Research Article

A Natural Diffusion Distance and Equivalence of Local Convergence and Local Equicontinuity for a General Symmetric Diffusion Semigroup

Maxim J. Goldberg [ID][1] **and Seonja Kim** [ID][2]

[1]*Theoretical and Applied Sci., Ramapo College of NJ, Mahwah, NJ 07430, USA*
[2]*Mathematics Dept., Middlesex County College, Edison, NJ 08818, USA*

Correspondence should be addressed to Maxim J. Goldberg; mgoldber@ramapo.edu

In this paper, we consider a general symmetric diffusion semigroup $\{T_t f\}_{t\geq 0}$ on a topological space X with a positive σ-finite measure, given, for $t > 0$, by an integral kernel operator: $T_t f(x) \triangleq \int_X \rho_t(x, y) f(y) dy$. As one of the contributions of our paper, we define a diffusion distance whose specification follows naturally from imposing a reasonable Lipschitz condition on diffused versions of arbitrary bounded functions. We next show that the mild assumption we make, that balls of positive radius have positive measure, is equivalent to a similar, and an even milder looking, geometric demand. In the main part of the paper, we establish that local convergence of $T_t f$ to f is equivalent to local equicontinuity (in t) of the family $\{T_t f\}_{t\geq 0}$. As a corollary of our main result, we show that, for $t_0 > 0$, $T_{t+t_0} f$ converges locally to $T_{t_0} f$, as t converges to 0^+. In the Appendix, we show that for very general metrics $\mathscr{D}$ on X, not necessarily arising from diffusion, $\int_X \rho_t(x, y) \mathscr{D}(x, y) dy \longrightarrow 0$ a.e., as $t \longrightarrow 0^+$. R. Coifman and W. Leeb have assumed a quantitative version of this convergence, uniformly in x, in their recent work introducing a family of multiscale diffusion distances and establishing quantitative results about the equivalence of a bounded function f being Lipschitz, and the rate of convergence of $T_t f$ to f, as $t \longrightarrow 0^+$. We do not make such an assumption in the present work.

1. Introduction

Diffusion semigroups play an important role in analysis, both theoretical and applied. Diffusion semigroups include the heat semigroup and, more generally, as discussed in, e.g., [1], arise from considering large classes of elliptic second-order (partial) differential operators on domains in Euclidean space or on manifolds. For examples of theoretical results involving diffusion semigroups, the interested reader may refer to Sturm [2] and Wu [3]. Some recent applications of diffusion semigroups to dimensionality reduction, data representation, multiscale analysis of complex structures, and the definition and efficient computation of natural diffusion distances can be found in, e.g., [4–11].

A particular important issue in harmonic analysis is to connect the smoothness of a function with the speed of convergence of its diffused version to itself, in the limit as time goes to zero. For the Euclidean setting, see, for example, [12, 13]. In order to consider the smoothness of diffusing functions in more general settings, a distance defined in terms of the diffusion itself seems particularly appropriate.

Defining diffusion distances is of interest in applications as well. As discussed in [5], dimensionality reduction of data and the concomitant issue of finding structures in data are highly important objectives in the fields of information theory, statistics, machine learning, sampling theory, etc. It is often useful to organize the given data as nodes in a weighted graph, where the weights reflect local interaction between data points. Random walks, or diffusion, on graphs may then help understand the interactions among the data points at increasing distance scales. To even consider different distance scales, it is necessary to define an appropriate diffusion distance on the constructed data graph.

In this paper, we consider a general symmetric diffusion semigroup $\{T_t f\}_{t \geq 0}$ on a topological space X with a positive σ-finite measure (i.e., X is a countable union of measurable sets with finite measure), given, for $t > 0$, by an integral kernel operator: $T_t f(x) \triangleq \int_X \rho_t(x, y) f(y) dy$. As part of their work in [7, 11], Coifman and Leeb introduce a family of multiscale diffusion distances and establish quantitative results about the equivalence of a bounded function f being Lipschitz, and the rate of convergence of $T_t f$ to f, as $t \longrightarrow 0^+$ (we are discussing some of their results using a continuous time t for convenience; most of Coifman's and Leeb's derivations are done for dyadically discretized times. Moreover, most of the authors' results are in fact established without the assumption of symmetry and under the weaker condition than positivity of the kernel, namely, an appropriate L_1 integrability statement (see [11])). To prove the implication that Lipschitz implies an appropriate estimate on the rate of convergence, Coifman and Leeb make a quantitative assumption about the decay of

$$\sup_x \int_X |\rho_t(x, y)| \, d(x, y) \, dy, \quad \text{as } t \longrightarrow 0^+, \qquad (1)$$

for their distances d, namely, that

$$\sup_x \int_X |\rho_t(x, y)| \, d(x, y) \, dy \leq C t^\alpha, \qquad (2)$$

for some $\alpha > 0$. The authors show that their decay assumption holds for semigroups arising in many different settings (for which suitable decay and continuity assumptions are made on diffusion kernels relative to an intrinsic metric D of the underlying space), and even for some examples of nonsymmetric diffusion kernels. Coifman and Leeb also establish that (2) above, in the case of positive diffusion kernels, is in fact *equivalent* to their conclusion about the rate of convergence of $T_t f$ to f, as $t \longrightarrow 0^+$, for a Lipschitz function f. Additionally, Coifman and Leeb show that, in some of the settings they consider (with decay and continuity assumptions on the diffusion kernels relative to an intrinsic metric), their multiscale diffusion distance is equivalent to (localized) $D(x, y)^\alpha$, where $D(x, y)$ is the intrinsic metric of the underlying space and α is a positive number strictly less than 1. The authors emphasize that α cannot be taken to equal 1.

In the present paper, we introduce a new family of diffusion distances generated by the diffusion semigroup $\{T_t f\}_{t \geq 0}$. We provide several reasons as to why we think our definition is natural; in particular, we show that, for a convolution diffusion kernel on $\mathbb{R}^n$, we achieve $\alpha = 1$ in the discussion just above; i.e., we can recover (local) Euclidean distance to the "full" power 1.

The implication established in [7, 11] that smoothness of f implies control of the speed of convergence of $T_t f$ to f seems to us to be a more notable result than the converse (which the authors establish without assuming the decay of (1)). However, if f is Lipschitz for the multiscale diffusion distance introduced in [7, 11], as the authors themselves point out their assumed estimate (2) almost tautologically leads to the desired estimate for the speed of convergence of $T_t f$ to f.

The main reason for our current work is that we wish to avoid making any assumptions about the decay of (1) and still establish a correspondence between some version of smoothness of a function f and convergence of $T_t f$ to f, as $t \longrightarrow 0^+$. Our main contribution is to establish, under almost no assumptions, that local equicontinuity (in t) is equivalent to local convergence; i.e., local control of the differences $T_t f(x) - T_t f(y)$ for all t small is equivalent to local control of the differences $T_t f(x) - f(x)$ for all small t. Here "local" is defined relative to a representative of our family of proposed diffusion distances.

Our paper is organized as follows. Following a notation and assumptions section (Section 2), we define our version of a natural diffusion distance D_g in Section 3:

$$D_g(x, y) \equiv \sup_{0 < t \leq 1} g(t) \left\| \rho_t(x, \cdot) - \rho_t(y, \cdot) \right\|_{L_1}, \qquad (3)$$

for g a bounded, nonnegative, increasing function on $(0, 1]$, with $\lim_{t \to 0^+} g(t) = 0$. We are led to our definition by requiring that a diffusion distance has the property that, for all functions f bounded in magnitude by 1, $T_t(f)$ be Lipschitz with respect to the distance, independent of the particular f (of course, we expect the Lipschitz constant to grow as t goes to 0). This requirement arises from the intuitively reasonable demand that diffusion be smoothing in some sense. We then discuss some other reasons why our resulting distance is natural. In particular, for diffusion semigroups with convolution kernels on $\mathbb{R}^n$ (this class includes the Poisson and heat kernels), our distance is equivalent to (local) Euclidean or sub-Euclidean distances for certain choices of the function g.

In Section 4, we make the assumption that balls of positive radius with respect to the distance D_g have positive measure. We show there is an equivalent topology, which does not depend on the function g, for which a corresponding statement about positive measure is equivalent to our assumption. The latter requirement, in turn, seems to be a mild and reasonable one.

In the main section, Section 5, we define our version of local convergence of $T_t(f)$ to f, as well as local equicontinuity of the family $\{T_t f\}_{t \geq 0}$. Both definitions use our distance D_g. We then establish that local convergence is equivalent to local equicontinuity. We next prove a corollary which extends an a.e. convergence result of Stein in [1]: for $t_0 > 0$, $T_{t+t_0} f$ converges locally to $T_{t_0} f$, as t converges to 0^+.

In the Appendix, we show that, for very general metrics $\mathscr{D}$ on X, not necessarily arising from diffusion,

$$\int_X \rho_t(x, y) \mathscr{D}(x, y) \, dy \longrightarrow 0 \text{ a.e.,} \quad \text{as } t \longrightarrow 0^+. \quad (4)$$

This result is clearly a weaker statement than (2), but has the advantage of holding under virtually no assumptions.

2. Notation and Assumptions

Let X be a topological space equipped with a positive σ-finite measure. For $t > 0$, $\rho_t(x, y)$ will denote a *symmetric* kernel on

$X \times X$, with $\rho_t(x, y) \geq 0$ for all x, y. We assume that ρ satisfies the semigroup property:

$$\int_X \rho_t(x, u)\, \rho_s(u, y)\, du = \rho_{t+s}(x, y), \qquad (5)$$

for all $x, y \in X$, and $s, t > 0$. In addition, we assume

$$\int_X \rho_t(x, y)\, dy = 1, \qquad (6)$$

for all $x \in X$ and all $t > 0$. We will refer to a kernel ρ_t satisfying the conditions above as a symmetric diffusion kernel (at time t). A typical example for ρ_t is the heat kernel on a Riemannian manifold (see [14], for example).

For a function f, say in $L_2(X)$ (or more generally, for any f where the following definition makes sense), we define the symmetric diffusion operator T_t, for $t > 0$, by

$$T_t f(x) \triangleq \int_X \rho_t(x, y)\, f(y)\, dy. \qquad (7)$$

We define T_0 to be the identity map. Note that, for all t, $\|T_t f\|_{L_1} \leq \|f\|_{L_1}$, by Fubini's theorem, that clearly $\|T_t f\|_{L_\infty} \leq \|f\|_{L_\infty}$, and hence $\|T_t f\|_{L_p} \leq \|f\|_{L_p}$, for $1 \leq p \leq \infty$, by interpolation.

To avoid degeneracy, e.g., each T_t being the averaging operator on a space of finite mass, we make an additional assumption: $T_t(f) \longrightarrow f$ in L_2, as $t \longrightarrow 0^+$.

The symmetric diffusion operator T_t has the following properties of a *symmetric diffusion semigroup*:

(i) T_0 is the identity

(ii) $T_{t+s} = T_t \circ T_s$, for all $s, t \geq 0$

(iii) $\|T_t(f)\|_{L_p} \leq \|f\|_{L_p}$, for $1 \leq p \leq \infty$

(iv) T_t is a self-adjoint operator on $L_2(X)$

(v) $T_t(f) \longrightarrow f$ in L_2, as $t \longrightarrow 0^+$

(vi) $T_t(f) \geq 0$ if $f \geq 0$

(vii) $T_t(1) = 1$

See Stein's book [1], in which the author derives various harmonic analysis results for symmetric diffusion semigroups without explicitly using kernels.

3. A Natural Diffusion Distance

We now define our diffusion distance.

Definition 1. For a bounded, nonnegative, increasing function g on $(0, 1]$, with $\lim_{t \to 0^+} g(t) = 0$, and g strictly positive on the interval $(0, 1]$, define the distance D_g by

$$D_g(x, y) \equiv \sup_{0 < t \leq 1} g(t)\, \|\rho_t(x, \cdot) - \rho_t(y, \cdot)\|_{L_1}. \qquad (8)$$

It is clear that the distance D_g satisfies the triangle inequality. Note that the restriction that g is bounded in the above supremum has the effect of making all "large" distances comparable to a constant, but this is not a drawback for smoothness considerations.

We would now like to discuss why we are using this particular diffusion distance and why we think it is a natural choice. Our starting point is the desire that, for a reasonable diffusion distance $d(\cdot, \cdot)$, $T_t(f)$ should be "smooth" for $t > 0$, even for "rough" functions f. This intuitive requirement is suggested by the idea that a diffusion semigroup be smoothing, in some sense. It would further be natural that the smoothness decays, for a general f, as $t \longrightarrow 0^+$. We are thus led to impose a Lipschitz-like requirement, namely, that, for a diffusion distance $d(\cdot, \cdot)$, and for $t > 0$,

$$c(t) \equiv \sup_{x \neq y, \|f\|_{L_\infty} = 1} \frac{|T_t(f)(x) - T_t(f)(y)|}{d(x, y)} < \infty. \qquad (9)$$

It is easy to see that

$$c(t) = \sup_{x \neq y} \frac{\|\rho_t(x, \cdot) - \rho_t(y, \cdot)\|_{L_1}}{d(x, y)}. \qquad (10)$$

Note that, for any x and y, $\|\rho_t(x, \cdot) - \rho_t(y, \cdot)\|_{L_1}$ is decreasing in t, since, for $h > 0$,

$$\|\rho_{t+h}(x, \cdot) - \rho_{t+h}(y, \cdot)\|_{L_1}$$

$$= \int_X |\rho_{t+h}(x, u) - \rho_{t+h}(y, u)|\, du$$

$$= \int_X \left| \int_X (\rho_t(x, v)\, \rho_h(v, u) - \rho_t(y, v)\, \rho_h(v, u))\, dv \right| du$$

$$\leq \int_X |\rho_t(x, v) - \rho_t(y, v)| \left(\int_X \rho_h(v, u)\, du \right) dv$$

$$= \int_X |\rho_t(x, v) - \rho_t(y, v)|\, dv$$

$$= \|\rho_t(x, \cdot) - \rho_t(y, \cdot)\|_{L_1}, \qquad (11)$$

using (5) and (6). Letting $g(t) = 1/c(t)$ we thus see that g is increasing, and from (10) we conclude that

$$d(x, y) \geq g(t)\, \|\rho_t(x, \cdot) - \rho_t(y, \cdot)\|_{L_1}, \quad t > 0. \qquad (12)$$

This last inequality motivates our Definition 1 of D_g. The restriction to $t \leq 1$ is to ensure that $D_g(x, y)$ is finite for all x and y and is not stringent, due to the fact that $\|\rho_t(x, \cdot) - \rho_t(y, \cdot)\|_{L_1}$ is decreasing in t and that for smoothness purposes we need to only concentrate on points x and y which are near each other.

A further indication of the naturality of our proposed diffusion distance D_g is that the L_1 norm of the difference of two probability densities, $\|\rho_t(x, \cdot) - \rho_t(y, \cdot)\|_{L_1}$, occurring in the definition of D_g, is the (scaled) total variation distance between the probability distributions $\rho_t(x, \cdot)$ and $\rho_t(y, \cdot)$, i.e.,

$$\frac{1}{2} \|\rho_t(x, \cdot) - \rho_t(y, \cdot)\|_{L_1} = \sup_A |\mu_{t,x}(A) - \mu_{t,y}(A)|. \qquad (13)$$

Here, $\mu_{t,x}$ is the measure given by $\mu_{t,x}(A) = \int_A \rho_t(x, u)\, du$, and $\mu_{t,y}$ is the measure given by $\mu_{t,y}(A) = \int_A \rho_t(y, u)\, du$

for measurable $A \subseteq X$; the supremum is taken over all measurable $A \subseteq X$ (see Chapter 4 of [15]).

As a final argument for the naturality of our proposed diffusion distance, we calculate D_g for a special case considered by the authors of [7] (for their own version of diffusion distances). We take $X = \mathbb{R}^n$, $g(t) = t^\alpha$, and assume that the diffusion kernel has the form $\rho_t(x, y) = t^{-n\beta}\phi(t^{-\beta}(x - y))$. Here, $\alpha, \beta > 0$ and ϕ is a nonnegative radial L_1 function whose gradient is also in L_1. The case $\beta = 1/2$ is for the heat kernel (with the appropriate ϕ), and the case $\beta = 1$ is for the Poisson kernel (with the appropriate ϕ). Now,

$$
\begin{aligned}
&\left\| \rho_t(x, \cdot) - \rho_t(y, \cdot) \right\|_{L_1} \\
&= \int_{\mathbb{R}^n} t^{-n\beta} \left| \phi\left(t^{-\beta}(x - u) \right) - \phi\left(t^{-\beta}(y - u) \right) \right| du \\
&= \int_{\mathbb{R}^n} \left| \phi\left(w + t^{-\beta}(y - x) \right) - \phi(w) \right| dw,
\end{aligned}
\tag{14}
$$

where we made the change of variables $w = t^{-\beta}(x - u)$. Let $h(z) = \int_{\mathbb{R}^n} |\phi(w + z) - \phi(w)| dw$. Then it is easy to see that h is radial and, for a "generic" ϕ, we have the estimates: $h(z) \sim |z|$ if $|z| \leq 1$, and $h(z) \sim 1$ if $|z| > 1$. Here, $|\cdot|$ is the usual Euclidean norm. Using this observation, and (14), we obtain the following (for this special case).

Proposition 2. *For $|x - y| \leq 1$, $D_g(x, y) \sim |x - y|^{\alpha/\beta}$ if $\alpha \leq \beta$, and $D_g(x, y) \sim |x - y|$ if $\alpha \geq \beta$. For $|x - y| > 1$, $D_g(x, y) \sim 1$.*

Proof. Using the notation for the special case above, we need to estimate $\sup_{0 < t \leq 1} t^\alpha h(t^{-\beta}(y - x))$.

Let us first consider the situation when $|x - y| > 1$. Then, for $0 < t \leq 1$, $t^{-\beta}|y - x| \geq 1$, so $\sup_{0 < t \leq 1} t^\alpha h(t^{-\beta}(y - x)) \sim 1$ using the estimate for h mentioned before the proposition.

Next, consider the situation when $|x - y| \leq 1$. Let $t_0 = |x - y|^{1/\beta}$. Note that $0 < t_0 \leq 1$.

When $t_0 \leq t \leq 1$, we have that $t^{-\beta}|x - y| \leq 1$, so

$$
t^\alpha h\left(t^{-\beta}(y - x) \right) \sim t^{\alpha - \beta} |x - y|.
\tag{15}
$$

If $\alpha \leq \beta$, the maximum of the right hand side occurs at $t = t_0$ and equals

$$
t_0^\alpha = |x - y|^{\alpha/\beta}.
\tag{16}
$$

If $\alpha \geq \beta$, the maximum of the right hand side occurs at $t = 1$ and equals $|x - y|$.

When $0 < t \leq t_0$, we have that $t^{-\beta}|x - y| \geq 1$, so

$$
t^\alpha h\left(t^{-\beta}(y - x) \right) \sim t^\alpha,
\tag{17}
$$

and the maximum of the right hand side occurs at $t = t_0$ and equals

$$
t_0^\alpha = |x - y|^{\alpha/\beta}.
\tag{18}
$$

Note that if $\alpha \geq \beta$, since $|x - y| \leq 1$, $|x - y|^{\alpha/\beta} \leq |x - y|$.

Combining the above discussions for the two ranges of values of t, the result follows. $\qquad\square$

Thus, for this special case of $X = \mathbb{R}^n$, $g(t) = t^\alpha$, and $\rho_t(x, y) = t^{-n\beta}\phi(t^{-\beta}(x - y))$, which includes both the heat kernel and the Poisson kernel, our definition of diffusion distance gives (local) Euclidean or sub-Euclidean distance (depending on the relative sizes of α and β). This result seems appropriate.

4. A Geometric Assumption about the Measure on X

We make the following reasonable assumption about our distance D_g: for any $x_0 \in X$ and any $\epsilon > 0$,

$$
B(x_0, \epsilon) \equiv \left\{ x : D_g(x_0, x) < \epsilon \right\},
\tag{19}
$$

the ball of radius ϵ and center x_0, has positive measure.

To justify the statement that this assumption is indeed reasonable, we first define another family of subsets of X. For any $x_0 \in X$, $t > 0$, and $\epsilon > 0$, let

$$
N(x_0, t, \epsilon) \equiv \left\{ x : \left\| \rho_t(x, \cdot) - \rho_t(x_0, \cdot) \right\|_{L_1} < \epsilon \right\}.
\tag{20}
$$

We then have the following equivalence of topologies induced by the sets $B(x_0, \epsilon)$ and $N(x_0, t, \epsilon)$:

Proposition 3. *For any $x_0 \in X$ and any $\epsilon > 0$, there exist $t > 0$ and $\delta > 0$ such that $N(x_0, t, \delta) \subseteq B(x_0, \epsilon)$. Conversely, for any $x_0 \in X$, $t > 0$, and $\epsilon > 0$, there exists a $\delta > 0$ such that $B(x_0, \delta) \subseteq N(x_0, t, \epsilon)$.*

Proof. Fix an $x_0 \in X$ and an $\epsilon > 0$. We first show that there exist $t > 0$ and $\delta > 0$ such that $N(x_0, t, \delta) \subseteq B(x_0, \epsilon)$.

Since we made the assumption that $\lim_{t \to 0^+} g(t) = 0$ for the function g used in defining the distance D_g, there exists a $0 < t < 1$ with $g(t) < \epsilon/4$. Let $\delta = \epsilon/(2M)$, where $M = \sup_{0 < s \leq 1} g(s) = g(1)$. Now, pick an arbitrary $x \in N(x_0, t, \delta)$.

For $0 < s \leq t$, since g in increasing, we see that

$$
g(s) \left\| \rho_s(x, \cdot) - \rho_s(x_0, \cdot) \right\|_{L_1} \leq 2g(t) < \frac{\epsilon}{2},
\tag{21}
$$

Using the fact that the L_1 norm of $\rho_s(x, \cdot)$ is 1 for any s and x.

Now consider the case when $t \leq s \leq 1$. Note that, by definition of $N(x_0, t, \delta)$, we have that $\left\| \rho_t(x, \cdot) - \rho_t(x_0, \cdot) \right\|_{L_1} < \delta = \epsilon/(2M)$. Then, for this range of s, we observe that

$$
\begin{aligned}
&g(s) \left\| \rho_s(x, \cdot) - \rho_s(x_0, \cdot) \right\|_{L_1} \\
&\leq M \left\| \rho_t(x, \cdot) - \rho_t(x_0, \cdot) \right\|_{L_1} < \frac{M\epsilon}{(2M)} = \frac{\epsilon}{2},
\end{aligned}
\tag{22}
$$

where we have used that $\left\| \rho_s(x, \cdot) - \rho_s(x_0, \cdot) \right\|_{L_1}$ is decreasing in s; see (11).

We conclude (see (8)) that

$$
D_g(x, x_0) = \sup_{0 < s \leq 1} g(s) \left\| \rho_s(x, \cdot) - \rho_s(y, \cdot) \right\|_{L_1} < \epsilon,
\tag{23}
$$

and hence $x \in B(x_0, \epsilon)$.

For the converse, fix $x_0 \in X$, $t > 0$ and $\epsilon > 0$. We will show that there exists a $\delta > 0$ such that $B(x_0, \delta) \subseteq N(x_0, t, \epsilon)$.

Since, for any x, $\|\rho_s(x,\cdot) - \rho_s(x_0,\cdot)\|_{L_1}$ is decreasing in s (see (11)), we clearly have that $N(x_0, s_1, \epsilon) \subseteq N(x_0, s_2, \epsilon)$ for any $0 < s_1 < s_2$. Thus, we may assume $0 < t < 1$. Let $\delta = \epsilon g(t)$. Then, for any $x \in B(x_0, \delta)$, we have that $D_g(x, x_0) < \epsilon g(t)$. Hence, using Definition 1 of the distance D_g, we obtain

$$g(t) \left\| \rho_t(x, \cdot) - \rho_t(x_0, \cdot) \right\|_{L_1} \leq D_g(x, x_0) < \epsilon g(t). \quad (24)$$

Thus, $\|\rho_t(x, \cdot) - \rho_t(x_0, \cdot)\|_{L_1} < \epsilon$, and we have that $x \in N(x_0, t, \epsilon)$. □

Returning to our assumption that, for any $x_0 \in X$ and any $\epsilon > 0$, $B(x_0, \epsilon)$ has positive measure, Proposition 3 shows that it is equivalent to require the following: for any $x_0 \in X$, $t > 0$, and $\epsilon > 0$, the set $N(x_0, t, \epsilon)$ has positive measure. Note that the definition of the sets $N(x_0, t, \epsilon)$ is more "universal" than that of the balls $B(x_0, \epsilon)$, since the former do not involve the function g.

The assumption that, for any $x_0 \in X$, $t > 0$, and $\epsilon > 0$, the set $N(x_0, t, \epsilon)$ has positive measure appears to us to be a very natural, and mild, one. In words, this requirement is saying that, for any time $t > 0$ and any $\epsilon > 0$, the set of points in our space X which have not diffused more than ϵ away (in the L_1 sense) from the diffused point x_0, at time t, is not "thin" with respect to the underlying measure on X. This assumption seems reasonable in both the discrete case (each point has positive mass, and $x = x_0$ is "enough") and the continuous case (every point x_0 has "many" arbitrarily close points in the sense of diffusion).

5. Local Convergence Is Equivalent to Local Equicontinuity

In this section, we define local convergence and local equicontinuity for our situation and show that the two concepts are equivalent under our assumptions.

In what follows, T_t is a symmetric diffusion operator as defined in Section 2.

Definition 4. Let $f \in L_p$, $1 \leq p \leq \infty$. Note that f is actually an equivalence class of functions on the space X. Suppose there exists a particular representative of this equivalence class, which we will also call f, such that this representative f is defined at every point of X, and for every $\epsilon > 0$, there exist $t_0 > 0$ and $\delta > 0$ so that $|T_t f(x) - f(x)| < \epsilon$, for all t with $0 < t \leq t_0$ and all $x \in B(x_0, \delta)$. We then say $T_t f$ converges to f locally at x_0.

We also make the following.

Definition 5. Let $f \in L_p$, $1 \leq p \leq \infty$. Suppose there exists a particular representative of the equivalence class specified by f and which we will also call f, such that this representative f is defined at every point of X, and for every $\epsilon > 0$, there exist $t_0 > 0$ and $\delta > 0$ with the property that, for all $x \in B(x_0, \delta)$, we have $|f(x) - f(x_0)| < \epsilon$ and for all t with $0 < t \leq t_0$, $|T_t f(x) - T_t f(x_0)| < \epsilon$. We then say the family $\{T_t f\}_{t \geq 0}$ is locally equicontinuous (in t) at x_0.

Our main result is the following.

Proposition 6. *For $f \in L_2 \cap L_\infty$ and any $x_0 \in X$, the following are equivalent:*

(i) *$T_t f$ converges to (the representative) f locally at x_0*

(ii) *The family $\{T_t f\}_{t \geq 0}$ is locally equicontinuous at x_0*

Moreover, if a representative f satisfies one of these statements, the same representative satisfies the other statement.

Proof. We first show that local convergence at x_0 implies local equicontinuity at x_0. We thus begin by assuming that $T_t f$ converges to a representative f locally at x_0.

First, we establish continuity of this representative f at x_0. Fix $\epsilon > 0$. By the assumption, there exist $1 \geq t_0 > 0$ and $\delta > 0$ such that $|T_{t_0} f(x) - f(x)| < \epsilon/3$, for all t with $0 < t \leq t_0$ and all $x \in B(x_0, \delta)$. Then, for any $x \in B(x_0, \delta)$, using the definition of the distance D_g, we see that

$$\begin{aligned}
&\left| f(x) - f(x_0) \right| \\
&\leq \left| f(x) - T_{t_0} f(x) \right| + \left| T_{t_0} f(x) - T_{t_0} f(x_0) \right| \\
&\quad + \left| T_{t_0} f(x_0) - f(x_0) \right| \\
&\leq \frac{2\epsilon}{3} + \left\| \rho_{t_0}(x, \cdot) - \rho_{t_0}(x_0, \cdot) \right\|_{L_1} \|f\|_{L_\infty} \\
&= \frac{2\epsilon}{3} + \frac{g(t_0) \left\| \rho_{t_0}(x, \cdot) - \rho_{t_0}(x_0, \cdot) \right\|_{L_1} \|f\|_{L_\infty}}{g(t_0)} \\
&\leq \frac{2\epsilon}{3} + \frac{D_g(x, x_0)}{g(t_0)} \|f\|_{L_\infty}.
\end{aligned} \quad (25)$$

Since we assumed that $g(t) > 0$ if $t > 0$, we have that $\delta' = \min(\delta, (\epsilon/3)(g(t_0)/\|f\|_{L_\infty})) > 0$. Thus, if $x \in B(x_0, \delta')$,

$$\left| f(x) - f(x_0) \right| < \frac{2\epsilon}{3} + \frac{\epsilon}{3} \frac{g(t_0)}{\|f\|_{L_\infty}} \frac{\|f\|_{L_\infty}}{g(t_0)} = \epsilon, \quad (26)$$

and continuity of f at x_0 is shown.

Next, note that

$$\begin{aligned}
\left| T_t f(x) - T_t f(x_0) \right| &\leq \left| T_t f(x) - f(x) \right| \\
&\quad + \left| f(x) - f(x_0) \right| \\
&\quad + \left| f(x_0) - T_t f(x_0) \right|.
\end{aligned} \quad (27)$$

Let t_0 and $\delta > 0$ be as above, i.e., $|T_t f(x) - f(x)| < \epsilon/3$, for all t with $0 < t \leq t_0$ and all $x \in B(x_0, \delta)$. Since we have already shown that f is continuous at x_0, there exists a $\delta'' > 0$ such that $|f(x) - f(x_0)| < \epsilon/3$ for $x \in B(x_0, \delta'')$. Let $\delta''' = \min(\delta, \delta'')$. Then, for $x \in B(x_0, \delta''')$ and $0 < t \leq t_0$, we see that $|T_t f(x) - T_t f(x_0)| < \epsilon$. Hence, the local equicontinuity of the family $\{T_t f\}_{t \geq 0}$ at x_0 follows.

Conversely, we now show that local equicontinuity at x_0 implies local convergence at x_0. We thus begin by assuming that the family $\{T_t f\}_{t \geq 0}$ is equicontinuous at x_0.

Fix $\epsilon > 0$. By the assumption, there exist $1 \geq t_0 > 0$ and $\delta > 0$ such that, for the representative f, $|f(x) - f(x_0)| < \epsilon/5$

and $|T_t f(x) - T_t f(x_0)| < \epsilon/5$, for all $x \in B(x_0, \delta)$ and all t with $0 < t \le t_0$. In Section 4, we made the assumption that all balls of positive radius have positive measure. Using Stein's Maximal Theorem (see Chapter III, §3 in [1]), $\lim_{t \to 0^+} T_t f = f$ a.e. So there is a $y_0 \in B(x_0, \delta)$ such that $\lim_{t \to 0^+} T_t f(y_0) = f(y_0)$. Now, for $x \in B(x_0, \delta)$,

$$
\begin{aligned}
\left|T_t f(x) - f(x)\right| &\le \left|T_t f(x) - T_t f(y_0)\right| \\
&\quad + \left|T_t f(y_0) - f(y_0)\right| \\
&\quad + \left|f(y_0) - f(x)\right|.
\end{aligned}
\tag{28}
$$

We estimate the first term on the right hand side of the above inequality as follows:

$$
\begin{aligned}
\left|T_t f(x) - T_t f(y_0)\right| &\le \left|T_t f(x) - T_t f(x_0)\right| \\
&\quad + \left|T_t f(x_0) - T_t f(y_0)\right| < \frac{2\epsilon}{5},
\end{aligned}
\tag{29}
$$

for all $0 < t \le t_0$, since $x, y_0 \in B(x_0, \delta)$. For the second term, we use that $\lim_{t \to 0^+} T_t f(y_0) = f(y_0)$: there exists a $t_1 > 0$ such that $|T_t f(y_0) - f(y_0)| < \epsilon/5$, for all t satisfying $0 < t \le t_1$. Finally, for the third term, we see that

$$
\begin{aligned}
\left|f(y_0) - f(x)\right| &\le \left|f(y_0) - f(x_0)\right| \\
&\quad + \left|f(x_0) - f(x)\right| < \frac{2\epsilon}{5},
\end{aligned}
\tag{30}
$$

since $x, y_0 \in B(x_0, \delta)$.

Thus, for all t with $0 < t \le \min(t_0, t_1)$, and for any $x \in B(x_0, \delta)$, we obtain that $|T_t f(x) - f(x)| < \epsilon$, which concludes the proof of the converse. □

In the proof above, we used Stein's Maximal Theorem (see Chapter III, §3 in [1]) to state that $\lim_{t \to 0^+} T_t f = f$ a.e. Stein's a.e. convergence result, for $f \in L_2$ say, is the main place in our paper where the symmetry of the operators T_t is needed: Stein requires symmetry to prove his Maximal Theorem.

We immediately have the following.

Corollary 7. *Let $f \in L_2 \cap L_\infty$. Fix $t_0 > 0$. Then for any $x_0 \in X$, $T_t(T_{t_0} f) = T_{t+t_0} f$ converges locally to $T_{t_0} f$ at x_0.*

Proof. By Proposition 6, it suffices to show that $\{T_{t+t_0} f\}_{t \ge 0}$ is locally equicontinuous at x_0. Fix $\epsilon > 0$. Let $G(t) = g(t)$ for $0 < t \le 1$ and $G(t) = g(1)$ for $t > 1$. For any $t \ge 0$, we have that

$$
\begin{aligned}
&\left|T_{t+t_0} f(x) - T_{t+t_0} f(x_0)\right| \\
&\quad \le \left\|\rho_{t+t_0}(x, \cdot) - \rho_{t+t_0}(x_0, \cdot)\right\|_{L_1} \|f\|_{L_\infty} \\
&\quad = \frac{G(t+t_0) \left\|\rho_{t+t_0}(x, \cdot) - \rho_{t+t_0}(x_0, \cdot)\right\|_{L_1}}{G(t+t_0)} \|f\|_{L_\infty} \\
&\quad \le \frac{D_g(x, x_0)}{G(t+t_0)} \|f\|_{L_\infty} \le \frac{D_g(x, x_0)}{G(t_0)} \|f\|_{L_\infty},
\end{aligned}
\tag{31}
$$

using the definition of the distance D_g and the function G, that G is increasing, and inequality (11). Then, for $\delta = (G(t_0)/\|f\|_{L_\infty})\epsilon$, we see that $x \in B(x_0, \delta)$ implies that $|T_{t+t_0} f(x) - T_{t+t_0} f(x_0)| \le \epsilon$, and we have shown local equicontinuity at x_0. □

Using our notation, Stein in [1] mentions that $\lim_{t \to 0^+} T_{t+t_0} f(x) = T_{t_0} f(x)$ for almost all x, since he proves that $T_t f$ is a real-analytic function of $t > 0$ for almost all x. Corollary 7 extends Stein's result (under our assumption discussed in Section 4) to show local convergence with respect to the distance D_g.

6. Conclusions and Future Work

In this paper, we have defined a diffusion distance which is natural if one imposes a reasonable Lipschitz condition on diffused versions of arbitrary bounded functions. We have next shown that the mild assumption that balls of positive radius have positive measure is equivalent to a similar, and an even milder looking, geometric demand. In the main part of the paper, we establish that local convergence of $T_t f$ to (a representative) f at a point is equivalent to local equicontinuity of the family $\{T_t f\}_{t \ge 0}$ at that point.

It may well be useful to have a quantitative estimate on the rate of convergence of $T_t f$ to f under the assumption that f is Lipschitz, say, with respect to some distance d (where d may be our D_g). As essentially pointed out in the papers [7, 11], a key issue is whether, and how rapidly,

$$
\sup_x \int_X \rho_t(x, y) d(x, y) dy \longrightarrow 0, \quad \text{as } t \longrightarrow 0^+.
\tag{32}
$$

In the Appendix below, we show that, for very general metrics $\mathscr{D}$ on X, not necessarily arising from diffusion,

$$
\int_X \rho_t(x, y) \mathscr{D}(x, y) dy \longrightarrow 0 \text{ a.e.}, \quad \text{as } t \longrightarrow 0^+.
\tag{33}
$$

This result is certainly far from establishing the convergence in (32), much less a quantitative estimate.

We plan to continue exploring for which (diffusion) distances the convergence in (32) holds and an estimate can be obtained.

Appendix

Proposition 8. *Let $\mathscr{D}$ be a metric on X with the following properties:*

(1) $\sup_{x, y \in X} \mathscr{D}(x, y) < \infty$

(2) *X is separable with respect to the metric $\mathscr{D}$, i.e., it contains a countable dense subset*

(3) *There exists a $\delta > 0$ so that $m[B(x, \delta)] < \infty$ for every $x \in X$ (the bound need not be uniform in x). Here, $m[B(x, \delta)]$ denotes the measure of the ball $B(x, \delta) \equiv \{y : \mathscr{D}(x, y) < \delta\}$*

Then,

$$
\int_X \rho_t(x, y) \mathscr{D}(x, y) dy \longrightarrow 0 \text{ a.e.}, \quad \text{as } t \longrightarrow 0^+.
\tag{A.1}
$$

To prove the proposition, we first establish the following.

Lemma 9. *For any $x_0 \in X$, if $r > 0$ is such that $m[B(x_0, r)] < \infty$, then*

$$\int_X \rho_t(x, y) \mathscr{D}(x_0, y) \, dy \longrightarrow \mathscr{D}(x_0, x),$$

$$\text{as } t \longrightarrow 0^+, \quad \text{(A.2)}$$

for almost all $x \in B(x_0, r)$.

Proof. Let $f(y) = \mathscr{D}(x_0, y)\chi(y)$, where $\chi(y)$ is the characteristic function of the ball $B(x_0, r)$. Since $m[B(x_0, r)] < \infty$, we see that $f \in L_2(X)$. Using Stein's Maximal Theorem (see Chapter III, §3 in [1]), we conclude that

$$\int_X \rho_t(x, y) f(y) \, dy \longrightarrow f(x) \text{ a.e. in } X,$$

$$\text{as } t \longrightarrow 0^+, \text{ i.e.,} \quad \text{(A.3)}$$

$$\int_{B(x_0, r)} \rho_t(x, y) \mathscr{D}(x_0, y) \, dy \longrightarrow$$

$$\mathscr{D}(x_0, x) \chi(x) \text{ a.e. in } X.$$

In particular, for some set $C \subseteq B(x_0, r)$, with $m[B(x_0, r) \setminus C] = 0$,

$$x \in C \implies \int_{B(x_0, r)} \rho_t(x, y) \mathscr{D}(x_0, y) \, dy \longrightarrow$$

$$\mathscr{D}(x_0, x), \quad \text{as } t \longrightarrow 0^+. \quad \text{(A.4)}$$

We would be done if the integration were over all of X, not just $B(x_0, r)$.

To this end, we apply Stein's Maximal Theorem to the L_2 function $\chi(y)$ to see that there is a set $D \subseteq B(x_0, r)$, with $m[B(x_0, r) \setminus D] = 0$, so that

$$x \in D \implies \int_{B(x_0, r)} \rho_t(x, y) \, dy \longrightarrow \chi(x) = 1,$$

$$\text{as } t \longrightarrow 0^+. \quad \text{(A.5)}$$

For $x \in D$, since

$$1 = \int_X \rho_t(x, y) \, dy$$

$$= \int_{B(x_0, r)} \rho_t(x, y) \, dy + \int_{B^c(x_0, r)} \rho_t(x, y) \, dy, \quad \text{(A.6)}$$

we conclude that

$$\int_{B^c(x_0, r)} \rho_t(x, y) \, dy \longrightarrow 0, \quad \text{as } t \longrightarrow 0^+, \quad \text{(A.7)}$$

where $B^c(x_0, r)$ is the complement of $B(x_0, r)$.

Since we assumed that $\sup_{x, y \in X} \mathscr{D}(x, y) < \infty$, we obtain that, for every $x \in D$,

$$\int_{B^c(x_0, r)} \rho_t(x, y) \mathscr{D}(x_0, y) \, dy$$

$$\leq c \int_{B^c(x_0, r)} \rho_t(x, y) \, dy \longrightarrow 0. \quad \text{(A.8)}$$

Combining (A.4) and (A.8), we conclude that, for $x \in C \cap D \subseteq B(x_0, r)$,

$$\int_X \rho_t(x, y) \mathscr{D}(x_0, y) \, dy$$

$$= \int_{B(x_0, r)} \rho_t(x, y) \mathscr{D}(x_0, y) \, dy$$

$$+ \int_{B^c(x_0, r)} \rho_t(x, y) \mathscr{D}(x_0, y) \, dy \longrightarrow \quad \text{(A.9)}$$

$$\mathscr{D}(x_0, x) + 0 = \mathscr{D}(x_0, x), \quad \text{as } t \longrightarrow 0^+.$$

Note that the set $C \cap D$ is of full measure in $B(x_0, r)$:

$$m[B(x_0, r) \setminus (C \cap D)]$$

$$\leq m[B(x_0, r) \setminus C] + m[B(x_0, r) \setminus D] = 0. \quad \text{(A.10)}$$

The lemma is proved. $\qquad\square$

We now turn to proving the proposition. Choose a positive integer N so that $1/N < \delta$, where $\delta > 0$ is such that $m[B(x, \delta)] < \infty$ for every $x \in X$ (Assumption (3)). Clearly, for every $n \geq N$ and $x \in X$,

$$m\left[B\left(x, \frac{1}{n}\right)\right] \leq m[B(x, \delta)] < \infty. \quad \text{(A.11)}$$

Let $\{x_k\}_{k=1}^\infty$ be a countable, dense subset of X relative to the distance $\mathscr{D}$ (Assumption (2)). For every $k = 1, 2, \ldots$, and $n = N, N+1, \ldots$, apply Lemma 9 to obtain a set $E(k, n)$ such that $E(k, n) \subseteq B(x_k, 1/n)$, $m[B(x_k, 1/n) \setminus E(k, n)] = 0$, and, for $x \in E(k, n)$,

$$\int_X \rho_t(x, y) \mathscr{D}(x_k, y) \, dy \longrightarrow \mathscr{D}(x_k, x),$$

$$\text{as } t \longrightarrow 0^+. \quad \text{(A.12)}$$

Let

$$E_n = \bigcup_{k=1}^\infty E(k, n). \quad \text{(A.13)}$$

Since $\{x_k\}_{k=1}^\infty$ is dense in X,

$$\bigcup_{k=1}^\infty B\left(x_k, \frac{1}{n}\right) = X. \quad \text{(A.14)}$$

Hence,

$$X = \bigcup_{k=1}^\infty \left[E(k, n) \cup \left(B\left(x_k, \frac{1}{n}\right) \setminus E(k, n)\right)\right]$$

$$= \left[\bigcup_{k=1}^\infty E(k, n)\right] \cup \left[\bigcup_{k=1}^\infty \left(B\left(x_k, \frac{1}{n}\right) \setminus E(k, n)\right)\right] \quad \text{(A.15)}$$

$$= E_n \cup \left[\bigcup_{k=1}^\infty \left(B\left(x_k, \frac{1}{n}\right) \setminus E(k, n)\right)\right].$$

Since

$$m\left[\bigcup_{k=1}^{\infty}\left(B\left(x_k,\frac{1}{n}\right)\setminus E\left(k,n\right)\right)\right]$$

$$\leq \sum_{k=1}^{\infty} m\left[\left(B\left(x_k,\frac{1}{n}\right)\setminus E\left(k,n\right)\right)\right]=0, \tag{A.16}$$

we see that

$$m\left[E_n^c\right]=0. \tag{A.17}$$

Now, let

$$E=\bigcap_{n=N}^{\infty}E_n. \tag{A.18}$$

Since $E^c=\bigcup_{n=N}^{\infty}E_n^c$ and $m[E_n^c]=0$, we see that $m[E^c]=0$.

To finish the proof of the proposition, we will show that, for every $x\in E$, hence a.e.,

$$\int_X \rho_t\left(x,y\right)\mathscr{D}\left(x,y\right)dy\longrightarrow 0,\quad\text{as }t\longrightarrow 0^+. \tag{A.19}$$

Fix $x\in E$. Choose any $\epsilon>0$. Let $K\geq N$ be such that $1/K<\epsilon$. Since $x\in E=\bigcap_{n=N}^{\infty}E_n$, we see that $x\in E_K=\bigcup_{k=1}^{\infty}E(k,K)$. Thus, $x\in E(k_0,K)\subseteq B(x_{k_0},1/K)$ for some k_0. But then, by definition of the set $E(k_0,K)$,

$$\int_X \rho_t\left(x,y\right)\mathscr{D}\left(x_{k_0},y\right)dy\longrightarrow \mathscr{D}\left(x_{k_0},x\right), \tag{A.20}$$

$$\text{as }t\longrightarrow 0^+.$$

Now, using the triangle inequality,

$$\int_X \rho_t\left(x,y\right)\mathscr{D}\left(x,y\right)dy$$

$$\leq \int_X \rho_t\left(x,y\right)\mathscr{D}\left(x,x_{k_0}\right)dy$$

$$+\int_X \rho_t\left(x,y\right)\mathscr{D}\left(x_{k_0},y\right)dy \tag{A.21}$$

$$=\mathscr{D}\left(x,x_{k_0}\right)+\int_X \rho_t\left(x,y\right)\mathscr{D}\left(x_{k_0},y\right)dy.$$

Hence,

$$\limsup_{t\longrightarrow 0^+}\int_X \rho_t\left(x,y\right)\mathscr{D}\left(x,y\right)dy\leq 2\mathscr{D}\left(x,x_{k_0}\right)$$

$$<\frac{2}{K}<2\epsilon. \tag{A.22}$$

Since $\epsilon>0$ is arbitrary,

$$\limsup_{t\longrightarrow 0^+}\int_X \rho_t\left(x,y\right)\mathscr{D}\left(x,y\right)dy=0, \tag{A.23}$$

and we are done.

Data Availability

Conflicts of Interest

Acknowledgments

References

[1] E. M. Stein, *Topics in Harmonic Analysis Related to the Littlewood-Paley Theory*, vol. 63 of *Annals of Mathematics Studies*, Princeton University Press, Princeton, NJ, USA, 3rd edition, 1985.

[2] K. T. Sturm, "Diffusion processes and heat kernels on metric spaces," *Annals of Probability*, vol. 26, no. 1, pp. 1–55, 1998.

[3] H.-T. Wu, "Embedding Riemannian manifolds by the heat kernel of the connection Laplacian," *Advances in Mathematics*, vol. 304, pp. 1055–1079, 2017.

[4] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[5] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006.

[6] R. R. Coifman, S. Lafon, A. B. Lee et al., "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 21, pp. 7426–7431, 2005.

[7] R. R. Coifman and W. E. Leeb, "Earth Mover's distance and equivalent metrics for spaces with semigroups," Tech. Rep. YALEU/DCS/TR-1481, 2013.

[8] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 53–94, 2006.

[9] M. J. Goldberg and S. Kim, "Some Remarks on Diffusion Distances," *Journal of Applied Mathematics*, vol. 2010, Article ID 464815, 17 pages, 2010.

[10] M. J. Goldberg and S. Kim, "An efficient tree-based computation of a metric comparable to a natural diffusion distance," *Applied and Computational Harmonic Analysis*, vol. 33, no. 2, pp. 261–281, 2012.

[11] W. Leeb and R. Coifman, "Hölder-Lipschitz norms and their duals on spaces with semigroups, with applications to earth mover's distance," *Journal of Fourier Analysis and Applications*, vol. 22, no. 4, pp. 910–953, 2016.

[12] P. L. Butzer and H. Berens, *Semi-Groups of Operators and Approximation*, Springer, Berlin, Germany, 1967.

[13] H. Triebel, *Theory of Function Spaces*, vol. 78 of *Monographs in Mathematics*, Birkhäuser, Boston, Mass, USA, 1983.

[14] I. Chavel, "Eigenvalues in Riemannian geometry," in *Pure and Applied Mathematics #115*, Academic Press, Orlando, Fla, USA, 1984.

[15] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*, American Mathematical Society, Providence, RI, USA, 2009.