Progress in Probability, Vol. 36

ON THE GEOMETRY DEFINED BY DIRICHLET FORMS

KARL-THEODOR STURM

ABSTRACT. Every regular, local Dirichlet form on a locally compact, separable space X defines in an intrinsic way a pseudo metric ρ on the state space. Assuming that this is actually a complete metric (compatible with the original topology), we prove that (X, ρ) is a geodesic space. That is, any two points in X are joined by a minimal geodesic.

Also analogues of the Hopf–Rinow Theorem and of the Cartan–Hadamard Theorem are obtained. The latter requires the notion of curvature which is defined by means of the CAT-inequality.

§1 THE DIRICHLET SPACE $(\mathcal{E}, \mathcal{F})$ AND THE ENERGY MEASURE Γ

The basic object for the sequel is a fixed regular Dirichlet form $\mathcal{E}$ with domain $\mathcal{F} = \mathcal{F}(X)$ on a real Hilbert space $L^2(X, m)$ with norm $\|u\| = (\int_X u^2\, dm)^{1/2}$. $\mathcal{F}$ is again a real Hilbert space with norm $\|u\|_{\mathcal{F}} := \sqrt{\mathcal{E}(u,u) + \|u\|}$.

The underlying topological space X is a locally compact, separable Hausdorff space and m is a positive Radon measure with $\mathrm{supp}[m] = X$. The initial Dirichlet form $\mathcal{E}$ is always assumed to be *symmetric* (i. e. $\mathcal{E}(u,v) = \mathcal{E}(v,u)$) and of *diffusion type* (i. e. $\mathcal{E}(u,v) = 0$ whenever $u \in \mathcal{F}$ is constant on a neighborhood of the support of $v \in \mathcal{F}$ or, in other words, $\mathcal{E}$ has no killing measure and no jumping measure). The selfadjoint operator associated with the initial form $\mathcal{E}$ is denoted by L.

Any such form can be written as

$$\mathcal{E}(u,v) = \int_X d\Gamma(u,v)$$

where Γ is a positive semidefinite, symmetric bilinear form on $\mathcal{F}$ with values in the signed Radon measures on X (the so-called *energy measure*). It can be defined by the formulae

$$\begin{aligned}\int_X \phi\, d\Gamma(u,u) &= \mathcal{E}(u, \phi u) - \frac{1}{2}\mathcal{E}(u^2, \phi) \\ &= \lim_{t \to 0} \frac{1}{2t} \int_X \int_X \phi(x) \cdot [u(x) - u(y)]^2\ T_t(x, dy)\, m(dx)\end{aligned}$$

1991 *Mathematics Subject Classification.* 60J60, 31C25, 58G32, 35J70.

Key words and phrases. Dirichlet forms, symmetric diffusion, Markov process, Carathéodory metric, geodesic space.

for every $u \in \mathcal{F}(X) \cap L^\infty(X,m)$ and every $\phi \in \mathcal{F}(X) \cap \mathcal{C}_0(X)$. Since $\mathcal{E}$ is assumed to be of diffusion type, the energy measure Γ is local and satisfies the Leibniz rule as well as the chain rule, cf. [8], [13], [2], [3] and [18]. As usual we extend the quadratic forms $u \mapsto \mathcal{E}(u,u)$ and $u \mapsto \Gamma(u,u)$ to the whole spaces $L^2(X,m)$ resp. $L^2_{\mathrm{loc}}(X,m)$ in such a way that $\mathcal{F}(X) = \{\, u \in L^2(X,m) : \mathcal{E}(u,u) < \infty \,\}$ and $\mathcal{F}_{\mathrm{loc}}(X) = \{\, u \in L^2_{\mathrm{loc}}(X,m) : \Gamma(u,u)$ is a Radon measure$\,\}$.

§2 The Intrinsic Metric ρ and Assumption (A)

The energy measure Γ defines in an intrinsic way a pseudo metric ρ on X by

$$\rho(x,y) = \sup\{\, u(x) - u(y) : u \in \mathcal{F}_{\mathrm{loc}}(X) \cap \mathcal{C}(X),\ d\Gamma(u,u) \le dm \text{ on } X \,\}, \qquad (1)$$

called intrinsic metric or Carathéodory metric (cf. [2], [3], [5], [22]). The condition $d\Gamma(u,u) \le dm$ in (1) means that the energy measure $\Gamma(u,u)$ is absolutely continuous w. r. t. the reference measure m with Radon–Nikodym derivative $\frac{d}{dm}\Gamma(u,u) \le 1$ on X (m-almost everywhere). The density $\frac{d}{dm}\Gamma(u,u)(z)$ should be interpreted as the square of the (length of the) gradient of u at $z \in X$. In general, ρ may be degenerate (i. e. $\rho(x,y) = \infty$ or $\rho(x,y) = 0$ for some $x \neq y$). Throughout this paper we make the

Assumption (A). *ρ is a complete metric on X which is compatible with the original topology on X.*

This assumption in particular implies that ρ is non-degenerate and that for any $y \in X$ the function $x \mapsto \rho(x,y)$ is continuous on X. It is discussed in more details in the paper [18]. There we also compared it with the weaker

Assumption (A'). *The topology induced by ρ is equivalent to the original topology.*

Note that under (A') the following assertions are equivalent:

- ρ is a metric (i.e. it is non-degenerate),
- $\rho(x,y) < \infty$ for all $x, y \in X$,
- X is connected.

In [18] we proved that under (A') the following basic property of the distance function holds true.

Lemma 1. *For every $y \in X$ the distance function $\rho_y : x \mapsto \rho(x,y)$ satisfies $\rho_y \in \mathcal{F}_{\mathrm{loc}}(X) \cap \mathcal{C}(X)$ and*

$$d\Gamma(\rho_y, \rho_y) \le dm. \qquad (2)$$

Hence, the distance function ρ_y can be used to construct cut-off functions on intrinsic balls $B_r(y)$ of the form $\rho_{y,r} : x \mapsto (r - \rho(x,y))_+$.

Let us list up some important facts on intrinsic balls (which need not be true in general metric spaces).

Proposition 1. *Under (A') the following properties are satisfied for any ball* $B_r(x) = \{\, y \in X : \rho(x,y) < r \,\}$ *resp. its closure* $\overline{B_r}(x)$*:*

(i) $B_r(x)$ *is connected;*

(ii) $\overline{B_r}(x) = \{\, y \in X : \rho(x,y) \le r \,\}$.

Proof. (i) Put $B = B_r(x)$. Assume that B is not connected. Let A be a nonempty subset of B which is open and closed in B. Put $C = B \setminus A$. Without restriction $x \in C$. Define the function ψ to be $\equiv 0$ on $X \setminus C$ and to be $r - \rho(x,\cdot)$ on C. Then ψ satisfies $\psi \in \mathcal{F}_{\mathrm{loc}}(X) \cap \mathcal{C}(X)$ and $d\Gamma(\psi,\psi) \le dm$. Moreover, $\psi(x) - \psi(y) = r$ for all $y \in A$. Hence, by the definition of ρ we must have $\rho(x,y) \ge r$ for all $y \in A$, i. e. $A \cap B = \emptyset$.

(ii) Put $K_r(x) = \{\, y \in X : \rho(x,y) \le r \,\}$. Obviously, $K_r(x)$ is closed and contains $\overline{B_r}(x)$. Assume that $K_r(x) \ne \overline{B_r}(x)$. Then there exists $y \in K_r(x)$ and $\epsilon > 0$ such that $B_\epsilon(y) \cap B_r(x) = \emptyset$. Now consider the following function ψ: on $B_r(x)$ we put $\psi = r - \rho(x,\cdot)$, on $B_\epsilon(y)$ we put $\psi = \rho(y,\cdot) - \epsilon$ and on the rest we put $\psi = 0$. Then ψ satisfies $\psi \in \mathcal{F}_{\mathrm{loc}}(X) \cap \mathcal{C}(X)$ and $d\Gamma(\psi,\psi) \le dm$. Moreover, $\psi(x) - \psi(y) = r + \epsilon$. Hence, by the definition of ρ this is a contradiction to $\rho(x,y) = r$. $\square$

In Theorem 2 we shall see that under Assumption (A) all balls $B_r(x)$ are relatively compact.

§3 Examples

The main examples which we have in mind are:

- L is the Laplace–Beltrami operator on a Riemannian manifold and m is the Riemannian volume; in this case, ρ is just the Riemannian distance. More generally, L can be chosen to be a uniformly elliptic operator on a Riemannian manifold (cf. [16]).
- L is a uniformly elliptic operator with a nonnegative weight ϕ on $\mathbb{R}^N$, i. e. $\mathcal{E}(u,v) = \sum_{i,j=1}^N \int a_{ij} \cdot \frac{\partial}{\partial x_i} u \cdot \frac{\partial}{\partial x_j} v \cdot \phi \, dx$ and $(u,v) = \int uv\phi \, dx$ with (a_{ij}) symmetric and uniformly elliptic and ϕ as well as $\phi^{-1} \in L^1_{\mathrm{loc}}(\mathbb{R}^N, dx)$; in this case, ρ is equivalent to the Euclidean distance (cf. [3], [14], [21]).
- L is a subelliptic operator on $\mathbb{R}^N$, i. e.

$$\mathcal{E}(u,v) = \sum_{i,j=1}^{N} \int a_{ij} \cdot \frac{\partial}{\partial x_i} u \cdot \frac{\partial}{\partial x_j} v \, dx$$

  and $(u,v) = \int uv \, dx$ with (a_{ij}) symmetric and such that $\mathcal{E}(u,u) \ge \delta \cdot \|u\|^2_{H^\epsilon} - \|u\|^2$ for some $\delta, \epsilon > 0$; in this case, ρ is equal to the metric used e. g. by Fefferman/Phong [6], Fefferman/Sanchez-Calle [7], Jerison [11], Jerison/Sanchez-Calle [12], Nagel/Stein/Wainger [15]; it can locally be estimated by the Euclidean distance $|\cdot|$ as follows

$$\frac{1}{C} \cdot |x-y| \le \rho(x,y) \le C \cdot |x-y|^\epsilon.$$

This includes also Hörmander type operators with bounded measurable coefficients in the sense of Saloff–Coste/Stroock [17].

§4 Construction of Geodesics

We always take Assumption (A) for granted. A *curve* (or *path*) on X is a continuous map $\gamma : I \to X$ where I is an interval (=connected set) in $\mathbb{R}$. The *length* $L(\gamma)$ of a curve $\gamma : I \to X$ is defined as

$$L(\gamma) = \sup\left\{\sum_{i=1}^{n} \rho\big(\gamma(t_i), \gamma(t_{i-1})\big) : n \in \mathbb{N},\, I \ni t_0 < t_1 < \cdots < t_n \in I\right\}. \qquad (3)$$

Obviously, the length of a curve $\gamma : [a,b] \to X$ dominates the distance of its endpoints, i. e.

$$L(\gamma) \geq \rho\big(\gamma(a), \gamma(b)\big)$$

and the length of a composed curve $\gamma : [a,b] \cup [b,c] \to X$ is the sum of the lengths of the parts $\gamma_1 = \gamma|_{[a,b]}$ and $\gamma_2 = \gamma|_{[b,c]}$, i. e.

$$L(\gamma) = L(\gamma_1) + L(\gamma_2).$$

Let us state (without proof) the following elementary properties of curves on X.

Lemma 2. *For a curve $\gamma : I \to X$ the following properties are equivalent:*

(i) *for all $r, s, t \in I$ with $r < s < t$*

$$\rho\big(\gamma(r), \gamma(t)\big) = \rho\big(\gamma(r), \gamma(s)\big) + \rho\big(\gamma(s), \gamma(t)\big); \qquad (4)$$

(ii) *for all compact intervals $J = [s,t] \subset I$*

$$L(\gamma|_{[s,t]}) = \rho\big(\gamma(s), \gamma(t)\big); \qquad (5)$$

(iii) *for an increasing sequence of compact intervals $J_n = [s_n, t_n]$ with $I = \bigcup_{n\in\mathbb{N}} J_n$ (e. g. for $J_n \equiv I$ if I itself is compact) property (5) holds true;*

(iv) *there exists a reparametrization $\tilde{\gamma} : \tilde{I} \to X$ with $\tilde{\gamma}(\tilde{I}) = \gamma(I)$ and*

$$\rho\big(\tilde{\gamma}(s), \tilde{\gamma}(t)\big) = |t - s| \quad \text{for all } s, t \in I. \qquad (6)$$

Definitions.

(i) *A minimal geodesic on X is a curve $\gamma : I \to X$ with one (hence all) of the properties mentioned in the Lemma 2.*

(ii) *A subunit curve on X is a curve $\gamma : I \to X$ with $|\dot{\gamma}| \leq 1$ on I, where the speed $|\dot{\gamma}|$ of γ is defined by $|\dot{\gamma}(t)| = \limsup_{\epsilon\to 0} \frac{\rho(\gamma(t),\gamma(t\pm\epsilon))}{\epsilon}$.*

Property (6) in Lemma 2 states that the reparametrized curve $\tilde{\gamma}$ is an isometry from $\tilde{I}$ to X. Such a curve is also called minimal geodesic *parametrized by arc length.* A *geodesic* on X is a curve $\gamma : I \to X$ with the property that for every $t \in I$ there exists an $\epsilon > 0$ such that the restriction of γ on $I \cap]t-\epsilon, t+\epsilon[$ is a minimal geodesic. Obviously, every geodesic which is parametrized by arc length is a subunit curve.

Lemma 3. *For any points $x, y \in X$ with $\rho(x,y) = R$ and for any $r \in]0, R[$ there exists an intermediate point $z \in X$ with*

$$\rho(x,z) = r \quad \textit{and} \quad \rho(z,y) = R - r.$$

In general, this intermediate point z between x and y is of course not unique.

Proof. If there exists a point z with $\rho(x,z) \leq r$ and $\rho(z,y) \leq R - r$ then by the triangle inequality this point z already must satisfy $\rho(x,z) = r$ and $\rho(z,y) = R - r$. (Otherwise $\rho(x,y) < R$!)

If $\rho(x,y) = 0$ then $x = y$ and thus one can choose $z = x$. Assume that $0 < \rho(x,y) < \infty$ and that there exists no such point z. That is, $K_x = \overline{B_r}(x)$ and $K_y = \overline{B_{R-r}}(y)$ are disjoint closed sets and K_x is compact. In this case, these two sets must have a strictly positive distance, say $\rho(K_x, K_y) \geq 3\delta > 0$. But then also the larger sets $L_x = \overline{B_{r+\delta}}(x)$ and $L_y = \overline{B_{R-r+\delta}}(y)$ have a strictly positive distance $\rho(L_x, L_y) \geq \delta > 0$.

Now let us consider the following function

$$\psi_0 = \begin{cases} \rho(x,\cdot) - (r+\delta) & \text{in } L_x, \\ 0 & \text{in } X \setminus (L_x \cup L_y), \\ -\rho(y,\cdot) + R - r + \delta & \text{in } L_y. \end{cases}$$

From Lemma 1 and the truncation property it follows that $\psi_0 \in \mathcal{C}(X) \cap \mathcal{F}_{\text{loc}}(X)$ with

$$d\Gamma(\psi_0, \psi_0) = 1_{L_x}\, d\Gamma(\rho_x, \rho_x) + 0 + 1_{L_y}\, d\Gamma(\rho_y, \rho_y) \leq dm.$$

Moreover, one obviously has $\psi_0(y) - \psi_0(x) = R + 2\delta > R$. But this is a contradiction to $R = \sup\{\, \psi(y) - \psi(x) : \psi \in \mathcal{C}(X) \cap \mathcal{F}_{\text{loc}}(X)$ with $d\Gamma(\psi,\psi) \leq dm \,\}$. $\square$

Theorem 1. *(X, ρ) is a geodesic space. That is, any two points $x, y \in X$ are joined by a minimal geodesic. In other words, there exists a continuous map $\gamma : [0,1] \to X$ with $\gamma(0) = x$, $\gamma(1) = y$ and*

$$\rho\big(\gamma(r), \gamma(t)\big) = \rho\big(\gamma(r), \gamma(s)\big) + \rho\big(\gamma(s), \gamma(t)\big) \quad \textit{for all } 0 \leq r < s < t \leq 1.$$

Proof. (i) We fix points $x, y \in X$ with $\rho(x,y) = R > 0$. By Assumption (A), the metric space (X, ρ) is complete and locally compact. Hence, for the point $x_0 = x$ there exists a maximal radius $R_0 \in]0, R]$ with the property that $B_r(X_0)$ is relatively compact for all $r < R_0$. First of all, we construct the desired geodesic γ on the interval $[0, R_0]$ such that its graph lies inside the closed ball $\overline{B_{R_0}}(x_0)$.

(ii) According to Lemma 3, there exists an intermediate point $\gamma(R_0/2)$ between $\gamma(0) = x_0$ and $\gamma(R) = y$ with the properties $\rho(\gamma(0), \gamma(R_0/2)) = R_0/2$ and $\rho(\gamma(R_0/2), \gamma(R)) = R - R_0/2$. Applying the same argument to the pairs of points $\gamma(0), \gamma(R_0/2)$ and $\gamma(R_0/2), \gamma(R)$ one obtains intermediate points $\gamma(R_0/4)$ (between $\gamma(0)$ and $\gamma(R_0/2)$) and $\gamma(3/4R_0)$ (between $\gamma(R_0/2)$ and $\gamma(R)$). Doing this iteratively, one finally gets a countable set of points $\{\gamma(\alpha R_0) : \alpha \in [0,1[\text{ dyadic}\}$ with the property

$$\rho\big(x, \gamma(\alpha R_0)\big) = \alpha R_0 = R - \rho\big(\gamma(\alpha R_0), y\big) \quad \text{for all dyadic numbers } \alpha \in [0,1[.$$

The closure of this countable set is the minimal geodesic γ restricted to $[0, R_0]$.

(iii) Denote the point $\gamma(R_0)$ on $\partial B_{R_0}(x_0)$ by x_1. Either $R_0 = R$ then $x_1 = y$ and we are finished, or $R_0 < R$.

In the latter case, similarly as in (i), we fix a maximal radius $R_1 \in]0, R - R_0]$ such that $B_r(x_1)$ is relatively compact for all $r < R_1$, and (as in (ii)) we construct the geodesic γ on the interval $[R_0, R_0 + R_1]$. The graph of this part of γ lies in $\overline{B_{R_1}}(x_1) \subset \overline{B_{R_0+R_1}}(x_0)$. Its endpoint $\gamma(R_0 + R_1)$ on $\partial B_{R_1}(x_1)$ (which lies also on $\partial B_{R_0+R_1}(x_0)$) will be denoted by x_2.

Doing this successively, one obtains sequences $\{x_n\}_n$ of points in X and $\{R_n\}_n$ of radii in $]0, R]$. Put $S_n = \sum_{k<n} R_k$ and $S_\infty = \sup_n S_n$. Then $x_n = \gamma(S_n)$. The above construction yields the desired geodesic γ restricted to the interval $[0, S_\infty] \subset [0, R]$.

(iv) If $S_\infty = R$ we are already finished. Hence, suppose $S_\infty < R$. Then $\{R_n\}_n$ converges to 0 and $\{x_n\}_n$ is a Cauchy sequence in $B_R(x)$ which converges to the point $x_\infty := \gamma(S_\infty)$. By Assumption (A), there exists a radius $R_\infty \in]0, R - S_\infty]$ such that $B_{R_\infty}(x_\infty)$ is relatively compact. For finally all $n \in \mathbb{N}$, the balls $B_{2R_n}(x_n)$ are contained in that ball $B_{R_\infty}(x_\infty)$ which implies that they are also relatively compact. This, however, contradicts the maximality of R_n. Therefore, S_∞ must equal R and we are finished. $\square$

Remark. Assume that instead of (A) only (A') is satisfied and that $\rho(x, y) < \infty$ for two points under consideration. Instead of requiring that (X, ρ) is complete it suffices to assume in Lemma 3 and Theorem 1 that there exists a complete (w. r. t. the metric ρ) subset $Y \subset X$ containing the convex hull of x and y. The latter means that all $z \in X$ satisfying the equality $\rho(x, z) + \rho(z, y) = \rho(x, y)$ lie in Y.

For instance, the closed ball $\overline{B_R}(x)$ contains the convex hull of x and $y \in \partial B_R(x)$. Thus, in Lemma 3 and Theorem 1 it suffices to assume that either $\overline{B_R}(x)$ or $\overline{B_R}(y)$ is complete.

Let us again assume that (A) is satisfied.

Corollary 1. *The following distances on X coincide:*

- $\rho(x, y) = \sup\{\psi(x) - \psi(y) : \psi \in \mathcal{C}(X) \cap \mathcal{F}_{\text{loc}}(X), d\Gamma(\psi, \psi) \le dm\}$,
- $\rho_0(x, y) = \sup\{\psi(x) - \psi(y) : \psi \in \mathcal{C}_0(X) \cap \mathcal{F}(X), d\Gamma(\psi, \psi) \le dm\}$,
- $\rho_1(x, y) = \inf\{ L(\gamma)$*: γ is a geodesic in X joining x and y*$\}$,
- $\rho_2(x, y) = \inf\{ L(\gamma)$*: γ is a curve in X joining x and y*$\}$,
- $\rho_3(x, y) = \inf\{ R > 0 :$ *there exists a subunit curve* $\gamma : [0, R] \to X$ *with* $\gamma(0) = x$ *and* $\gamma(R) = y\}$.

Proof. The equality "$\rho = \rho_0$" was already proven in [18]. For the inequality "$\rho \ge \rho_1$" note that the previous Theorem states that $\rho(x, y)$ is the length of a suitable (minimal) geodesic joining x and y. The inequalities "$\rho_1 \geq \rho_2$" and "$\rho_2 \geq \rho$" are obvious.

In order to see "$\rho_1 \geq \rho_3$", note that in the definition of ρ_1 we may replace "geodesic" by "geodesic parametized by arc length" and that for the latter the length equals the length of its defining interval. Finally, note that every geodesic parametrized by arc length is a subunit curve.

For the remaining inequality "$\rho_3 \leq \rho$" note that every subunit curve $\gamma : [0, R] \to X$ satisfies $\rho(\gamma(s), \gamma(t)) \leq |t - s|$ for all $s, t \in [0, R]$. In particular, $\rho(\gamma(0), \gamma(R)) \leq R$. □

We close this section with some analogue to the Hopf–Rinow Theorem.

Theorem 2. *Under Assumption (A') and assuming that X is connected the following are equivalent:*

(i) *the metric space (M, ρ) is complete;*
(ii) *every ball $B_r(x)$ is relatively compact;*
(iii) *every (minimal) geodesic $\gamma : I \to X$ defined on an interval $I \subset \mathbb{R}$ can be extended to a (minimal) geodesic defined on $\bar{I}$.*

Proof. The implication (i) $\Rightarrow$ (ii) was proved by Gromov ([10], Thme. 1.10, cf. also [9], Lemme III. 18). It is based on the result of Theorem 1 saying that under (i) the metric space (X, ρ) is a geodesic space. The implications (ii) $\Rightarrow$ (i) and (i) $\Rightarrow$ (iii) are trivial. In order to prove the converse implication, assume that (X, ρ) is not complete, say, the ball $B_{R_0}(x)$ is not relatively compact for some fixed $x \in X$ and some $R_0 < \infty$. Let $R = R(x) = \inf\{\, r > 0 : B_r(x) \text{ is not relatively compact}\,\}$. By assumption (A'), $B_R(x)$ is not relatively compact whereas $B_r(x)$ is relatively compact for all $r < R$.

Now consider a Cauchy sequence $\{y_n\}_n$ in $(B_R(x), \rho)$ which does not converge in $\overline{B_R}(x)$ but to an abstract point y. Take $\hat{B} = \overline{B_R}(x) \cup \{y\}$. Lemma 3 allows to construct a midpoint z between x and y. Namely, for any $n \in \mathbb{N}$ there exists a midpoint z_n between x and y_n satisfying $\rho(x, z_n) = \rho(z_n, y_n) = 1/2\rho(x, y_n)$. Obviously, $|\rho(x, z_n) - 1/2\rho(x, y)| = 1/2|\rho(x, y_n) - \rho(x, y)| \leq 1/2\rho(y_n, y) \to 0$ and $|\rho(z_n, y) - 1/2\rho(x, y)| \leq |\rho(z_n, y_n) - 1/2\rho(x, y_n)| + 3/2\rho(y_n, y) = 3/2\rho(y_n, y) \to 0$ (for $n \to \infty$). All these points z_n, $n \in \mathbb{N}$, lie in the compact set $\overline{B_{r/2}}(x)$. Hence, there exists a cluster point z of $\{z_n\}_n$ in $\overline{B_{r/2}}(x)$ which (by continuiuty of ρ) must satisfy $\rho(x, z) = 1/2\rho(x, y)$ and $\rho(z, y) = 1/2\rho(x, y)$.

Therefore, a slight modification of Theorem 1 allows to construct a (minimal) geodesic $\gamma : [0, R] \to \hat{B}$ from x to y. The curve $\gamma : [0, R[\to B_R(x)$ is a (minimal) geodesic in X defined on the parameter interval $[0, R[$ which can not be extended to a curve in X defined on $[0, R]$. □

Even if the state space (X, ρ) is complete, it is in general not possible to extend a geodesic $\gamma : I \to X$ to a geodesic defined on the whole interval $\mathbb{R}$. For instance, it is not possible if X is the closed unit ball in $\mathbb{R}^N$ and $\mathcal{E}$ is the classical Dirichlet form with Neumann boundary conditions (i. e. $\mathcal{F} = H^1(B_1(0))$). In this case, no (!) geodesic can be extended to a geodesic defined on the whole $\mathbb{R}$.

Remark. Let (A) be satisfied. The following two conditions are equivalent:

- every geodesic $\gamma : I \to X$ can be extended to a geodesic defined on $\mathbb{R}$;
- every point $x \in X$ has a neighborhood Y such that for every $y \in Y$ there exists a point $z \in X \setminus \{x\}$ satisfying

$$\rho(y, z) = \rho(y, x) + \rho(x, z).$$

§5 An Alternative Approach to the Intrinsic Metric

Up to now we can measure the distance ρ between two points x and y on X only by means of functions ψ which are defined on the whole space. Our aim is to localize this procedure. In particular, we want to measure distances by moving along paths.

Of course, this can already be done according to Corollary 1. Note, however, that our previous definition of the length of a curve requires the knowledge of the distance of points. In the sequel, we will reverse the procedure. We define an alternative notion L^* of length of curves without referring to the metric ρ and by means of L^* we define an alternative notion ρ^* of distance of points.

Definitions.

(i) For a curve $\gamma : [a, b] \to X$ we define

$$L^*(\gamma) = \sup\{\, u(\gamma(a)) - u(\gamma(b)) : \, Y \text{ is an open neighborhood of } \gamma([a,b]) \subset X,\ u \in \mathcal{F}_{\text{loc}}(Y) \cap \mathcal{C}(Y),\ d\Gamma(u,u) \le dm \text{ on } Y \,\}$$

and, generally, for a curve $\gamma : I \to X$ we define

$$L^*(\gamma) = \limsup_{n \in \mathbb{N}} L^*(\gamma|_{I \cap [-n,n]}).$$

(ii) For $x, y \in X$ we define

$$\rho^*(x, y) = \inf\{\, L^*(\gamma) : \ \gamma \text{ is a curve joining } x \text{ and } y \,\}.$$

Note that this notion L^* of length ignores loops. In particular, the L^*-length of a closed curve is 0.

The main advantage of this alternative definition $L^*(\gamma)$ is that it depends only on quantities in an arbitrary close neighborhood of $\gamma(I)$. Obviously, $L^*(\gamma) \ge \rho(\gamma(a), \gamma(b))$ for any curve $\gamma : [a, b] \to X$ and thus

$$\rho^* \ge \rho. \tag{7}$$

This holds true without Assumption (A) or (A').

Lemma 4. *Let $\gamma : [a,c] \to X$ be a curve without selfintersections and let $b \in]a,c[$. Define $\gamma_1 = \gamma|_{[a,b]}$ and $\gamma_2 = \gamma|[b,c]$. Then*

$$L^*(\gamma) = L^*(\gamma_1) + L^*(\gamma_2).$$

Proof. Let $x = \gamma(a)$, $y = \gamma(b)$ and $z = \gamma(c)$. The inequality "$\leq$" is obvious since every function u used in the definition of $L^*(\gamma)$ can simultaneously be used in the definition of $L^*(\gamma_1)$ and $L^*(\gamma_2)$ and since $u(x)-u(z) = [u(x)-u(y)]+[u(y)-u(z)]$.

For the converse inequality, choose $\epsilon > 0$. We have to construct a neighborhood Y of $\gamma([a,c])$ and a suitable function u on it with $u(x)-u(z) \geq L^*(\gamma_1)+L^*(\gamma_2)-4\epsilon$ and $d\Gamma(u,u) \leq dm$ on Y. For $i = 1,2$ there exist neighborhoods Y_i of the graphs of γ_i and functions $u_i \in \mathcal{F}_{\mathrm{loc}}(Y_i) \cap \mathcal{C}(Y_i)$ with $u_1(x) - u_1(y) \geq L^*(\gamma_1) - \epsilon$ and $u_2(y) - u_2(z) \geq L^*(\gamma_2) - \epsilon$ and $d\Gamma(u_i,u_i) \leq dm$ on Y_i.

Let $U \subset Y_1 \cap Y_2$ be a neighborhood of y such that $u_1 < u_1(y) + \epsilon$ on U and $u_2 > u_2(y) - \epsilon$. Since γ is without selfintersections, we may assume without restriction that $U = Y_1 \cap Y_2$. Now choose an open neighborhood Y of the graph of γ with $Y \subset Y_1 \cup Y_2$. Define

$$u = \begin{cases} (u_1 - u_1(y) - \epsilon) \vee 0 & \text{in } Y \cap Y_1, \\ (u_2(y) - u_2 + \epsilon) \wedge 0 & \text{in } Y \cap Y_2. \end{cases}$$

Then $u(x) - u(z) \geq L^*(\gamma_1) + L^*(\gamma_2) - 4\epsilon$. Moreover, $u \equiv 0$ in U. Therefore (by means of the truncation property), we deduce that $u \in \mathcal{F}_{\mathrm{loc}}(Y) \cap \mathcal{C}(Y)$ with $d\Gamma(u,u) \leq dm$ on Y. $\square$

Theorem 3. *Assume (A') and let $\gamma : I \to X$ be a curve without selfintersections and such that $\gamma(I)$ is relatively compact in (X,ρ). Then*

$$L^*(\gamma) = L(\gamma).$$

Proof. Let $L^*(\gamma) = L^*$ and, without restriction, $I = [a,b]$. Choose $\epsilon > 0$, an open neighborhood Y of $\gamma(I)$ and an admissible function u on Y with $u(\gamma(a))-u(\gamma(b)) \geq L^* - \epsilon$. (Here and below we call a function u on an open set $Y \subset X$ admissible on Y if $u \in \mathcal{F}_{\mathrm{loc}}(Y) \cap \mathcal{C}(Y)$ and $d\Gamma(u,u) \leq dm$ on Y.)

Let $\delta = 1/4 \cdot \rho(\gamma(I), X \setminus Y)$. The relative compactness of $\gamma(I)$ implies $\delta > 0$. Choose $a = t_0 < t_1 < \ldots < t_n = b$ with $\delta_i := \rho(\gamma(t_i), \gamma(t_{i-1})) \leq \delta$. Then for every $i = 1, \ldots, n$ the function u is defined and admissible on the whole ball $B_{4\delta_i}(\gamma(t_i))$. Hence, $\tilde{v}_i = [3\delta_i - \rho(\gamma(t_i), \cdot)] \wedge [u - u(\gamma(t_{i-1}))]$ is defined and admissible on $B_{4\delta_i}(\gamma(t_i))$. It immediately follows that $\tilde{v}_i \leq 0$ on $B_{4\delta_i}(\gamma(t_i)) \setminus B_{3\delta_i}(\gamma(t_i))$. Hence,

$$v_i = \begin{cases} \tilde{v}_i \vee 0, & \text{on } B_{3\delta_i}(\gamma(t_i)), \\ 0, & \text{else}, \end{cases}$$

is defined and admissible on the whole space X. From the definition of ρ it follows now that

$$v_i\big(\gamma(t_i)\big) - v_i\big(\gamma(t_{i-1})\big) \leq \rho\big(\gamma(t_i), \gamma(t_{i-1})\big) = \delta_i$$

and thus

$$u(\gamma(t_i)) - u(\gamma(t_{i-1})) \le \delta_i.$$

This implies

$$\begin{aligned} L^*(\gamma) - \epsilon \le u(\gamma(a)) - u(\gamma(b)) &= \sum_{i=1}^{n} u(\gamma(t_i)) - u(\gamma(t_{i-1})) \\ &\le \sum_{i=1}^{n} \rho(\gamma(t_i), \gamma(t_{i-1})) \le L(\gamma) \end{aligned}$$

and, hence, $L^*(\gamma) = L(\gamma)$. □

Corollary 2. *Assume that (X, ρ) satisfies Assumption (A). Then*

$$\rho^* = \rho. \tag{8}$$

§6 Curvature and the Cartan-Hadamard Theorem

In the following section we briefly sketch some ideas and results from the general theory of geodesic spaces and the method of "comparison of geometries". This approach goes back to A. D. Aleksandrov (cf. [1]) and was further elaborated among many others by M. Gromov (cf. [10]).

We again make Assumption (A) which implies that (X, ρ) is a locally compact, complete, geodesic metric space. The basic idea is to define upper bounds for the "curvature" on X by comparing geodesic triangles in X with isometric triangles in spaces of constant sectional curvature.

For $\kappa \in \mathbb{R}$, we denote by H_κ the two-dimensional complete, simply connected Riemannian manifold of constant sectional curvature κ. For $\kappa = 0$ this is the Euclidean plane, for $\kappa > 0$ it is a two-dimensional sphere of radius $1/\sqrt{\kappa}$ and for $\kappa < 0$ it is the two-dimensional hyperbolic plane (homothetic to the Poincaré disc).

A geodesic triangle T in X consists of three points in X and three minimal geodesics connecting them. A comparison triangle T_κ for T in H_κ is a geodesic triangle in H_κ with the same edge lenghts as T. It is clear that T_κ is unique up to an isometry of H_κ and such a triangle T_κ exists if $\kappa \le 0$ or if $\kappa > 0$ and the perimeter of T is less than $2\pi/\sqrt{\kappa}$. There is a unique map $T \to T_\kappa$ which takes each edge of T isometrically onto the corresponding edge of T_κ. For each $x \in T$ let x_κ denote its image in T_κ under this map.

Definitions.

(i) A triangle T in X (of perimeter $\le 2\pi/\sqrt{\kappa}$ if $\kappa > 0$) *satisfies the CAT(κ)-inequality* iff for any vertex $y \in T$ and any point $x \in T$ on the side opposite to y we have

$$\rho(x, y) \le \rho_\kappa(x_\kappa, y_\kappa), \tag{9}$$

where x_κ and y_κ are the corresponding points of a comparison triangle in H_κ and ρ_κ denotes the distance in H_κ. (The abbreviation CAT comes from C=comparison, A=Aleksandrov, T=Toponogov.)

(ii) The geodesic space X *has curvature* $\leq \kappa$ iff any $x \in X$ has a neighborhood Y such that any geodesic triangle in Y (of perimeter $\leq 2\pi/\sqrt{\kappa}$ in the case $\kappa > 0$) satisfies the CAT(κ)-inequality.

Remark (cf. [1], [9]). Let (M, g) be a complete Riemannian manifold. Then M has curvature $\leq \kappa$ (in the sense of the above definition) if and only if it has sectional curvature $K \leq \kappa$.

Having at hand the notion of curvature, one can derive for our geodesic spaces many properties which are known to hold for complete Riemannian manifolds with sectional curvature $K \leq \kappa$. We pick out one of these results, namely the famous Cartan–Hadamard Theorem. A complete proof of this result in the full generality of geodesic spaces was given by W. Ballmann ([9], Chap. 10, Thm. 14).

Theorem 4. *If X is simply connected and has curvature ≤ 0, then it is contractible. Any two points $x, y \in X$ are joined by exactly one geodesic and this geodesic is minimal and depends continuously on x and y.*

Acknowledgement. The author is grateful to Professor T. Lyons for stimulating discussions on the subject of this paper.

References

1. A. D. Aleksandrov, V. N. Berestovvskii and I. G. Nikolaev, *Generalized Riemannian spaces*, Russian Math. Surveys **41, 3** (1986), 1–54.
2. M. Biroli and U. Mosco, *Formes de Dirichlet et estimations structurelles dans les milieux discontinus*, C. R. Acad. Sci. Paris **313** (1991), 593–598.
3. M. Biroli and U. Mosco, *A Saint-Venant Principle for Dirichlet forms on discontinuous media*, Elenco Preprint, Rome, 1992.
4. E. A. Carlen, S. Kusuoka and D. W. Stroock, *Upper bounds for symmetric Markov transition functions*, Ann. Inst. H. Poincaré **2** (1987), 245–287.
5. E. B. Davies, *Heat kernels and spectral theory*, Cambridge University Press, 1989.
6. C. L. Fefferman and D. Phong, *Subelliptic eigenvalue problems*, in: Conf. on Harmonic Analysis, Chicago (W. Beckner et al., ed.), Wadsworth, 1981, pp. 590–606.
7. C. L. Fefferman and A. Sanchez-Calle, *Fundamental solutions for second order subelliptic operators*, Ann. of Math. **124** (1986), 247–272.
8. M. Fukushima, *Dirichlet forms and Markov processes*, North Holland and Kodansha, Amsterdam, 1980.
9. E. Ghys and P. de la Harpe, *Sur les groupes hypberboliques d'aprés Mikhael Gromov*, Prog. Math., vol. 83, Birkhäuser, Boston, 1990.
10. M. Gromov, *Structures métriques pour les variétés riemanniennes*, Rédigé par J. Lafontaine et P. Pansu., Cedic/F.Nathan, 1981.
11. D. Jerison, *The Poincaré inequality for vector fields satisfying an Hörmander's condition*, Duke J. Math. **53** (1986), 503–523.
12. D. Jerison and A. Sanchez-Calle, *Estimates for the heat kernel for a sum of squares of vector fields*, Indiana Univ. J. of Math. **35** (1986), 835–854.

13. Y. Lejan, *Mesures associées a une forme de Dirichlet. Applications*, Bull. Soc. Math. France **106** (1978), 61–112.
14. Z. Ma and M. Röckner, *Introduction to the theory of (non-symmetric) Dirichlet forms*, Springer Universitext, 1992.
15. A. Nagel, E. Stein and S. Wainger, *Balls and metrics defined by vector fields. I: Basic properties*, Acta Math. **155** (1985), 103–147.
16. L. Saloff-Coste, *Uniformly elliptic operators on Riemannian manifolds*, J. Diff. Geom. **36** (1992), 417–450.
17. L. Saloff-Coste and D. W. Stroock, *Operateurs uniformement sous-elliptiques sur des groupes de Lie*, J. Funct. Anal. **98** (1991), 97–121.
18. K.-Th. Sturm, *Analysis on local Dirichlet spaces – I. Recurrence, conservativeness and L^p-Liouville properties* (to appear in J. Reine Angew. Math.).
19. K.-Th. Sturm, *Analysis on local Dirichlet spaces – II. Upper Gaussian estimates for the fundamental solutions of parabolic equations* (to appear in Osaka J. Math.).
20. K.-Th. Sturm, *Analysis on local Dirichlet spaces – III. The parabolic Harnack inequality*, Preprint, Erlangen, 1994.
21. K.-Th. Sturm, *Sharp estimates for capacities and applications to symmetric diffusions*, Preprint, Erlangen, 1994.
22. N. Varopoulos, L. Saloff-Coste and T. Coulhon, *Analysis and geometry on groups*, Cambridge University Press, 1992.

K.-T. Sturm, Mathematisches Institut, Universität Erlangen-Nürnberg, Bismarckstrasse 1 1/2, D-91054 Erlangen