathbf{R}_+)$.

Proof. This follows from (2.2) and the Parseval formula. □

Lemma 2.1 is the first corner-stone of BC method. We let

$$S(t,\lambda) = \frac{\sin(\sqrt{\lambda}t)}{\sqrt{\lambda}}, \quad \widetilde{S}(t,s,\lambda) = S(t,\lambda)S(s,\lambda), \tag{2.4}$$

and use the notation in Chap. 5, §3 to rewrite the right-hand side of (2.3) as

$$\sum_i \int_0^t \int_0^s dt'ds'\widetilde{S}(t-t', s-s', \lambda_i)\left(\delta_\Gamma^* P_i \delta_\Gamma f(t'), h(s')\right). \tag{2.5}$$

This implies the following corollary.

Corollary 2.2. *The inner product* $(u^f(t), u^h(s))$ *is written only by BSP.*

This is also true when $-\Delta_g$ has the continuous spectrum. Recall that in §4 of Chap. 5, the Laplace-Beltrami operator on Ω^c admits the spectral representation

$\mathcal{F}_c^{(+)}$. In this case, to modify the formula (2.3), we have only to add the integral of $\mathcal{F}_c^{(+)}(k)^*\mathcal{F}_c^{(+)}(k)$ to the right-hand side of (2.5):

$$\begin{aligned}(2.6)\qquad &\int_0^\infty dk \int_0^t\int_0^s dt'ds'\, \widetilde{S}(t-t's-s',k^2)\left(\delta_\Gamma^*\mathcal{F}_c^{(+)}(k)^*\mathcal{F}_c^{(+)}(k)\delta_\Gamma f(t'),h(s')\right)\\ &+\sum_i\int_0^t\int_0^s dt'ds'\widetilde{S}(t-t',s-s',\lambda_i)\,(\delta_\Gamma^* P_i^c\delta_\Gamma f(t'),h(s'))\,.\end{aligned}$$

Again $(u^f(t),u^h(s))$ is written only by BSP.

Let us remark that in [**77**], p. 214, Lemma 4.9, it is shown that one can construct BSD from BSP up to a multiplication factor if $\mathcal{N}$ is compact.

3. Geodesics

Let us recall some basic notions from Riemannian geometry. The distance of two points x,y of a Riemannian manifold $\mathcal{N}$, denoted by $d(x,y)$, is defined by the infimum of length of piecewise smooth curves joining x and y. This makes $\mathcal{N}$ a metric space. If $\mathcal{N}$ is complete in this metric, it is said to be *metrically complete.* When $\partial\mathcal{N}=\emptyset$, by the theorem of Hopf-Rinow (see e.g. [**36**], pp. 94, 95), it is equivalent to that $\mathcal{N}$ is *geodesically complete*, i.e. any solution of the equation of geodesics can be extended onto the whole line $\mathbf{R}$. In this case, again by the theorem of Hopf-Rinow, any two points in $\mathcal{N}$ can be joined by the minimal geodesic (i.e. the shortest curve).

In local coordinates, the equation of geodesics is written as

$$(3.1)\qquad \frac{d^2x^k}{dt^2}+\Gamma^k_{ij}(x)\frac{dx^i}{dt}\frac{dx^j}{dt}=0.$$

Let $x(t,y,v)$ be the solution of (3.1) satisfying

$$x(0,y,v)=y,\quad \partial_t x(0,y,v)=v\in T_y(\mathcal{N}),$$

where $\partial_t=d/dt$ and $T_y(\mathcal{N})$ is the tangent space at y. Let $|v|_g$ be the length of $v\in T_y(\mathcal{N})$. Then the map defined by

$$(3.2)\qquad \exp_y(v):T_y(\mathcal{N})\ni v\to x(1,y,v)=x(|v|_g,y,\hat{v})\in\mathcal{N},\quad \hat{v}=v/|v|_g$$

is called the *exponential map.* Using this exponential map, we define the *Riemannian normal coordinates* centered at y in the following way. Let $B_{y,\rho}=\{v\in T_y(\mathcal{N});|v|_g<\rho\}$. Then for ρ sufficiently small, the map

$$\exp_y:B_{y,\rho}\ni v\to\exp_y(v)\in\exp(B_{y,\rho})\subset\mathcal{N}$$

is a diffeomorphism. Hence $v=(v_1,\cdots,v_n)$ can be used as local coordinates on $\exp_y(B_{y,\rho})$. Note that (3.2) implies that, when dealing with geodesics $x(t,y,v)$, we can always parametrize them so that $|v|_g=1$. This parametrization is called the arclength parametrization and will be always used in this chapter.

Almost all of the notions from Riemannian geometry can be extended to the manifold with boundary by obvious changes. The problem of the existence of the shortest curves, however, is delicate. Think of, for example, non-convex domains in $\mathbf{R}^n$. However, for any $x,y\in\mathcal{N}$, there exists a shortest curve, which is C^1-smooth. See e.g. [**4**]. Moreover, the segments of this curve lying inside $\mathcal{N}$ are (minimal) geodesics in $\mathcal{N}$, while the segments of this curve lying on $\partial\mathcal{N}$ are minimal geodesics on $\partial\mathcal{N}$.

The following lemma is easy to prove. Let $d(x, y)$ be the distance between x and y with respect to the Riemannian metric g, and for a subset $S \subset \mathcal{N}$, $d(x, S) = \inf\{d(x, y)\,;\, y \in S\}$.

Lemma 3.1. *For any $x \in \mathcal{N}$, there exists $z \in \partial\mathcal{N}$ such that $d(x, z) = d(x, \partial\mathcal{N})$. Moreover $x = \gamma_z(s)$, where γ_z is the geodesic starting from z with initial direction the inner unit normal to $\partial\mathcal{N}$, and $s = d(x, z)$.*

4. Controllabilty and observability

Two notions in the title of this section are fundamental concepts in control theory. They are related to properties of solution operators of dynamical problems.

4.1. Domains of influence. For any set $A \subset \mathcal{N}$ and $t_0 > 0$, we define the *domain of influence* of A (at time t_0) by

$$\mathcal{N}(A, t_0) = \{x \in \mathcal{N}\,;\, d(x, A) \le t_0\}.$$

We introduce the forward, $D_+(A, t_0)$, backward, $D_-(A, t_0)$, and double cones, $D(A, t_0)$, of dependence by

$$\begin{aligned} D_\pm(A, t_0) &= \{(x, t)\,;\, x \in \mathcal{N}(A, t_0 \mp t),\ 0 \le \pm t \le t_0\}, \\ D(A, t_0) &= D_+(A, t_0) \cup D_-(A, t_0). \end{aligned}$$

Lemma 4.1. *Take $t_0 > 0$ and a bounded open set $A \subset \mathcal{N}$ arbitrarily. Let u be a solution to the initial boundary value problem*

$$\begin{cases} \partial_t^2 u = \Delta_g u, & \text{in} \quad \mathcal{N} \times \mathbf{R}, \\ u = \partial_t u = 0, & \text{on} \quad \mathcal{N}(A, t_0) \quad \text{at} \quad t = 0, \\ \partial_\nu u = 0, & \text{on} \quad D(A, t_0) \cap (\partial\mathcal{N} \times \mathbf{R}). \end{cases} \tag{4.1}$$

Then $u = 0$ in $D(A, t_0)$.

Proof. We prove this lemma in the case when $\mathcal{N}$ is a domain in $\mathbf{R}^n$ and, due to symmetry $t \to -t$, for $t > 0$. The general case can be proved in the same way by taking local coordinates.

First we recall the well-known energy inequality. Note the identity:

$$\begin{aligned} &\frac{1}{2}\partial_t\left((\partial_t v)^2 + g^{ij}\partial_i v \partial_j v\right) - \frac{1}{\sqrt{g}}\partial_i\left(\sqrt{g} g^{ij}\partial_j v \partial_t v\right) \\ &= \Big(\partial_t^2 v - \frac{1}{\sqrt{g}}\partial_i(\sqrt{g} g^{ij}\partial_j v)\Big)\partial_t v, \end{aligned} \tag{4.2}$$

where $\partial_t = \partial/\partial t$, $\partial_i = \partial/\partial x^i$. Take a time interval I = [0,T], a family of connected open sets $A(t) \subset \mathbf{R}^n$ $(t \in I)$ and consider a domain $D(T) \subset \mathbf{R}^n \times \mathbf{R}^1$ such that

$$D(T) = \{(x, t)\,;\, t \in I,\ x \in A(t)\}.$$

Then $\partial D(t)$ consists of 3 parts:

$$\partial D(T) = A(T) \cup A(0) \cup S,$$

where the lateral boundary S consists of 2 parts:

$$S = S_\partial \cup S_r, \quad S_\partial = \overline{D(T)} \cap \big(\partial\mathcal{N} \times [0, T]\big), \quad S_r = S \setminus S_\partial. \tag{4.3}$$

Assume that S_r is piecewise smooth and its unit normal $n = (n_1, \cdots, n_n, n_t)$, with respect to the Euclidean metric, has the property

$$n_t \geq \left(g^{ij} n_i n_j\right)^{1/2}, \quad \text{on} \quad S_r. \tag{4.4}$$

Suppose that a real-valued function $v = v(x,t)$ satisfies the wave equation

$$\begin{cases} \partial_t^2 v - \dfrac{1}{\sqrt{g}} \partial_i (\sqrt{g} g^{ij} \partial_j v) = 0, & \text{in} \quad D(T), \\ \partial_\nu v = 0, \quad \text{on} \quad S_r. \end{cases} \tag{4.5}$$

Mutilplying (4.2) by $\sqrt{g}$ and integrating on $D(T)$, we have

$$\begin{aligned} &\frac{1}{2}\Big[\int_{A(t)} \left((\partial_t v)^2 + g^{ij}\partial_i v \partial_j v\right) \sqrt{g}\, dx\Big]_{t=0}^{t=T} \\ &= -\frac{1}{2}\int_{S_r} n_t \left((\partial_t v)^2 + g^{ij}\partial_i v \partial_j v\right)\sqrt{g}\, dS + \int_{S_r} n_i g^{ij} \partial_j v \partial_t v \sqrt{g}\, dS, \end{aligned} \tag{4.6}$$

where the integral over S_∂ disappears due to the boundary condition in (4.5) and $n_t = 0$ on S_∂. The right-hand side is non-positive by (4.4), estimate

$$|n_i g^{ij} \partial_j v \partial_t v| \leq |\partial_t v| (g^{ij} n_i n_j)^{1/2} (g^{ij} \partial_i \partial_j)^{1/2} \leq n_t |\partial_t v| (g^{ij} \partial_i \partial_j)^{1/2}$$

and the Cauchy-Schwarz inequality. This implies

$$\int_{A(T)} \left((\partial_t v)^2 + g^{ij}\partial_i v \partial_j v\right)\sqrt{g}\, dx \leq \int_{A(0)} \left((\partial_t v)^2 + g^{ij}\partial_i v \partial_j v\right)\sqrt{g}\, dx.$$

This holds with T relplaced by $\tau \in (0,T)$. Therefore, if $v\big|_{t=0} = \partial_t v\big|_{t=0} = 0$ on $A(0)$, we have $\nabla v\big|_{t=\tau} = 0, \partial_t v\big|_{t=\tau} = 0$ on $A(\tau)$, hence $v = 0$ on $D(T)$.

We turn to the proof of Lemma 4.1. In the following, C_0 and C denote constants independent of small $\epsilon > 0$ and $j = (j_1, \cdots, j_n) \in \mathbf{Z}^n$.

For a small $\epsilon > 0$, we take lattice points $P(j,\epsilon) = (j_1\epsilon/C_0, \cdots, j_n\epsilon/C_0)$, where C_0 is a large constant. We extend $\left(g^{\alpha\beta}(x)\right)$ smoothly outside $\mathcal{N}$, and put

$$G^{(j,\epsilon)} = \Big(g^{\alpha\beta}(P(j,\epsilon))\Big) + \epsilon C_0 I_n, \tag{4.7}$$

I_n being the $n \times n$ identity matrix. Letting $d_{j,\epsilon}(\cdot,\cdot)$ be the distance defined by the Riemannian metric $G_{j,\epsilon} = (G^{(j,\epsilon)})^{-1}$, we put

$$B(j,\epsilon) = \{x \in \mathbf{R}^n \,;\, d_{j,\epsilon}(x, P(j,\epsilon)) \leq \epsilon\}.$$

We also let

$$\mathcal{N}_\epsilon(A,t_0) = \{x \in \mathcal{N}(A,t_0)\,;\, d(x, \partial\mathcal{N}(A,t_0)) > \epsilon\},$$

where $d(\cdot,\cdot)$ is the distance defined by the Riemannian metric $\left(g_{\alpha\beta}(x)\right)$. Then $\mathcal{N}_\epsilon(A,t_0) \subset \mathcal{N}(A,t_0)$ and $\mathcal{N}_\epsilon(A,t_0) \to \mathcal{N}(A,t_0)$ as $\epsilon \to 0$.

We now consider a finite set

$$J(\epsilon) = \{j\,;\, P(j,\epsilon) \in \mathcal{N}_\epsilon(A,t_0)\},$$

and for $j \in J(\epsilon)$, we put

$$D(j,\epsilon) = \Big\{(x,t)\,;\, x \in \mathcal{N},\ d_{j,\epsilon}(x, P(j,\epsilon)) \leq \epsilon + t,\ 0 \leq t \leq \epsilon/C_0\Big\}.$$

As above, its lateral boundary consists of 2 parts like (4.3). We show that the condition (4.4) is satisfied on S_r.

For the sake of simplicity, we assume that $P(j,\epsilon) = 0$. The lateral boundary is defined as the zeros of

$$\varphi(x,t) = \epsilon + t - (G_{j,\epsilon}x, x)^{1/2}.$$

Since the Euclidean normal unit of the lateral boundary is given by $(\nabla_x\varphi, \partial_t\varphi)/(|\nabla_x\varphi|^2 + (\partial_t\varphi)^2)^{1/2}$, we have only to show that for any x on the lateral boundary

$$1 \geq (G(x)^{-1}\nabla_x\varphi, \nabla_x\varphi), \quad G(x) = \big(g_{\alpha\beta}(x)\big). \tag{4.8}$$

Let $G_0 = G(P(j,\epsilon))$. Then $G_{j,\epsilon} = (G_0^{-1} + \epsilon C_0)^{-1}$ and $G(x) = G_0 + O(\epsilon)$. Since $\nabla_x\varphi = -G_{j,\epsilon}x/(G_{j,\epsilon}x, x)^{1/2}$, we have

$$(G(x)^{-1}\nabla_x\varphi, \nabla_x\varphi) = \frac{(G_0^{-1}G_{j,\epsilon}x, G_{j,\epsilon}x)}{(G_{j,\epsilon}x, x)} + \frac{(O(\epsilon)G_{j,\epsilon}x, G_{j,\epsilon}x)}{(G_{j,\epsilon}x, x)}. \tag{4.9}$$

In the right-hand side, G_0 and $G_{j,\epsilon}$ are positive definite, and $O(\epsilon)$ is symmetric. Noting that

$$\sqrt{G_{j,\epsilon}}\, O(\epsilon) \sqrt{G_{j,\epsilon}} \leq \epsilon C_1$$

for some constant $C_1 > 0$, we see that

$$\frac{(O(\epsilon)G_{j,\epsilon}x, G_{j,\epsilon}x)}{(G_{j,\epsilon}x, x)} \leq \epsilon C_1. \tag{4.10}$$

To compute the 1st term of the right-hand side of (4.9), we first note $G_0^{-1}G_{j,\epsilon} = (1 + \epsilon C_0 G_0)^{-1}$. Letting λ_1 be the smallest eignvalue of G_0, we have

$$(1 + \epsilon C_0 G_0)^{-1} \leq (1 + \epsilon C_0 \lambda_1)^{-1}.$$

Then, letting $y = \sqrt{G_{j,\epsilon}}x$, and noting that G_0 and $G_{j,\epsilon}$ commute, we can estimate the 1st term as

$$\frac{((1 + \epsilon C_0 G_0)^{-1}y, y)}{(y, y)} \leq \frac{1}{1 + \epsilon C_0 \lambda_1}. \tag{4.11}$$

In view of (4.10) and (4.11), taking C_0 large enough, we see that (4.8) is satisfied.

We now put

$$D_1(\epsilon) = \bigcup_{j \in J(\epsilon)} D(j,\epsilon), \tag{4.12}$$

and apply the energy inequality to have

$$u = 0, \quad \text{in} \quad D_1(\epsilon). \tag{4.13}$$

Let $D(A, t_0, \tau)$ be the section of $D(A, t_0)$ at time $t = \tau$. We also let $\Sigma_1^{high}(\tau)$ be the boundary of the section of $D_1(\epsilon)$ at time $t = \tau$, and $\Sigma_1^{low}(\tau)$ be the surface such that

$$\begin{cases} \Sigma_1^{low}(\tau) \ \sqsupset \ \Sigma_1^{high}(\tau), \\ d(\Sigma_1^{low}(\tau), \Sigma_1^{high}(\tau)) = 2\epsilon + C\epsilon\tau, \end{cases} \tag{4.14}$$

where for 2 compact surfaces S_1 and S_2, $S_1 \sqsupset S_2$ (or $S_2 \sqsubset S_1$) means that S_2 is contained in the bounded domain with boundary S_1, and where C is chosen large enough.

The meaning of (4.14) is as follows. At time $t = 0$, we take the surface $\Sigma_1^{high}(0)$ and $\Sigma_1^{low}(0)$ inside and outside of $\partial D(A, t_0)$ with distance ϵ. We then develop them by speeds higher or lower than that of waves. At time t, the distance between $\Sigma_1^{high}(t)$ and $\Sigma_1^{low}(t)$ will increase at most by $C\epsilon t$.

Let $\Sigma(\tau)$ be the boundary of $D(A, t_0, \tau)$. Then we have

$$\Sigma_1^{high}(t) \sqsubset \Sigma(t) \sqsubset \Sigma_1^{low}(t), \quad 0 \le t \le \epsilon/C_0. \tag{4.15}$$

The next step starts from the time $t = \epsilon/C_0$ instead of $t = 0$, and $D_1(\epsilon)$ instead of $D(A, t_0)$. One can then construct $D_2(\epsilon)$ and $\Sigma_2^{high}(t)$ as above for $\epsilon/C_0 \le t \le 2\epsilon/C_0$. Then by the energy inequality

$$u = 0, \quad \text{in} \quad D_2(\epsilon), \tag{4.16}$$

for the time interval $\epsilon/C_0 \le t \le 2\epsilon/C_0$. The surface $\Sigma_2^{low}(\tau)$ is defined by

$$\begin{cases} \Sigma_2^{low}(\tau) \sqsupset \Sigma_2^{high}(\tau), \\ d(\Sigma_2^{low}(\tau), \Sigma_2^{high}(\tau)) = 2\epsilon + \dfrac{C}{C_0}\epsilon^2 + C\epsilon(\tau - \dfrac{\epsilon}{C_0}). \end{cases} \tag{4.17}$$

We continue this procedure. In the k-th step, we obtain

$$u = 0, \quad \text{in} \quad D_k(\epsilon), \tag{4.18}$$

in the time interval $(k-1)\epsilon/C_0 \le t \le k\epsilon/C_0$, and

$$\begin{cases} \Sigma_k^{low}(\tau) \sqsupset \Sigma_k^{high}(\tau), \\ d(\Sigma_k^{low}(\tau), \Sigma_k^{high}(\tau)) = 2\epsilon + \dfrac{C}{C_0}(k-1)\epsilon^2 + C\epsilon(\tau - \dfrac{(k-1)}{C_0}\epsilon). \end{cases} \tag{4.19}$$

Now, with a given time $t_0 > 0$ and a large number N, we take ϵ as $N\epsilon/C_0 = t_0$. We put

$$D_N = \cup_{j=1}^N D_k(\epsilon).$$

Then, by the above consideration,

$$u = 0, \quad \text{in} \quad D_N.$$

By our construction, $D_N \subset D(A, t_0)$. When $N \to \infty$, $D(N)$ tends to $D(A, t_0)$. In fact, by (4.19) and $N\epsilon = t_0$,

$$d(\Sigma_k^{low}(\tau), \Sigma_k^{high}(\tau)) \le (2 + Ct_0)\epsilon \to 0.$$

This proves Lemma 4.1. □

In the proof this lemma, we follow the basic steps of Theorem IV 2.2 of [**89**], making them more precise by taking into the account the variable velocity of the wave propagation.

Using Lemma 4.1, we can describe the support of the waves generated by the Neumann boundary sources, namely the solution u^f of the IBVP,

$$\begin{cases} \partial_t^2 u = \Delta_g u, \quad \text{in} \quad \mathcal{N} \times (0, \infty) \\ u\big|_{t=0} = \partial_t u\big|_{t=0} = 0, \quad \text{on} \quad \mathcal{N}, \\ \partial_\nu u\big|_{\partial\mathcal{N} \times (0,\infty)} = f \in C_0^\infty(\partial\mathcal{N} \times (0, \infty)). \end{cases} \tag{4.20}$$

To this end, for any subset $A \subset \mathcal{N}$, we introduce the forward, $C_+(A)$, backward, $C_-(A)$, and the double, $C(A)$, cones of influence

$$\begin{aligned} C_\pm(A) &= \{(x, t)\,;\, d(x, A) \le \pm t,\ \pm t > 0\}, \\ C(A) &= C_+(A) \cup C_-(A). \end{aligned} \tag{4.21}$$

Corollary 4.2. *Let u^f be the solution to IBVP (4.20). Let, in addition, $\operatorname{supp} f \subset S \times (0,\infty)$, where $S \subset \partial\mathcal{N}$ is open. Then*

$$\operatorname{supp} u^f \subset C_+(S).$$

Proof. Let $t_0 > 0$ and $(y_0, t_0) \notin C_+(S)$, then for small $r > 0$,

$$\{(x,t)\,;\ x \in \mathcal{N},\ d(x,S) \le t,\ 0 \le t \le t_0\} \cap D(B_r(y_0), t_0) = \emptyset,$$

$B_r(y_0)$ being the ball of radius $r > 0$ centered at y_0. Applying Lemma 6.4.1, we have $u^f(y_0, t_0) = 0$. To complete the proof, just note that for $t < 0$, $u^f(x,t) = 0$. □

4.2. Unique continuation and controllabilty. Next we describe the properties of $u^f(\cdot, t)$ in $\mathcal{N}(S,t)$, when $\operatorname{supp} f \subset S \times (0,\infty)$. We start with the following global uniqueness theorem which is essentially due to Tataru ([**125**]).

Theorem 4.3. *Let $u \in H^1_{loc}(\mathcal{N} \times (-t_0, t_0))$ satisfies*

$$\begin{cases} \partial_t^2 u = \Delta_g u \quad \text{in} \quad \mathcal{N} \times (-t_0, t_0), \\ \partial_\nu u\big|_{\partial\mathcal{N} \times (-t_0,t_0)} = 0, \quad u\big|_{S \times (-t_0,t_0)} = 0. \end{cases} \tag{4.22}$$

Then $u = 0$ in $D(S, t_0)$.

For a measurabe subset $D \subset \mathcal{N}$ and $v \in L^2(D)$, we define $v = 0$ on $\mathcal{N} \setminus D$ and regard $L^2(D)$ as a closed subspace of $L^2(\mathcal{N})$.

Corollary 4.4. *Assume v satisfies*

$$\begin{cases} \partial_t^2 v = \Delta_g v \quad \text{in} \quad \mathcal{N} \times \mathbf{R}, \\ v\big|_{t=t_0} = 0, \quad \partial_t v\big|_{t=t_0} =: \psi \in L^2(\mathcal{N}(S,t_0)), \\ \partial_\nu v\big|_{\partial\mathcal{N} \times (0,t_0)} = 0, \quad v\big|_{S \times (0,t_0)} = 0. \end{cases} \tag{4.23}$$

Then $\partial_t v\big|_{t=t_0} = 0$.

Proof. We extend $v(t)$ on the time interval $(t_0, 2t_0)$ by $v(t) = -v(2t_0 - t)$, and put $w(t) = v(t - t_0)$. Then w satisfies the conditions in Theorem 4.3. □

Corollary 4.4 shows the usefulness of the notion of the *observability operator*,

$$\mathcal{O}^S_{t_0} : L^2(\mathcal{N}(S,t_0)) \ni \psi \to v^\psi\big|_{S \times (0,t_0)} \in L^2(S \times (0,t_0)),$$

where v^ψ satisfies

$$\begin{cases} \partial_t^2 v = \Delta_g v \quad \text{in} \quad \mathcal{N} \times \mathbf{R}, \\ v\big|_{t=t_0} = 0, \quad \partial_t v\big|_{t=t_0} = \psi \in L^2(\mathcal{N}(S,t_0)), \\ \partial_\nu v\big|_{\partial\mathcal{N} \times (0,t_0)} = 0. \end{cases} \tag{4.24}$$

Note that $v^\psi\big|_{\partial\mathcal{N} \times \mathbf{R}} \in C(\mathbf{R}; H^{1/2}(\partial\mathcal{N}))$, and

$$\|\mathcal{O}^S_{t_0}\psi\|_{L^2(S \times (0,t_0))} \le C\|\psi\|_{L^2(\mathcal{N})}, \tag{4.25}$$

where $C = C_{t_0}$ is a constant.

Corollary 4.4 is equivalent to the following fact, called the *observability.*

Corollary 4.5. *For any open set $S \subset \partial\mathcal{N}$ and $t_0 > 0$, we have*

$$\operatorname{Ker} \mathcal{O}^S_{t_0} = \{0\}.$$

We consider now, the map $\mathcal{C}_{t_0}^S$ defined by

$$\mathcal{C}_{t_0}^S : L^2(S \times (0,t_0)) \ni f \to u^f\big|_{t=t_0} \in L^2(\mathcal{N}(S,t_0)).$$

The crucial fact about $\mathcal{C}_{t_0}^S$ is the following theorem.

Theorem 4.6. $\overline{\mathrm{Ran}\,(\mathcal{C}_{t_0}^S)} = L^2(\mathcal{N}(S,t_0))$.

Proof. Due to Corollary 4.5, it is sufficient to show

$$\mathcal{C}_{t_0}^S = -\big(\mathcal{O}_{t_0}^S\big)^*, \tag{4.26}$$

i.e.

$$(C_{t_0}^S f, \psi)_{L^2(\mathcal{N}(S,t_0))} = -(f, \mathcal{O}_{t_0}^S \psi)_{L^2(S\times(0,t_0))}, \tag{4.27}$$

for $f \in L^2(S\times(0,t_0))$, $\psi \in L^2(\mathcal{N}(S,t_0))$. Clearly, we can take $f \in C_0^\infty(S\times(0,t_0))$ with both f and ψ being real-valued. By integration by parts, we have

$$\begin{aligned} 0 &= \int_{\mathcal{N}}\int_0^{t_0} \big((\partial_t^2 u^f - \Delta_g u^f)v^\psi - u^f(\partial_t^2 v^\psi - \Delta_g v^\psi)\big)\, dt dV_g \\ &= \int_{\mathcal{N}} \big[(\partial_t u^f)v^\psi - u^f(\partial_t v^\psi)\big]_{t=0}^{t=t_0} dV_g \\ &\quad - \int_{\partial\mathcal{N}}\int_0^{t_0} \big((\partial_\nu u^f)v^\psi - u^f(\partial_\nu v^\psi)\big)\, dt dS_g. \end{aligned}$$

By the initial conditions, $u^f\big|_{t=0} = \partial_t u^f\big|_{t=0} = 0$, and $v^\psi\big|_{t=t_0} = 0$, $\partial_t v^\psi\big|_{t=t_0} = \psi$. By the boundary condition, $\partial_\nu v^\psi\big|_{\partial\mathcal{N}\times\mathbf{R}} = 0$, and $\partial_\nu u^f\big|_{\partial\mathcal{N}\times\mathbf{R}} = f$. We then have

$$\int_{\mathcal{N}} \mathcal{C}_{t_0}^S f \psi dV_g = -\int_{\partial\mathcal{N}}\int_0^{t_0} f v^\psi dt dS_g.$$

Since f is supported in $S\times(0,t_0)$, the right-hand side is rewritten as

$$-\int_S\int_0^{t_0} f v^\psi\big|_{S\times(0,t_0)} dt dS_g = -(f, \mathcal{O}_{t_0}^S\psi)_{L^2(S\times(0,t_0))},$$

which proves the lemma. □

By this theorem, for any $\epsilon > 0$ and $a \in L^2(\mathcal{N})$ such that $\mathrm{supp}\, a \subset \mathcal{N}(S,t_0)$, there exists $f = f_{\epsilon,a} \in C_0^\infty(S\times(0,t_0))$ satisfying $\|u^f(\cdot,t_0) - a\|_{L^2(\mathcal{N})} < \epsilon$. Therefore the property described in Theorem 4.6 should be called *approximate controllability.*

4.3. Further results on uniqueness. Results of the type of Theorem 4.3 (Holmgren-John type uniqueness theorems) have a long story, starting from the classical result by Holmgren:

Theorem 4.7. *Let u be a classical, i.e. C^2, solution to the partial differential equation $P(x,D_x)u = 0$ with analytic coeffcients. If $u = 0$ in one side of a non-characteristic surface Σ, then* $\mathrm{supp}\, u \cap \Sigma = \emptyset$*, i.e. $u = 0$ near Σ.*

For the proof, see e.g. [**55**] Vol 1, p. 309 and [**101**] p. 250. Recall that for a differential operator $P(x,D_x) = \sum_{|\alpha|\le m} p_\alpha(x) D_x^\alpha$ defined on an open set U in $\mathbf{R}^n$, its *principal part* is defined by $P_m(x,\xi) = \sum_{|\alpha|=m} p_\alpha(x)\xi^\alpha$. A surface Σ of co-dimension 1 in U is said to be *non-characteristic* to $P(x,D_x)$, if $P_m(x,\nu_x) \neq 0$ for any $x \in \Sigma$ and normal ν_x to Σ at x. Theorem 4.6 was first proved by E. Holmgren in 1901 [**54**] and extended by F. John in 1949 [**72**]. This theorem has been tried to be extended to the C^∞-coefficient case by Robbiano [**116**] or Hörmander [**56**], and finally Tataru [**125**] succeeded in obtaining the result in full generality (see also [**77**], p. 117). The importance of *non-analyticity* should largely be emphasized in applications to inverse problems. We formulate Tataru's local uniqueness theorem in the form convenient for future applications.

Theorem 4.8. *Let $u \in H^1_{loc}(\Omega)$, $\Omega \subset \widetilde{\mathcal{N}} \times \mathbf{R}$, be a weak solution to the wave equation $\partial_t^2 u = \Delta_{\tilde{g}} u$, where $(\widetilde{\mathcal{N}}, \tilde{g})$ is a Riemannian manifold. Let $\Sigma \subset \Omega$ be a non-characteristic surface. If $u = 0$ on one side of Σ, then* $\operatorname{supp} u \cap \Sigma = \emptyset$.

Actually, this theorem implies Theorem 4.3 due to the fact that we can continue by 0 untill we hit the chracteristic surface giving rise to the double cone of dependence. Note also that this theorem implies more general version of Theorem 4.3 where condition $\partial_\nu u\big|_{\partial\mathcal{N}\times(-t_0,t_0)} = 0$ is changed to $\partial_\nu u\big|_{S\times(-t_0,t_0)} = 0$.

5. Topological reconstruction of $\mathcal{N}$ by $R(\mathcal{N})$

5.1. Reconstruction from boundary distance functions. The key idea of the geometric BC-method is to reconstruct the *boundary distance function*, $r_x(z)$, defined as follows: For any $x \in \mathcal{N}$, r_x is defined by

$$r_x(z) = d(x, z), \quad z \in \partial\mathcal{N}, \tag{5.1}$$

$d(x, y)$ being the distance of $x, y \in \mathcal{N}$. We define the map R by

$$R : \mathcal{N} \ni x \to r_x(\cdot) \in C(\partial\mathcal{N}).$$

If $\partial\mathcal{N}$ is compact, $R(\mathcal{N})$ becomes a metric space by the distance

$$d_\infty(r_1, r_2) = \|r_1(\cdot) - r_2(\cdot)\|_{L^\infty(\partial\mathcal{N})},$$

and the following inclusion relation hold

$$R(\mathcal{N}) \subset C^{0,1}(\partial\mathcal{N}) \subset L^\infty(\partial\mathcal{N}),$$

where $C^{0,1}(\partial\mathcal{N})$ is the space of Lipschitz continuous functions on $\partial\mathcal{N}$. The utility of the boundary distance function is seen in the following lemma.

Lemma 5.1. *If $\partial\mathcal{N}$ is compact, $(R(\mathcal{N}), d_\infty)$ is homeomorphic to $(\mathcal{N}, d)$.*

Proof. By the triangle inequality, for any $z \in \partial\mathcal{N}$, $|d(x, z) - d(y, z)| \le d(x, y)$. Hence $\max_{z\in\partial\mathcal{N}} |d(x, z) - d(y, z)| \le d(x, y)$. This implies

$$d_\infty(r_x, r_y) \le d(x, y). \tag{5.2}$$

Both of $(R(\mathcal{N}), d_\infty)$ and $(\mathcal{N}, d)$ are complete metric spaces. By (5.2), the map $R : (\mathcal{N}, d) \to (R(\mathcal{N}), d_\infty)$ is continuous. Let us show that R is injective. Assume $r_x(z) = r_y(z)$, $\forall z \in \partial\mathcal{N}$. Let z_m be a point of minimum of r_x and r_y. Then x lies on the geodesic normal to $\partial\mathcal{N}$ from z_m at the arclength $r_x(z_m)$, but also y lies on the geodesic normal at arclength $r_y(z_m) = r_x(z_m)$. Then $x = y$.

We show that R^{-1} is continuous. Suppose $r_{x_n}(\cdot)$ converges to $r_x(\cdot)$ uniformly on $\partial\mathcal{N}$. Then $\sup_n \|r_{x_n}\|_{L^\infty} < \infty$. Since $\min r_{x_n} = d(x_n, \partial\mathcal{N})$, and $\partial\mathcal{N}$ is compact, this means that $\{x_n\}$ is in a compact subset in $\mathcal{N}$. Therefore, for any subsequence of $\{x_n\}$, one can select a sub-subsequnce $\{x'_n\}$ such that x'_n converges to some point $y \in \mathcal{N}$. By (5.2), $r_{x'_n}(\cdot)$ converges uniformly to $r_y(\cdot)$. However, since $r_{x_n}(\cdot)$ converges to $r_x(\cdot)$, we have $r_x(\cdot) = r_y(\cdot)$. Therefore $x = y$. Since every subsequence of $\{x_n\}$ contains a sub-sub sequence which converges to one and the same limit x, x_n converges to x. This proves the lemma. $\square$

5.2. Metrics on $R(\mathcal{N})$. $R(\mathcal{N})$ is a set of functions indexed by the points $x \in \mathcal{N}$. However in the inverse problem we are now considering, we know neither $\mathcal{N}$ nor x, since they are the objects we are trying to reconstruct. So, changing the notation, we let $r_1 = r_x, r_2 = r_y$, where $x, y \in \mathcal{N}$. Now we ask a question: *Does $d_\infty(r_1, r_2)$ determine $d(x, y)$?* If it is true, it becomes a mile stone for our inverse problem.

Assume we can find new distance $\widehat{d}(r_1, r_2)$ from $d_\infty(r_1, r_2)$ so that $\widehat{d}(r_1, r_2) = d(x, y)$ for x, y such that $r_1 = r_x$, $r_2 = r_y$. Then $(R(\mathcal{N}), \widehat{d})$ becomes isometric, as a metric space, to $(\mathcal{N}, d)$. By the Myers-Steenrod theorem [**108**] (see e.g. [**24**], p. 175), this implies that there is a unique Riemannian manifold structure on $R(\mathcal{N})$ such that $R : \mathcal{N} \to R(\mathcal{N})$ is isometry. In the following, we give a direct way of reconstructing the Riemannian manifold structure on $R(\mathcal{N})$ to make R a Riemannian isometry from $\mathcal{N}$ to $R(\mathcal{N})$, without leaning over the abstract nature of the Myers-Steenrod theorem.

To find an isometry from $R(\mathcal{N})$ to $\mathcal{N}$, perhaps the simplest case is the *simple manifold*. By definition (in the strong sense) simple manifold means that any $x, y \in \mathcal{N}$ are connected by a unique shortest geodesic which continues to both directions to $\partial\mathcal{N}$ as the shortest geodesic, and $\partial\mathcal{N}$ is geodesically convex.

Proposition 5.2. *If $\mathcal{N}$ is simple, then $d_\infty(r_x, r_y) = d(x, y)$.*

Proof. Recall (5.2). Let z be the point on $\partial\mathcal{N}$ lying on the continuation of the geodesic from x to y. Then $d(x, z) - d(y, z) = d(x, y)$. This proves the proposition. □

Remark 5.3. It is known that even in the case of non-simple manifold, there exists a constant $0 < C \leq 1$ such that

$$Cd(x, y) \leq d_\infty(r_x, r_y) \leq d(x, y).$$

Remark 5.4. Let $\partial\mathcal{N}_1 = \partial\mathcal{N}_2$, and compare $R(\mathcal{N}_1)$ and $R(\mathcal{N}_2)$. To this end, we can take the Hausdorff distance $d_H(R(\mathcal{N}_1), R(\mathcal{N}_2))$. Let us recall that if $\mathcal{N}$ be a metric space, $S_1, S_2 \subset \mathcal{N}$, then the Hausdorff distance is defined by

$$d_H(S_1, S_2) = \max\{\sup_{x \in S_1} d(x, S_2), \sup_{y \in S_2} d(y, S_1)\}.$$

A natural question is, if $d_H(R(\mathcal{N}_1), R(\mathcal{N}_2))$ is small, does it mean that $\mathcal{N}_1$ and $\mathcal{N}_2$ are close and which sense?

In general, the answer is "No", which is the manifestation of well-known *ill-posedness* of the inverse problem. However, we can add some a-priori conditions, e.g. in terms of Gromov compactness on manifolds $(\mathcal{N}, g)$, to obtain a positive answer. See e.g. [**3**]

6. Boundary cut locus

In this and the next sections, we devote ourselves to geometric preliminaries. For a Riemannian manifold $\mathcal{N}$, let $T_x(\mathcal{N})$ be the tangent space at $x \in \mathcal{N}$. Recall that for $\xi, \eta \in T_x(\mathcal{N})$, the inner product and the length are defined by

$$g_x(\xi, \eta) = g_{ij}(x)\xi^i\eta^j = \sum_{i,j=1}^n g_{ij}(x)\xi^i\eta^j, \quad |\xi|_g = \sqrt{g_x(\xi, \xi)}$$

Put $S_x(\mathcal{N}) = \{\xi \in T_x(\mathcal{N})\,;\, |\xi|_g = 1\}$. Let $T(\mathcal{N})$ and $T^*(\mathcal{N})$ be the tangent bundle and the cotangent bundle of $\mathcal{N}$, respectively.

We are dealing with the manifold $\mathcal{N}$ with boundary. To consider the differential at $z \in \partial\mathcal{N}$ of a map defined on $\mathcal{N}$, we can extend the manifold $\mathcal{N}$ to a bigger manifold $\widetilde{\mathcal{N}}$ of the same dimension so that z is in the interior of $\widetilde{\mathcal{N}}$. This defines the tangent space $T_z(\mathcal{N})$ at z which is independent of the choice of $\widetilde{\mathcal{N}}$. When we consider the tangent space of $\partial\mathcal{N}$ at $z \in \partial\mathcal{N}$, we denote it by $T_z(\partial\mathcal{N})$. Note that $T_z(\partial\mathcal{N})$ is canonically identified with the subspace of codimension 1 in $T_z(\mathcal{N})$ whose unit normal is the unit normal to $\partial\mathcal{N}$ at z.

6.1. Variation and Jacobi fields. Let $c(t)$ be a curve on $\mathcal{N}$. For a vector field $X(t)$ on $\mathcal{N}$ along $c(t)$, with components $(X^1(t), \cdots, X^n(t))$ in local coordinates, the *covariant differential* $\dfrac{D}{dt}X(t)$ along $c(t)$ is defined by

$$\nabla_{\dot{c}}X(t) = \frac{D}{dt}X^k(t) = \dot{X}^k(t) + \Gamma^k_{ij}(c(t))\dot{c}^i(t)X^j(t),$$

where we used the abbreviation $\dot{f}(t) = \dfrac{df(t)}{dt}$. Note that $\nabla_{\dot{c}}X(t)$ is independent of local coordinates. A vector field $Z(t)$ is said to be *parallel* along $c(t)$ if it satisfies $\dfrac{D}{dt}Z(t) = 0$. In particular, $c(t)$ is a geodesic if and only if $\dot{c}(t)$ is parallel along $c(t)$. For any C^∞-curve $c(t)$ and vector fields $\xi(t)$ and $\eta(t)$ along $c(t)$, we have

$$\frac{d}{dt}\, g_{c(t)}\big(\xi(t), \eta(t)\big) = g_{c(t)}\left(\frac{D}{dt}\xi(t), \eta(t)\right) + g_{c(t)}\left(\xi(t), \frac{D}{dt}\eta(t)\right).$$

The *energy* of a curve $c(t)$ is defined by

$$E(c) = \frac{1}{2}\int_a^b g_{c(t)}(\dot{c}(t), \dot{c}(t))dt, \tag{6.1}$$

and the *(arc)length* of $c(t)$ is defined by

$$L(c) = \int_a^b \sqrt{g_{c(t)}(\dot{c}(t), \dot{c}(t))}dt. \tag{6.2}$$

Then by the Cauchy-Schwarz inequality, we have

$$L(c)^2 \le 2(b-a)E(c), \tag{6.3}$$

where the equality holds only when the speed $\sqrt{g_{c(t)}(\dot{c}(t), \dot{c}(t))}$ is constant.

A C^∞-map : $[a,b] \times (-\epsilon, \epsilon) \ni (t,s) \to H(t,s) \in \mathcal{N}$ is said to be a *variation* of $c(t)$ if $H(t,0) = c(t)$ $(a \le t \le b)$. It is said to be a *geodesic variation* if for each s, the curve : $t \to H(t,s)$ is a geodesic.

For $p \in \mathcal{N}$ and $v \in T_p(\mathcal{N})$, let $c_p(t,v)$ be the geodesic such that $c_p(0,v) = p$, $\dot{c}_p(0,v) = v$. The *exponential map* is defined by

$$\exp_p(v) = c_p(1,v).$$

For any $v \in T_p(\mathcal{N})$, the curve : $t \to \exp_p(tv)$ is a geodesic.

The *curvature tensor* R is defined by

$$(R(X,Y)Z)^l = R^l_{ijk}X^iY^jZ^k,$$

$$R^l_{ijk} = \frac{\partial\Gamma^l_{jk}}{\partial x^i} - \frac{\partial\Gamma^l_{ik}}{\partial x^j} + \Gamma^l_{ir}\Gamma^r_{jk} - \Gamma^l_{jr}\Gamma^r_{ik},$$

where X, Y, Z are vector fields on $\mathcal{N}$. Note that although we use coordinates to define R^l_{ijk}, this is actually a $(1,3)$ tensor. It satisfies

$$R(X,Y)Z = \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z. \tag{6.4}$$

Lemma 6.1. *Let $H(t,s)$ be a variation of $c(t)$, and put $c_s(t) = H(t,s)$. We define the vector field $Y(t)$ along $c(t)$ by*

$$Y(t) = \frac{\partial}{\partial s}H(t,s)\Big|_{s=0}.$$

Then the following formulae hold.
(1) The 1st variation formula:

$$\frac{d}{ds}E(c_s)\Big|_{s=0} = g_{c(b)}(Y(b),\dot c(b)) - g_{c(a)}(Y(a),\dot c(a)) - \int_a^b g_{c(t)}\Big(Y(t),\frac{D}{dt}\dot c(t)\Big)dt,$$

where D/dt is the covariant differential along $c(t)$.
(2) The 2nd variation formula:

$$\begin{aligned}\frac{d^2}{ds^2}E(c_s)\Big|_{s=0} &= g_{c(b)}(S(b),\dot c(b)) - g_{c(a)}(S(a),\dot c(a))\\ &\quad + \int_a^b \Big\{g_{c(t)}\Big(\frac{D}{dt}Y(t),\frac{D}{dt}Y(t)\Big) - g_{c(t)}\big(R(Y(t),\dot c(t))\dot c(t),Y(t)\big)\\ &\qquad - g_{c(t)}\Big(S(t),\frac{D}{dt}\dot c(t)\Big)\Big\}dt,\end{aligned}$$

where, letting D/ds be the covariant differential along the curve $C_t(s) : s \to H(t,s)$,

$$S(t) = \frac{D}{ds}\frac{\partial H(t,s)}{\partial s}\Big|_{s=0}. \tag{6.5}$$

.

For the proof of above lemma, see e.g. [**36**], Chap. 3

Lemma 6.2. *Let $c(t)$ $(a \le t \le b)$ be a geodesic on $\mathcal{N}$, and $H(t,s)$ its geodesic variation. Then $Y(t) = \partial H(t,s)/\partial s\big|_{s=0}$ satisfies*

$$\Big(\frac{D}{dt}\Big)^2 Y + R(Y,\dot c)\dot c = 0, \quad a \le t \le b, \tag{6.6}$$

where D/dt is the covariant differential along $c(t)$. Conversely, if a vector field $Y(t)$ along the geodesic $c(t)$ satisfies the equation (6.6), there is a geodesic variation $H(t,s)$ such that $H(t,0) = c(t)$ and $Y(t) = \partial H(t,s)/\partial s\big|_{s=0}$.

Proof. Direct computation shows that

$$\frac{D}{ds}\frac{\partial}{\partial t}H(t,s) = \frac{D}{dt}\frac{\partial}{\partial s}H(t,s).$$

Therefore by (6.4),

$$\frac{D}{dt}\frac{D}{dt}\frac{\partial H}{\partial s} = \frac{D}{\partial t}\frac{D}{\partial s}\frac{\partial H}{\partial t} = \Big(\frac{D}{\partial s}\frac{D}{\partial t} + R(\frac{\partial H}{\partial t},\frac{\partial H}{\partial s})\Big)\frac{\partial H}{\partial t}.$$

Since $c_s(t)$ are geodesics, $D(\partial H(t,s)/\partial t)/dt = 0$. Thus, letting $s = 0$, we obtain $(D/dt)^2 Y = R(\dot c, Y)\dot c$, which proves (6.6).

Conversely, suppose $Y(t)$ satisfies (6.6). Take a curve $z(s)$ such that $z(0) = c(a)$, $\dot{z}(0) = Y(a)$. Let $X_0(s)$, $X_1(s)$ are vector fields which are parallel along $z(s)$, and satisfy $X_0(0) = \dot{c}(a)$, $X_1(0) = (DY/dt)(a)$. We put

$$H(t,s) = \exp_{z(s)}\left((t-a)\big(X_0(s) + sX_1(s)\big)\right).$$

Then the curve : $t \to H(t,s)$ is a geodesic for each s, and $H(t,0) = c(t)$. Let $Z(t) = \partial H(t,s)/\partial s\big|_{s=0}$. Then, as has been shown above, $Z(t)$ satisfies (6.6). Moreover, $Z(a) = \dot{z}(0) = Y(a)$. Then

$$\begin{aligned}\frac{DZ}{dt}(a) &= \frac{D}{dt}\frac{\partial H}{\partial s}\Big|_{t=a,s=0} = \frac{D}{ds}\frac{\partial H}{\partial t}\Big|_{t=a,s=0} \\ &= \frac{D}{ds}\left(X_0(s) + sX_1(s)\right)\big|_{s=0} \\ &= X_1(0) = \frac{DY}{dt}(a),\end{aligned}$$

where in the last step, we use $X_0(s)$, $X_1(s)$ are parallel along $z(s)$. Therefore $Y(t) = Z(t)$ by the uniqueness for solutions of differential equations. $\square$

A solution $Y(t)$ of (6.6) is called *Jacobi field* along $c(t)$.

6.2. Focal point. In the following, we consider the *boundary normal geodesic*, denoted by $\gamma_z(t)$ or $\exp_{\partial\mathcal{N}}(z,t)$, starting from $z \in \partial\mathcal{N}$ with initial direction being the inner unit normal at z. Explicitly, take local coordinates $z = (z_1, \cdots, z_{n-1})$ on $\partial\mathcal{N}$, and $(z_1, \cdots, z_{n-1}, x_n)$, where $x_n = 0$ is a defining equation of $\partial\mathcal{N}$, as local coordinates in $\mathcal{N}$. Coinsider the equation of geodesics

$$\begin{cases}\dfrac{d^2x^k}{dt^2} + \Gamma^k_{ij}(x(t))\dfrac{dx^i}{dt}\dfrac{dx^j}{dt} = 0, \\ x(0) = (z,0), \quad \dfrac{dx}{dt}(0) = \nu(z),\end{cases}$$

where $\nu(z)$ is the unit normal at the boundary. Then, the map $\gamma_z(t) : (z,t) \to x(t,z)$ is a diffeomorphism near $\partial\mathcal{N}$, and we use (z,t) as boundary normal coordinates in $\mathcal{N}$ near $\partial\mathcal{N}$.

Proposition 6.3. *In the boundary normal coordinates, the Riemannian metric is written as*

$$ds^2 = (dt)^2 + \sum_{i,j=1}^{n-1} h_{ij}(z,t)dz^i dz^j.$$

Proof. Since $x(t)$ is a geodesic, we have

$$g_{nn} = g\Big(\frac{\partial x}{\partial t}, \frac{\partial x}{\partial t}\Big) = 1.$$

For $1 \leq i \leq n-1$, we have

$$\begin{aligned}\frac{d}{dt}g_{ni} &= \frac{d}{dt}g\Big(\frac{\partial x}{\partial t}, \frac{\partial x}{\partial z^i}\Big) = g\Big(\frac{\partial x}{\partial t}, \frac{D}{dt}\frac{\partial x}{\partial z^i}\Big) \\ &= g\Big(\frac{\partial x}{\partial t}, \frac{D}{\partial z^i}\frac{\partial x}{\partial t}\Big) = \frac{1}{2}\frac{\partial}{\partial z^i}g\Big(\frac{\partial x}{\partial t}, \frac{\partial x}{\partial t}\Big) = 0.\end{aligned}$$

Since $\dfrac{dx}{dt}(0) = \nu(z)$ is normal to $\partial\mathcal{N}$, $g_{ni}(z,0) = 0$. Therefore, $g_{ni} = 0$, and the proof is completed. $\square$

Fixing t, we define the map $\exp_{\partial\mathcal{N}}(\cdot, t)$ by

$$\exp_{\partial\mathcal{N}}(\cdot, t) : \partial\mathcal{N} \ni z \to \gamma_z(t) \in \mathcal{N}.$$

Let $d_{\partial\mathcal{N}} \exp_{\partial\mathcal{N}}(z_0, t) : T_{z_0}(\partial\mathcal{N}) \to T_{\gamma_{z_0}(t)}(\mathcal{N})$ be the differential of $\exp_{\partial\mathcal{N}}(\cdot, t)$ evaluated at z_0.

Definition 6.4. Let $\gamma_{z_0}(t)$ be the boundary normal geodesic starting from $z_0 \in \partial\mathcal{N}$. The point $\gamma_{z_0}(t_0) = \exp_{\partial\mathcal{N}}(z_0, t_0)$ is called a *focal point* along $\gamma_{z_0}(t)$ if

$$\operatorname{rank}\left(d_{\partial\mathcal{N}} \exp_{\partial\mathcal{N}}(z_0, t_0)\right) < n - 1.$$

Lemma 6.5. *Let $\gamma_{z_0}(t)$ $(0 \le t \le t_0)$ be a boundary normal geodesic starting from $z_0 \in \partial\mathcal{N}$. If $\gamma_{z_0}(t_1)$ is a focal point along γ_{z_0} for some $0 < t_1 < t_0$, then $\tau = d(\gamma_{z_0}(t_0), \partial\mathcal{N}) < t_0$ and there exist $w \in \partial\mathcal{N}$ such that $\gamma_w(\tau) = \gamma_{z_0}(t_0)$.*

Note that this lemma is a particular case of Fermi coordinates associated with k-dimensional submanifold in $\mathcal{N}$, where $k < n$. See [**24**], §3.6. See [**17**], p. 232, or [**119**], Chap. 3, Lemma 2.11 for the complete proof.

We prove this lemma under the following additional assumption.

Condition (TG) : In a neighborhood of z_0, we can extend $\mathcal{N}$ to a bigger manifold $\widetilde{\mathcal{N}}$ so that, in a neighborhood of z_0, $\partial\mathcal{N}$ is a totally geodesic submanifold of $\widetilde{\mathcal{N}}$.

Let us recall that, given a Riemannian manifold $\widetilde{\mathcal{N}}$, its submanifold $\mathcal{S}$ is said to be *totally geodesic* if any geodesic of $\widetilde{\mathcal{N}}$ starting from a point $z \in \mathcal{S}$ in a direction tangential to $\mathcal{S}$ lies in $\mathcal{S}$. Note that, if $\dim(\mathcal{S}) = n-1$, which is the case of $\mathcal{S} = \partial\mathcal{N}$, this condition is equivalent to the fact that the second fundamental form (the shape operator) of $\mathcal{S}$ vanishes. In turn, this is equivalent to the fact that $\nu(z)$ is parallel along $\mathcal{S}$.

For example, if for some $\epsilon > 0$, $\widetilde{\mathcal{N}} = \mathcal{S} \times (-\epsilon, \epsilon)$, and the metric of $\widetilde{\mathcal{N}}$ is of product form:

$$ds^2 = (dt)^2 + h(\omega, d\omega),$$

where $h(\omega, d\omega)$ is the positive definite metric on $\mathcal{S}$ induced from that of $\widetilde{\mathcal{N}}$, then $\mathcal{S}$ is totally geodesic.

Proof of Lemma (6.5). By the assumption, there exists $0 \neq \xi \in T_{z_0}(\partial\mathcal{N})$ such that

$$\left(d_{\partial\mathcal{N}} \exp_{\partial\mathcal{N}}(z_0, t_1)\right) \xi = 0. \tag{6.7}$$

Let $z(s)$ be a geodesic in $\widetilde{\mathcal{N}}$ such that $z(0) = z_0$, $\dot{z}(0) = \xi$. By the condition (TG), $z(s)$ is also a geodesic in $\partial\mathcal{N}$. We put

$$\widetilde{H}(t, s) = \left(\exp_{\partial\mathcal{N}}(t)\right)(z(s)) = \gamma_{z(s)}(t),$$

$$\widetilde{Y}(t) = \frac{\partial \widetilde{H}(t, s)}{\partial s}\Big|_{s=0}.$$

Then, by Lemma 6.2, $\widetilde{Y}(t)$ is a Jacobi field along $c(t)$ and satisfies

$$\widetilde{Y}(0) = \xi, \quad \widetilde{Y}(t_1) = 0. \tag{6.8}$$

These facts follow from $\widetilde{H}(0, s) = z(s)$, (6.7), and

$$\frac{\partial}{\partial s} \widetilde{H}(t_1, s)\big|_{s=0} = \frac{\partial}{\partial s} \exp_{\partial N}(t_1)(z(s))\big|_{s=0} = \left(d_{\partial\mathcal{N}} \exp_{\partial\mathcal{N}}(z_0, t_1)\right) \xi.$$

Take a parallel vector field $Z(t)$ satisfying

$$\begin{cases} \dfrac{D}{dt}Z(t) = 0, \quad \text{for} \quad 0 < t < t_0, \\ Z(t_1) = -\dfrac{D}{dt}\widetilde{Y}(t_1). \end{cases} \tag{6.9}$$

Pick $f(t) \in C_0^\infty((0,t_0))$ such that $f(t_1) = 1$, and put for $\alpha \in \mathbf{R}$

$$V_\alpha(t) = \begin{cases} \widetilde{Y}(t) + \alpha f(t)Z(t), \quad 0 \le t \le t_1, \\ \alpha f(t)Z(t), \quad t_1 \le t \le t_0. \end{cases} \tag{6.10}$$

Note that at $t = t_1$, $V_\alpha(t)$ is continuous by (6.8), however, $\frac{D}{dt}V_\alpha(t)$ is discontinuous. As a variation of $c(t) = \gamma_{z_0}(t)$, we consider

$$H_\alpha(t,s) = \exp_{c(t)}(sV_\alpha(t)). \tag{6.11}$$

Let $c_{\alpha,s}(t)$ be the curve : $t \to H_\alpha(t,s)$. Then $c_{\alpha,0}(t) = c(t)$ for all α. Define the energy of $c_{\alpha,s}(t)$ by (6.1). We can then prove the following formula.

Proposition 6.6. *For small $|\alpha|$, we have*

$$\frac{d^2}{ds^2}E(c_{\alpha,s})\Big|_{s=0} = -2\alpha\, g_{c(t_1)}\Big(\frac{D\widetilde{Y}}{dt}(t_1), \frac{D\widetilde{Y}}{dt}(t_1)\Big) + O(\alpha^2). \tag{6.12}$$

Granting this proposition for the moment, we complete the proof of Lemma 6.5. We have $\dfrac{D\widetilde{Y}}{dt}(t_1) \neq 0$. In fact, if this vanishes, since $\widetilde{Y}(t_1) = 0$ and $\widetilde{Y}(t)$ is a solution of the 2nd order differential equation, $\widetilde{Y}(t)$ vanishes identically. Proposition 6.6 then yields

$$(d/ds)^2E(c_{\alpha,s})\big|_{s=0} < 0, \tag{6.13}$$

if $\alpha > 0$ is chosen small enough. Letting

$$Y_\alpha(t) = \partial H_\alpha(t,s)/\partial s\big|_{s=0} = V_\alpha(t),$$

and using $Y_\alpha(0) = \widetilde{Y}(0) = \xi, Y_\alpha(t_0) = 0$, we have by Lemma 6.1 (1),

$$(d/ds)E(c_{\alpha,s})\big|_{s=0} = 0.$$

This, combined with (6.13), implies $E(c_{\alpha,s}) < E(c)$, for $0 < s < \epsilon$, if $\epsilon > 0$ is small enough. For $0 < s < \epsilon$, we have, by the Cauchy-Schwarz inequality (6.3),

$$L(c_{\alpha,s})^2 \le 2t_0E(c_{\alpha,s}) < 2t_0E(c) = L(c)^2,$$

where in the last step we use the fact $c_0(t)$ is a unit speed geodesic. Therefore, $d(\gamma_{z_0}(t_0), \partial\mathcal{N}) < t_0$, which implies an existence of $w \in \partial\mathcal{N}$ with desired property. This proves Lemma 6.5. □

Now we prove Proposition 6.6. We split energy into 2 parts:

$$\begin{aligned} E(c_{\alpha,s}) &= \frac{1}{2}\int_0^{t_1} g_{c_{\alpha,s}}(t)(\dot{c}_{\alpha,s}(t), \dot{c}_{\alpha,s}(t))dt + \frac{1}{2}\int_{t_1}^{t_0} g_{c_{\alpha,s}}(t)(\dot{c}_{\alpha,s}(t), \dot{c}_{\alpha,s}(t))dt \\ &=: E_1(c_{\alpha,s}) + E_2(c_{\alpha,s}). \end{aligned}$$

Let $S_\alpha(t)$ be defined by (6.5). Then, by Lemma 6.1 (2),

$$\frac{d^2}{ds^2}E_1(c_{\alpha,s})\Big|_{s=0} = g_{c(t_1)}(S_\alpha(t_1), \dot c(t_1)) - g_{c(0)}(S_\alpha(0), \dot c(0)) + \int_0^{t_1}\Big\{g\Big(\frac{D}{dt}V_\alpha, \frac{D}{dt}V_\alpha\Big) - g\big(R(V_\alpha, \dot c)\dot c, V_\alpha\big)\Big\}dt.$$

Since $DZ/dt = 0$, the integral in the right-hand side is equal to

$$\begin{aligned}&\int_0^{t_1}\Big\{g\Big(\frac{D\widetilde Y}{dt} + \alpha\dot f Z, \frac{D\widetilde Y}{dt} + \alpha \dot f Z\Big) - g\big(R(\widetilde Y + \alpha f Z, \dot c)\dot c, \widetilde Y + \alpha f Z\big)\Big\}dt\\ &= \int_0^{t_1}\Big\{g\Big(\frac{D\widetilde Y}{dt}, \frac{D\widetilde Y}{dt}\Big) - g\big(R(\widetilde Y, \dot c)\dot c, \widetilde Y\big)\Big\}dt\\ &+ 2\alpha\int_0^{t_1}\Big\{g\Big(\dot f Z, \frac{D\widetilde Y}{dt}\Big) - g\big(R(\widetilde Y, \dot c)\dot c, f Z\big)\Big\}dt + O(\alpha^2).\end{aligned}$$

Since $\widetilde Y$ is a Jacobi field, it satisfies (6.6). This imples

$$\begin{aligned}\frac{d^2}{ds^2}E_1(c_s)\Big|_{s=0} &= g_{c(t_1)}(S_\alpha(t_1), \dot c(t_1)) - g_{c(0)}(S_\alpha(0), \dot c(0))\\ &+ \int_0^{t_1}\Big\{g\Big(\frac{D\widetilde Y}{dt}, \frac{D\widetilde Y}{dt}\Big) + g\big(\frac{D^2\widetilde Y}{dt^2}, \widetilde Y\big)\Big\}dt\\ &+ 2\alpha\int_0^{t_1}\Big\{g\Big(\dot f Z, \frac{D\widetilde Y}{dt}\Big) + g\big(\frac{D^2\widetilde Y}{dt^2}, f Z\big)\Big\}dt + O(\alpha^2).\end{aligned} \tag{6.14}$$

Then two integrals of the right-hand side are computed as

$$\begin{aligned}&\int_0^{t_1}\frac{d}{dt}g\Big(\frac{D\widetilde Y}{dt}, \widetilde Y\Big)dt + 2\alpha\int_0^{t_1}\frac{d}{dt}g\big(\frac{D\widetilde Y}{dt}, fZ\big)dt\\ &= g_{c(t_1)}\Big(\frac{D\widetilde Y}{dt}(t_1 - 0), \widetilde Y(t_1)\Big) - g_{c(0)}\Big(\frac{D\widetilde Y}{dt}(0), \widetilde Y(0)\Big)\\ &+ 2\alpha\Big\{g_{c(t_1)}\Big(\frac{D\widetilde Y}{dt}(t_1), f(t_1)Z(t_1)\Big) - g_{c(0)}\Big(\frac{D\widetilde Y}{dt}(0), f(0)Z(0)\Big)\Big\}.\end{aligned} \tag{6.15}$$

Recall that $\widetilde Y(t_1) = 0$. We also note that the curve : $s \to H(t,s) = \exp_{c(t)}(sV_\alpha(t))$ is a geodesic for $t \ge 0$. Then we have

$$S_\alpha(t) = \frac{D}{ds}\frac{\partial H(t,s)}{\partial s}\Big|_{s=0} = 0, \quad t \ge 0. \tag{6.16}$$

We show that $\frac{D\widetilde Y}{dt}(0) = 0$. In fact, since

$$\frac{D}{dt}\widetilde Y(0) = \frac{D}{dt}\frac{\partial \widetilde H}{\partial s}\Big|_{s=t=0} = \frac{D}{ds}\frac{\partial \widetilde H}{\partial t}\Big|_{s=t=0} = \frac{D}{ds}\nu(z(s))\Big|_{s=0} = 0. \tag{6.17}$$

where the last equation follows from vanishing of the second fundamental form in z_0. Plugging (6.14) $\sim$ (6.17), we obtain

$$\frac{d^2}{ds^2}E_1(c_{\alpha,s})\Big|_{s=0} = 2\alpha g_{c(t_1)}\Big(\frac{D\widetilde Y}{dt}(t_1), Z(t_1)\Big) + O(\alpha^2). \tag{6.18}$$

We turn to $E_2(c_{\alpha,s})$. As above,

$$\frac{d^2}{ds^2}E_2(c_{\alpha,s})\Big|_{s=0} = g_{c(t_0)}(S_\alpha(t_0),\dot c(t_0)) - g_{c(t_1)}(S_\alpha(t_1),\dot c(t_1))$$
$$+\int_{t_1}^{t_0}\Big\{g\Big(\frac{D}{dt}V_\alpha,\frac{D}{dt}V_\alpha\Big) - g\big(R(V_\alpha,\dot c)\dot c, V_\alpha\big)\Big\}dt.$$

We compute in the same way as for $E_1(c_{\alpha,s})$. Since $\widetilde{Y}$ does not appear in this case, we have

$$\frac{d^2}{ds^2}E_2(c_{\alpha,s})\Big|_{s=0} = O(\alpha^2) \tag{6.19}$$

In view of (6.9), (6.18) and (6.19), we have completed the proof. □

Remark 6.7. The above proof can be immediately extended to the case when the second fundamental form of $\partial\mathcal{N}$ vanishes just at the point z_0. Indeed, the above proof shows that, for sufficiently small $\alpha > 0$ and $|s|$,

$$d(z(s),\gamma_{z_0}(t_0)) < t_0 - c\alpha s^2.$$

Since $d(z(s),\partial\mathcal{N}) = O(|s|^3)$, the result follows.

6.3. Boundary cut point. Let $\gamma_z(\cdot)$ be the boundary normal geodesic starting from $z\in\partial\mathcal{N}$. A point $\gamma_z(t)$ is said to be *uniquely minimizing* along the geodesic $\gamma_z(\cdot)$ if $t = d(\gamma_z(t),\partial\mathcal{N})$ and $t < d(\gamma_z(t),w)$ for any $w\in\partial\mathcal{N}$ such that $w\neq z$. Thus, $\{\gamma_z(s)\,;\,0\le s\le t\}$ is a unique shortest geodesic from $\partial\mathcal{N}$ to $\gamma_z(t)$.

Lemma 6.8. *Let $\gamma_z(t)$ $(0\le t\le t_0)$ be the boundary normal geodesic starting from $z\in\partial\mathcal{N}$. If $\gamma_z(t_1)$ is not uniquely minimizing for some $0<t_1<t_0$, then $d(\gamma_z(t_0),\partial\mathcal{N}) < t_0$.*

Proof. Since $\gamma_z(t_1)$ is not uniquely minimizing, there exists $w\in\partial\mathcal{N}$ such that $\gamma_w(t)=\gamma_z(t_1)$, $t\le t_1$. Consider a once broken geodesics $c(s)=\gamma_w([0,t])\cup\gamma_z([t_1,t_0])$. Here, for any curve $c(s)$, by $c([a,b])$ we denote the piece of $c(s)$ for $s\in[a,b]$. Then $\gamma_z(t_0)=c(s)$, $s=t_0+(t-t_1)$. This proves the lemma when $t<t_1$.

For $t=t_1$, consider a curve $c(s)$ which consists of 3 parts: the geodesic $\gamma_w(s), 0\le s\le t-\epsilon$, the minimizing geodesic $c'(\tau)$ connecting $\gamma_w(t-\epsilon)$ and $\gamma_z(t_1+\epsilon)$, and the piece of geodesic $\gamma_z(s)$ for $t_1+\epsilon\le s\le t_0$. Note that, by the short-cut arguments, $L(c')<2\epsilon$. Therefore,

$$L(c) = (t-\epsilon) + L(c') + (t_0-(t_1+\epsilon)) < t_0-(t_1-t) = t_0,$$

which proves the lemma. □

By the above lemma, if $\gamma_z(t)$ is uniquely minimizing along $\gamma_z(\cdot)$, then so is $\gamma_z(s)$ for any $0<s<t$. We put

$$\tau(z) = \sup\{t\,;\,\gamma_z(t) \text{ is uniquely minimizing}\}. \tag{6.20}$$

We then have

$$d(\gamma_z(t),\partial\mathcal{N}) < t, \quad \text{for} \quad \tau(z) < t.$$

In fact, we have only to take $\tau(z)<t_1<t$ and apply Lemma 6.8.

Definition 6.9. The function $\tau(z)$ defined by (6.20) is called the *boundary cut function*, and the point $\gamma_z(\tau(z))$ for $\tau(z)<\infty$ is called *boundary cut point* of z along γ_z. If $\tau(z)=\infty$, we say that there is no boundary cut point along the boundary normal geodesic γ_z.

Lemma 6.10. *For $z_0 \in \partial\mathcal{N}$, let $\tau(z_0)$ be as in Definition 6.9. At the boundary cut point,*

$$d(\gamma_{z_0}(\tau(z_0)), z_0) = \tau(z_0),$$

and at least one (possibly both) of the following statements holds:
(a) $\gamma_{z_0}(\tau(z_0))$ is an ordinary boundary cut point, i.e. there is $w \in \partial\mathcal{N}$ such that $w \neq z_0$ and $\gamma_{z_0}(\tau(z_0)) = \gamma_w(\tau(z_0))$.
(b) $\gamma_{z_0}(\tau(z_0))$ is the first focal point along γ_{z_0}, i.e.

$$\begin{aligned} &\mathrm{rank}\,(d_{\partial\mathcal{N}} \exp_{\partial\mathcal{N}}(z_0, t)) = n-1 \quad \text{if} \quad t < \tau(z_0), \\ &\mathrm{rank}\,(d_{\partial\mathcal{N}} \exp_{\partial\mathcal{N}}(z_0, t)) < n-1 \quad \text{if} \quad t = \tau(z_0). \end{aligned}$$

Proof. By definition, we have $d(\gamma_{z_0}(s), \partial\mathcal{N}) = s$ for $s < \tau(z_0)$. Letting $s \to \tau(z_0)$, we have $d(\gamma_{z_0}(\tau(z_0)), \partial\mathcal{N}) = \tau(z_0)$. This implies, by Lemma 6.5, $\gamma_{z_0}(s)$ is not a focal point for $0 < s < \tau(z_0)$.

There exists $\delta > 0$ such that the geodesic $\gamma_{z_0}(t)$ exists in the interval $[0, \tau(z_0) + \delta]$. Take a sequence $\delta > \epsilon_1 > \epsilon_2 \cdots \to 0$ and put $t_n = \tau(z_0) + \epsilon_n$. Then, by the definition of $\tau(z_0)$, there exists $w_n \in \partial\mathcal{N}$, $w_n \neq z_0$, and $s_n < t_n$ such that $\gamma_{w_n}(s_n) = \gamma_z(t_n)$. Since $\partial\mathcal{N}$ is compact, there exists a subsequence $\{w_n, s_n\}$, such that $w_n \to w \in \partial\mathcal{N}$, $s_n \to s$, where $0 \leq s \leq \tau(z_0)$. Then $\gamma_w(s) = \gamma_{z_0}(\tau(z_0))$, which implies $s = \tau(z_0)$. This gives rise to ordinary boundary cut point if $w \neq z_0$.

Suppose $w = z_0$. Let us show that $\gamma_{z_0}(\tau_{z_0})$ is the first focal point along γ_{z_0}. Assume that $\mathrm{rank}\,(d_{\partial\mathcal{N}} \exp_{\partial\mathcal{N}}(z_0, \tau(z_0))) = n-1$. Take a small neighborhood V of z_0 in $\partial\mathcal{N}$ and small $\epsilon > 0$. Then the map : $V \times (\tau(z_0) - \epsilon, \tau(z_0) + \epsilon) \ni (z, t) \to \gamma_z(t)$ is a diffeomorphsim. Therefore, in a small neighborhood U of $\gamma_{z_0}(\tau(z_0))$, $\gamma_z(t)^{-1}$ is a diffeomorphism. Since $w_n \to z_0$ and $s_n \to \tau(z_0)$, $\gamma_{w_n}(s_n) \in U$. However, $\gamma_{z_0}(t_n) \in U$, and $\gamma_{z_0}(t_n) = \gamma_{w_n}(s_n)$. We thus arrive at the contradiction. By Lemma 6.5, for $t < \tau_0$, $\gamma_{z_0}(t)$ is not a focal point. $\square$

We introduce a topology in $\mathbf{R}_+ \cup \infty$ by taking intervals (a, b) and $(a, \infty] = (a, \infty) \cup \infty$ as basis for the open sets.

Lemma 6.11. *The function $\tau(z)$ in Definition 6.9 is continuous from $\partial\mathcal{N}$ to $\mathbf{R}_+ \cup \infty$.*

Proof. Suppose $\tau(z)$ is not continuous at $\overline{z} \in \partial\mathcal{N}$, and let $z_k \in \partial\mathcal{N}$ be such that $z_k \to \overline{z}$ and $\lim \tau(z_k) \neq \tau(\overline{z})$. Set $\tau_k = \tau(z_k)$, $\tau_\infty = \lim \tau(z_k)$ and $\overline{\tau} = \tau(\overline{z})$.

We first consider the case $\overline{\tau} > \tau_\infty$. Since $\overline{\tau} = \tau(\overline{z}) > \tau_\infty$, then $\tau_\infty < \infty$ and by Lemma 6.5, $\exp_{\partial\mathcal{N}}(\tau_\infty, \overline{z})$ is not a focal point along the boundary normal geodesic $\gamma_{\overline{z}}(t)$. Therefore, $\mathrm{rank}(d_{\partial\mathcal{N}} \exp_{\partial\mathcal{N}}(\overline{z}, \tau_\infty)) = n-1$. Then, there is a neighborhood V of $\overline{z}$ in $\partial\mathcal{N}$ and $\epsilon > 0$ such that the map $V \times (\tau_\infty - \epsilon, \tau_\infty + \epsilon) \ni (z, t) \to \exp_{\partial\mathcal{N}}(t, z)$ is a diffeomorphism. Since $z_k \to \overline{z}$, $\tau_k \to \tau_\infty$, we have $(z_k, \tau_k) \in V \times (\tau_\infty - \epsilon, \tau_\infty + \epsilon)$ for large k. Therefore, $\mathrm{rank}(d_{\partial\mathcal{N}} \exp_{\partial\mathcal{N}}(z_k, \tau_k)) = n-1$ for large k. Then by Lemma 6.10, $\exp_{\partial\mathcal{N}}(z_k, \tau_k)$ is not the focal point along the boundary normal geodesic $\exp_{\partial\mathcal{N}}(z_k, t)$, but the ordinary boundary cut point, i.e. there exists $w_k \in \partial\mathcal{N}$ such that $w_k \neq z_k$ and $\exp_{\partial\mathcal{N}}(w_k, \tau_k) = \exp_{\partial\mathcal{N}}(z_k, \tau_k)$. We see that $w_k \notin V$, since $\exp_{\partial\mathcal{N}}$ is a diffeomorphism on $V \times (\tau_\infty - \epsilon, \tau_\infty + \epsilon)$. By taking a subsequence if necessary, we can assume that w_k converges to $w \in \partial\mathcal{N}$. By shrinking V if necessary, we have $w \notin V$. We than have

$$\begin{aligned} \exp_{\partial\mathcal{N}}(w, \tau_\infty) &= \lim \exp_{\partial\mathcal{N}}(w_k, \tau_k) = \lim \exp_{\partial\mathcal{N}}(z_k, \tau_k) \\ &= \exp_{\partial\mathcal{N}}(\overline{z}, \tau_\infty). \end{aligned}$$

This contradicts Lemma 6.8 and the definition of $\overline{\tau} = \tau(\overline{z})$.

Next we assume $\overline{\tau} < \tau_\infty$. Take $\overline{\tau} < \tau < \infty$. Then, there is $w \in \partial\mathcal{N}$ and $s < \tau$ such that $\gamma_{\overline{z}}(\tau) = \gamma_w(s)$. Since $z_k \to \overline{z}$, $\gamma_{z_k}(\tau) \to \gamma_{\overline{z}}(\tau)$. By the triangle inequlaity,

$$d(w, \gamma_{z_k}(\tau)) \leq d(w, \gamma_{\overline{z}}(\tau)) + d(\gamma_{\overline{z}}(\tau), \gamma_{z_k}(\tau)) = s + d(\gamma_{\overline{z}}(\tau), \gamma_{z_k}(\tau)).$$

Since $s < \tau$, taking k large enough, we see that $d(w, \gamma_{z_k}(\tau)) < \tau$. Since $\tau < \tau_\infty$, so that $\tau < \tau(z_k)$ for large k, we get the contradiction. □

6.4. Boundary cut locus. Boundary normal coordinates.

Definition 6.12. The *boundary cut locus* ω is defined by

$$\omega = \{\gamma_z(\tau(z))\,;\, z \in \partial\mathcal{N}\},$$

where $\gamma_z(\tau(z))$ is the boundary cut point of z along the boundary normal geodesic $\gamma_z(t) = \exp_{\partial\mathcal{N}}(z,t)$ in Definition 6.8.

Recall that by Lemma 6.10, we have $d(\gamma_z(\tau(z)), z) = \tau(z)$. Let us investigate the structure of ω. We put

$$B(\mathcal{N}) = \bigcup_{z \in \partial\mathcal{N}} \{\gamma_z(t)\,;\, 0 \leq t < \tau(z)\}.$$

Lemma 6.13. *(1) $\mathcal{N} = B(\mathcal{N}) \cup \omega$, $B(\mathcal{N}) \cap \omega = \emptyset$.*
(2) ω is a closed set of measure 0. In particular, it has no interior points.
(3) $B(\mathcal{N})$ is an open set.

Proof. For any $x \in \mathcal{N}$, there exists $z_x \in \partial\mathcal{N}$ such that $d(x, z_x) = d(x, \partial\mathcal{N}) := s(x)$. Therefore $x = \gamma_{z_x}(s(x))$ (see Lemma 3.1). Let us prove $s(x) \leq \tau(z_x)$, where $\tau(z)$ is boundary cut function, see Definition 6.9. Indeed, if $s(x) > \tau(z_x)$, there exists $w \in \partial\mathcal{N}$ such that $d(x,w) < s(x)$, which is a contradiction, since $s(x) = d(x, \partial\mathcal{N})$.

Therefore, we have shown that, for any $x \in \mathcal{N}$, there exists $z_x \in \partial\mathcal{N}$ such that $x = \exp_{\partial\mathcal{N}}(z_x, d(x, \partial\mathcal{N}))$ and $d(x, \partial\mathcal{N}) \leq \tau(z_x)$. This proves $\mathcal{N} = B(\mathcal{N}) \cup \omega$.

The disjointness of B and ω is obvious. Since $\tau(z)$ is continuous, $U := \{(z, \tau(z))\,;\, z \in \partial\mathcal{N}\} \subset \partial\mathcal{N} \times \mathbf{R}_+$ has measure 0. Since $\exp_{\partial\mathcal{N}}(z,t)$ is continuous, $\omega = \exp_{\partial\mathcal{N}}(U)$ has measure 0. This implies that ω has no interior points and, since $\partial\mathcal{N}$ is compact, ω is compact. □

Example 6.14. (1) Let $\mathcal{N} = B^1 = \{|x| < 1\}$ equipped with the Euclidean metric. Then $\omega = \{0\}$, which is both an ordinary boundary cut point and the first focal point.
(2) Let $\mathcal{N}$ be the inside of an ellipse : $\mathcal{N} = \{(x,y) \in \mathbf{R}^2; x^2/a^2 + y^2/b^2 < 1\}$, $(a > b > 0)$ equipped with the Euclidean metric. Then $\omega = \{(x,0); |x| \leq (a^2 - b^2)/a\}$. The end points $(\pm(a^2 - b^2)/a, 0)$ are focal points, and all the points in the open interval $\{(x,0); |x| \leq (a^2 - b^2)/a\}$ are ordinary boundary cut points.

Based upon Lemma 6.13, we make the following definition.

Definition 6.15. The *boundary normal coordinates* is the map,

$$B(\mathcal{N}) = \mathcal{N} \setminus \omega \ni x \to (z(x), s(x)) \in \partial\mathcal{N} \times \mathbf{R}_+, \tag{6.21}$$

where $s(x)$ is the distance from x to $\partial\mathcal{N}$ and $z(x)$ is the unique point on $\partial\mathcal{N}$ which is the closest to x, i.e. $x = \gamma_{z(x)}(s(x))$.

7. Boundary distance coordinates

7.1. Conjugate point. The boundary cut locus is different from the standard notion of cut locus on the manifold without boundary. Therefore, we shall assume in this section that the manifold $\mathcal{N}$ is embedded in a complete manifold of the same dimension $\widetilde{\mathcal{N}}$, where $\widetilde{\mathcal{N}}$ has no boundary. Note that we can always construct $\widetilde{\mathcal{N}}$ taking it to be the Hopf double of $\mathcal{N}$ equipped with metric which is a smooth Seeley-Borel continuation across $\partial\mathcal{N}$.

Definition 7.1. Let $c(t)$ $(a \le t \le b)$ be a geodesic on $\widetilde{\mathcal{N}}$. Two points $c(a)$ and $c(b)$ are said to be *conjugate* along $c(t)$ if there exists a non-trivial Jacobi field $Y(t)$ along $c(t)$ such that $Y(a) = 0, Y(b) = 0$. We also say that $c(b)$ is conjugate to $c(a)$ along $c(t)$.

For $y \in \widetilde{\mathcal{N}}$, let $\gamma_{(y,v)}(t) = \exp_y(tv)$ be the unit speed geodesic starting from y with initial direction $v \in S_y(\widetilde{\mathcal{N}})$, where $S_y(\widetilde{\mathcal{N}}) = \{v \in T_y(\widetilde{\mathcal{N}})\,;\, |v|_g = 1\}$.

Lemma 7.2. *Let $c(t) = \gamma_{(y,v)}(t)$ be a unit speed geodesic on $\widetilde{\mathcal{N}}$. Then $c(t_0)$ is conjugate to y along $c(t)$ if and only if there exists $0 \neq \xi \in T_y(\widetilde{\mathcal{N}}) = T_{t_0v}(T_y(\widetilde{\mathcal{N}}))$ such that*

$$d\exp_y\Big|_{t_0v}\xi = 0.$$

For the proof, see e.g. [**9**], p. 17, or [**24**], Theorem 2.16.

Lemma 7.3. *Let $c(t)$ $(a \le t \le b)$ be a geodesic on $\widetilde{\mathcal{N}}$. If there exists $a < \tau < b$ such that $c(\tau)$ is conjugate to $c(a)$ along $c(t)$, there is another geodesic with end points $c(a)$ and $c(b)$ which is strictly shorter than the arclength, $b-a$, of the geodesic $c(t)$, $a \le t \le b$.*

For the proof, see e.g. [**24**], Theorem 2.11, or [**82**], p. 87.

Similary to the boundary cut function $\tau(z)$, we introduce (Riemannian) cut function, τ^R,

Definition 7.4. The *(Riemannian) cut function* $\tau^R, : S(\widetilde{\mathcal{N}}) \to \mathbf{R}_+$ is given by

$$\tau^R(y,v) = \sup_{t\ge 0}\left\{t\,;\, d(\gamma_{(y,v)}(t), y) = t\right\}. \tag{7.1}$$

Note that $d(y, \gamma_{(y,v)}(\tau^R(y,v)) = \tau^R(y,v)$. The point $\gamma_{(y,v)}(\tau^R(y,v))$ is called the *cut point* for y along the geodesic $\gamma_{(y,v)}(\cdot)$. This should not be confused with the boundary cut point of Definition 6.9, where we considered the distance to $\partial\mathcal{N}$.

Remark 7.5. Assume that $\mathcal{N} = \widetilde{\mathcal{N}} \setminus B(x_0,a)$, where $B(x_0,a)$ is a ball of radius $a > 0$ centered at x_0. Let

$$a < \min_{v\in S_{x_0}(\widetilde{\mathcal{N}})} \tau^R(x_0,v).$$

Parametrize the points on $\partial\mathcal{N} = \partial B(x_0,a)$ by v and observe that the normal geodesics to $\partial\mathcal{N}$, i.e. $\gamma_v(t)$ are actually the continuations of the geodesics $\gamma_{x_0,v}(t)$, namely, $\gamma_v(t) = \gamma_{x_0,v}(t+a)$. Therefore, the focal and boundary cut points along γ_v are actully the conjugate and Riemannian cut points along $\gamma_{x_0,v}$. This implies, due to Lemma 7.3, the validity of Lemma 6.5 for $\partial\mathcal{N} = \partial B(x_0,a)$.

Lemma 7.6. *The mapping $\tau^R(y,v) : S(\widetilde{\mathcal{N}}) \to \mathbf{R}_+ \cup \infty$ is continuous.*

This is proven in the same way as Lemma 6.11. See e.g. [**24**], Theorem 3.1, or [**82**], p. 98.

Lemma 7.7. *Let $z \in \partial\mathcal{N}$, and ν be the inner unit normal to $\partial\mathcal{N}$ at z. Then $\tau^R(z,\nu) > \tau(z)$.*

Proof. Assume that for some $z \in \partial\mathcal{N}$, $\tau^R(z,\nu) \le \tau(z)$. Note that, following our notations for the boundary normal geodesics and geodesics starting at z, we have $\gamma_z(t) = \gamma_{(z,\nu)}(t)$ for $t > 0$. Take $x = \gamma_{(z,\nu)}(\tau^R(z,\nu))$ and $\xi = -\dot\gamma_{(z,\nu)}(t)$ at $t = \tau^R(z,\nu)$. By duality, $\tau^R(x,\xi) = \tau^R(z,\nu)$. We extend $\gamma_{(x,\xi)}(t)$ on the interval $[0, \tau^R(x,\xi)+\delta]$ with $\delta > 0$. Since $\dot\gamma_{(x,\xi)}(\tau^R(z,\nu)) = -\nu$, by choosing $\delta > 0$ small enough, we can assume that, if $\tau^R(x,\xi) < s < \tau^R(x,\xi)+\delta$, $\gamma_{(x,\xi)}(s)$ is outside the original $\mathcal{N}$. Let $y(t) = \gamma_{(x,\xi)}(t + \tau^R(x,\xi))$. Then, for small t, $d(y(t), z) = d(y(t), \partial\mathcal{N}) = t$.

Note that, by the definition of τ^R, for $t > 0$ $d(y(t),x) < t + \tau^R(x,\xi)$. Therefore, there is a shortest geodesic $\mu(s)$ from $y(t)$ to x with $\mu(\overline{s}) = x$ and $\overline{s} < \tau^R(x,\xi) + t$. Let w be the last point on μ where μ crosses $\partial\mathcal{N}$.

By triangle inequality,

$$\overline{s} \ge d(y(t), w) + d(\omega, x) \ge t + d(w,x) \ge t + \tau^R(z,\nu),$$

where in the last step we use the assumption $\tau^R(z,\nu) \le \tau(z)$. This is a contradiction. $\square$

Let $z \in \partial\mathcal{N}$ and γ_z be the boundary normal geodesic from z. Then, by Lemma 7.7, there exists $\epsilon > 0$ such that for $t < \tau(z) + \epsilon$, $\gamma_z(\cdot)$ is still the shortest geodesic (lying inside $\mathcal{N}$) from z to $\gamma_z(t)$.

7.2. Hamilton's equation. Let $(g^{ij}) = (g_{ij})^{-1}$ be the contravariant metric tensor, and define a C^∞-function on $T^*(M)$ by $H(x,\xi) = \frac{1}{2} g^{ij}(x)\xi_i\xi_j$. As has been mentioned in Subsection 1.4 in Chap. 1, the equation of geodesic can be rewritten as Hamiltons's canonical equation

$$\begin{cases} \dfrac{dx^i}{dt} = \dfrac{\partial H}{\partial \xi_i} = g^{ij}(x)\xi_j, \\ \dfrac{d\xi_i}{dt} = -\dfrac{\partial H}{\partial x^i} = -\dfrac{1}{2}\left(\dfrac{\partial g^{kl}(x)}{\partial x^i}\right)\xi_k\xi_l. \end{cases} \tag{7.2}$$

Fix a point $y \in \mathcal{N}$ and let $x(t)$, $\xi(t)$ be the solution to (7.2) with initial data $x(0) = y$, $\xi(0) = \xi_0$, where ξ_0 satisfies $g^{ij}(y)\xi_{0i}\xi_{0j} = 1$. Then, by the energy conservation law,

$$g^{ij}(x(t))\xi_i(t)\xi_j(t) = 1. \tag{7.3}$$

Let $v^i(t) = dx^i(t)/dt = g^{ij}(x(t))\xi_j(t)$, and put $v(t) = (v^1(t), \cdots, v^n(t))$, $v_0 = v(0)$. Then $x(t)$ is a geodesic starting from y with initial direction v_0. Assume that, for $U \subset S_y(\mathcal{N})$, $0 < t_1 < t_2$, the map : $U \times (t_1,t_2) \ni (v_0,t) \to x(t)$ is a diffeomorphism. Then t and v_0 become smooth functions of x depending (smoothly) on the parameter y : $t = t(x,y)$, $v_0 = v_0(x,y)$. Hence, so is $\xi = \xi(x,y)$. Since $t(x,y) = \int_y^x \xi_i dx^i$, we have

$$\frac{\partial t(x,y)}{\partial x^i} = \xi_i(x,y). \tag{7.4}$$

This equality can be rewritten as

$$(\mathrm{grad}_x t(x,y))^i = g^{ij}(x)\frac{\partial t}{\partial x^j}(x,y) = v^i(x,y). \tag{7.5}$$

Note also that, if $t_2 < \tau^R(y, v_0)$ and U is a small neighborhood of v_0, the above map is, indeed, a diffeomorphism and $t(x,y) = d(x,y)$.

7.3. Boundary distance coordinates. Near the cut locus, we cannot use the boundary normal coordinates. However, the *boundary distance coordinates* constructed below can be used everywhere on $\mathcal{N}^{int} = \mathcal{N} \setminus \partial\mathcal{N}$.

Lemma 7.8. *For any $x_0 \in \mathcal{N}^{int}$, there exist points $z_1, \cdots, z_n \in \partial\mathcal{N}$ such that the functions $(\rho_1(x), \cdots, \rho_n(x))$, where $\rho_i(x) = d(x, z_i)$, give local coordinates in a small neighborhhood of x_0.*

Proof. Let $z_0 \in \partial\mathcal{N}$ be a point nearest to x_0, i.e. $x_0 = \gamma_{z_0}(s_0)$, where $s_0 = d(x_0, z_0) = d(x_0, \partial\mathcal{N})$. If there are several such points, one can take any of them. Let $v_0 = -\dot\gamma_{z_0}(t)|_{t=t_0} \in S_{x_0}(\mathcal{N})$ so that $\gamma_{(x_0,v_0)}(s_0) = z_0$. By Lemma 7.7, we have $s_0 < \tau^R(z_0, \nu(z_0)) = \tau^R(x_0, v_0)$. By Lemma 7.2, $d\exp_{x_0}\big|_{s_0v_0} : T_{s_0v_0}(T_{x_0}(\mathcal{N})) = T_{x_0}(\mathcal{N}) \to T_{z_0}(\mathcal{N})$ is non-singular.

Consider curves $z_i(t), i = 1, \cdots, n-1$, in $\partial\mathcal{N}$ such that $z_i(0) = z_0$ and the vectors $\dot z_i(0)$, $i = 1, \cdots, n-1$, form an orthonormal basis of $T_{z_0}(\partial\mathcal{N})$. Let $v_i = (d\exp_{x_0}\big|_{s_0v_0})^{-1}\dot z_i(0)$ for $i = 1, \cdots, n-1$, and $v_n = v_0$. Then v_i, $i = 1, \cdots, n$, form a basis of $T_{x_0}(\mathcal{N})$. Furthermore, $c_i(s) := (\exp_{x_0})^{-1}(z_i(s)) \in T_{x_0}(\mathcal{N})$, $i = 1, \cdots, n-1$, satisfy $c_i(0) = s_0v_0$ and $\dot c_i(0) = v_i$. For $i = 1, \cdots, n-1$, let $z_i = z_i(\epsilon)$ for a sufficiently small ϵ and $z_n = z_0$. We define $\rho_i(x) = d(x, z_i)$, $i = 1, \cdots, n$. Then, by (7.5), $\mathrm{grad}_x\rho_i(x_0) = -\dot c_i(\epsilon)/|\dot c_i(\epsilon)|_g$, $i = 1, \ldots, n$, are linearly independent. The inverse function theorem completes the proof. $\square$

Example 7.9. Let $\mathcal{N}$ be a Euclidean sphere : $\mathcal{N} = \{|x| \le 1\}$. Then the boundary normal coordinates are essentially polar coordinates with center at the with $r \to 1 - r$, $r \le 1$. The center is the cut locus. To define the local coordinate around the origin, we have only to take n points $w_1, \cdots, w_n$ on $\partial\mathcal{N}$ which are linearly independent, and $\rho_i(x) = |x - w_i|$.

7.4. Reconstruction of the metric. The following lemma is a key trick to reconstruct the Riemannian metric.

Lemma 7.10. *Let $x_0 \in \mathcal{N}$. Then we can recover the metric tensor $g_{ij}(x)$ from the boundary distance functions $\partial\mathcal{N} \ni w \to d(x, w)$.*

Proof. For $x_0 \in \mathcal{N}$, let $z_0 \in \partial\mathcal{N}$ be such that $d(x_0, z_0) = d(x_0, \partial\mathcal{N})$. Then there is a small open cone of directions $C \subset S_{x_0}(\mathcal{N})$ such that the geodesic starting from x_0 with initial direction in C hits $\partial\mathcal{N}$ transversally in a neighborhood W_0 of z_0. Using the proof of Lemma 7.8, this means that the directions of the shortest geodesics from $z \in W_0$ to x_0 form the cone $-C$ in $S_{x_0}(\mathcal{N})$.

Let U be a small neighborhood of x_0. For $x \in U$ and $z \in W_0$, we consier $d(x, z)$. Passing to Hamilton's equation, we have $d(x, z) = t(x, z)$, where $t(x, z)$ is defined in Subsection 7.2. By (7.3), we have

$$g^{ij}(x_0)\xi_i(x_0, z)\xi_j(x_0, z) = 1.$$

We can compute $\xi_i(x_0, z)$ from (7.4): $\xi_i(x_0, z) = \dfrac{\partial d}{\partial x^i}(x_0, z)$. Let z vary on W_0. Then, since $\xi(x_0, z)$ varies over an open set in $S^*_{x_0}(\mathcal{N})$, the unit sphere in the cotangent space $T^*_{x_0}(\mathcal{N})$, we can recover the contravariant metric tensor $g^{ij}(x_0)$. □

8. Reconstruction of $R(\mathcal{N})$ from BSP

In this section, we shall prove that if two manifolds $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$ have the same BSP, the space of boundary distance functions $R(\mathcal{N}^{(1)})$ and $R(\mathcal{N}^{(2)})$ coincide. We use the expression "BSP determines the quantity A" to mean the following: Let $A^{(1)}$ and $A^{(2)}$ be the quantities associated to the manifolds $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$, respectively. Then if $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$ have the same BSP, $A^{(1)} = A^{(2)}$ holds.

8.1. Projection to the domain of influence. Recall that, for a subset $\Gamma \subset \partial\mathcal{N} \subset \mathcal{N}$ and $\tau > 0$, we put

$$\mathcal{N}(\Gamma, \tau) = \{x \in \mathcal{N}\,;\, d(x, \Gamma) \le \tau\}.$$

We also define for $z \in \partial\mathcal{N}$

$$\mathcal{N}(z, \tau) = \{x \in \mathcal{N}\,;\, d(x, z) \le \tau\}.$$

Let $\chi_{\mathcal{N}(\Gamma,\tau)}(x)$ be the characteristic function of $\mathcal{N}(\Gamma, \tau)$. We define a projection on $L^2(\mathcal{N})$ by

$$P_{\Gamma,\tau} f(x) = \chi_{\mathcal{N}(\Gamma,\tau)}(x) f(x) \in L^2(\mathcal{N}(\Gamma, \tau)), \quad f \in L^2(\mathcal{N}). \tag{8.1}$$

Let $u^f(t)$ be the solution to IBVP (2.1).

Lemma 8.1. *Let $f \in C_0^\infty(\partial\mathcal{N} \times (0, \infty))$ and $\tau, t > 0$. Let $\Gamma \subset \partial\mathcal{N}$ be an open set. Then one can choose a sequence $f_j \in C_0^\infty(\Gamma \times (0, \tau))$ satisfying $u^{f_j}(t) \to P_{\Gamma,\tau} u^f(t)$ by using only BSP.*

Proof. Let us recall an elementary fact on the projection in a Hilbert space $\mathcal{H}$. Let P be a projection onto a closed subspace S of $\mathcal{H}$. For $u \in \mathcal{H}$, take $v_n \in S$ such tht $\lim_{n\to\infty} \|u - v_n\| = \inf_{v\in S} \|u - v\| = \|(1 - P)u\|$. Then $v_n \to Pu$.

Using Theorem 4.6, we have

$$\begin{aligned} \|u^f(t)\|^2 - \|P_{\Gamma,\tau} u^f(t)\|^2 &= \|(1 - P_{\Gamma,\tau}) u^f(t)\|^2 \\ &= \inf_{\eta \in C_0^\infty(\Gamma\times(0,\tau))} \|u^f(t) - u^\eta(\tau)\|^2. \end{aligned} \tag{8.2}$$

Noting that

$$\|u^f(t) - u^\eta(\tau)\|^2 = \|u^f(t)\|^2 - 2\mathrm{Re}(u^f(t), u^\eta(\tau)) + \|u^\eta(\tau)\|^2,$$

one can compute the right-hand side of (8.2) by Corollary 2.2. We then choose a sequence $f_j \in C_0^\infty(\Gamma \times (0, \tau))$ which attains the infimum of (8.2). Then $u^{f_j}(\tau) \to P_{\Gamma,\tau} u^f(t)$. This procedure depends only on BSP. □

Lemma 8.2. *Let $f, h \in C_0^\infty(\partial\mathcal{N} \times (0, \infty))$ and $\tau_1, \tau_2, t, s > 0$.*
(1) Let $\Gamma_1, \Gamma_2 \subset \partial\mathcal{N}$ be open sets. Then BSP determines the inner product

$$\left(P_{\Gamma_1,\tau_1} u^f(t), P_{\Gamma_2,\tau_2} u^h(s)\right)_{L^2(\mathcal{N})}.$$

(2) Let $z_1, z_2 \in \partial\mathcal{N}$. Then BSP determines the inner product

$$\left(P_{z_1,\tau_1} u^f(t), P_{z_2,\tau_2} u^h(s)\right)_{L^2(\mathcal{N})}.$$

Proof. (1) is an obvious consequence of Lemma 8.1. Taking open sets $\Gamma_1, \Gamma_2 \subset \partial\mathcal{N}$ shrinking to $z_1, z_2 \in \partial\mathcal{N}$, and applying Lebesgue's convergence theorem, we obtain (2). $\square$

8.2. Domain of influence and $R(\mathcal{N})$. Following [**78**], we can identify the boundary normal geodesic from BSP.

Lemma 8.3. *Let $\gamma_z(\cdot)$ be the boundary normal geodesic starting from $z \in \partial\mathcal{N}$, and $s > 0$. Then the following 3 assertions are equivalent.*
(1) $d(\gamma_z(s), z) = d(\gamma_z(s), \partial\mathcal{N})$.
(2) For any $\epsilon > 0$ and any neighborhood $\Gamma \subset \partial\mathcal{N}$ of z, the interior of $\big(\mathcal{N}(\Gamma, s) \setminus \mathcal{N}(\partial\mathcal{N}, s-\epsilon)\big) \neq \emptyset$.
(3) For any neighborhood $\Gamma \subset \partial\mathcal{N}$ of z, there exists $h \in C_0^\infty(\Gamma \times (0, s))$ such that $\|u^h(s)\| > \|P_{\partial\mathcal{N}, s-\epsilon} u^h(s)\|$.

Proof. Suppose (1) holds, and consider the open ball $B_{\epsilon/2}(x_\epsilon)$, where $x_\epsilon = \gamma_z(s - \epsilon/2)$. Clearly $B_{\epsilon/2}(x_\epsilon) \subset \mathcal{N}(\Gamma, s)$. Let us show $B_{\epsilon/2}(x_\epsilon) \cap \mathcal{N}(\partial\mathcal{N}, s-\epsilon) = \emptyset$. Indeed, if there exists $x \in B_{\epsilon/2}(x_\epsilon) \cap \mathcal{N}(\partial\mathcal{N}, s-\epsilon)$, Then

$$d(x_\epsilon, \partial\mathcal{N}) \leq d(x_\epsilon, x) + d(x, \partial\mathcal{N}) < \epsilon/2 + (s - \epsilon) = s - \epsilon/2,$$

which contradicts (1). Hence (2) holds.

Suppose (2) holds. Take a sequence $\epsilon_n \to 0$ and a neighborhood $\Gamma_n \subset \partial\mathcal{N}$ of z of $\mathrm{diam}\,(\Gamma_n) < \epsilon_n$. There exists a sequence $x_n, \delta_n \in (0, \epsilon_n/2)$ such that $B_{\delta_n}(x_n) \subset \mathcal{N}(\Gamma_n, s) \setminus \mathcal{N}(\partial\mathcal{N}, s-\epsilon_n)$. Up to taking a subsequence, $x_n \to \overline{x} \in \mathcal{N}$. Since $s - \epsilon_n < d(x_n, \partial\mathcal{N}) \leq d(x_n, \Gamma_n) \leq s$, we have $d(\overline{x}, \partial\mathcal{N}) = d(\overline{x}, z) = s$. This implies that $\overline{x} = \gamma_z(s)$, hence (1) holds.

Suppose (2) holds. Let χ be the characteristic function of $\mathcal{N}(\Gamma, s) \setminus \mathcal{N}(\partial\mathcal{N}, s-\epsilon)$. Then $\|\chi\|_{L^2(\mathcal{N})} > 0$. Approximating χ by $u^h(s)$, where $h \in C_0^\infty(\Gamma \times (0, s))$, we get (3).

Evidently, (3) implies (2). $\square$

Lemma 8.4. *Let $\gamma_w(\cdot)$ be the boundary normal geodesic starting from $w \in \partial\mathcal{N}$, and $s > 0$ be such that $d(\gamma_w(s), w) = d(\gamma_w(s), \partial\mathcal{N})$. Let $z \in \partial\mathcal{N}$ and $t > 0$. Then the following 3 assertions are equivalent.*
(1) $t > d(\gamma_w(s), z)$.
(2) There exist a neighborhood $\Gamma \subset \partial\mathcal{N}$ of w and $\epsilon > 0$ such that

$$\mathcal{N}(\Gamma, s) \subset \mathcal{N}(\partial\mathcal{N}, s-\epsilon) \cup \mathcal{N}(z, t-\epsilon).$$

(3) There exist a neighborhood $\Gamma \subset \partial\mathcal{N}$ of w and $\epsilon > 0$ such that for any $h \in C_0^\infty(\Gamma \times (0, s))$

$$\|u^h(s)\|^2 = \|P_{\partial\mathcal{N}, s-\epsilon} u^h(s)\|^2 + \|P_{z, t-\epsilon} u^h(s)\|^2 - (P_{\partial\mathcal{N}, s-\epsilon} u^h(s), P_{z, t-\epsilon} u^h(s)).$$

Proof. Assume (1) holds. If (2) does not hold, there exist a sequence $\Gamma_n \subset \partial\mathcal{N}$ shrinking to $\{w\}$ and $\epsilon_n \to 0$, such that $\mathcal{N}(\Gamma_n, s) \not\subset \mathcal{N}(\partial\mathcal{N}, s-\epsilon_n) \cup \mathcal{N}(z, t-\epsilon_n)$. Then there exists $x_n \in \mathcal{N}$ such that $d(x_n, \partial\mathcal{N}) > s - \epsilon_n$, $d(x_n, z) > t - \epsilon_n$, and $d(x_n, \Gamma_n) \leq s$. Then, up to subsequence, $x_n \to \overline{x}$, with $d(\overline{x}, \partial\mathcal{N}) = d(\overline{x}, w) = s$, and $d(\overline{x}, z) \geq t$. Therefore $\overline{x} = \gamma_w(s)$, which by (1) implies $d(\gamma_w(s), z) = d(\overline{x}, z) < t$. This contradiction shows that (1) implies (2).

Suppose (2) holds. Since the condition $d(\gamma_w(s), w) = d(\gamma_w(s), \partial\mathcal{N})$ implies that $\gamma_w(s) \notin \mathcal{N}(\partial\mathcal{N}, s-\epsilon)$, then $\gamma_w(s) \in \mathcal{N}(z, t-\epsilon)$. Thus, $d(\gamma_w(s), z) \leq t - \epsilon$, proving (1).

Let $P = P_{\partial\mathcal{N},s-\epsilon}$, $Q = P_{z,t-\epsilon}$. Using (8.1), we see that $R = P + Q - PQ$ is a projection onto $L^2(\mathcal{N}(\partial\mathcal{N}, s-\epsilon) \cup \mathcal{N}(z, t-\epsilon))$. Then (2) is equivalent to

$$u^h(s) = Ru^h(s), \quad \forall h \in C_0^\infty(\Gamma \times (0, s)).$$

Since R is a projection, this is equivalent to

$$\|u^h(s)\|^2 = \|Ru^h(s)\|^2, \quad \forall h \in C_0^\infty(\Gamma \times (0, s)).$$

which is equivalent to (3). □

8.3. Main theorem. We are now in a position to prove the following theorem.

Theorem 8.5. *Let $(\mathcal{N}, g)$ be a connected Riemannian manifold with compact boundary. Suppose we are given the boundary spectral projections of the Neumann Laplacian on $\mathcal{N}$. Then these data determine $(\mathcal{N}, g)$ uniquely.*

Proof. We take $w \in \partial\mathcal{N}$. By Lemma 8.2 and Lemma 8.3 (3), we can determine, by using BSP, whether or not $\gamma_w([0, s])$ is a shortest geodesic to $\partial\mathcal{N}$. In particular, this detemines the boundary cut function $\tau(w)$.

By Lemma 8.4, for $s \le \tau(w)$, we can compute, by using BSP, $d(\gamma_w(s), z)$ for any $z \in \partial\mathcal{N}$. Thus, for any $w \in \partial\mathcal{N}$ and $s \le \tau(w)$, we associate, using BSD, a function $r^{(w,s)}(\cdot) \in C(\partial\mathcal{N})$:

$$r^{(w,s)}(z) = d(\gamma_w(s), z), \quad z \in \partial\mathcal{N}.$$

Note, see (5.1), that $r^{(w,s)}(\cdot)$ is the boundary distance function corresponding to $x = \gamma_w(s)$.

Lemma 6.13 shows that, when w runs over $\partial\mathcal{N}$ and s runs over $[0, \tau(w)]$, then $r^{(w,s)}(z)$ runs over the whole $R(\mathcal{N}) \subset C(\partial\mathcal{N})$. Thus, BSP determines $R(\mathcal{N})$.

We then recover the topology of $\mathcal{N}$ by Lemma 5.1. By Lemma 7.10, we recover the metric by BSP. □

We note that the uniqueness in the above Theorem means "up to an isometry". We have used the generalized Fourier transform to represent BSP. However, in the above proof, we have actually used the hyperbolic Neumann-to-Dirichlet map and this can be controlled under milder assumptions. In fact, the BC-method also works for the manifold of bounded geometry, i.e. with the assumption of uniform injective radius of Riemannian normal coordinates, and the boundedness of curvature tensor. See [**78**].

9. Wave fronts and $R(\mathcal{N})$

As has been seen above, the construction of boundary distance functions from BSP is the step where the geodesic is traced using Blagovestchenski identity for the solutions to IBVP, providing an interplay between geometry and partial differential equations. Therefore, it is of interest to try other ideas. In this section, we explain the method which deals with the wave front of solution $u^f(t)$ to IBVP (4.1).

(i) *Controlled subspaces.* By the finite propagation property, we have

$$\operatorname{supp} u^f(\cdot, t) \subset \mathcal{N}(\Gamma, t) := \{x \in \mathcal{N}\,;\, d\,(x, \Gamma) \le t\}.$$

Recall that the closure in $L^2(\mathcal{N})$ of $\{u^f(\cdot, t)\,;\, f \in C_0^\infty(\Gamma \times (0, t)\}$ is $L^2(\mathcal{N}(\Gamma, t))$.

We define a unitary operator

$$\mathcal{F} = (\mathcal{F}_c^{(+)}, \mathcal{F}_p) : L^2(\mathcal{N}) \to L^2((0, \infty); \mathbf{h}; dk) \oplus \mathbf{C}^d,$$

where $\mathcal{F}_c^{(+)}$ is the generalized Fourier transform, and $\mathcal{F}_p$ is the spectral representation associated with the point spectrum for H:

$$\mathcal{F}_p : L^2(\mathcal{N}) \ni u = \sum_i a_i\varphi_i(x) \to (a_1, a_2, \cdots) \in \mathbf{C}^d,$$

where d is the dimension of the point spectral subspace of H. If $d = \infty$, $\mathbf{C}^d = l^2$. If $\mathcal{N}$ is compact, $\mathcal{F}_c^{(+)}$ is absent.

(ii) *Projections.* Let $P_{\Gamma,t}$ be the orthogonal projection

$$P_{\Gamma,t} : L^2(\mathcal{N}) \ni u \to \chi_{\mathcal{N}(\Gamma,t)}(x)u(x) \in L^2(\mathcal{N}(\Gamma,t)),$$

$\chi_{\mathcal{N}(\Gamma,t)}(x)$ being the characteristsic function of the set $\mathcal{N}(\Gamma,t)$. Passing to the Fourier transform, we have

$$\mathcal{F}P_{\Gamma,t} = \mathcal{P}_{\Gamma,t}\mathcal{F},$$

where $\mathcal{P}_{\Gamma,t}$ is the orthogonal projection :

$$\mathcal{P}_{\Gamma,t} : L^2((0,\infty); \mathbf{h}; dk) \oplus \mathbf{C}^d \to \mathcal{L}^2(\Gamma,t).$$

(iii) *Layers.* It is obvious that

$$L^2(\mathcal{N}(\Gamma,t_-)) \subset L^2(\mathcal{N}(\Gamma,t_+)), \quad 0 \le t_- < t_+,$$
$$\mathcal{L}^2(\Gamma,t_-) \subset \mathcal{L}^2(\Gamma,t_+), \quad 0 \le t_- < t_+.$$

Take $\mathcal{L}^2(\Gamma,t_+,t_-) = \mathcal{L}^2(\Gamma,t_+) \ominus \mathcal{L}^2(\Gamma,t_-)$, which are the Fourier transforms of functions with support in the *shell type layer* or *approximate wave front*

$$\mathcal{N}(\Gamma,t^+) \setminus \mathcal{N}(\Gamma,t^-) := \mathcal{S}h(\Gamma,t^+,t^-).$$

Take (Γ_1,t_1^+,t_1^-) and (Γ_2,t_2^+,t_2^-). Then

$$\begin{aligned} &\mathcal{L}^2(\Gamma_1,t_1^+,t_1^-) \cap \mathcal{L}^2(\Gamma_2,t_2^+,t_2^-) \\ &= \mathcal{F}\{a\,;\, \operatorname{supp} a \subset \mathcal{S}h(\Gamma_1,t_1^+,t_1^-) \cap \mathcal{S}h(\Gamma_2,t_2^+,t_2^-)\}. \end{aligned} \tag{9.1}$$

(iv) *Approximate distance functions.* We take Γ_i, $t_i^\pm$, $i = 1,\cdots,N$, and consider $\cap_{i=1}^N \mathcal{L}^2(\Gamma_i,t_1^+,t_1^-)$, which is the Fourier image of functions with support in the intersection of layers. If the intersection of layers has measure 0, then $\cap_{i=1}^N \mathcal{L}^2(\Gamma_i,t_1^+,t_1^-) = \{0\}$. If this intersection has positive measure, then $\dim\big(\cap_{i=1}^N \mathcal{L}^2(\Gamma_i,t_1^+,t_1^-)\big) = \infty$. In particular, there is $x \in \mathcal{N}$ such that $t_i^- \le d(x,\Gamma_i) \le t_i^+$.

Divide $\partial\mathcal{N}$ into a large number, which is denoted by $N(\epsilon)$, of Γ_i with $\operatorname{diam}\Gamma_i < \epsilon$. For any vector $\mathbf{n} = (n_1,\cdots,n_{N(\epsilon)}) \in \mathbf{Z}_+^{N(\epsilon)}$, put $t_i^- = (n_i-1)\epsilon$, $t_i^+ = n_i\epsilon$. Construct $\cap_i\mathcal{L}^2(\Gamma_i,t_i^+,t_i^-)$. We call $\mathbf{n}$ admissible, if $\cap_i\mathcal{L}^2(\Gamma_i,t_i^+,t_i^-) \ne \{0\}$. For any admissible $\mathbf{n}$, we associate a function

$$\kappa_{\mathbf{n}} \in L^\infty(\partial\mathcal{N}), \quad \kappa_{\mathbf{n}}(z) = n_i\epsilon, \quad \text{for} \quad z \in \Gamma_i.$$

Take all these $\kappa_{\mathbf{n}}(z)$ for all admissible $\mathbf{n}$, and get a finite number of $L^\infty(\partial\mathcal{N})$ functions. They are roughly distances from various points in $\mathcal{N}$ to $\partial\mathcal{N}$. Let us denote the set of these functions as $R^\epsilon(\mathcal{N})$.

(v) *Boundary distance representation of* $\mathcal{N}$. Recall that, see §5.1, for any $x \in \mathcal{N}$, there is the boundary distance function $r_x(z)$, $z \in \partial\mathcal{N}$,

$$r_x(z) = d(x,z).$$

This defines the map

$$R : \mathcal{N} \to C^{0,1}(\partial\mathcal{N}) \subset L^\infty(\partial\mathcal{N}), \quad R(x) = r_x(\cdot).$$

Let $R(\mathcal{N})$ be the image of $\mathcal{N}$ by this map. Then the Hausdorff distance in $L^\infty(\partial\mathcal{N})$ between $R(\mathcal{N})$ and $R^\epsilon(\mathcal{N})$ is estimated as

$$d_H(R(\mathcal{N}), R^\epsilon(\mathcal{N})) < 3\epsilon. \tag{9.2}$$

In fact, since $(n_i-1)\epsilon \le d(x,\Gamma_i) \le n_i\epsilon$ and $\operatorname{diam}\Gamma_i \le \epsilon$, we have

$$|d(x,z) - n_i\epsilon| \le 2\epsilon, \quad z \in \Gamma_i,$$

for all $x \in \cap\mathcal{S}h(\Gamma_i, n_i\epsilon, (n_i-1)\epsilon)$. As, for any $x \in \mathcal{N}$, there is $\tilde{x} \in \cap\mathcal{S}h(\Gamma_i, n_i\epsilon, (n_i-1)\epsilon)$ with $d(x,\tilde{x}) < \epsilon$, this proves (9.2).

In summary, we have shown the following lemma.

Lemma 9.1. *For any $\epsilon > 0$, we can construct, from BSP, a finite set $R^\epsilon(\mathcal{N}) \subset L^\infty(\partial\mathcal{N})$, such that $d_H(R(\mathcal{N}), R^\infty(\mathcal{N})) < 3\epsilon$. Taking $\epsilon \to 0$, we obtain the boundary distance representation $R(\mathcal{N})$ of $\mathcal{N}$.*

10. Propagation of singularities and $R(\mathcal{N})$

The singularities of solutions to the wave equation on Riemannian manifolds propagate along the geodesics. Using this property, we can determine the boundary distance function from BSP. The tool we use is the Gaussian beams which are complex valued asymptotic solutions to the wave equation in $\mathcal{N} \times \mathbf{R}$ having the following property: A Gaussian beam is concentrated near a light ray $(\gamma(t), t)$, where $\gamma(t)$ is a unit speed geodesic. For any t, the profile of the Gaussian beam is close to Gaussian, with its peak at $x = \gamma(t)$. Therefore, it is a wave packet moving along the geodesic. Since whole procedure requires long computations, we only give the sketch here. The details can be found in [**77**]. The exposition of [**114**] is a good introduction to the theory of Gaussian beams.

The *Gaussian beam* is an asymptotic solution to the wave equation of the form

$$U_\epsilon(x,t) = (\pi\epsilon)^{-n/4} \exp\left(-\frac{\theta(x,t)}{i\epsilon}\right) \sum_{j=0}^{\infty} (i\epsilon)^j u_j(x,t), \tag{10.1}$$

where the phase function has the following property:

$$\operatorname{Im}\theta(\gamma(t),t) = 0, \quad \operatorname{Im}\theta(x,t) \ge C_0 d(x,\gamma(t))^2, \tag{10.2}$$

where $\gamma(t)$ is a geodesic associated with U_ϵ. The fact that U_ϵ is an asymptotic solution means that, if we take a finite sum,

$$U_\epsilon^{(N)}(x,t) = (\pi\epsilon)^{-n/4} \exp\left(-\frac{\theta(x,t)}{i\epsilon}\right) \sum_{j=0}^{N} (i\epsilon)^j u_j(x,t),$$

then, for any given time interval $[0,T]$, there exists a constant $C_T > 0$ such that $U_\epsilon^{(N)}(x,t)$ satisfies

$$\left|(\partial_t^2 - \Delta_g) U_\epsilon^{(N)}(x,t)\right| \le C_T \epsilon^{\alpha(N)}, \quad \text{on} \quad \mathcal{N} \times [0,T], \tag{10.3}$$

$$\alpha(N) \to \infty, \quad \text{for} \quad N \to \infty.$$

Fixing boundary normal coordinates, we consider in the half-space $\mathbf{R}^n_+ = \{x = (z, x_n)\,;\, z \in \mathbf{R}^{n-1}, x_n > 0\}$. For $z_0 \in \mathbf{R}^{n-1}$ and $t_0 > 0$, and we put the following highly oscillatory data on the boundary:

$$f_\epsilon(z,t) = (\pi\epsilon)^{-n/4} \chi_0(z,t) \exp\left(-\frac{\Theta(z,t)}{i\epsilon}\right), \tag{10.4}$$

where $\epsilon > 0$ is a small parameter, $\chi_0(z,t)$ is a smooth cut-off function near (z_0, t_0) and

$$\Theta(z,t) = -(t - t_0) + \frac{1}{2}\big(H_0(z - z_0), z - z_0\big) + \frac{i}{2}(t - t_0)^2, \tag{10.5}$$

$(\ , \ )$ being the Euclidean inner product, H_0 a complex symmetric matrix with a positive definite imaginary part.

Since we are taking boundary normal coordinates, the Riemannian metric becomes $ds^2 = g_{ij}(x)dz^i dz^j + (dx^n)^2$, and the boundary normal geodesic emanating from z_0 at time $t = t_0$ is $\gamma_{z_0}(t) = (z_0, t - t_0)$. Then for any given z_0, t_0, H_0 and V, one can construct the Gaussian beam (10.1) as follows:

(i) Let $l(z_0)$ be the time when the normal geodesic starting from z_0 at time 0 hits the boundary. Then the Gaussian beam is constructed on the time interval $I(z_0) = [0, t_0 + l(z_0))$.
(ii) It concentrates along the geodesic $\gamma_{z_0}(t) = (z_0, t - t_0)$, i.e. (10.2) is satisfield for $\gamma(t) = \gamma_{z_0}(t)$ on $I(z_0)$.
(iii) Its phase function and the amplitude functions satisfy

$$\theta(z, 0, t) \approx \Theta(z, 0), \quad u_j(z, 0, t) \approx \delta_{j0},$$

where $f(z) \approx g(z)$ means $\partial_z^\alpha(f(z) - g(z)) = 0$, $\forall \alpha$, at $z = z_0$, and

$$(\partial_t\theta)^2 - g_{ij}(x)(\partial_i\theta)(\partial_j\theta) \asymp 0,$$

$$L_\theta u_n \asymp (\partial_t^2 - \Delta_g)u_{n-1}, \quad u_{-1} = 0,$$

where $L_\theta = 2(\partial_t\theta)\partial_t - 2g^{ij}(\partial_i\theta)\partial_j + (\partial_t^2 - \Delta_g)\theta$, $\partial_j = \partial/\partial x^j$, and $f(x) \asymp g(x)$ means $\partial_x^\alpha(f(x) - g(x)) = 0$, $\forall\alpha$, at $x = \gamma_{z_0}(t)$ on $I(z_0)$.

Let $u_\epsilon(t)$ be the solution to IBVP (4.1) with f repalced by $f_\epsilon(z,t)$ of (10.4). Then as can be checked easily

$$\|u_\epsilon(t) - U_\epsilon^{(N)}(t)\| \leq C_N \epsilon^{\alpha(N)}.$$

Using this Gaussian beam one can prove the following lemma (see Corollary 3.25 of [**77**]).

Lemma 10.1. *For any $z_0 \in \partial\mathcal{N}$, $t_0 < t < t_0 + l(z_0)$ and $\tau > 0$, we have*

$$\lim_{\epsilon\to 0}(P_{y,\tau}u_\epsilon(t), u_\epsilon(t)) = \begin{cases} \alpha(t), & \text{if} \quad d(\gamma_{z_0}(t), y) < \tau, \\ 0, & \text{if} \quad d(\gamma_{z_0}(t), y) > \tau, \end{cases}$$

where $\alpha(t) > 0$.

Therefore we can compute $d(\gamma_{z_0}(t), y)$ from BSP.

11. Eigenfunction coordinates

11.1. Regularity of the metric. Let us discuss regularity problems for the metric. For the details, see [**3**]. If $g_{ij} \in C^{k,\alpha}$, the distance is locally $C^{k-1,\alpha}$. Then g_{ij} in distance coordinates is only $C^{k-2,\alpha}$, since the Jacobian is involved. As regard to this regularity loss problem, a nice choice is the *harmonic coordinates* $X^i(x)$, $i = 1, \cdots, n$, such that $\Delta_g X^i = 0$. The feature of these harmonic coordinates is that they are the best possible for smoothness. In fact, assume that, in some coordinates $(x^1, \cdots, x^n)$, g_{ij} is $C^{k,\alpha}$. Then $X^j(x)$, $j = 1, \ldots, n$, are $C^{k+1,\alpha}$, which implies that g_{ij} is $C^{k,\alpha}$ in the coordinates $(X^1, \cdots, X^n)$. Another important feature is that, in the harmonic coordinates, the following equation holds:

$$\Delta_g g_{ij} = -2\mathrm{Ric}_{ij} + \mathcal{F}_{ij}(g, \nabla g),$$

where Ric_{ij} is the Ricci curvature. For the proof, see [**28**], Lemma 4.1. See also [**48**] for harmonic coordinates.

We should also remark that eigenfunctions of Δ_g are good candidates of coordinates. In this section, we only consider the case of compact manifold.

Lemma 11.1. *Let $\varphi_j(x)$, $j = 1, 2, \cdots$, be a complete orthonormal system of eigenfunctions of Δ_g with Neumann boundary condition. Then, for any $x_0 \in \mathcal{N}^{int}$, there exists a neighborhood of x_0 and $j_1, \cdots, j_n$ such that $\varphi_{j_1}(x), \cdots, \varphi_{j_n}(x)$ form local coordinates on U.*

Proof. By the Fourier expansion for any $a \in C_0^\infty(\mathcal{N})$, $a(x) = \sum a_k\varphi_k(x)$, where the series converges in $C^\infty(\mathcal{N})$. From this one can show that, for any $x_0 \in \mathcal{N}^{int}$, $\mathrm{Sp}\{\nabla\varphi_k(x_0)\}_{k=1}^\infty = T_{x_0}(\mathcal{N}) := \mathbf{R}^n$, where $\mathrm{Sp}(A)$ means the linear span of the set A. In fact, take some local coordinates near x_0 and let $a(x)$ be a smooth function which is linear around x_0. Then $\nabla a(x) = \sum a_k \nabla\varphi_k(x)$ near x_0. This means that the direction $\nabla a(x_0)$ is approximated by a linear combination of $\nabla\varphi_k(x_0)$. Therefore, one can choose n functions $\varphi_{j_i}(x), i = 1, \cdots, n$, such that $\mathrm{Sp}\{\nabla\varphi_k(x_0); k = j_1, \cdots, j_n\} = \mathbf{R}^n$. $\square$

Note that, since $\Delta_g\varphi_k = \lambda_k\varphi_k$, we have, by elliptic regularity, that $\varphi_k \in C^{k+1,\alpha}$ if $g_{ij} \in C^{k,\alpha}$.

Suppose we can find $\varphi_k(x), k = 1, 2, \cdots$, in $R(\mathcal{N})$. Then, we can reconstruct the distance on $\mathcal{N}$ by looking at the heat kernel

$$h(x, y, t) = \sum e^{-\lambda_k t}\varphi_k(x)\varphi_k(y).$$

In fact, we have as $t \to 0$

$$h(x, y, t) \sim \frac{C_n}{t^{n/2}} e^{-\frac{d^2(x,y)}{4t}}.$$

Therefore,

$$\Big(-\lim_{t\to 0} 4t \log h(x, y, t)\Big)^{1/2} = d(x, y).$$

This is another way of reconstructing the distance on $R(\mathcal{N})$.

11.2. Spectral map. From $R(\mathcal{N})$, we have reconstructed the differential structure of $\mathcal{N}$ by finding boundary normal coordinates and boundary distance coordinates. However, the distance coordinates have the disadvantage that we lose 2 orders of regularity, say, of g_{ij}. As for the regularity problem, the best choice is the coordinate system made of eigenfunctions. Let

$$\mu_1, \mu_2, \mu_3 \cdots \quad \text{and} \quad \psi_1(x), \psi_2(x), \psi_3(x) \cdots$$

be the eigenvalues and eigenfunctions of Dirichlet problem, and

$$\lambda_0, \lambda_1, \lambda_2, \cdots \quad \text{and} \quad \varphi_0(x), \varphi_1(x), \varphi_2(x), \cdots$$

those of Neumann problem.

Lemma 11.2. *Having BSD for, say, Neumann problem, we can find BSD for Dirichlet proplem.*

Proof. Let Δ^N and Δ^D be Neumann and Dirichlet Laplacians on $\mathcal{N}$, and $\{\lambda_i, \varphi_i\big|_{\partial\mathcal{N}}; i = 0, 1, 2, \cdots\}$ and $\{\mu_i, \partial\psi_i/\partial\nu\big|_{\partial\mathcal{N}}; i = 1, 2, \cdots\}$ be the boundary spectral data for Neumann and Dirichlet problem, respectively. Take $z \notin \sigma(-\Delta^N) \cup \sigma(-\Delta^D)$. The Neumann-to-Dirichlet map is defined to be $R^N(z) : f \to u\big|_{\partial M}$, where

$$\begin{cases} (-\Delta_g - z)u = 0 \quad \text{in} \quad \mathcal{N}, \\ \dfrac{\partial u}{\partial \nu} = f \quad \text{on} \quad \partial\mathcal{N}. \end{cases}$$

and the Dirichlet-to-Neumann map is defined to be $R^D(z) : f \to \partial v/\partial\nu\big|_{\partial\mathcal{N}}$, where

$$\begin{cases} (-\Delta_g - v)u = 0 \quad \text{in} \quad \mathcal{N}, \\ v = f \quad \text{on} \quad \partial\mathcal{N}. \end{cases}$$

As is seen before, $R^N(z)$ has an integral kernel

$$R^N(z; x, y) = \sum_{i=0}^{\infty} \frac{\varphi_i(x)\varphi_i(y)}{z - \lambda_i}, \quad x, y \in \partial\mathcal{N}.$$

By definition, one can easily see that $(R^N(z))^{-1} = R^D(z)$, and $R^N(z)$ is determined by the Neumann spectral data. Therefore, $R^D(z)$ is determined by the Neumann spectral data. Now $R^D(z)$ has the following formal integral kernel

$$R^D(z; x, y) = \sum_{i=1}^{\infty} \frac{\partial_\nu\psi_i(x)\partial_\nu\psi_i(y)}{z - \mu_i}, \quad x, y \in \partial\mathcal{N}.$$

Actually this sum does not converge. However, $R^D(z)$ is known to be an operator-valued meromorphic function of z with simple poles at $z = \mu_i$ and its residue is given by $\sum_{\mu_k = \mu_i} \partial_\nu\psi_{\mu_k}(x)\partial_\nu\psi_{\mu_k}(y)$, which proves the lemma. $\square$

By the same argument as in the proof of Lemma 11.1, one can show the following lemma.

Lemma 11.3. *Let $x \in \partial\mathcal{N}$. Then there are $n-1$ eigenfunctions of Neumann problem, and one eigenfunction of the Dirichlet problem such that $\{\varphi_{i_1}, \cdots, \varphi_{i_{n-1}}, \psi_{i_n}\}$ form a coordinate system near x.*

Now we define the spectral map $S : \mathcal{N} \to \mathbf{R}^\infty$ by

$$S(x) = \{\varphi_0(x), \psi_1(x), \varphi_1(x), \psi_2(x), \varphi_2(x), \cdots\}.$$

Since these eigenfunctions satisfy $-\Delta_g\varphi_i = \lambda_i\varphi_i$, $-\Delta_g\psi_i = \mu_i\psi_i$, they can be used to find coefficients of Δ_g in "eigenfunction coordinates", i.e. the metric tensor. This is now an well-known idea in geometry, see e.g [**16**], [**75**].

The problem is how to find these eigenfunction coordinates.

Lemma 11.4. *BSD determines $S(\mathcal{N}) \subset \mathbf{R}^\infty$.*

Proof. Let us recall the slicing procedure in §9. There, by solving the initial boundary value problem for the wave equation, we have constructed a layer $Sh(\Gamma, t^+, t^-)$. By taking the intersection of these layers in a generic position, we can find a region of positive measure in $\mathcal{N}$. Let us call it "a pixel", and denote by P_X. Passing to the Fourier transforms $\mathcal{F}^N$ (Neumann case) or $\mathcal{F}^D$ (Dirichlet case), we then find

$$l^{2,N}(P_X) := \mathcal{F}^N(L^2(P_X)), \quad l^{2,D}(P_X) := \mathcal{F}^D(L^2(P_X)).$$

Observe that

$$\mathcal{F}^D\psi_i = e_i = (0, \cdots, 0, 1, 0, \cdots, 0, \cdots),$$
$$\mathcal{F}^N\varphi_i = f_i = (0, \cdots, 0, 1, 0, \cdots, 0, \cdots),$$

Let

$$Q^D(P_X) : l^2 \to l^{2,D}(P_X),$$
$$Q^N(P_X) : l^2 \to l^{2,N}(P_X)$$

be the associated orthogonal projections. We then have

$$(Q^D(P_X)e_i, e_j) = \int_{P_X} \psi_i(x)\psi_j(x)dV,$$
$$(Q^N(P_X)f_0, f_0) = \frac{1}{\mathrm{Vol}((\mathcal{N})} \int_{P_X} dV.$$

We now let P_X shrink to a point : $P_X \to \{x\}$. Then we have

$$\frac{(Q^D e_i, e_j)}{(Q^N f_0, f_0)} \to \mathrm{Vol}(\mathcal{N})\psi_i(x)\psi_j(x),$$
$$\frac{(Q^N f_i, f_0)}{(Q^N f_0, f_0)} \to \mathrm{Vol}^{1/2}(\mathcal{N})\varphi_i(x), \quad \mathrm{Vol}^{-1/2}(\mathcal{N}) = \varphi_0|_{\partial\mathcal{N}}.$$

We thus find a map

$$\widetilde{S} : \mathcal{N} \ni x \to \{\varphi_0(x), \psi_1(x)^2, \varphi_1(x), \psi_2(x)\psi_1(x), \cdots\}.$$

Since $\psi_1(x) > 0$, one can find $\psi_1(x)$ from $\psi_1(x)^2$ on $\mathcal{N}$. Therefore by dividing by $\psi_1(x)$, we get $\{\varphi_0(x), \psi_1(x), \varphi_1(x), \psi_2(x), \cdots\} = S(x)$. $\square$