

# Hölder–Lipschitz Norms and Their Duals on Spaces with Semigroups, with Applications to Earth Mover’s Distance

William Leeb<sup>1,2</sup> · Ronald Coifman<sup>1</sup>

Received: 26 January 2015  
© Springer Science+Business Media New York 2015

**Abstract** We introduce a family of bounded, multiscale distances on any space equipped with an operator semigroup. In many examples, these distances are equivalent to a snowflake of the natural distance on the space. Under weak regularity assumptions on the kernels defining the semigroup, we derive simple characterizations of the Hölder–Lipschitz norm and its dual with respect to these distances. As the dual norm of the difference of two probability measures is the Earth Mover’s Distance (EMD) between these measures, our characterizations give simple formulas for a metric equivalent to EMD. We extend these results to the mixed Hölder–Lipschitz norm and its dual on the product of spaces, each of which is equipped with its own semigroup. Additionally, we derive an approximation theorem for mixed Lipschitz functions in this setting.

**Keywords** Semigroup · Heat kernel · Earth Mover’s Distance · Dominating mixed derivatives · Hölder · Lipschitz

**Mathematics Subject Classification** 46E27 · 46E15 · 26B35 · 26A16

---

Communicated by Massimo Fornasier.

---

✉ William Leeb  
wleeb@math.princeton.edu  
Ronald Coifman  
coifman@math.yale.edu

<sup>1</sup> Department of Mathematics, Yale University, New Haven, CT 06511, USA

<sup>2</sup> Present Address: Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA

## 1 Introduction

Operator semigroups are ubiquitous objects in pure and applied mathematics. It is well-known that many function spaces, such as the Besov spaces on  $\mathbb{R}^n$ , can be characterized using, for example, the heat kernel [3, 31]. Recent work has generalized these results to function spaces on more general domains; for instance, see [2, 17]. More abstractly, a substantial amount of classical harmonic analysis in  $\mathbb{R}^n$  can be pushed through to the setting of a measure space equipped with a diffusion semigroup, as found in Stein's book [29]. A limiting aspect of the theory developed in this book is the absence of an explicit geometry; though the statements of maximal theorems, Littlewood–Paley theorems and interpolation theorems make sense in this general setting, basic notions such as Lipschitz functions cannot be defined, since these require a metric on the underlying measure space.

A natural way of introducing a geometry in such an abstract setting is to use the semigroup itself to define a distance. This is the approach taken in the theory of *diffusion maps* [6] and related work [15, 16]. If the kernel of the semigroup at time  $t$  is denoted by  $a_t(x, y)$ , then the *diffusion distance* at time  $t$  is defined as  $\|a_t(x, \cdot) - a_t(y, \cdot)\|_{L^2(d\mu)}$ , for an appropriate measure  $\mu$ . This conceptually meaningful distance has found wide application in machine learning, where the kernel  $a_t(x, y)$  is a power of an affinity matrix measuring the relationship between two points in a data set.

The idea of defining distances using semigroups is the starting point for the present work. However, the ground distances  $D_\alpha(x, y)$  we introduce in Sect. 2 are not defined at a fixed scale, but incorporate all scales at once. Though the parameter  $\alpha > 0$  controls the weight placed at each scale, all scales are present. In a variety of examples, and for appropriate ranges of  $\alpha$ , we will show that this distance is equivalent to a “snowflake” of the intrinsic distance  $\rho(x, y)$  on the underlying space; that is, the distance  $\rho(x, y)$  raised to a power less than 1 [18]. In the examples we consider, this power is a constant times  $\alpha$ , and it follows that we must restrict  $\alpha$  to be less than 1.

Next, we consider the space  $\Lambda_\alpha$  of functions that are Lipschitz with respect to the distance  $D_\alpha(x, y)$ . In the examples where  $D_\alpha(x, y)$  is a snowflake of an intrinsic distance  $\rho(x, y)$ , Lipschitz functions are Hölder with respect to  $\rho(x, y)$ . We will therefore call  $\Lambda_\alpha$  the *Hölder–Lipschitz space*. This space arises in many areas of applications. For instance, in nonparametric statistics Hölder–Lipschitz functions arise naturally as a model for unknown functions corrupted by noise. Simple equivalent formulas for the Hölder–Lipschitz norm in the Euclidean setting, such as those derived from wavelet theory [23], have been used for signal recovery [11].

In Sect. 4, we give simple characterizations of the Hölder–Lipschitz norm using the semigroup itself. As in the Euclidean setting, where similar results are well-known [3, 29, 31], the fundamental observation is that the size of a function's variation across scales is equivalent to the size of its variation in space.

In Sect. 5, we study the space  $\Lambda_\alpha^*$  of measures that can be integrated against Hölder–Lipschitz functions—that is, the space dual to  $\Lambda_\alpha$ . In particular, we give simple characterizations of the norm on this space. This is of interest in many applications, as the dual norm of the difference between two probability measures is equal to their *Earth Mover's Distance (EMD)*. We will recall the definition of EMD and prove a basic property of it in Sect. 6. EMD is a popular tool in machine learning that suffers

from high computational cost. The equivalent metrics we develop provide, in many situations, a fast approximation to it.

We impose certain regularity conditions on the semigroup. The conditions are highly non-restrictive, and in Sect. 3 we show that they hold for a broad class of semigroups, such as those considered in [17]. Examples include heat kernels on closed Riemannian manifolds, heat kernels on certain fractals, and subordinated heat kernels in  $\mathbb{R}^n$  (including the Poisson kernel), as well as the non-symmetric example of shifted heat kernels on  $\mathbb{R}^n$ . In addition, we will show that if the theory holds for some finite collection of semigroups on different spaces, then it holds for their product on the cross-product of these spaces. In all the examples we consider, the parameter  $\alpha$  defining the distance  $D_\alpha(x, y)$  must lie between 0 and 1, and sometimes must be bounded away from 1.

In Sects. 7 and 8, we generalize our results to the product of measure spaces, each equipped with its own semigroup. We will focus on the example of two spaces for concreteness, though all results hold for arbitrarily many. In this setting, the natural measure of a function's regularity is not the supremum of its difference quotients, but rather of its mixed difference quotients. We derive equivalent formulas for the norms on the space  $\Lambda_{\alpha, \beta}$  of mixed Hölder–Lipschitz functions and its dual  $\Lambda_{\alpha, \beta}^*$ , where  $\alpha$  and  $\beta$  are the parameters used to define the distances on the two spaces.

Product spaces arise naturally in applications. In signal processing, for example, the spectrogram of a signal is a function on the product of the time and Fourier domains. By assuming that the spectrogram lies in a Sobolev space with dominating mixed derivatives—akin to the mixed Hölder–Lipschitz space  $\Lambda_{\alpha, \beta}$ —one can develop effective estimators for recovering a spectrogram corrupted by noise [24]. Furthermore, the norm dual to Hölder–Lipschitz is a robust distance between two spectrograms. The equivalent dual norms we derive in this paper provide distances for comparing measures on any database with a product structure where each axis has its own semigroup, and hence its own geometry.

## 1.1 Notation

By “ $A \lesssim B$ ” or “ $B \gtrsim A$ ” we mean inequalities up to positive constants; that is, there is a constant  $C > 0$  such that  $A \leq C \cdot B$ . Similarly, by “ $A \simeq B$ ” we mean there are constants  $c, C > 0$  such that  $c \cdot A \leq B \leq C \cdot A$ . What is meant by  $C$  being a “constant” will be clear in each instance.

We will encounter a variety of norms and seminorms throughout the paper. We will use  $\|\cdot\|$ , augmented with appropriate subscripts and superscripts, to denote norms, while we will use capital letters and parentheses, e.g.  $V(\cdot)$ , augmented with appropriate subscripts and superscripts, to denote seminorms.

## 2 Multiscale Diffusion Distance

Our setting throughout the paper will be a sigma-finite measure space  $\mathcal{X}$ . We will not need to explicitly refer to the  $\sigma$ -algebra or the measure. We suppose that  $\mathcal{X}$  is equipped

with a family of kernels  $a_t(x, y)$ ,  $t > 0$ , in  $L^1$ . Defining the operators

$$A_t f(x) = \int_{\mathcal{X}} a_t(x, y) f(y) dy$$

we assume the following conditions:

- (S) (The semigroup property) For all  $t, s > 0$ ,  $A_t A_s = A_{t+s}$ . This property can be expressed in terms of the kernels  $a_t(x, y)$  as

$$a_{t+s}(x, y) = \int_{\mathcal{X}} a_t(x, w) a_s(w, y) dw.$$

- (C) (The conservation property) If  $\mathbf{1}$  is the constant function 1 on  $\mathcal{X}$ , then for all  $t > 0$ ,  $A_t \mathbf{1} = \mathbf{1}$ . This property can be expressed in terms of the kernels  $a_t(x, y)$  as

$$\int_{\mathcal{X}} a_t(x, y) dy = 1.$$

- (I) (The integrability property) There is a constant  $C > 0$  such that for all  $t > 0$  and  $x \in \mathcal{X}$ ,

$$\int_{\mathcal{X}} |a_t(x, y)| dy \leq C.$$

- (R) (The regularity property) There are constants  $C > 0$  and  $\alpha > 0$  such that for every  $1 \geq s \geq t > 0$  and every  $x \in \mathcal{X}$ ,

$$\int_{\mathcal{X}} |a_t(x, y)| \cdot \|a_s(x, \cdot) - a_s(y, \cdot)\|_1 dy \leq C \left(\frac{t}{s}\right)^\alpha.$$

We will actually only require a slightly weaker version of condition (R), namely the same condition restricted to dyadic times  $t = 2^{-k}$  and  $s = 2^{-l}$ , with  $k$  and  $l$  non-negative integers. Later in this section we will give an alternate characterization of (R), namely condition (G) below, that reveals its geometric content with respect to the distance  $D_\alpha(x, y)$  defined by formula (1). In Sect. 3, we will show that condition (R) holds for a wide variety of spaces  $\mathcal{X}$  and kernels  $a_t(x, y)$ . In Sect. 4 we will show that this condition also implies convergence to the identity for the class of Hölder–Lipschitz functions that we will define there. Note too that in every example we discuss  $\alpha$  will be strictly less than 1, and sometimes will be bounded away from 1.

In contrast to [2, 17], we do not assume the existence of a metric on the space  $\mathcal{X}$ . Rather, we will use the kernels  $a_t(x, y)$  to *define* a metric from scratch. This approach is inspired by the papers [6, 16] and related work. In the theory of diffusion maps [6], for example, each time  $t$  defines a *diffusion distance*, namely the  $L^2$  distance between  $a_t(x, \cdot)$  and  $a_t(y, \cdot)$ . Each such distance captures the geometry of the space at a particular scale. These distances also have the feature that they can be approximately embedded into a low-dimensional Euclidean space.

As in [16], our starting point is the  $L^1$  distance between kernels at each scale, and not the  $L^2$  distance. We also consider a single distance that incorporates all scales at once, rather than a family of distances. Though there are no Euclidean embeddings of the distance we define (as with an  $L^2$  diffusion distance), for the application areas we have in mind there will usually be no need to explicitly compute our distance for all pairs of points; see Sect. 6.

We will be concerned with dyadic scales  $t \in (0, 1]$ ; that is, scales  $t = 2^{-k}$ ,  $k \geq 0$ . To this end, define

$$P_k = A_{2^{-k}}$$

and

$$p_k(x, y) = a_{2^{-k}}(x, y).$$

Also, we define

$$D_k(x, y) = \|p_k(x, \cdot) - p_k(y, \cdot)\|_1.$$

Define the multiscale distance

$$D_\alpha(x, y) = \sum_{k \geq 0} 2^{-k\alpha} D_k(x, y). \quad (1)$$

Note that condition (I) guarantees not only that  $D_\alpha(x, y)$  is finite, but that it is uniformly bounded for all  $x$  and  $y$ .

In Sect. 3 we will compute the distance  $D_\alpha(x, y)$  for many examples of semigroups. Before doing so, however, it will be convenient to turn our attention to the regularity condition (R) we impose on the kernels  $a_t(x, y)$ . We reformulate condition (R) in geometric terms, where the geometry is defined by the distance  $D_\alpha(x, y)$ . To that end, define the geometric condition (G) by

(G) (The geometric property) There are constants  $C > 0$  and  $\alpha > 0$  such that for all  $k \geq 0$  and  $x \in \mathcal{X}$ ,

$$\int_{\mathcal{X}} |p_k(x, y)| \cdot D_\alpha(x, y) dy \leq C 2^{-k\alpha}.$$

We show that conditions (R) and (G) are essentially equivalent. The following lemma will be convenient.

**Lemma 1** *Suppose there are constants  $C > 0$  and  $\alpha > 0$  such that for all non-negative integers  $k, l \geq 0$  and all  $x \in \mathcal{X}$ ,*

$$\int |p_k(x, y)| \sum_{l=0}^k 2^{-l\alpha} \|p_l(x, \cdot) - p_l(y, \cdot)\|_1 dy \leq C 2^{-k\alpha}.$$

Then (G) holds, for the same choice of  $\alpha$  and a possibly different constant  $C$ .

*Proof* By the integrability condition (I) the integrals  $\int_{\mathcal{X}} |p_l(x, y)| dy$  are uniformly bounded. Therefore

$$\sum_{l \geq k+1} 2^{-l\alpha} \|p_l(x, \cdot) - p_l(y, \cdot)\|_1 \lesssim 2^{-k\alpha}$$

and so

$$\int_{\mathcal{X}} |p_k(x, y)| \sum_{l \geq k+1} 2^{-l\alpha} \|p_l(x, \cdot) - p_l(y, \cdot)\|_1 dy \lesssim 2^{-k\alpha} \int_{\mathcal{X}} |p_k(x, y)| dy \lesssim 2^{-k\alpha}.$$

Recalling that  $D_l(x, y) = \|p_l(x, \cdot) - p_l(y, \cdot)\|_1$ , we can therefore write

$$\begin{aligned} \int_{\mathcal{X}} |p_k(x, y)| \cdot D_\alpha(x, y) dy &= \int_{\mathcal{X}} |p_k(x, y)| \sum_{l=0}^k 2^{-l\alpha} D_l(x, y) dy \\ &\quad + \int_{\mathcal{X}} |p_k(x, y)| \sum_{l=k+1}^{\infty} 2^{-l\alpha} D_l(x, y) dy \\ &\lesssim 2^{-k\alpha} \end{aligned}$$

completing the proof.  $\square$

**Proposition 1** Suppose that (R) holds for some  $\alpha > 0$  and all dyadic times  $s = 2^{-l}, t = 2^{-k}$ , where  $0 \leq l \leq k$ . Then (G) holds for all  $k \geq 0$  and for any  $0 < \alpha' < \alpha$ .

*Proof* For all  $x$ , condition (R) implies

$$\int_{\mathcal{X}} |p_k(x, y)| 2^{-l\alpha'} \|p_l(x, \cdot) - p_l(y, \cdot)\|_1 \lesssim 2^{-k\alpha} 2^{l(\alpha-\alpha')}.$$

Summing over  $l = 0, \dots, k$  gives

$$\begin{aligned} \int_{\mathcal{X}} |p_k(x, y)| \sum_{l=0}^k 2^{-l\alpha'} \|p_l(x, \cdot) - p_l(y, \cdot)\|_1 dy &\lesssim 2^{-k\alpha} \sum_{l=0}^k 2^{l(\alpha-\alpha')} \\ &\lesssim 2^{-k\alpha} 2^{k(\alpha-\alpha')} = 2^{-k\alpha}. \end{aligned}$$

By Lemma 1, we are done.  $\square$

**Proposition 2** Suppose condition (G) holds for some  $\alpha > 0$ . Then (R) holds for all dyadic times  $s = 2^{-l}, t = 2^{-k}$ , and for all  $0 < \alpha' \leq \alpha$ . In other words, for all  $0 < \alpha' \leq \alpha$  there is a constant  $C$  such that for all  $0 \leq l \leq k$  and  $x \in \mathcal{X}$ ,

$$\int_{\mathcal{X}} |p_k(x, y)| \cdot \|p_l(x, \cdot) - p_l(y, \cdot)\|_1 dy \leq C 2^{-(k-l)\alpha'}.$$

*Proof* Since  $2^{-l\alpha} \|p_l(x, \cdot) - p_l(y, \cdot)\|_1 \leq D_\alpha(x, y)$  for all  $l \geq 0$ , we have

$$\int_{\mathcal{X}} |p_k(x, y)| \cdot \|p_l(x, \cdot) - p_l(y, \cdot)\|_1 dy \leq 2^{l\alpha} \int_{\mathcal{X}} |p_k(x, y)| D_\alpha(x, y) dy \lesssim 2^{-(k-l)\alpha}.$$

Since  $\alpha' \leq \alpha$ , the result follows.  $\square$

We will find condition (G) to be a more useful statement of regularity than (R) going forward. Note too that, as stated earlier, to recover (G) we need only assume (R) for dyadic times  $s$  and  $t$  between 0 and 1.

### 3 Examples of Kernels Satisfying Our Conditions

In this section, we show that the conditions (S), (C), (I) and (R) (and equivalently, by Propositions 1 and 2, condition (G) as well) we impose on the kernels  $a_t(x, y)$  hold for semigroups arising in many different settings. Specifically, we consider two very general conditions (conditions 1 and 2 formulated below) that are found in the paper [17]. These conditions assume the existence of another metric  $\rho(x, y)$  on  $\mathcal{X}$  and posit that the kernels  $a_t(x, y)$  exhibit a certain regularity with respect to  $\rho(x, y)$ . We show that we can recover the four conditions (S), (C), (I) and (R) from conditions 1 and 2.

Furthermore, we also obtain an upper bound on the distance  $D_\alpha(x, y)$  given by Eq. (1) in terms of the distance  $\rho(x, y)$ . More precisely, we show that  $D_\alpha(x, y)$  is bounded above by  $\rho(x, y)$  raised to a power less than 1. Such a distance—that is, a distance of the form  $\rho(x, y)^\delta$ , where  $0 < \delta < 1$ —is called a *snowflake* of  $\rho(x, y)$  [18]. By imposing even stronger regularity on  $a_t(x, y)$  in the form of condition 3 below (also found in [17]), we will show that the distance  $D_\alpha(x, y)$  is in fact equivalent to a thresholded snowflake of  $\rho(x, y)$ . We will use our analysis to establish conditions (S), (C), (I) and (R) for many examples of semigroups and compute the distance  $D_\alpha(x, y)$  for these examples.

Throughout this section, we will always assume  $0 < \alpha < 1$ ; as we will see, it will often be necessary to impose an even tighter upper bound on  $\alpha$ .

#### 3.1 Hölder Continuous Kernels with Decay

Suppose there is a metric  $\rho(x, y)$  on  $\mathcal{X}$  and a measure  $\mu$  such that  $\mu(B(x, r)) \lesssim r^n$ , where  $n > 0$  is fixed. In addition to the conservation property (C) and the uniform  $L^1$  bound (I) that we already assume, the kernel  $a_t(x, y)$  is assumed to be symmetric, and the following two regularity conditions are imposed:

1. An upper bound on the kernel: there is a non-negative, monotonic decreasing function  $\Phi : [0, \infty) \rightarrow \mathbb{R}$  and a number  $\beta > 0$  such that for any  $\gamma < \beta$ ,

$$\int_0^\infty \tau^{n+\gamma} \Phi(\tau) \frac{d\tau}{\tau} < \infty$$

and

$$|a_t(x, y)| \leq \frac{1}{t^{n/\beta}} \Phi\left(\frac{\rho(x, y)}{t^{1/\beta}}\right).$$

2. The Hölder continuity estimate: there is some constant  $\Theta > 0$  sufficiently small such that for all  $t \in (0, 1]$ , all  $x$  and  $y$  in  $\mathcal{X}$  with  $\rho(x, y) \leq t^{1/\beta}$ , and all  $u \in \mathcal{X}$ ,

$$|a_t(x, u) - a_t(y, u)| \leq \left(\frac{\rho(x, y)}{t^{1/\beta}}\right)^{\Theta} \frac{1}{t^{n/\beta}} \Phi\left(\frac{\rho(x, u)}{t^{1/\beta}}\right).$$

These conditions are found in [17]. As discussed there, examples of semigroups satisfying these estimates include the subordinated heat kernels in  $\mathbb{R}^n$ , the heat kernel on certain Riemannian manifolds, the heat kernel on a variety of fractals such as the unbounded Sierpinski Gasket, and the heat kernel of the semigroup  $e^{-tL}$  for certain elliptic operators  $L$  on  $\mathbb{R}^n$ .

We will show that if we assume conditions 1 and 2, then our geometric condition (G) is satisfied for all  $0 < \alpha < \min\{1, \Theta/\beta\}$ . The first step in showing this is to prove that our distance  $D_\alpha(x, y)$  defined from the semigroup is bounded above by a power of the distance  $\rho(x, y)$ .

**Lemma 2** *For any  $0 \leq \eta < 1$ , there is a finite constant  $C > 0$  such that for every  $0 < t \leq 1$  and every  $x \in \mathcal{X}$ ,*

$$\int_{\mathcal{X}} \rho(x, y)^{\beta\eta} \frac{1}{t^{n/\beta}} \Phi\left(\frac{\rho(x, y)}{t^{1/\beta}}\right) dy \leq C t^\eta.$$

*Proof* Let  $V_k = B(x, 2^{k+1}t^{1/\beta}) \setminus B(x, 2^k t^{1/\beta})$ . The upper bound on the kernel from condition 1 yields the following inequality:

$$\begin{aligned} & \int_{\mathcal{X}} \rho(x, y)^{\beta\eta} \frac{1}{t^{n/\beta}} \Phi\left(\frac{\rho(x, y)}{t^{1/\beta}}\right) dy \\ &= \frac{1}{t^{n/\beta}} \left\{ \int_{B(x, t^{1/\beta})} + \sum_{k=0}^{\infty} \int_{V_k} \right\} \rho(x, y)^{\beta\eta} \Phi\left(\frac{\rho(x, y)}{t^{1/\beta}}\right) dy \\ &\lesssim t^\eta t^{-n/\beta} \left\{ \Phi(0) \mu(B(x, t^{1/\beta})) \right. \\ &\quad \left. + \sum_{k=0}^{\infty} 2^{k\eta\beta} \Phi(2^k) \mu(B(x, 2^{k+1}t^{1/\beta})) \right\} \\ &\lesssim t^\eta t^{-n/\beta} \left\{ \Phi(0) t^{n/\beta} + \sum_{k=0}^{\infty} \Phi(2^k) 2^{k(n+\eta\beta)} t^{n/\beta} \right\} \\ &\lesssim t^\eta \left\{ \Phi(0) + \int_1^\infty \tau^{n+\eta\beta} \Phi(\tau) \frac{d\tau}{\tau} \right\} \lesssim t^\eta. \end{aligned}$$

We used that  $\eta < 1$  and condition 1 to conclude that the last integral is finite. This is the desired result.  $\square$

**Proposition 3** *For every  $0 < \alpha < \min\{1, \Theta/\beta\}$ , there is a constant  $C > 0$  such that  $D_\alpha(x, y) \leq C \min\{1, \rho(x, y)^{\alpha\beta}\}$ .*

*Proof* Since  $D_\alpha(x, y)$  is uniformly bounded, we need only consider the case when  $\rho(x, y) \leq 1$ . Condition 2 and Lemma 2 above with  $\eta = 0$  imply that whenever  $\rho(x, y) \leq t^{1/\beta}$ ,

$$\|a_t(x, \cdot) - a_t(y, \cdot)\|_1 \leq \left( \frac{\rho(x, y)}{t^{1/\beta}} \right)^\Theta \frac{1}{t^{n/\beta}} \int_{\mathcal{X}} \Phi \left( \frac{\rho(x, u)}{t^{1/\beta}} \right) du \lesssim \left( \frac{\rho(x, y)}{t^{1/\beta}} \right)^\Theta.$$

Consequently, if we define  $K$  so that  $2^{-K} \leq \rho(x, y)^\beta < 2^{-K+1}$ , then

$$\begin{aligned} D_\alpha(x, y) &\lesssim \rho(x, y)^\Theta \sum_{k=0}^K 2^{-k\alpha} 2^{k\Theta/\beta} + \sum_{k=K+1}^{\infty} 2^{-k\alpha} \lesssim \rho(x, y)^\Theta 2^{K(\Theta/\beta-\alpha)} + 2^{-K\alpha} \\ &\lesssim \rho(x, y)^{\alpha\beta}. \end{aligned}$$

We used that  $\alpha < \Theta/\beta$  for the upper bound on the first sum.  $\square$

With this upper bound on  $D_\alpha(x, y)$ , it is now straightforward to show that our geometric condition (G) holds for a range of  $\alpha$ .

**Theorem 1** *Under conditions 1 and 2, condition (G) holds for all  $0 < \alpha < \min\{1, \Theta/\beta\}$ .*

*Proof* From Proposition 3, we have the upper bound  $D_\alpha(x, y) \lesssim \rho(x, y)^{\alpha\beta}$ . Consequently, taking  $\eta = \alpha$  in Lemma 2 yields

$$\int_{\mathcal{X}} |a_t(x, y)| D_\alpha(x, y) dy \lesssim \int_{\mathcal{X}} \rho(x, y)^{\alpha\beta} \frac{1}{t^{n/\beta}} \Phi \left( \frac{\rho(x, y)}{t^{1/\beta}} \right) dy \lesssim t^\alpha$$

which is the desired result.  $\square$

### 3.2 The Distance $D_\alpha(x, y)$ for Kernels with a Matching Lower Bound

Having established conditions (G) and (R) from the upper bound  $D_\alpha(x, y) \lesssim \min\{1, \rho(x, y)^{\alpha\beta}\}$  for all  $0 < \alpha < \min\{1, \Theta/\beta\}$  under the continuity and decay conditions 1 and 2 of the previous section, we now formulate general conditions under which we can prove a corresponding lower bound,  $D_\alpha(x, y) \gtrsim \min\{1, \rho(x, y)^{\alpha\beta}\}$ . We will then study several examples where both conditions are satisfied.

Note that the lower bound is not necessary for the general results of our paper to hold; in particular, our primary concern was to prove Theorem 1, which establishes condition (G) for a large class of examples. We only prove lower bounds on  $D_\alpha(x, y)$

to show that it is equivalent to a snowflake of the “natural” distance  $\rho(x, y)$  in many cases.

In this section we will assume a stronger relation between the measure  $\rho(x, y)$  and the metric  $\mu$ , namely the two-sided estimate  $\mu(B(x, r)) \simeq r^n$ . We suppose too that in addition to conditions 1 and 2 of the previous section, we also have the following condition:

3. A local lower bound on the kernel: there is a monotonic decreasing function  $\Psi : [0, \infty) \rightarrow \mathbb{R}$  and  $R > 0$  such that for all  $t \in (0, 1]$  and all  $\rho(x, y) < R$

$$|a_t(x, y)| \geq \frac{1}{t^{n/\beta}} \Psi\left(\frac{\rho(x, y)}{t^{1/\beta}}\right).$$

We will show:

**Proposition 4** *Under the conditions 1,2 and 3,  $D_\alpha(x, y) \gtrsim \min\{1, \rho(x, y)^{\alpha\beta}\}$ .*

We will deduce the result from the following lemmas.

**Lemma 3** *There is a constant  $A > 1$  and a constant  $\epsilon > 0$  such that whenever  $x, y \in \mathcal{X}$  and  $t \in (0, 1]$  satisfy  $At^{1/\beta} \leq \rho(x, y) < R$ , we have*

$$\|a_t(x, \cdot) - a_t(y, \cdot)\|_1 > \epsilon.$$

*Proof* Temporarily fix any  $A > 1$  and suppose  $At^{1/\beta} \leq \rho(x, y) < R$ . Then for any  $u \in B(x, t^{1/\beta})$ , the triangle inequality implies

$$\rho(y, u) \geq \rho(x, y) - \rho(x, u) \geq (A - 1)t^{1/\beta}.$$

From the monotonicity of  $\Phi$  it follows that  $\Phi(\rho(y, u)/t^{1/\beta}) \leq \Phi(A - 1)$ . Consequently, using the upper and lower bounds on  $|a_t(x, y)|$  and the fact that  $\mu(B(x, r)) \simeq r^n$ , we have

$$\begin{aligned} & \|a_t(x, \cdot) - a_t(y, \cdot)\|_1 \\ & \geq \int_{B(x, t^{1/\beta})} |a_t(x, u)| du - \int_{B(x, t^{1/\beta})} |a_t(y, u)| du \\ & \geq \frac{1}{t^{n/\beta}} \int_{B(x, t^{1/\beta})} \Psi\left(\frac{\rho(x, u)}{t^{1/\beta}}\right) du - \frac{1}{t^{n/\beta}} \int_{B(x, t^{1/\beta})} \Phi\left(\frac{\rho(y, u)}{t^{1/\beta}}\right) du \\ & \geq \frac{1}{t^{n/\beta}} \int_{B(x, t^{1/\beta})} \Psi\left(\frac{\rho(x, u)}{t^{1/\beta}}\right) du - \frac{1}{t^{n/\beta}} \int_{B(x, t^{1/\beta})} \Phi(A - 1) du \\ & \geq C_1 \Psi(1) - C_2 \Phi(A - 1) \end{aligned}$$

for some constants  $C_1, C_2 > 0$ . Since  $\Phi$  is decreasing, by choosing  $A$  large enough we can guarantee  $\epsilon \equiv C_1 \Psi(1) - C_2 \Phi(A - 1) > 0$ , yielding the desired result.  $\square$

**Corollary 1** Let  $R$  be as in condition 3. Then for all  $\rho(x, y) < R$ ,

$$D_\alpha(x, y) \gtrsim \rho(x, y)^{\alpha\beta}.$$

*Proof* By the previous lemma,  $\|a_t(x, \cdot) - a_t(y, \cdot)\|_1 > \epsilon > 0$  whenever  $At^{1/\beta} \leq \rho(x, y)$ . Take  $L$  so that  $2^{-L} \leq \rho(x, y)^\beta / A^\beta < 2^{-L+1}$ . Then

$$D_\alpha(x, y) \geq \sum_{k=L}^{\infty} 2^{-k\alpha} \|p_k(x, \cdot) - p_k(y, \cdot)\|_1 \geq \epsilon \sum_{k=L}^{\infty} 2^{-k\alpha} \simeq 2^{-L\alpha} \simeq \rho(x, y)^{\alpha\beta}.$$

□

**Lemma 4** Let  $R$  be as in condition 3. There are constants  $C > 0$  and  $\delta > 0$  such that whenever  $\rho(x, y) \geq R$  and  $t^{1/\beta} < \delta R$ ,

$$\|a_t(x, \cdot) - a_t(y, \cdot)\|_1 \geq C.$$

*Proof* Since  $\rho(x, y) \geq R$ , the balls  $B(x, R/2)$  and  $B(y, R/2)$  are disjoint. Consequently

$$\begin{aligned} \|a_t(x, \cdot) - a_t(y, \cdot)\|_1 &\geq \int_{B(x, R/2)} |a_t(x, u)| du - \int_{B(x, R/2)} |a_t(y, u)| du \\ &\geq \int_{B(x, R/2)} |a_t(x, u)| du - \int_{B(y, R/2)^c} |a_t(y, u)| du \\ &\geq 1 - \int_{B(x, R/2)^c} |a_t(x, u)| du - \int_{B(y, R/2)^c} |a_t(y, u)| du. \end{aligned}$$

The result follows from Lemma 5 below. □

**Lemma 5** Fix any  $r > 0$  and  $\epsilon > 0$ . Then there exists  $\delta > 0$  sufficiently small so that whenever  $0 < t^{1/\beta} < \delta r$ ,

$$\int_{B(x, r)^c} |a_t(x, u)| du \leq \epsilon.$$

*Proof* Temporarily fix  $\delta > 0$  and suppose  $0 < t^{1/\beta} < \delta r$ . Then if we let  $V_k = B(x, 2^{k+1}r) \setminus B(x, 2^k r)$ , we have

$$\begin{aligned} \int_{B(x, r)^c} |a_t(x, u)| du &\leq \frac{1}{t^{n/\beta}} \int_{B(x, r)^c} \Phi\left(\frac{\rho(x, y)}{t^{1/\beta}}\right) du = \frac{1}{t^{n/\beta}} \sum_{k \geq 0} \int_{V_k} \Phi\left(\frac{\rho(x, y)}{t^{1/\beta}}\right) du \\ &\lesssim \frac{1}{t^{n/\beta}} \sum_{k \geq 0} \Phi\left(\frac{2^k r}{t^{1/\beta}}\right) (2^{k+1}r)^n \lesssim \frac{r^n}{t^{n/\beta}} \int_1^\infty \tau^n \Phi\left(\frac{\tau r}{t^{1/\beta}}\right) \frac{d\tau}{\tau} \\ &= \frac{r^n}{t^{n/\beta}} \int_{rt^{-1/\beta}}^\infty t^{n/\beta} r^{-n} s^n \Phi(s) \frac{ds}{s} \leq \int_{\delta^{-1}}^\infty s^n \Phi(s) \frac{ds}{s}. \end{aligned}$$

By taking  $\delta$  small enough, the integral can be made as small as desired, completing the proof.  $\square$

**Corollary 2** *There is a constant  $B > 0$  such that whenever  $\rho(x, y) \geq R$ ,  $D_\alpha(x, y) \geq B$ .*

*Proof* Take  $C$  and  $\delta$  from Lemma 4. Let  $K = \lfloor \log_2(1/(\delta^\beta R^\beta)) \rfloor$ . Then  $2^{-K} \leq \delta^\beta R^\beta$ , and so

$$D_\alpha(x, y) \geq \sum_{k=K}^{\infty} 2^{-k\alpha} C \simeq C \delta^{\alpha\beta} R^{\alpha\beta} > 0$$

which completes the proof.  $\square$

Corollaries 1 and 2 easily imply Proposition 4. Furthermore, Propositions 3 and 4 yield the following theorem:

**Theorem 2** *If all the conditions 1, 2 and 3 on  $a_t(x, y)$  hold, and if  $\mu(B(x, r)) \simeq r^n$ , then for  $0 < \alpha < \min\{1, \Theta/\beta\}$  the distance  $D_\alpha(x, y)$  is equivalent to the thresholded snowflake distance  $\min\{1, \rho(x, y)^{\alpha\beta}\}$ .*

### 3.3 Heat Kernel on a Riemannian Manifold

We illustrate Theorems 1 and 2 on some selected examples. We consider first the case where  $\mathcal{X}$  is a closed (compact, without boundary) Riemannian manifold of dimension  $n$ , and  $a_t(x, y)$  is its heat kernel. This section may be of interest to those in the machine-learning community, as approximations to the heat kernel on data sampled from submanifolds of  $\mathbb{R}^n$  are widely-used in the analysis of many data sets [1, 21].

Since  $\mathcal{X}$  is compact, it is true that  $\mu(B(x, r)) \simeq r^n$ . Furthermore, the following two lemmas can be easily derived from the parametrix construction of the heat kernel given in [5, Chapter VI, Section 4]. Here,  $\rho(x, y)$  is the geodesic distance on the manifold.

**Lemma 6** *There are positive constants  $A, B$  such that*

$$a_t(x, y) \leq \frac{A}{t^{n/2}} e^{-B\rho(x,y)^2/t}$$

for all  $t \in (0, 1]$ .

**Lemma 7** *There are positive constants  $C, D$*

$$\frac{C}{t^{n/2}} e^{-D\rho(x,y)^2/t} \leq a_t(x, y)$$

whenever  $t \in (0, 1]$  and  $\rho(x, y)$  is sufficiently small.

Taking  $\beta = 2$ , from Lemma 6 we see that the heat kernel  $a_t(x, y)$  satisfies condition 1 with  $\Phi(\tau) \simeq e^{-B\tau^2}$  and times  $0 < t \leq 1$ , which is the only range of  $t$  we made use of in Sect. 3.1; and from Lemma 7 the heat kernel satisfies condition 3 with  $\Psi(\tau) \simeq e^{-D\tau^2}$ . To apply Theorem 2, it remains to show condition 2. We will deduce the continuity estimate from the following gradient bound:

**Lemma 8** *There are constants  $E, F > 0$  such that for all  $t \in (0, 1]$  and for all  $x$  and  $y$  in  $\mathcal{X}$ ,*

$$\|\nabla_x a_t(x, y)\| \leq \frac{E}{\sqrt{t}} \frac{e^{-F\rho(x, y)^2/t}}{t^{n/2}}$$

where  $\nabla_x$  denotes the gradient with respect to the first variable.

*Proof* Using the asymptotic expansion of  $a_t(x, y)$  in [5, Chapter VI, Section 4], it is easy to show a Gaussian upper bound on the time derivative of  $a_t(x, y)$ , namely

$$\left| \frac{\partial a_t}{\partial t}(x, y) \right| \leq \frac{b}{t} \frac{e^{-c\rho(x, y)^2/t}}{t^{n/2}}$$

for some positive constants  $b, c$ . Since the curvature of  $\mathcal{X}$  is bounded (because  $\mathcal{X}$  is compact) we can apply Theorem 1.4 from [20], which states that there are constants  $A_1, A_2, A_3$  such that

$$\|\nabla_x a_t(x, y)\|^2 \leq \left( A_1 + \frac{A_2}{t} \right) a_t(x, y)^2 + A_3 a_t(x, y) \frac{\partial a_t}{\partial t}(x, y).$$

For  $t \in (0, 1]$ , it follows from the Gaussian estimates on  $a_t(x, y)$  and  $\partial_t a_t(x, y)$  that

$$\|\nabla_x a_t(x, y)\|^2 \leq \frac{\tilde{A}}{t} \frac{e^{-2B\rho(x, y)^2/t}}{t^n} + \frac{b}{t} \frac{e^{-c\rho(x, y)^2/t}}{t^{n/2}} \frac{A}{t^{n/2}} e^{-B\rho(x, y)^2/t} \lesssim \frac{1}{t} \frac{e^{-\tilde{B}\rho(x, y)^2/t}}{t^n}$$

for sufficiently small  $\tilde{B} > 0$ , from which the result follows.  $\square$

We next prove a consequence of the mean value theorem that will be useful.

**Lemma 9** *If  $x, y \in \mathcal{X}$  are sufficiently close, then for any smooth function  $h : \mathcal{X} \rightarrow \mathbb{R}$  there is a point  $\tilde{x}$  lying on the minimal geodesic from  $x$  to  $y$  such that*

$$|h(x) - h(y)| \leq \|\nabla h(\tilde{x})\| \rho(x, y).$$

*Proof* Suppose  $r \equiv \rho(x, y)$  is less than the injectivity radius of the manifold  $M$  (which is positive, since  $M$  is compact). Let  $\gamma(t)$  be the unit speed geodesic connecting  $x$  to  $y$ . Then  $\gamma(r) = y$ , and  $\gamma(0) = x$ . For details, see, for instance, [10, Chapter 13, Section 2].

Consider the function  $\hat{h}(t) = h(\gamma(t))$ . Observe that  $\hat{h}(0) = h(x)$  and  $\hat{h}(r) = h(y)$ . By the mean value theorem, there is some point  $t_1$  between 0 and  $r$  such that

$$\frac{h(y) - h(x)}{\rho(x, y)} = \frac{\hat{h}(r) - \hat{h}(0)}{r} = \hat{h}'(t_1) = \frac{d}{dt} h(\gamma(t)) \Big|_{t=t_1} = \langle \nabla h(\gamma(t_1)), \gamma'(t_1) \rangle$$

Consequently, since  $\gamma$  has unit speed, the Cauchy–Schwarz inequality gives

$$|h(y) - h(x)| = |\langle \nabla h(\gamma(t_1)), \gamma'(t_1) \rangle| \rho(x, y) \leq \|\nabla h(\gamma(t_1))\| \rho(x, y).$$

Consequently, if we let  $\tilde{x} = \gamma(t_1)$ , then  $\tilde{x}$  lies on the minimal geodesic connecting  $x$  and  $y$ , and

$$|h(x) - h(y)| \leq \|\nabla h(\tilde{x})\| \rho(x, y).$$

□

**Corollary 3** *There are positive constants  $G, H$  such that whenever  $t \in (0, 1]$  and  $\rho(x, y) \leq t^{1/2}$ ,*

$$|a_t(x, u) - a_t(y, u)| \leq G \frac{\rho(x, y)}{\sqrt{t}} \frac{e^{-H\rho(x, u)^2/t}}{t^{n/2}}.$$

*Proof* From Lemma 9 and the gradient estimate from Lemma 8, we have the bound

$$|a_t(x, u) - a_t(y, u)| \leq \rho(x, y) \frac{E}{\sqrt{t}} \frac{e^{-F\rho(u, \tilde{x})^2/t}}{t^{n/2}}.$$

where  $\tilde{x}$  is some point on the minimal geodesic connecting  $x$  and  $y$ . Since  $\rho(x, y) \leq t^{1/2}$ , it is also true that  $\rho(x, \tilde{x}) \leq t^{1/2}$ . Consequently, we have

$$\rho(u, x)^2 \leq 2\rho(u, \tilde{x})^2 + 2\rho(\tilde{x}, x)^2 \leq 2\rho(u, \tilde{x})^2 + 2t$$

and so

$$\begin{aligned} |a_t(x, u) - a_t(y, u)| &\leq \rho(x, y) \frac{E}{\sqrt{t}} \frac{e^{-F\rho(u, \tilde{x})^2/t}}{t^{n/2}} \leq \rho(x, y) \frac{E}{\sqrt{t}} \frac{e^{-F(\rho(u, x)^2 - 2t)/2t}}{t^{n/2}} \\ &\leq \rho(x, y) \frac{E e^{-F/2\rho(u, x)^2/t}}{\sqrt{t}} \end{aligned}$$

which is the desired result. □

Corollary 3 gives us condition 2. We can therefore apply Theorems 1 and 2 to conclude that for all  $0 < \alpha < 1/2$ ,  $D_\alpha(x, y) \simeq \rho(x, y)^{2\alpha}$ , and condition (G) is satisfied.

We note that to establish (G), we only made use of the Gaussian upper bound from Lemma 6, the continuity estimate from Corollary 3, and the upper bound  $\mu(B(x, r)) \lesssim r^n$ . Condition (G) will therefore hold for the heat kernel on any manifold where these estimates are true, and not just closed manifolds. For example, as discussed in [17], the Gaussian upper bounds and continuity estimates hold for the heat kernel on any geodesically complete Riemannian manifold with non-negative curvature.

### 3.4 Subordinated Heat Kernels with Shifts on $\mathbb{R}^n$

Next we consider the case in which  $a_t(x, y) = K_t(x - y)$ , where  $K_t(u)$  is a radial kernel, i.e.  $K_t(x) = K_t(y)$  if  $|x| = |y|$ , satisfying the following scaling property:

$$K_t(x) = t^{-n/\beta} K_1(t^{-1/\beta} x)$$

where  $0 < \beta \leq 2$ . For details on the construction of such kernels in one dimension, the reader can refer to the book [33]. These kernels are known as *subordinated heat kernels* on  $\mathbb{R}^n$ , and can be expressed as an average of the Gaussian heat kernel at different scales. Concretely, when  $0 < \beta < 2$  (that is,  $\beta \neq 2$ ),  $K_t(x)$  is of the form

$$K_t(x) = \int_0^\infty \eta_t(s) g_s(x) ds$$

where  $g_s$  is the Gaussian kernel at time  $s$ , and for each  $t$  the function  $\eta_t(s)$  is a probability density on  $(0, \infty)$ , known as the *subordinator*. In fact,  $\eta_t$  satisfies the identity

$$\exp(-t\lambda^{\beta/2}) = \int_0^\infty \eta_t(s) e^{-s\lambda} ds$$

for all  $\lambda \geq 0$ , from which it easily follows that the Fourier transform of  $K_t$  is

$$\hat{K}_t(\xi) = \exp(-t|\xi|^\beta).$$

For example, when  $\beta = 2$ ,  $K_t$  is equal to the Gaussian heat kernel, and when  $\beta = 1$  it is equal to the Poisson kernel.

It is shown in [17] that any subordinated heat kernel satisfies conditions 1, 2 and 3. More precisely,

$$a_t(x, y) \simeq \frac{1}{t^{n/\beta}} \left(1 + \frac{|x - y|}{t^{1/\beta}}\right)^{-(n+\beta)}$$

and

$$|a_t(x, u) - a_t(y, u)| \lesssim \frac{|x - y|}{t^{1/m}} \frac{1}{t^{n/m}} \left(1 + \frac{|x - y|}{t^{1/m}}\right)^{-(n+m)}.$$

From Theorems 1 and 2, it follows immediately that whenever  $0 < \alpha < 1/\beta$ , the distance  $D_\alpha(x, y)$  with respect to the kernel  $a_t(x, y) = K_t(x - y)$  is equivalent to  $\min\{1, |x - y|^{\alpha\beta}\}$  and that condition (G) holds. Note that our use of the parameter  $\beta$  in the definition of the subordinated heat kernel coincides with its use in the conditions 1, 2 and 3.

This leads us to a family of examples of *non-symmetric* semigroups for which condition (G) holds, namely the subordinated heat kernels with shifts. Take  $\beta \in [1, 2]$ . Then for a fixed parameter  $\theta \in \mathbb{R}$ , define

$$a_t(x, y) = t^{-n\beta} K_1(t^{-1/\beta}(x - \theta t - y)).$$

It is easy to check from the semigroup property for the non-shift case  $\theta = 0$  that  $a_t(x, y)$  is also a semigroup. Furthermore, when  $0 < \alpha < 1/\beta$  we still have  $D_\alpha(x, y) \simeq \min\{1, |x - y|^{\alpha\beta}\}$ . Therefore, we can verify condition (G) directly by writing

$$\begin{aligned} \int_{\mathbb{R}^n} a_t(x, y) D_\alpha(x, y) dy &\lesssim \int_{\mathbb{R}^n} t^{-n/\beta} K_1(t^{-1/\beta}(x - \theta t - y)) \min\{1, |x - y|^{\alpha\beta}\} dy \\ &= \int_{\mathbb{R}^n} t^{-n/\beta} K_1(t^{-1/\beta}(x - y)) \min\{1, |x - y + \theta t|^{\alpha\beta}\} dy \\ &\leq \int_{\mathbb{R}^n} t^{-n/\beta} K_1(t^{-1/\beta}(x - y)) \min\{1, |x - y|^{\alpha\beta}\} dy \\ &\quad + \int_{\mathbb{R}^n} t^{-n/\beta} K_1(t^{-1/\beta}(x - y)) |\theta t|^{\alpha\beta} dy \\ &\lesssim t^\alpha + t^{\alpha\beta}. \end{aligned}$$

The last line follows from condition (G) in the case  $\theta = 0$ . Since  $\beta \geq 1$ , condition (G) is satisfied. Note that this range of  $\beta$  includes both the heat kernel ( $\beta = 2$ ) and the Poisson kernel ( $\beta = 1$ ).

### 3.5 Products of Kernels and Anisotropic Distances

Suppose that  $a_t(x_1, x_2)$  and  $b_t(y_1, y_2)$  are two semigroups on spaces  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively, for which the conditions (S), (C), (I) and (R) (and thus (G)) hold. We define their product by  $c_t((x_1, y_1), (x_2, y_2)) = a_t(x_1, x_2) \cdot b_t(y_1, y_2)$ . It is easy to check that the kernel  $c_t$  defines a semigroup on  $\mathcal{X} \times \mathcal{Y}$ , and that the three conditions (S), (C), and (I) all hold. We will check that (G) holds as well. Since we have three semigroups, we augment our notation to distinguish between the distances each one induces. Fixing the distance parameter  $\alpha$ , we will write  $D_\alpha^a(x_1, x_2)$  for the distance induced by  $a_t$ , and similarly for  $b_t$  and  $c_t$ . We then have:

**Proposition 5** *The distance  $D_\alpha^c((x_1, y_1), (x_2, y_2))$  on  $\mathcal{X} \times \mathcal{Y}$  is equivalent to  $D_\alpha^a(x_1, x_2) + D_\alpha^b(y_1, y_2)$ .*

*Proof* This follows immediately from the following lemma.  $\square$

**Lemma 10** For every  $z_1 = (x_1, y_1), z_2 = (x_2, y_2)$  in  $\mathcal{X} \times \mathcal{Y}$ , we have

$$\|c_t(z_1, \cdot) - c_t(z_2, \cdot)\|_1 \simeq \|a_t(x_1, \cdot) - a_t(x_2, \cdot)\|_1 + \|b_t(y_1, \cdot) - b_t(y_2, \cdot)\|_1.$$

*Proof* First, we prove that

$$\|c_t(z_1, \cdot) - c_t(z_2, \cdot)\|_1 \lesssim \|a_t(x_1, \cdot) - a_t(x_2, \cdot)\|_1 + \|b_t(y_1, \cdot) - b_t(y_2, \cdot)\|_1.$$

To see this, observe that if  $x_1 = x_2$  then

$$\begin{aligned} \|c_t((x_1, y_1), \cdot) - c_t((x_1, y_2), \cdot)\|_1 &= \int_{\mathcal{X}} \int_{\mathcal{Y}} |a_t(x_1, x)| |b_t(y_1, y) - b_t(y_2, y)| dy dx \\ &\lesssim \|b_t(y_1, \cdot) - b_t(y_2, \cdot)\|_1 \end{aligned}$$

where we have used condition (I) on the kernels  $a_t$ . Similarly,

$$\|c_t((x_1, y_2), \cdot) - c_t((x_2, y_2), \cdot)\|_1 \lesssim \|a_t(x_1, \cdot) - a_t(x_2, \cdot)\|_1.$$

We therefore have

$$\begin{aligned} &\|c_t((x_1, y_1), \cdot) - c_t((x_2, y_2), \cdot)\|_1 \\ &\leq \|c_t((x_1, y_1), \cdot) - c_t((x_1, y_2), \cdot)\|_1 \\ &\quad + \|c_t((x_1, y_2), \cdot) - c_t((x_2, y_2), \cdot)\|_1 \\ &\lesssim \|a_t(x_1, \cdot) - a_t(x_2, \cdot)\|_1 + \|b_t(y_1, \cdot) - b_t(y_2, \cdot)\|_1, \end{aligned}$$

as desired.

For the other direction, observe that

$$\begin{aligned} \|c_t(z_1, \cdot) - c_t(z_2, \cdot)\|_1 &= \int_{\mathcal{X}} \int_{\mathcal{Y}} |a_t(x_1, x)b_t(y_1, y) - a_t(x_2, x)b_t(y_2, y)| dy dx \\ &\geq \int_{\mathcal{X}} \left| \int_{\mathcal{Y}} [a_t(x_1, x)b_t(y_1, y) - a_t(x_2, x)b_t(y_2, y)] dy \right| dx \\ &= \int_{\mathcal{X}} \left| a_t(x_1, x) \int_{\mathcal{Y}} b_t(y_1, y) dy - a_t(x_2, x) \int_{\mathcal{Y}} b_t(y_2, y) dy \right| dx \\ &= \int_{\mathcal{X}} |a_t(x_1, x) - a_t(x_2, x)| dx = \|a_t(x_1, \cdot) - a_t(x_2, \cdot)\|_1. \end{aligned}$$

where we have used condition (C) in the last equality. Similarly,

$$\|c_t(z_1, \cdot) - c_t(z_2, \cdot)\|_1 \geq \|b_t(y_1, \cdot) - b_t(y_2, \cdot)\|_1$$

from which it follows

$$\|c_t(z_1, \cdot) - c_t(z_2, \cdot)\|_1 \geq \frac{1}{2} (\|a_t(x_1, \cdot) - a_t(x_2, \cdot)\|_1 + \|b_t(y_1, \cdot) - b_t(y_2, \cdot)\|_1)$$

completing the proof.  $\square$

From Proposition 5 we can easily deduce that condition (G) holds for  $c_t$  if it holds for  $a_t$  and  $b_t$ .

**Proposition 6** *If condition (G) holds for  $a_t$  and  $b_t$ , then it holds for their product  $c_t$  as well.*

*Proof* We have, using condition (I) for both  $a_t$  and  $b_t$ ,

$$\begin{aligned} & \int_{\mathcal{X} \times \mathcal{Y}} |c_t(z_1, z_2)| D_\alpha^c(z_1, z_2) dz_2 \\ & \lesssim \int_{\mathcal{X}} \int_{\mathcal{Y}} |a_t(x_1, x_2) \cdot b_t(y_1, y_2)| (D_\alpha^a(x_1, x_2) + D_\alpha^b(y_1, y_2)) dx_2 dy_2 \\ & \lesssim \int_{\mathcal{X}} |a_t(x_1, x_2)| D_\alpha^a(x_1, x_2) dx_2 + \int_{\mathcal{Y}} |b_t(y_1, y_2)| D_\alpha^b(y_1, y_2) dy_2 \\ & \lesssim t^\alpha \end{aligned}$$

which is the desired result.  $\square$

Of course, these results hold for the product of any number of kernels, not just two, and the proofs are similar. A natural example is the product of subordinated heat kernels on  $\mathbb{R}^n$ . Suppose that  $n = n_1 + \dots + n_l$  and that on each space  $\mathbb{R}^{n_i}$  we have a subordinated heat kernel  $a_t^{(i)}(x_i, y_i)$  with scaling parameter  $\beta_i$ , as in Sect. 3.4. Then as long as  $0 < \alpha < \min\{1, 1/\beta_1, \dots, 1/\beta_l\}$ , for  $x = (x_1, \dots, x_l)$ ,  $y = (y_1, \dots, y_l)$ ,  $x_i, y_i \in \mathbb{R}^{n_i}$ , the kernel

$$a_t(x, y) = \prod_{i=1}^l a_t^{(i)}(x_i, y_i)$$

generates the distance

$$D_\alpha(x, y) \simeq \min\{1, |x - y|_{\text{AN}}\}$$

where

$$|x - y|_{\text{AN}} = \sum_{i=1}^l |x_i - y_i|^{\alpha \beta_i}$$

is an example of an anisotropic distance on  $\mathbb{R}^n$ ; see, for example, [9, 13, 27] for work on function spaces built from such distances.

## 4 The Space of Hölder–Lipschitz Functions

We now turn to characterizing functions that are Lipschitz with respect to the distance  $D_\alpha(x, y)$ , for a fixed  $\alpha \in (0, 1)$ . We assume that  $\alpha$  is chosen so that condition (G)

holds; in particular, by Proposition 1 if the kernel satisfies condition (R) for some  $\alpha'$ , we take any  $0 < \alpha < \alpha'$ . As we have seen in Sect. 3, the distance  $D_\alpha(x, y)$  is, in many cases of interest, of the form  $\rho(x, y)^{\alpha\beta}$ , where  $\rho(x, y)$  is some other distance on  $\mathcal{X}$  and  $0 < \alpha\beta < 1$ . In classical analysis, functions that are Lipschitz with respect to a snowflake metric are called *Hölder*. We will therefore refer to the space of Lipschitz functions with respect to  $D_\alpha(x, y)$  as the *Hölder–Lipschitz space*.

More formally, for a function  $f$  on  $\mathcal{X}$  define its maximum variation seminorm as

$$V(f) = \sup_{x \neq y} \frac{|f(x) - f(y)|}{D_\alpha(x, y)}.$$

We then define the Hölder–Lipschitz norm of a function  $f$  on  $\mathcal{X}$  to be

$$\|f\|_{\Lambda_\alpha} = \sup_x |f(x)| + V(f)$$

and the Hölder–Lipschitz space  $\Lambda_\alpha$  to be the collection of all functions on  $\mathcal{X}$  for which this norm is finite.

We will define two alternate norms on  $\Lambda_\alpha$  and show that they are equivalent to  $\|f\|_{\Lambda_\alpha}$ . We define the difference operators

$$\Delta_k = P_{k+1} - P_k, \quad \delta_k = I - P_k.$$

We also define the seminorms

$$V^{(1)}(f) = \sup_{k \geq 0} \sup_x 2^{k\alpha} |\Delta_k f(x)|$$

and

$$V^{(2)}(f) = \sup_{k \geq 0} \sup_x 2^{k\alpha} |\delta_k f(x)|.$$

The alternate norms can now be defined as

$$\|f\|_{\Lambda_\alpha}^{(1)} = \sup_x |f(x)| + V^{(1)}(f)$$

and

$$\|f\|_{\Lambda_\alpha}^{(2)} = \sup_x |f(x)| + V^{(2)}(f).$$

We immediately see the use of condition (R) and its equivalent condition (G) in the following result:

**Proposition 7** *For all  $f \in \Lambda_\alpha$ ,  $V^{(2)}(f) \lesssim V(f)$ .*

*Proof* Take any  $k \geq 0$ . Since  $p_k(x, \cdot)$  has integral 1 for every  $x$ , we have

$$\begin{aligned} |f(x) - P_k f(x)| &= \left| f(x) - \int_{\mathcal{X}} p_k(x, y) f(y) dy \right| = \left| \int_{\mathcal{X}} p_k(x, y) (f(x) - f(y)) dy \right| \\ &\leq V(f) \int_{\mathcal{X}} |p_k(x, y)| D_\alpha(x, y) dy \lesssim V(f) 2^{-k\alpha} \end{aligned}$$

from which the desired inequality follows trivially.  $\square$

**Corollary 4** For all  $f \in \Lambda_\alpha$ ,  $\|f\|_{\Lambda_\alpha}^{(2)} \lesssim \|f\|_{\Lambda_\alpha}$ .

Next, we make the following simple observation about uniform convergence to the identity:

**Lemma 11** If  $\|f\|_{\Lambda_\alpha}^{(2)} < \infty$ , then  $P_k f$  converges to  $f$  uniformly as  $k \rightarrow \infty$ .

*Proof* This is clear from the definition of  $\|f\|_{\Lambda_\alpha}^{(2)}$  (more specifically, the definition of  $V^{(2)}(f)$ ).  $\square$

Since  $\|f\|_{\Lambda_\alpha}^{(2)} \lesssim \|f\|_{\Lambda_\alpha}$ , it follows that:

**Lemma 12** For all  $f \in \Lambda_\alpha$ ,  $P_k f$  converges to  $f$  uniformly as  $k \rightarrow \infty$ .

We now prove:

**Proposition 8** The seminorms  $V^{(1)}(f)$  and  $V^{(2)}(f)$  are equivalent for  $f \in \Lambda_\alpha$ .

*Proof* First, write  $f$  as a telescopic series:

$$f - P_0 f = \sum_{l=0}^{\infty} [P_{l+1} f - P_l f]$$

where the series converges uniformly by Lemma 12.

Similarly we can write  $P_k f$  as a telescopic series

$$P_k f - P_0 f = \sum_{l=0}^{k-1} [P_{l+1} f - P_l f]$$

and subtracting the two series gives:

$$\begin{aligned} |f - P_k f| &= \left| \sum_{l=0}^{\infty} (P_{l+1} f - P_l f) - \sum_{l=0}^{k-1} (P_{l+1} f - P_l f) \right| \\ &= \left| \sum_{l=k}^{\infty} (P_{l+1} f - P_l f) \right| \leq \sum_{l=k}^{\infty} |P_{l+1} f - P_l f| \\ &\leq V^{(1)}(f) \sum_{l=k}^{\infty} 2^{-l\alpha} = V^{(1)}(f) \frac{1}{1 - 2^{-\alpha}} 2^{-k\alpha} \end{aligned}$$

and consequently

$$2^{k\alpha} \sup_x |\delta_k f(x)| \leq \frac{1}{1 - 2^{-\alpha}} V^{(1)}(f).$$

Taking the supremum over all  $k \geq 0$  shows  $V^{(2)}(f) \lesssim V^{(1)}(f)$ .

For the other direction, we simply observe

$$|P_k f - P_{k+1} f| \leq |(P_k - I)f| + |(P_{k+1} - I)f| \leq 2V^{(2)}(f)2^{-k\alpha}$$

implying

$$2^{k\alpha} \sup_x |\Delta_k f(x)| \leq 2V^{(2)}(f).$$

Taking the supremum over all  $k \geq 0$  gives the result.  $\square$

**Corollary 5** *The norms  $\|f\|_{\Lambda_\alpha}^{(1)}$  and  $\|f\|_{\Lambda_\alpha}^{(2)}$  are equivalent on  $\Lambda_\alpha$ .*

Now we turn to proving the main result of this section, namely that  $\|f\|_{\Lambda_\alpha}^{(1)}$  and  $\|f\|_{\Lambda_\alpha}^{(2)}$  are equivalent to  $\|f\|_{\Lambda_\alpha}$ . The following simple observation will be useful:

**Lemma 13**  $(P_{k+1} + P_k)\Delta_k = \Delta_{k-1}$ .

*Proof* This is a simple algebraic computation:

$$\begin{aligned} (P_{k+1} + P_k)\Delta_k &= (P_{k+1} + P_k)(P_{k+1} - P_k) = P_{k+1}P_{k+1} - P_{k+1}P_k + P_kP_{k+1} - P_kP_k \\ &= A_{2-(k+1)}A_{2-(k+1)} - A_{2-k}A_{2-k} = A_{2-(k+1)+2-(k+1)} - A_{2-k+2-k} \\ &= A_{2-k} - A_{2-(k-1)} = P_k - P_{k-1} = \Delta_{k-1}. \end{aligned}$$

$\square$

**Lemma 14** *Suppose  $f$  is bounded. Then for all  $x, y \in \mathcal{X}$ ,*

$$|P_k f(x) - P_k f(y)| \leq \sup_{x'} |f(x')| \cdot D_k(x, y)$$

*Proof*

$$\begin{aligned} |P_k f(x) - P_k f(y)| &= \left| \int_{\mathcal{X}} p_k(x, u) f(u) du - \int_{\mathcal{X}} p_k(y, u) f(u) du \right| \\ &= \left| \int_{\mathcal{X}} [p_k(x, u) - p_k(y, u)] f(u) du \right| \\ &\leq \sup_{x'} |f(x')| \cdot \|p_k(x, \cdot) - p_k(y, \cdot)\|_1 \\ &= \sup_{x'} |f(x')| \cdot D_k(x, y). \end{aligned}$$

$\square$

**Proposition 9** For all  $f \in \Lambda_\alpha$ ,  $\|f\|_{\Lambda_\alpha} \lesssim \|f\|_{\Lambda_\alpha}^{(1)}$ .

*Proof* Expand  $f$  in a telescopic series:

$$f - P_0 f = \sum_{k=0}^{\infty} [P_{k+1} f - P_k f] = \sum_{k=0}^{\infty} \Delta_k f(x) = \sum_{k=0}^{\infty} [(P_{k+1} + P_{k+2}) \Delta_{k+1} f]$$

where we have used Lemma 13. The series converges uniformly by Lemma 12.

For all  $x, y \in \mathcal{X}$ ,

$$\begin{aligned} |P_k \Delta_k f(x) - P_k \Delta_k f(y)| &= \left| \int_{\mathcal{X}} p_k(x, u) (\Delta_k f)(u) du - \int_{\mathcal{X}} p_k(y, u) (\Delta_k f)(u) du \right| \\ &= \left| \int_{\mathcal{X}} (p_k(x, u) - p_k(y, u)) (\Delta_k f)(u) du \right| \\ &\leq V^{(1)}(f) 2^{-k\alpha} D_k(x, y). \end{aligned} \quad (2)$$

Similarly,

$$|P_{k+1} \Delta_k f(x) - P_{k+1} \Delta_k f(y)| \leq V^{(1)}(f) 2^{-k\alpha} D_{k+1}(x, y). \quad (3)$$

For every  $x, y \in \mathcal{X}$

$$\begin{aligned} f(x) - f(y) &= \sum_{k=0}^{\infty} [(P_{k+1} + P_{k+2}) \Delta_{k+1} f](x) - \sum_{k=0}^{\infty} [(P_{k+1} + P_{k+2}) \Delta_{k+1} f](y) \\ &\quad + P_0 f(x) - P_0 f(y). \end{aligned}$$

From the inequalities (2) and (3) we get

$$\begin{aligned} &\left| \sum_{k=0}^{\infty} [(P_{k+1} + P_{k+2}) \Delta_{k+1} f](x) - \sum_{k=0}^{\infty} [(P_{k+1} + P_{k+2}) \Delta_{k+1} f](y) \right| \\ &\leq \left| \sum_{k=0}^{\infty} (P_{k+1} \Delta_{k+1} f(x) - P_{k+1} \Delta_{k+1} f(y)) \right| \\ &\quad + \left| \sum_{k=0}^{\infty} (P_{k+2} \Delta_{k+1} f(x) - P_{k+2} \Delta_{k+1} f(y)) \right| \\ &\leq \sum_{k=0}^{\infty} V^{(1)}(f) 2^{-(k+1)\alpha} D_{k+1}(x, y) + \sum_{k=0}^{\infty} V^{(1)}(f) 2^{-(k+1)\alpha} D_{k+2}(x, y) \\ &\leq V^{(1)}(f) (1 + 2^\alpha) D_\alpha(x, y). \end{aligned}$$

By Lemma 14, we also know  $|P_0 f(x) - P_0 f(y)| \leq \sup_{x'} |f(x')| D_\alpha(x, y)$ . Consequently, for every  $x, y \in \mathcal{X}$

$$|f(x) - f(y)| \leq (V^{(1)}(f)(1 + 2^\alpha) + \sup_{x'} |f(x')|) D_\alpha(x, y)$$

and so

$$\sup_{x \neq y} \frac{|f(x) - f(y)|}{D_\alpha(x, y)} \leq V^{(1)}(f)(1 + 2^\alpha) + \sup_{x'} |f(x')|.$$

Therefore

$$\begin{aligned} \|f\|_{\Lambda_\alpha} &= \sup_x |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{D_\alpha(x, y)} \leq V^{(1)}(f)(1 + 2^\alpha) + 2 \sup_x |f(x)| \\ &\leq 3 \|f\|_{\Lambda_\alpha}^{(1)}. \end{aligned}$$

□

Putting together Corollaries 4, 5 and Proposition 9, we have shown:

**Theorem 3** *The norms  $\|f\|_{\Lambda_\alpha}$ ,  $\|f\|_{\Lambda_\alpha}^{(1)}$  and  $\|f\|_{\Lambda_\alpha}^{(2)}$  are equivalent on  $\Lambda_\alpha$ .*

#### 4.1 The Necessity of Condition (G) for Positive Kernels

The only place in which we have made use of condition (G) so far—and indeed, the only place we will directly use it in the whole paper—was in establishing Proposition 7; in fact, assuming condition (G) made the proof of Proposition 7 almost tautological. While it is true that in Sect. 3 we showed that condition (G) holds for vastly many semigroups encountered throughout mathematics, the reader may still wonder if it is actually necessary to assume it, or if perhaps an even weaker condition could be imposed on the semigroup.

In this subsection, we give a partial answer to this question. Specifically, we show that if the kernels  $p_k(x, y)$  are non-negative for all  $k \geq 0$  and for all  $x, y \in \mathcal{X}$ , then condition (G) is equivalent to the result of Proposition 7. Consequently, in order for the Hölder–Lipschitz norm of a function to be equivalent in size to its scale variations, as defined by the norms  $\|f\|_{\Lambda_\alpha}^{(1)}$  and  $\|f\|_{\Lambda_\alpha}^{(2)}$ , condition (G) must hold. In fact, this statement is true not only for the distance  $D_\alpha(x, y)$  but for *any* distance  $D(x, y)$  on  $\mathcal{X}$ .

We now formalize this result:

**Proposition 10** *Suppose the kernels  $p_k(x, y)$  are non-negative for all  $k \geq 0$  and for all  $x, y \in \mathcal{X}$ . Let  $D(x, y)$  be any distance on  $\mathcal{X}$  and  $\alpha \in (0, 1)$ . Suppose there is a constant  $C > 0$  such that for any function  $f$  on  $\mathcal{X}$*

$$\sup_{k \geq 0} \sup_{x \in \mathcal{X}} 2^{k\alpha} |\delta_k f(x)| \leq C \sup_{x \neq y} \frac{|f(x) - f(y)|}{D(x, y)} \tag{4}$$

Then condition (G) must hold, with  $D(x, y)$  in place of  $D_\alpha(x, y)$ ; that is,

$$\int_{\mathcal{X}} p_k(x, y) \cdot D(x, y) dy \leq C 2^{-k\alpha}.$$

*Proof* Take any  $x_0 \in \mathcal{X}$ , and consider the function  $f(x) = D(x_0, x)$ . Then  $f(x_0) = 0$ , so that

$$|\delta_k f(x_0)| = |f(x_0) - P_k f(x_0)| = \int_{\mathcal{X}} p_k(x_0, x) D(x_0, x) dy.$$

Because

$$\sup_{x \neq y} \frac{|f(x) - f(y)|}{D(x, y)} = 1$$

assumption (4) gives

$$\int_{\mathcal{X}} p_k(x, y) \cdot D(x, y) dy = |\delta_k f(x_0)| \leq C 2^{-k\alpha}$$

completing the proof.  $\square$

## 5 The Space Dual to $\Lambda_\alpha$

We now turn to the space of  $L^1$  measures on  $\mathcal{X}$  (that is, measures with finite total variation; see, for instance, [14]). Since all Hölder–Lipschitz functions are in  $L^\infty$ , we can view every such measure as a distribution acting on  $\Lambda_\alpha$ . We will denote the action of such a distribution  $T$  on a function  $f \in \Lambda_\alpha$  by  $\langle f, T \rangle = \int_{\mathcal{X}} f(x) dT(x)$ . The dual norm to the Hölder–Lipschitz space  $\Lambda_\alpha$  is defined as:

$$\|T\|_{\Lambda_\alpha^*} = \sup_{\|f\|_{\Lambda_\alpha} \leq 1} \langle f, T \rangle.$$

The space  $\Lambda_\alpha^*$  is the space of  $L^1$  measures  $T$  equipped with the norm  $\|T\|_{\Lambda_\alpha^*}$ . In Sect. 6, we will give a well-known interpretation of the dual norm of the difference of two probability measures.

In this section we will define two other norms on  $\Lambda_\alpha^*$  that are more amenable to computation, and prove their equivalence to  $\|T\|_{\Lambda_\alpha^*}$ . In what follows, for a linear operator  $A$  acting on  $\Lambda$  we denote by  $A^*$  the dual operator acting on  $\Lambda^*$ .

We define the seminorms

$$W^{(1)}(T) = \sum_{k \geq 0} 2^{-k\alpha} \|\Delta_k^* T\|_1$$

and

$$W^{(2)}(T) = \sum_{k \geq 0} 2^{-k\alpha} \|d_k^* T\|_1$$

where

$$d_k = P_k - P_0.$$

Now we define the equivalent norms. The first is defined by

$$\|T\|_{\Lambda_\alpha^*}^{(1)} = \|P_0^* T\|_1 + W^{(1)}(T)$$

and the second is defined by

$$\|T\|_{\Lambda_\alpha^*}^{(2)} = \|P_0^* T\|_1 + W^{(2)}(T).$$

We show that all three norms  $\|T\|_{\Lambda_\alpha^*}$ ,  $\|T\|_{\Lambda_\alpha^*}^{(1)}$  and  $\|T\|_{\Lambda_\alpha^*}^{(2)}$  are equivalent on  $\Lambda_\alpha^*$ .

**Proposition 11** *The seminorms  $W^{(1)}(T)$  and  $W^{(2)}(T)$  are equivalent on  $\Lambda_\alpha^*$ .*

*Proof* We first show  $W^{(1)}(T) \leq (1 + 2^\alpha)W^{(2)}(T)$ .

$$\begin{aligned} W^{(1)}(T) &= \sum_{k=0}^{\infty} 2^{-k\alpha} \|\Delta_k^* T\|_1 = \sum_{k=0}^{\infty} 2^{-k\alpha} \|(P_k^* - P_{k+1}^*)T\|_1 \\ &\leq \sum_{k=0}^{\infty} 2^{-k\alpha} \|(P_k^* - P_0^*)T\|_1 + \sum_{k=0}^{\infty} 2^{-k\alpha} \|(P_{k+1}^* - P_0^*)T\|_1 \\ &\leq (1 + 2^\alpha)W^{(2)}(T). \end{aligned}$$

For the other direction, we write  $d_k^*$  as the telescopic sum

$$d_k^* T = P_k^* T - P_0^* T = \sum_{l=0}^{k-1} [(P_{l+1}^* T(x) - P_l^*)T] = \sum_{l=0}^{k-1} \Delta_l^* T.$$

Then  $\|d_k^* T\|_1 \leq \sum_{l=0}^{k-1} \|\Delta_l^* T\|_1$ , and consequently Fubini's theorem yields

$$\begin{aligned} W^{(2)}(T) &= \sum_{k=0}^{\infty} 2^{-k\alpha} \|d_k^* T\|_1 \leq \sum_{k=0}^{\infty} 2^{-k\alpha} \sum_{l=0}^{k-1} \|\Delta_l^* T\|_1 \\ &= \sum_{l=0}^{\infty} \|\Delta_l^* T\|_1 \sum_{k=l+1}^{\infty} 2^{-k\alpha} \\ &= \sum_{l=0}^{\infty} \|\Delta_l^* T\|_1 \frac{2^{-(l+1)\alpha}}{1 - 2^{-\alpha}} = \frac{2^{-\alpha}}{1 - 2^{-\alpha}} W^{(1)}(T) \end{aligned}$$

completing the proof.  $\square$

**Corollary 6** *The norms  $\|T\|_{\Lambda_\alpha^*}^{(1)}$  and  $\|T\|_{\Lambda_\alpha^*}^{(2)}$  are equivalent on  $\Lambda_\alpha^*$ .*

Next we turn to the main result of this section, namely that  $\|T\|_{\Lambda_\alpha^*}^{(1)}$  and  $\|T\|_{\Lambda_\alpha^*}^{(2)}$  are equivalent to  $\|T\|_{\Lambda_\alpha^*}$ .

**Proposition 12** *For all  $T \in \Lambda_\alpha^*$ ,  $\|T\|_{\Lambda_\alpha^*} \lesssim \|T\|_{\Lambda_\alpha^*}^{(2)}$ .*

*Proof* Suppose  $f$  is any function with  $\|f\|_{\Lambda_\alpha} \leq 1$ . Making use of Lemma 13 and the uniform convergence of  $P_k f$  to  $f$  as  $k \rightarrow \infty$  we can write

$$\begin{aligned} f - P_0 f &= \sum_{j=0}^{\infty} \Delta_j f = \sum_{j=1}^{\infty} P_j \Delta_j f + \sum_{j=1}^{\infty} P_{j+1} \Delta_j f \\ &= \sum_{j=1}^{\infty} (P_j - P_0) \Delta_j f + \sum_{j=1}^{\infty} (P_{j+1} - P_0) \Delta_j f + 2P_0(I - P_1)f. \end{aligned}$$

Therefore,

$$\begin{aligned} \langle f, T \rangle &= \sum_{j=1}^{\infty} \langle T, (P_j - P_0) \Delta_j f \rangle + \sum_{j=1}^{\infty} \langle T, (P_{j+1} - P_0) \Delta_j f \rangle \\ &\quad + \langle T, (3P_0 - 2P_0 P_1) f \rangle \\ &= \sum_{j=1}^{\infty} \langle (P_j^* - P_0^*) T, \Delta_j f \rangle + \sum_{j=1}^{\infty} \langle (P_{j+1}^* - P_0^*) T, \Delta_j f \rangle \\ &\quad + \langle P_0^* T, (3I - 2P_1) f \rangle. \end{aligned}$$

Consequently, we have

$$\begin{aligned} |\langle f, T \rangle| &\leq \sum_{j=1}^{\infty} \|(P_j^* - P_0^*) T\|_1 \sup_x |\Delta_j f(x)| + \sum_{j=1}^{\infty} \|(P_{j+1}^* - P_0^*) T\|_1 \sup_x |\Delta_j f(x)| \\ &\quad + \|P_0^* T\|_1 \sup_x |(3I - 2P_1) f(x)| \\ &\lesssim \sum_{j=1}^{\infty} 2^{-j\alpha} \|d_j^* T\|_1 + \sum_{j=1}^{\infty} 2^{-j\alpha} \|d_{j+1}^* T\|_1 + \|P_0^* T\|_1 \sup_x |f(x)| \end{aligned}$$

where in the last inequality we have used the equivalence of  $\|f\|_{\Lambda_\alpha}$  and  $\|f\|_{\Lambda_\alpha}^{(1)}$  from Theorem 3 and the fact that  $\sup_x |P_k f(x)| \lesssim \sup_x |f(x)|$  (a trivial consequence of condition (I) on the kernel). Since  $\sup_x |f(x)| \leq \|f\|_{\Lambda_\alpha} \leq 1$ , it follows immediately that

$$|\langle f, T \rangle| \lesssim \sum_{j=1}^{\infty} 2^{-j\alpha} \|d_j^* T\|_1 + \|P_0^* T\|_1 = \|T\|_{\Lambda_\alpha^*}^{(2)}.$$

Now take the supremum over all  $f$  with  $\|f\|_{\Lambda_\alpha} \leq 1$  to reach the desired conclusion.  $\square$

**Proposition 13** For all  $T \in \Lambda_\alpha^*$ ,  $\|T\|_{\Lambda_\alpha^*}^{(2)} \lesssim \|T\|_{\Lambda_\alpha^*}$ .

*Proof* Define the function  $f$  by

$$\begin{aligned} f(x) &= \sum_{k=1}^{\infty} 2^{-k\alpha} (P_k - P_0) \operatorname{sgn} [(P_k^* - P_0^*) T](x) + P_0 [\operatorname{sgn}(P_0^* T)](x) \\ &= \sum_{k=1}^{\infty} 2^{-k\alpha} P_k \operatorname{sgn} [(P_k^* - P_0^*) T](x) + P_0 F(x) \end{aligned}$$

where

$$F(x) = \operatorname{sgn}(P_0^* T)(x) - \sum_{k=1}^{\infty} 2^{-k\alpha} \operatorname{sgn} [(P_k^* - P_0^*) T](x).$$

Since

$$\sup_x |F(x)| \leq 1 + \sum_{k=1}^{\infty} 2^{-k\alpha} \leq 1 + \frac{2^{-\alpha}}{1+2^{-\alpha}}$$

therefore, by Lemma 14,

$$|P_0 F(x) - P_0 F(y)| \leq \left(1 + \frac{2^{-\alpha}}{1+2^{-\alpha}}\right) D_\alpha(x, y)$$

for all  $x, y \in \mathcal{X}$ . Furthermore, letting  $h_k = \operatorname{sgn}[(P_k^* - P_0^*) T]$ , Lemma 14 also implies that  $|P_k h_k(x) - P_k h_k(y)| \leq D_k(x, y)$ , and consequently

$$\begin{aligned} &\left| \sum_{k=1}^{\infty} 2^{-k\alpha} P_k \operatorname{sgn} [(P_k^* - P_0^*) T](x) - \sum_{k=1}^{\infty} 2^{-k\alpha} P_k \operatorname{sgn} [(P_k^* - P_0^*) T](y) \right| \\ &\leq \sum_{k=1}^{\infty} 2^{-k\alpha} |P_k h_k(x) - P_k h_k(y)| \\ &\leq \sum_{k=1}^{\infty} 2^{-k\alpha} D_k(x, y) \leq D_\alpha(x, y). \end{aligned}$$

We also have the estimate

$$\|f\|_\infty \lesssim \sum_{k=1}^{\infty} 2^{-k\alpha} + 1 \leq 1 + \frac{2^{-\alpha+1}}{1+2^{-\alpha}}.$$

It follows that  $\|f\|_{\Lambda_\alpha} \leq C(\alpha)$ , where  $C(\alpha)$  is a constant depending only on  $\alpha$  (in particular, not on  $T$  or  $f$ ). By the definition of  $f$ , we see

$$\begin{aligned}\langle f, T \rangle &= \sum_{k=1}^{\infty} 2^{-k\alpha} \langle (P_k - P_0) \operatorname{sgn}[(P_k^* - P_0^*)T], T \rangle + \langle P_0[\operatorname{sgn}(P_0^*T)], T \rangle \\ &= \sum_{k=1}^{\infty} 2^{-k\alpha} \langle \operatorname{sgn}[(P_k^* - P_0^*)T], (P_k^* - P_0^*)T \rangle + \langle \operatorname{sgn}(P_0^*T), P_0^*T \rangle \\ &= \sum_{k=1}^{\infty} 2^{-k\alpha} \|(P_k^* - P_0^*)T\|_1 + \|P_0^*T\|_1.\end{aligned}$$

Therefore,

$$\|T\|_{\Lambda_\alpha^*} \geq C(\alpha)^{-1} \langle f, T \rangle \simeq \sum_{k=1}^{\infty} 2^{-k\alpha} \|(P_k^* - P_0^*)T\|_1 + \|P_0^*T\|_1 = \|T\|_{\Lambda_\alpha^*}^{(2)}$$

which is the desired result.  $\square$

Putting together Corollary 6, Propositions 12 and 13 we have shown:

**Theorem 4** *The norms  $\|T\|_{\Lambda_\alpha^*}$ ,  $\|T\|_{\Lambda_\alpha^*}^{(1)}$  and  $\|T\|_{\Lambda_\alpha^*}^{(2)}$  are equivalent on  $\Lambda_\alpha^*$ .*

## 6 Application to Earth Mover's Distance

The dual norm  $\|T\|_{\Lambda_\alpha^*}$  has a natural interpretation when the distribution  $T$  is the difference of two probability measures  $P$  and  $Q$ . We will explain this in a more general setting. Suppose  $\Omega$  is any metric/measure space with metric  $\rho$ . A measure  $\pi$  on  $\Omega \times \Omega$  satisfies the equality-of-marginals condition with respect to  $P$  and  $Q$  if

$$\begin{aligned}\pi(\Omega, E) &= P(E) \\ \pi(E, \Omega) &= Q(E)\end{aligned}\tag{EM}$$

for all measurable sets  $E \subset \Omega$ . The Kantorovich–Rubinstein condition is the statement that:

$$\sup_{g: |g(x) - g(y)| \leq \rho(x, y)} \left\{ \int_{\Omega} gdP - \int_{\Omega} gdQ \right\} = \inf_{\pi: (EM) \text{ holds}} \int_{\Omega \times \Omega} \rho(x, y)d\pi(x, y).\tag{KR}$$

The following theorem gives conditions under which (KR) is true:

**Theorem 5** (Kantorovich–Rubinstein) *Suppose  $\Omega$  is a measure space that is separable with respect to the metric  $\rho$ . Let  $P$  and  $Q$  be two probability measures on  $\Omega$  such that the expected distance under  $P$  and  $Q$  from any point is finite. Then the equation (KR) is true.*

*Proof* The proof can be found in [12].  $\square$

The quantity on the right of (KR) is known as the Earth Mover's Distance between  $P$  and  $Q$ , denoted  $\text{EMD}(P, Q)$ . It has a physical interpretation. We view each measure  $\pi$  satisfying the equality-of-marginals condition (EM) with respect to  $P$  and  $Q$  as a transport between the measures  $Q$  and  $P$ ; that is, for any two measurable sets  $A, B \subset \Omega$ ,  $\pi(A, B)$  is interpreted as the amount of mass moved from set  $A$  to set  $B$ . The equality-of-marginals condition (EM) guarantees that the transport rearranges the mass distribution described by  $Q$  to end up with the distribution described by  $P$ . If  $\rho(x, y)$  is the cost-per-mass of moving mass from location  $x$  to location  $y$ , then  $\text{EMD}(P, Q)$  is the minimal cost over all transports; in other words, it is the cheapest way of rearranging mass distributed like  $Q$  to get mass distributed like  $P$ .

The quantity on the left of (KR) is equal to the norm of  $T = P - Q$  in the space dual to Lipschitz functions, except we do not require that the functions  $T$  is integrated against lie in  $L^\infty$ . However, when the diameter of the space is finite, as for the distances  $D_\alpha$  we have defined, and  $\int dP = \int dQ$ , then we can assume that all Lipschitz functions integrated against  $P - Q$  are uniformly bounded, and the two definitions are easily seen to be equivalent; that is, the norm  $\|P - Q\|_{\Lambda_\alpha^*}$  is equivalent in size to the left side of (KR).

Due to the way it exploits the geometry of the metric space on which the two probability distributions are defined, EMD has many desirable properties that make it a natural choice of metric for many problems in machine learning [22, 25, 26, 32]. We now describe one such property, which helps explain its robustness.

Suppose  $\Omega$  is a space with a metric  $\rho(x, y)$  and measure  $\mu$ . Suppose  $\Omega$  is separable with respect to  $\rho$ . Let  $p$  be a probability density on  $\Omega$  relative to  $\mu$ ; that is,  $p$  takes values in  $[0, \infty)$  and  $\int_\Omega p d\mu = 1$ . Let  $h : \Omega \rightarrow \Omega$  be a 1-1, absolutely continuous (with respect to  $\mu$ ) transformation satisfying

$$\rho(x, h(x)) \leq \epsilon \quad (5)$$

for all  $x \in \Omega$ . Let  $v$  be the measure induced by the change-of-variable  $h$ , that is,  $v(S) = \mu(h(S))$  for measurable subsets  $S \subset \Omega$ ; and let  $\frac{dv}{d\mu}$  denote the Radon–Nikodym derivative of  $v$  with respect to  $\mu$ . Then we define the probability density

$$q(x) = p(h(x)) \frac{dv}{d\mu}(x)$$

obtained from  $p$  by the change-of-variable  $h$ . We can think of  $q$  as a perturbation of  $p$ . The distance between  $p$  and  $q$  in some metrics may be quite large, even though the change-of-variable  $h$  is small. For example, if  $p$  is a narrow bump on  $\mathbb{R}$  of width less than  $\epsilon$ , and  $h$  is a shift of size  $\epsilon$ , then the supports of  $p$  and  $q$  will be disjoint and their  $L^1$  distance will be 2, the maximum possible. However, we now show that  $\text{EMD}(p, q)$  is no greater than the size of the perturbation  $h$  itself.

**Proposition 14** *Under the assumptions described above,  $\text{EMD}(p, q) \leq \epsilon$ .*

*Proof* We use the Kantorovich–Rubinstein Theorem (KR):

$$\text{EMD}(p, q) = \sup \left\{ \int_\Omega f(x)(p(x) - q(x))d\mu(x) : |f(x) - f(y)| \leq \rho(x, y) \right\}.$$

Take any  $f$  with  $|f(x) - f(y)| \leq \rho(x, y)$  for all  $x$  and  $y$ , and observe that

$$\begin{aligned}\int_{\Omega} f(x)q(x)d\mu(x) &= \int_{\Omega} f(x)p(h(x))\frac{d\nu}{d\mu}(x)d\mu(x) \\ &= \int_{\Omega} f(x)p(h(x))d\nu(x) \\ &= \int_{\Omega} f(h^{-1}(y))p(y)d\mu(y).\end{aligned}$$

Consequently,

$$\int_{\Omega} f(x)(p(x) - q(x))d\mu(x) = \int_{\Omega} p(x)(f(x) - f(h^{-1}(x)))d\mu(x).$$

By assumption (5) on  $h$ , we have  $\rho(x, h^{-1}(x)) \leq \epsilon$ ; hence, since  $f$  has Lipschitz constant 1, we have

$$|f(x) - f(h^{-1}(x))| \leq \rho(x, h^{-1}(x)) \leq \epsilon$$

and therefore

$$\begin{aligned}\int_{\Omega} f(x)(p(x) - q(x))d\mu(x) &= \int_{\Omega} p(x)(f(x) - f(h^{-1}(x)))d\mu(x) \\ &\leq \epsilon \int_{\Omega} p(x)d\mu(x) = \epsilon\end{aligned}$$

since  $p$  is a probability density. Taking the supremum over all Lipschitz  $f$  gives  $\text{EMD}(p, q) \leq \epsilon$ , as desired.  $\square$

In order to apply this theory to the setting of this paper, we need to check that condition (KR) holds when the space  $\mathcal{X}$  is given the metric  $D_\alpha(x, y)$ . Since  $D_\alpha(x, y)$  is uniformly bounded (so the expected distance from any point under a probability measure is automatically finite), to apply the Kantorovich–Rubinstein Theorem (Theorem 5) it remains to check that the resulting metric space is separable. We will prove separability by using our assumption that  $\mathcal{X}$  is sigma-finite.

**Lemma 15** *Under the metric  $D_\alpha(x, y)$ , balls in  $\mathcal{X}$  of positive radius have positive measure.*

*Proof* We deduce this from condition (G) as follows. Suppose that there were some ball  $B(x, r)$ ,  $r > 0$ , with measure zero. Then

$$1 = \int_{\mathcal{X}} p_k(x, y)dy \leq \int_{\mathcal{X}} |p_k(x, y)|dy = \int_{B(x, r)^c} |p_k(x, y)|dy$$

and consequently

$$r \leq \int_{B(x,r)^c} |p_k(x,y)| D_\alpha(x,y) dy \leq C 2^{-k\alpha}.$$

Since  $r > 0$ , taking  $k \rightarrow \infty$  yields a contradiction.  $\square$

**Proposition 15** *The metric  $D_\alpha(x, y)$  turns  $\mathcal{X}$  into a separable metric space; in particular, the Kantorovich–Rubinstein Theorem holds on  $\mathcal{X}$ .*

*Proof* Since we assume  $\mathcal{X}$  is sigma-finite, we can write  $\mathcal{X}$  as a countable union of finite measure sets. Without loss of generality, we can therefore assume that  $\mathcal{X}$  itself has finite measure. Take any positive integer  $n$ . Use Zorn's Lemma to find a maximal collection of points  $\{x_i^{(n)}\}_{i \in \mathcal{I}_n}$  so that  $D_\alpha(x_i^{(n)}, x_j^{(n)}) \geq 1/n$ , where  $\mathcal{I}_n$  is some index set. By maximality, every point in  $\mathcal{X}$  is within  $1/n$  of one of the points  $x_i^{(n)}$ .

We will show that  $\mathcal{I}_n$  is countable. Observe that the balls  $B(x_i^{(n)}, 1/2n)$  are pairwise disjoint and have positive measure. Since  $\mathcal{X}$  has finite measure, there can only be finitely many balls whose measure lies in the interval  $(2^{-k-1}, 2^{-k}]$ , for each  $k \in \mathbb{Z}$ . Since the measure of each ball must lie in one such interval, and there are countably many intervals, there are only countably many balls.

Consequently,  $\cup_{n=1}^{\infty} \{x_i^{(n)}\}_{i \in \mathcal{I}_n}$  is a countable dense subset of  $\mathcal{S}$ , and the proof is complete.  $\square$

In our setting of the space  $\mathcal{X}$  with the semigroup  $a_t(x, y)$ , the formulas for the norm  $\|T\|_{\Lambda_\alpha^*}$  from Sect. 5 provide an approximation to Earth Mover's Distance. From Theorem 4, and the Kantorovich–Rubinstein Theorem, the Earth Mover's Distance between two probability measures  $\mu$  and  $\nu$  is equivalent to the expressions

$$\|\mu - \nu\|_{\Lambda_\alpha^*}^{(1)} = \|P_0^*(\mu - \nu)\|_1 + \sum_{k \geq 0} 2^{-k\alpha} \|\Delta_k^*(\mu - \nu)\|_1 \quad (6)$$

and

$$\|\mu - \nu\|_{\Lambda_\alpha^*}^{(2)} = \|P_0^*(\mu - \nu)\|_1 + \sum_{k \geq 0} 2^{-k\alpha} \|d_k^*(\mu - \nu)\|_1. \quad (7)$$

In machine learning applications, these formulas can often be computed fast, and thus provide a fast approximation to Earth Mover's Distance. We only give a sketch of how this works, waving our hands regarding the issues that arise when using discrete data. We take  $\mathcal{X}$  to be a collection of  $n$  data points, and the operators  $P_k$  to be dyadic powers of a Markov matrix  $M$  on the data, as in the theory of diffusion maps [6].

If the mixing time of the walk is bounded by a power of  $n$ , then the series found in formulas (6) and (7) can be well approximated by the first  $O(\log n)$  terms. For many operators encountered in practice, all their dyadic powers can be applied in time  $O(n \log^k n)$ ; for instance, see [7, 8]. In this environment, the approximations to EMD given by (6) and (7) can be evaluated at cost  $O(n \log^k n)$  as well. Note that a

simplifying consideration in the case of the operators used for diffusion maps is that the Markov matrices  $M$  considered there are similar to a symmetric positive definite matrix (with maximum eigenvalue equal to 1), and so any algorithm that permits rapid application of all powers of such matrices will enable a fast approximation to EMD in our setting.

In more specialized cases similar formulas have been shown to approximate EMD as well. The work that most closely resembles this one is wavelet EMD [28]. Here, wavelets are used in place of the operators  $\Delta_k$  and  $d_k$ . The applicability of this method is limited to  $\mathbb{R}^n$ , where the ground distance is a snowflake of the Euclidean metric.

The reader can also refer to the papers by Charikar [4] and Indyk and Thaper [19]. Though the particulars are quite different than in the present work, the general spirit is the same; EMD can be approximated by a weighted sum of  $L^1$  norms of difference operators at different scales, whatever the notion of “scale” might mean for the geometry under consideration.

## 7 Mixed Hölder–Lipschitz Functions on Product Spaces

We now consider the setting where we have a product of spaces, each equipped with its own semigroup satisfying  $(S)$ ,  $(C)$ ,  $(I)$  and  $(R)$  so that the theory developed so far can be applied. For simplicity, we will consider only two spaces, which we will denote  $\mathcal{X}$  and  $\mathcal{Y}$ , each with a semigroup  $A_s$  and  $B_t$  with kernels  $a_s(x, x')$  and  $b_t(y, y')$ , respectively. All the results and their proofs can be extended to arbitrarily many semigroups. We define the dyadic discretizations for times between 0 and 1

$$P_k = A_{2^{-k}}, \quad p_k(x, x') = a_{2^{-k}}(x, x'), \quad k \geq 0$$

and

$$Q_l = B_{2^{-l}}, \quad q_l(y, y') = b_{2^{-l}}(y, y'), \quad l \geq 0$$

and the distances

$$D_{\mathcal{X},k}(x, x') = \|p_k(x, \cdot) - p_k(x', \cdot)\|_1$$

and

$$D_{\mathcal{Y},l}(y, y') = \|q_l(y, \cdot) - q_l(y', \cdot)\|_1.$$

For  $0 < \alpha, \beta < 1$ , such that the geometric condition  $(G)$  holds for  $P_k$  with respect to  $\alpha$  and  $(G)$  holds for  $Q_l$  with respect to  $\beta$ , we define metrics on  $\mathcal{X}$  and  $\mathcal{Y}$  by

$$D_{\mathcal{X},\alpha}(x, x') = \sum_{k \geq 0} 2^{-k\alpha} D_{\mathcal{X},k}(x, x')$$

and

$$D_{\mathcal{Y},\beta}(y, y') = \sum_{l \geq 0} 2^{-l\beta} D_{\mathcal{Y},l}(y, y').$$

For brevity, we will let  $D_{\mathcal{X}} = D_{\mathcal{X}, \alpha}$  and  $D_{\mathcal{Y}} = D_{\mathcal{Y}, \beta}$ .

We will define a regularity norm and its dual on the product space  $\mathcal{X} \times \mathcal{Y}$ . We first define the following quantities:

$$V_{\mathcal{X}}(f) = \sup_{y, x \neq x'} \frac{f(x, y) - f(x', y)}{D_{\mathcal{X}}(x, x')},$$

$$V_{\mathcal{Y}}(f) = \sup_{x, y \neq y'} \frac{f(x, y) - f(x, y')}{D_{\mathcal{Y}}(y, y')},$$

and

$$M(f) = \sup_{x \neq x', y \neq y'} \frac{f(x, y) - f(x, y') - f(x', y) + f(x', y')}{D_{\mathcal{X}}(x, x')D_{\mathcal{Y}}(y, y')}.$$

We then define the norm

$$\|f\|_{\Lambda_{\alpha, \beta}} \equiv M(f) + V_{\mathcal{X}}(f) + V_{\mathcal{Y}}(f) + \sup_x |f(x)|$$

and denote by  $\Lambda_{\alpha, \beta}$  the space of all functions  $f$  where  $\|f\|_{\Lambda_{\alpha, \beta}} < \infty$ . We will call  $\Lambda_{\alpha, \beta}$  the *mixed Hölder-Lipschitz space*, since the functions  $f \in \Lambda_{\alpha, \beta}$  must have bounded mixed difference quotients. The space  $\Lambda_{\alpha, \beta}$  is analogous to spaces of functions with dominating mixed derivatives on  $\mathbb{R}^n$ ; see [30].

**Lemma 16** *Taking*

$$\tilde{V}_{\mathcal{X}}(f) = \sup_{y, x \neq x'} \frac{(Q_0 f)(x, y) - (Q_0 f)(x', y)}{D_{\mathcal{X}}(x, x')},$$

$$\tilde{V}_{\mathcal{Y}}(f) = \sup_{x, y \neq y'} \frac{(P_0 f)(x, y) - (P_0 f)(x, y')}{D_{\mathcal{Y}}(y, y')}$$

in place of, respectively, the seminorms  $V_{\mathcal{X}}(f)$  and  $V_{\mathcal{Y}}(f)$  in the definition of  $\|f\|_{\Lambda_{\alpha, \beta}}$  yields an equivalent norm.

*Proof* From condition (I), it is immediate that  $\tilde{V}_{\mathcal{X}}(f) \lesssim V_{\mathcal{X}}(f)$  and  $\tilde{V}_{\mathcal{Y}}(f) \leq V_{\mathcal{Y}}(f)$ . For the other inequality, we can control  $V_{\mathcal{X}}(f)$  by  $\tilde{V}_{\mathcal{X}}(f)$  and  $M(f)$ , and control  $V_{\mathcal{Y}}(f)$  by  $\tilde{V}_{\mathcal{Y}}(f)$  and  $M(f)$ . To see this, observe that

$$\begin{aligned} & |(Q_0 f)(x, y) - (Q_0 f)(x', y) - f(x, y) + f(x', y)| \\ &= \left| \int_{\mathcal{X}} q_0(y, y')[f(x, y') - f(x', y') - f(x, y) + f(x', y)] dy' \right| \\ &\leq C D_{\mathcal{X}}(x, x') \operatorname{diam}(\mathcal{Y}) M(f) \\ &\lesssim M(f) D_{\mathcal{X}}(x, x') \end{aligned}$$

where we have used condition (I) in the second-to-last inequality. Consequently,  $V_{\mathcal{X}}(f) \lesssim \tilde{V}_{\mathcal{X}}(f) + M(f)$ ; similarly,  $V_{\mathcal{Y}}(f) \lesssim \tilde{V}_{\mathcal{Y}}(f) + M(f)$ . It follows that replacing  $V_{\mathcal{X}}(f)$  and  $V_{\mathcal{Y}}(f)$  by, respectively,  $\tilde{V}_{\mathcal{X}}(f)$  and  $\tilde{V}_{\mathcal{Y}}(f)$  in the definition of  $\|f\|_{\Lambda_{\alpha,\beta}}$  yields an equivalent norm.  $\square$

Of course, other minor variations in the definition of  $\|f\|_{\Lambda_{\alpha,\beta}}$  yielding equivalent norms are also possible. However, as in the case of a single space our primary goal is to give simpler characterizations of the norm  $\|f\|_{\Lambda_{\alpha,\beta}}$  involving the changes in the function's averages across scales. More precisely, the equivalent norms will measure the variations across all pairs of scales. In Sect. 8, we will use these to give simple characterizations of the norm on the space dual to  $\Lambda_{\alpha,\beta}$ .

We define the difference operators

$$\Delta_{P,k} = P_{k+1} - P_k, \quad \Delta_{Q,l} = Q_{l+1} - Q_l$$

as well as

$$\delta_{P,k} = I - P_k, \quad \delta_{Q,l} = I - Q_l.$$

We then define

$$V_{\mathcal{X}}^{(1)}(f) = \sup_{k \geq 0} \sup_{x,y} 2^{k\alpha} |\Delta_{P,k} f(x, y)|, \quad V_{\mathcal{Y}}^{(1)}(f) = \sup_{l \geq 0} \sup_{x,y} 2^{l\beta} |\Delta_{Q,l} f(x, y)|,$$

and

$$M^{(1)}(f) = \sup_{k \geq 0, l \geq 0} \sup_{x,y} 2^{k\alpha+l\beta} |\Delta_{P,k} \Delta_{Q,l} f(x, y)|.$$

Similarly, define

$$V_{\mathcal{X}}^{(2)}(f) = \sup_{k \geq 0} \sup_{x,y} 2^{k\alpha} |\delta_{P,k} f(x, y)|, \quad V_{\mathcal{Y}}^{(2)}(f) = \sup_{l \geq 0} \sup_{x,y} 2^{l\beta} |\delta_{Q,l} f(x, y)|,$$

and

$$M^{(2)}(f) = \sup_{k \geq 0, l \geq 0} \sup_{x,y} 2^{k\alpha+l\beta} |\delta_{P,k} \delta_{Q,l} f(x, y)|.$$

We can now define the equivalent regularity norms by

$$\|f\|_{\Lambda_{\alpha,\beta}}^{(1)} = M^{(1)}(f) + V_{\mathcal{X}}^{(1)}(f) + V_{\mathcal{Y}}^{(1)}(f) + \sup_{x,y} |f(x, y)|$$

and

$$\|f\|_{\Lambda_{\alpha,\beta}}^{(2)} = M^{(2)}(f) + V_{\mathcal{X}}^{(2)}(f) + V_{\mathcal{Y}}^{(2)}(f) + \sup_{x,y} |f(x, y)|.$$

We first show that  $\|f\|_{\Lambda_{\alpha,\beta}}$  controls  $\|f\|_{\Lambda_{\alpha,\beta}}^{(2)}$ . It will follow that on  $\Lambda_{\alpha,\beta}$  we have uniform convergence of the semigroups and their products to the identity.

**Proposition 16** *For any function  $f$ ,  $\|f\|_{\Lambda_{\alpha,\beta}}^{(2)} \lesssim \|f\|_{\Lambda_{\alpha,\beta}}$ .*

*Proof* Showing that  $V_{\mathcal{X}}^{(2)}(f)$  and  $V_{\mathcal{Y}}^{(2)}(f)$  are controlled by, respectively,  $V_{\mathcal{X}}(f)$  and  $V_{\mathcal{Y}}(f)$  is an immediate consequence of the one-dimensional result, Proposition 7. To show  $M^{(2)}(f) \lesssim M(f)$ , observe that Proposition 7 also gives

$$\begin{aligned} |2^{k\alpha} \delta_{P,k} 2^{l\beta} \delta_{Q,l} f(x, y)| &\lesssim \sup_{x \neq x'} \frac{2^{l\beta} \delta_{Q,l} f(x, y) - 2^{l\beta} \delta_{Q,l} f(x', y)}{D_{\mathcal{X}}(x, x')} \\ &= \sup_{x \neq x'} \frac{2^{l\beta} \delta_{Q,l}[f(x, \cdot) - f(x', \cdot)](y)}{D_{\mathcal{X}}(x, x')}. \end{aligned}$$

Now apply Proposition 7 again to the function  $y \mapsto f(x, y) - f(x', y)$  to obtain the bound

$$|2^{l\beta} \delta_{Q,l}[f(x, \cdot) - f(x', \cdot)](y)| \lesssim \sup_{y \neq y'} \frac{f(x, y) - f(x', y) - f(x, y') + f(x', y')}{D_{\mathcal{Y}}(y, y')}.$$

The result follows.  $\square$

It is easy to see that if  $\|f\|_{\Lambda_{\alpha,\beta}}^{(2)} < \infty$ , then

$$\lim_{k \rightarrow \infty, l \rightarrow \infty} P_k Q_l f = f$$

uniformly, where the limits can be taken in either order or simultaneously. Since  $\|f\|_{\Lambda_{\alpha,\beta}}^{(2)} \lesssim \|f\|_{\Lambda_{\alpha,\beta}}$ , the same convergence applies for any  $f \in \Lambda_{\alpha,\beta}$ .

We will next show that  $\|f\|_{\Lambda_{\alpha,\beta}}^{(1)}$  and  $\|f\|_{\Lambda_{\alpha,\beta}}^{(2)}$  are equivalent, and then that  $\|f\|_{\Lambda_{\alpha,\beta}} \lesssim \|f\|_{\Lambda_{\alpha,\beta}}^{(1)}$ . To that end:

**Lemma 17** *The seminorms  $V_{\mathcal{X}}^{(1)}(f)$  and  $V_{\mathcal{X}}^{(2)}(f)$  are equivalent on  $\Lambda_{\alpha,\beta}$ , as are the seminorms  $V_{\mathcal{Y}}^{(1)}(f)$  and  $V_{\mathcal{Y}}^{(2)}(f)$ .*

*Proof* This follows immediately from Proposition 8 for a single semigroup.  $\square$

**Lemma 18** *The seminorms  $M^{(1)}(f)$  and  $M^{(2)}(f)$  are equivalent on  $\Lambda_{\alpha,\beta}$ .*

*Proof* From Proposition 8, we have

$$\begin{aligned} |2^{k\alpha} \Delta_{P,k} 2^{l\beta} \Delta_{Q,l} f(x, y)| &\lesssim \sup_{x'} \sup_{k' \geq 0} 2^{k'\alpha} |\delta_{P,k'} 2^{l\beta} \Delta_{Q,l} f(x', y)| \\ &= \sup_{x'} \sup_{k' \geq 0} 2^{l\beta} |\Delta_{Q,l} 2^{k'\alpha} \delta_{P,k'} f(x', y)| \\ &\lesssim \sup_{x'} \sup_{k' \geq 0} \sup_{y'} \sup_{l'} 2^{l'\beta} |\delta_{Q,l} 2^{k'\alpha} \delta_{P,k'} f(x', y')| \end{aligned}$$

which proves  $M^{(1)}(f) \lesssim M^{(2)}(f)$ . The other direction is proved similarly.  $\square$

Combining Lemmas 17 and 18, we get:

**Proposition 17** *The norms  $\|f\|_{\Lambda_{\alpha,\beta}}^{(1)}$  and  $\|f\|_{\Lambda_{\alpha,\beta}}^{(2)}$  are equivalent on  $\Lambda_{\alpha,\beta}$ .*

To finish proving that all three norms are equivalent, we will show that  $\|f\|_{\Lambda_{\alpha,\beta}} \lesssim \|f\|_{\Lambda_{\alpha,\beta}}^{(1)}$ .

**Proposition 18** *For all  $f \in \Lambda_{\alpha,\beta}$ ,  $\|f\|_{\Lambda_{\alpha,\beta}} \lesssim \|f\|_{\Lambda_{\alpha,\beta}}^{(1)}$ .*

*Proof* First, it is trivial to deduce  $V_{\mathcal{X}}(f) \lesssim V_{\mathcal{X}}^{(1)}(f) + \sup_{x,y} |f(x, y)|$  and  $V_{\mathcal{Y}}(f) \lesssim V_{\mathcal{Y}}^{(1)}(f) + \sup_{x,y} |f(x, y)|$  from Proposition 9. Therefore, it remains to show  $M(f) \lesssim \|f\|_{\Lambda_{\alpha,\beta}}^{(1)}$ .

Fix any  $y, y' \in \mathcal{Y}$  and define

$$g(x) = \frac{f(x, y) - f(x, y')}{D_{\mathcal{Y}}(y, y')}.$$

From Proposition 9 again, we have that for all  $x \neq x'$ ,

$$\frac{f(x, y) - f(x, y') - f(x', y) + f(x', y')}{D_{\mathcal{X}}(x, x') D_{\mathcal{Y}}(y, y')} \lesssim \sup_{k \geq 0} \sup_{x''} 2^{k\alpha} |\Delta_{P,k} g(x'')| + \sup_{x''} |g(x'')|.$$

The supremum of  $g$  is bounded by

$$\sup_{x''} |g(x'')| \leq V_{\mathcal{Y}}(f) \lesssim V_{\mathcal{Y}}^{(1)}(f) + \sup_{x,y} |f(x, y)|.$$

Furthermore, we have

$$\begin{aligned} 2^{k\alpha} |\Delta_{P,k} g(x'')| &= 2^{k\alpha} \frac{|\Delta_{P,k} f(x'', y) - \Delta_{P,k} f(x'', y')|}{D_{\mathcal{Y}}(y, y')} \\ &\lesssim 2^{k\alpha} \sup_{l \geq 0} \sup_{y''} 2^{l\beta} |\Delta_{Q,l} \Delta_{P,k} f(x'', y'')| + 2^{k\alpha} \sup_{y''} |\Delta_{P,k} f(x'', y'')| \\ &\leq M^{(1)}(f) + V_{\mathcal{X}}^{(1)}(f). \end{aligned}$$

It follows that  $M(f) \lesssim M^{(1)}(f) + V_{\mathcal{Y}}^{(1)}(f) + V_{\mathcal{X}}^{(1)}(f) + \sup_{x,y} |f(x, y)| = \|f\|_{\Lambda_{\alpha,\beta}}^{(1)}$ , which completes the proof.  $\square$

Combining Propositions 16, 17, and 18, we have shown

**Theorem 6** *The norms  $\|f\|_{\Lambda_{\alpha,\beta}}$ ,  $\|f\|_{\Lambda_{\alpha,\beta}}^{(1)}$  and  $\|f\|_{\Lambda_{\alpha,\beta}}^{(2)}$  are equivalent on  $\Lambda_{\alpha,\beta}$ .*

## 7.1 Approximating Mixed Hölder–Lipschitz Functions

The reader will recall Proposition 7, which states that for a Hölder–Lipschitz function  $f$  on a single space  $\mathcal{X}$ ,

$$\sup_x |f(x) - P_L f(x)| \leq C V(f) 2^{-L\alpha}$$

for some constant  $C > 0$ . In other words, Hölder–Lipschitz functions  $f$  are well-approximated by their averages under the semigroup. We will derive a similar result for mixed Hölder–Lipschitz functions. For any positive integer  $L$ , define the operator  $\mathcal{P}_L$  by

$$\mathcal{P}_L f = \sum_{k,l:k+l \leq L} \Delta_{P,k} \Delta_{Q,l} f + \delta_{Q,L} P_0 f + \delta_{P,L} Q_0 f + P_0 Q_0 f.$$

Loosely speaking,  $\mathcal{P}_L f$  captures the activity of the function  $f$  at the pairs of scales  $(2^{-k}, 2^{-l})$  with  $2^{-(k+l)} \geq 2^{-L}$ ; in particular, the reciprocals  $(2^k, 2^l)$  lie in the hyperbolic cross  $\{(x, y) \in \mathbb{R}^2 : |xy| \leq 2^L\}$ . There is an extensive theory of hyperbolic cross approximations from classical analysis; for instance, see [30]. In our setting, we have the following result on uniformly approximating  $f$  by  $\mathcal{P}_L f$ :

**Proposition 19** *For any  $L \geq 0$ , we have*

$$\sup_{x,y} |\mathcal{P}_L f(x, y) - f(x, y)| \leq C \|f\|_{\Lambda_{\alpha,\beta}} \begin{cases} L 2^{-L\alpha}, & \text{if } \alpha = \beta \\ 2^{-L \min(\alpha, \beta)}, & \text{if } \alpha \neq \beta \end{cases}$$

where  $C > 0$  is a constant.

*Proof* We can write  $f$  as

$$\begin{aligned} f &= \sum_{k,l \geq 0} \Delta_{P,k} \Delta_{Q,l} f + \sum_{l \geq 0} \Delta_{Q,l} P_0 f + \sum_{k \geq 0} \Delta_{P,k} Q_0 f + P_0 Q_0 f \\ &= \sum_{k,l \geq 0} \Delta_{P,k} \Delta_{Q,l} f + \sum_{l \geq L} \Delta_{Q,l} P_0 f + \delta_{Q,L} P_0 f + \sum_{k \geq L} \Delta_{P,k} Q_0 f \\ &\quad + \delta_{P,L} Q_0 f + P_0 Q_0 f. \end{aligned}$$

Therefore,

$$f - \mathcal{P}_L f = \sum_{k,l:k+l>L} \Delta_{P,k} \Delta_{Q,l} f + \sum_{l \geq L} \Delta_{Q,l} P_0 f + \sum_{k \geq L} \Delta_{P,k} Q_0 f.$$

It is easy to see from condition (I) that  $V_{\mathcal{Y}}^{(1)}(P_0 f) \lesssim V_{\mathcal{Y}}^{(1)}(f)$  and  $V_{\mathcal{X}}^{(1)}(Q_0 f) \lesssim V_{\mathcal{X}}^{(1)}(f)$ . Using this, we have

$$\left| \sum_{l \geq L} \Delta_{Q,l} P_0 f \right| \lesssim \sum_{l \geq L} V_{\mathcal{Y}}^{(1)}(P_0 f) 2^{-l\beta} \lesssim \sum_{l \geq L} V_{\mathcal{Y}}^{(1)}(f) 2^{-l\beta} \lesssim V_{\mathcal{Y}}^{(1)}(f) 2^{-L\beta} \quad (8)$$

and similarly

$$\left| \sum_{l \geq L} \Delta_{Q,l} P_0 f \right| \lesssim V_{\mathcal{X}}^{(1)}(f) 2^{-L\alpha}. \quad (9)$$

Finally, we control the mixed difference term:

$$\begin{aligned} & \left| \sum_{k,l:k+l>L} \Delta_{P,k} \Delta_{Q,l} f \right| \\ & \leq \sum_{k=0}^L \sum_{l=L-k+1}^{\infty} |\Delta_{P,k} \Delta_{Q,l} f| + \sum_{k=L+1}^{\infty} \sum_{l=0}^{\infty} |\Delta_{P,k} \Delta_{Q,l} f| \\ & \leq M^{(1)}(f) \sum_{k=0}^L 2^{-k\alpha} \sum_{l=L-k+1}^{\infty} 2^{-l\beta} + M^{(1)}(f) \sum_{k=L+1}^{\infty} 2^{-k\alpha} \sum_{l=0}^{\infty} 2^{-l\beta} \\ & \leq \frac{M^{(1)}(f)}{1 - 2^{-\beta}} 2^{-(L+1)\beta} \sum_{k=0}^L 2^{-k(\alpha-\beta)} + \frac{M^{(1)}(f)}{1 - 2^{-\alpha}} \frac{1}{1 - 2^{-\alpha}} 2^{-(L+1)\alpha}. \end{aligned} \quad (10)$$

Now, if  $\alpha = \beta$ , then  $\sum_{k=0}^{L+1} 2^{-k(\alpha-\beta)} = L + 2$ , and (10) becomes

$$\left| \sum_{k,l:k+l>L} \Delta_{P,k} \Delta_{Q,l} f \right| \lesssim M^{(1)}(f) L 2^{-L\alpha}$$

which, combined with (8) and (9), gives the estimate  $\sup_{x,y} |\mathcal{P}_L f(x, y) - f(x, y)| \lesssim \|f\|_{\Lambda_{\alpha,\beta}}^{(1)} L 2^{-L\alpha}$ .

If  $\alpha < \beta$ , then  $\sum_{k=0}^{L+1} 2^{-k(\alpha-\beta)} \simeq 2^{L(\beta-\alpha)}$  so that (10) becomes

$$\left| \sum_{k,l:k+l>L} \Delta_{P,k} \Delta_{Q,l} f \right| \lesssim M^{(1)}(f) (2^{-L\beta} 2^{L(\beta-\alpha)} + 2^{-L\alpha}) \lesssim M^{(1)}(f) 2^{-L\alpha}$$

and combining this with (8) and (9) gives the estimate  $\sup_{x,y} |\mathcal{P}_L f(x, y) - f(x, y)| \lesssim \|f\|_{\Lambda_{\alpha,\beta}}^{(1)} 2^{-L\alpha}$ .

Finally, if  $\alpha > \beta$ , then  $\sum_{k=0}^{L+1} 2^{-k(\alpha-\beta)} \simeq 1$ , and so from (10) we get:

$$\left| \sum_{k,l:k+l>L} \Delta_{P,k} \Delta_{Q,l} f \right| \lesssim M^{(1)}(f) (2^{-L\beta} + 2^{-L\alpha}) \lesssim M^{(1)}(f) 2^{-L\beta}$$

from which the estimate  $\sup_{x,y} |\mathcal{P}_L f(x, y) - f(x, y)| \lesssim \|f\|_{\Lambda_{\alpha,\beta}}^{(1)} 2^{-L\beta}$  also follows.

Since  $\|f\|_{\Lambda_{\alpha,\beta}}^{(1)} \lesssim \|f\|_{\Lambda_{\alpha,\beta}}$  by Theorem 6, we are done.  $\square$

## 8 The Space Dual to $\Lambda_{\alpha,\beta}$

We now consider the space  $\Lambda_{\alpha,\beta}^*$  of  $L^1$  measures dual to the space  $\Lambda_{\alpha,\beta}$  of mixed Hölder–Lipschitz functions. We will derive two simpler norms that are equivalent to the canonical dual norm on  $\Lambda_{\alpha,\beta}^*$ , as we did for the case of a single semigroup in Sect. 5.

The dual norm of  $T \in \Lambda_{\alpha,\beta}^*$  is given by

$$\|T\|_{\Lambda_{\alpha,\beta}^*} = \sup_{\|f\|_{\Lambda_{\alpha,\beta}} \leq 1} \langle f, T \rangle.$$

Before defining the equivalent norms, we introduce some notation. Define

$$d_{P,k} = P_k - P_0, \quad d_{Q,l} = Q_k - Q_0$$

and

$$W_{\mathcal{X}}^{(1)}(T) = \sum_{k \geq 0} 2^{-k\alpha} \|\Delta_{P,k}^* Q_0^* f\|_1, \quad W_{\mathcal{Y}}^{(1)}(T) = \sum_{l \geq 0} 2^{-l\beta} \|\Delta_{Q,l}^* P_0^* f\|_1$$

and

$$W_{\mathcal{X}}^{(2)}(T) = \sum_{k \geq 0} 2^{-k\alpha} \|d_{P,k}^* Q_0^* T\|_1, \quad W_{\mathcal{Y}}^{(2)}(T) = \sum_{l \geq 0} 2^{-l\beta} \|d_{Q,l}^* P_0^* T\|_1$$

as well as

$$\begin{aligned} N^{(1)}(T) &= \sum_{k \geq 0, l \geq 0} 2^{-k\alpha} 2^{-l\beta} \|\Delta_{P,k}^* \Delta_{Q,l}^* T\|_1, \quad N^{(2)}(T) \\ &= \sum_{k \geq 0, l \geq 0} 2^{-k\alpha} 2^{-l\beta} \|d_{P,k}^* d_{Q,l}^* T\|_1. \end{aligned}$$

With these definitions, we define the two norms we will show are equivalent to  $\|T\|_{\Lambda_{\alpha,\beta}^*}$ . The first norm is defined by

$$\|T\|_{\Lambda_{\alpha,\beta}^*}^{(1)} = N^{(1)}(T) + W_{\mathcal{X}}^{(1)}(T) + W_{\mathcal{Y}}^{(1)}(T) + \|P_0^* Q_0^* T\|_1$$

and the second is defined by

$$\|T\|_{\Lambda_{\alpha,\beta}^*}^{(2)} = N^{(2)}(T) + W_{\mathcal{X}}^{(2)}(T) + W_{\mathcal{Y}}^{(2)}(T) + \|P_0^* Q_0^* T\|_1.$$

**Lemma 19** *The seminorms  $W_{\mathcal{X}}^{(1)}(T)$  and  $W_{\mathcal{X}}^{(2)}(T)$  are equivalent on  $\Lambda_{\alpha,\beta}^*$ , as are the seminorms  $W_{\mathcal{Y}}^{(1)}(T)$  and  $W_{\mathcal{Y}}^{(2)}(T)$ .*

*Proof* This follows immediately from Proposition 11. □

**Lemma 20** *The seminorms  $N^{(1)}(T)$  and  $N^{(2)}(T)$  are equivalent on  $\Lambda_{\alpha,\beta}^*$ .*

*Proof* We reduce the proof to the case of a single semigroup by applying Proposition 11 repeatedly. We have

$$\begin{aligned} N^{(1)}(T) &= \sum_{l \geq 0} 2^{-l\beta} \sum_{k \geq 0} 2^{-k\alpha} \|\Delta_{P,k}^* \Delta_{Q,l}^* T\|_1 \\ &= \int_{\mathcal{Y}} \sum_{l \geq 0} 2^{-l\beta} \sum_{k \geq 0} 2^{-k\alpha} \|\Delta_{P,k}^* \Delta_{Q,l}^* T(\cdot, y)\|_{L^1(X)} dy \\ &\simeq \int_{\mathcal{Y}} \sum_{l \geq 0} 2^{-l\beta} \sum_{k \geq 0} 2^{-k\alpha} \|d_{P,k}^* \Delta_{Q,l}^* T(\cdot, y)\|_{L^1(X)} dy \\ &= \int_{\mathcal{X}} \sum_{k \geq 0} 2^{-k\alpha} \sum_{l \geq 0} 2^{-l\beta} \|\Delta_{Q,l}^* d_{P,k}^* T(x, \cdot)\|_{L^1(Y)} dx \\ &\simeq \int_{\mathcal{X}} \sum_{k \geq 0} 2^{-k\alpha} \sum_{l \geq 0} 2^{-l\beta} \|d_{Q,l}^* d_{P,k}^* T(x, \cdot)\|_{L^1(X)} dx \\ &= \sum_{k \geq 0} 2^{-k\alpha} \sum_{l \geq 0} 2^{-l\beta} \|d_{Q,l}^* d_{P,k}^* T\|_1 = N^{(2)}(T). \end{aligned}$$

□

**Proposition 20** *The norms  $\|T\|_{\Lambda_{\alpha,\beta}^*}^{(1)}$  and  $\|T\|_{\Lambda_{\alpha,\beta}^*}^{(2)}$  are equivalent on  $\Lambda_{\alpha,\beta}^*$ .*

*Proof* This follows immediately from the preceding two lemmas. □

We will now prove that  $\|T\|_{\Lambda_{\alpha,\beta}^*}$  and  $\|T\|_{\Lambda_{\alpha,\beta}^*}^{(2)}$  are equivalent on  $\Lambda_{\alpha,\beta}^*$ . We will work formally; all manipulations can be easily justified by the fact that  $P_k Q_l f$  converges uniformly to  $f$  as  $k, l \rightarrow \infty$ , whenever  $f \in \Lambda_{\alpha,\beta}$ . Take any function  $f$  with  $\|f\|_{\Lambda_{\alpha,\beta}} \leq 1$ . Write

$$f = \sum_{k \geq 0, l \geq 0} \Delta_{P,k} \Delta_{Q,l} f + \sum_{k \geq 0} \Delta_{P,k} Q_0 f + \sum_{l \geq 0} \Delta_{Q,l} P_0 f + P_0 Q_0 f. \quad (11)$$

We want to show that  $|\langle f, T \rangle| \lesssim \|T\|_{\Lambda_{\alpha,\beta}^*}^{(2)}$ . We will deal with the inner product of  $f$  with each of the four terms on the right side of (11) separately. First, we have

$$|\langle P_0 Q_0 f, T \rangle| = |\langle f, P_0^* Q_0^* T \rangle| \leq \|P_0^* Q_0^* T\|_1 \quad (12)$$

since  $\|f\|_\infty \leq 1$ .

To control the inner product of  $T$  with  $\sum_{k \geq 0} \Delta_{P,k} Q_0 f$ , first observe that, using Lemma 13, we get

$$\begin{aligned} \Delta_{P,k} Q_0 &= \Delta_{P,k+1} (P_{k+2} + P_{k+1}) Q_0 \\ &= \Delta_{P,k+1} P_{k+2} Q_0 + \Delta_{P,k+1} P_{k+1} Q_0. \end{aligned} \quad (13)$$

Now, we have the identity

$$\begin{aligned}\Delta_{P,k+1} P_{k+1} Q_0 &= \Delta_{P,k+1}(P_{k+1} - P_0)Q_0 + \Delta_{P,k+1} P_0 Q_0 \\ &= \Delta_{P,k+1} d_{P,k+1} Q_0 + \Delta_{P,k+1} P_0 Q_0,\end{aligned}$$

from which it follows:

$$\begin{aligned}|\langle \Delta_{P,k+1} P_{k+1} Q_0 f, T \rangle| &\leq |\langle \Delta_{P,k+1} d_{P,k+1} Q_0 f, T \rangle| + |\langle \Delta_{P,k+1} P_0 Q_0 f, T \rangle| \\ &= |\langle \Delta_{P,k+1} f, d_{P,k+1}^* Q_0^* T \rangle| + |\langle \Delta_{P,k+1} f, P_0^* Q_0^* T \rangle| \\ &\leq \sup_{x,y} |\Delta_{P,k+1} f(x, y)| \left\{ \|d_{P,k+1}^* Q_0^* f\|_1 + \|P_0^* Q_0^* T\|_1 \right\} \\ &\lesssim 2^{-k\alpha} \|d_{P,k+1}^* Q_0^* T\|_1 + 2^{-k\alpha} \|P_0^* Q_0^* T\|_1.\end{aligned}\quad (14)$$

Similarly,

$$|\langle \Delta_{P,k+1} P_{k+2} Q_0 f, T \rangle| \leq 2^{-k\alpha} \|d_{P,k+2}^* Q_0^* T\|_1 + 2^{-k\alpha} \|P_0^* Q_0^* T\|_1. \quad (15)$$

Combining equation (13) with the estimates (14) and (15) yields

$$|\langle \Delta_{P,k} Q_0 f, T \rangle| \lesssim 2^{-k\alpha} \left\{ \|d_{P,k+1}^* Q_0^* T\|_1 + \|d_{P,k+2}^* Q_0^* T\|_1 + 2 \|P_0^* Q_0^* T\|_1 \right\}$$

and summing over  $k \geq 0$  then yields

$$\left| \left\langle \sum_{k \geq 0} \Delta_{P,k} Q_0 f, T \right\rangle \right| \lesssim W_{\mathcal{X}}^{(2)}(T) + \|P_0^* Q_0^* T\|_1. \quad (16)$$

A nearly identical proof shows that

$$\left| \left\langle \sum_{l \geq 0} \Delta_{Q,l} P_0 f, T \right\rangle \right| \lesssim W_{\mathcal{Y}}^{(2)}(T) + \|P_0^* Q_0^* T\|_1. \quad (17)$$

The only term left to control from (11) is the inner product of  $T$  with

$$\sum_{k \geq 0, l \geq 0} \Delta_{P,k} \Delta_{Q,l} f.$$

Using the identity

$$\begin{aligned}\Delta_{P,k-1} \Delta_{Q,l-1} &= \Delta_{P,k} (P_{k+1} + P_k) \Delta_{Q,l} (Q_{l+1} + Q_l) \\ &= \Delta_{P,k} P_{k+1} \Delta_{Q,l} Q_{l+1} + \Delta_{P,k} P_k \Delta_{Q,l} Q_{l+1} \\ &\quad + \Delta_{P,k} P_{k+1} \Delta_{Q,l} Q_l + \Delta_{P,k} P_k \Delta_{Q,l} Q_l\end{aligned}\quad (18)$$

it follows that we must control the inner product of  $T$  with each of the four terms on the right side of (18) (applied to  $f$ ). The argument is the same for each, so we will show it only for  $\Delta_{P,k} P_k \Delta_{Q,l} Q_l f = \Delta_{P,k} \Delta_{Q,l} P_k Q_l f$ .

From the easily-verified identity

$$P_k Q_l f = d_{P,k} d_{Q,l} f + d_{P,k} Q_0 f + d_{Q,l} P_0 f + P_0 Q_0 f$$

it follows that

$$\begin{aligned} \Delta_{P,k} \Delta_{Q,l} P_k Q_l f &= \Delta_{P,k} \Delta_{Q,l} d_{P,k} d_{Q,l} f + \Delta_{P,k} \Delta_{Q,l} d_{P,k} Q_0 f \\ &\quad + \Delta_{P,k} \Delta_{Q,l} d_{Q,l} P_0 f + \Delta_{P,k} \Delta_{Q,l} P_0 Q_0 f. \end{aligned} \quad (19)$$

We will bound the inner product of  $T$  with the sum over  $k \geq 0$  and  $l \geq 0$  of each of the four terms in (19) separately. First, we have

$$|\langle \Delta_{P,k} \Delta_{Q,l} P_0 Q_0 f, T \rangle| = |\langle \Delta_{P,k} \Delta_{Q,l} f, P_0^* Q_0^* T \rangle| \leq 2^{-k\alpha} 2^{-l\beta} \|P_0^* Q_0^* T\|_1$$

and summing over  $k$  and  $l$  gives the upper bound  $\|P_0^* Q_0^* T\|_1$ .

Next, observe that

$$|\langle \Delta_{P,k} \Delta_{Q,l} d_{Q,l} P_0 f, T \rangle| = |\langle \Delta_{P,k} \Delta_{Q,l} f, d_{Q,l}^* P_0^* T \rangle| \leq 2^{-k\alpha} 2^{-l\beta} \|d_{Q,l}^* P_0^* T\|_1$$

and summing over  $k$  and  $l$  gives the upper bound  $\sum_{l=0}^{\infty} 2^{-l\beta} \|d_{Q,l}^* P_0^* T\|_1 = W_{\mathcal{Y}}^{(2)}(T)$ . Similarly, the inner product of  $T$  with  $\Delta_{P,k} \Delta_{Q,l} d_{P,k} Q_0 f$  can be bounded above by  $W_{\mathcal{X}}^{(2)}(T)$ .

Finally, we have the upper bound

$$|\langle \Delta_{P,k} \Delta_{Q,l} d_{P,k} d_{Q,l} f, T \rangle| = |\langle \Delta_{P,k} \Delta_{Q,l} f, d_{P,k}^* d_{Q,l}^* T \rangle| \leq 2^{-k\alpha} 2^{-l\beta} \|d_{P,k}^* d_{Q,l}^* T\|_1$$

and summing over  $k$  and  $l$  gives the upper bound  $\sum_{k,l} 2^{-k\alpha} 2^{-l\beta} \|d_{P,k}^* d_{Q,l}^* T\|_1 = N^{(2)}(T)$ . Putting the four bounds together and applying Eq. (19) yields

$$\left| \left\langle \sum_{k,l} \Delta_{P,k} \Delta_{Q,l} P_k Q_l f, T \right\rangle \right| \lesssim \|T\|_{\Lambda_{\alpha,\beta}^*}^{(2)}$$

and the same estimate applied to each of the four terms on the right side of Eq. (18) gives

$$\left| \left\langle \sum_{k \geq 0, l \geq 0} \Delta_{P,k} \Delta_{Q,l} f, T \right\rangle \right| \lesssim \|T\|_{\Lambda_{\alpha,\beta}^*}^{(2)}. \quad (20)$$

Combining the estimates (12), (16), (17) and (20) with (11), we complete the proof that  $\|T\|_{\Lambda_{\alpha,\beta}^*} \lesssim \|T\|_{\Lambda_{\alpha,\beta}^*}^{(2)}$ .

To prove the reverse inequality, as in the proof of Proposition 13 for a single semigroup we define a function  $f$  such that  $\|f\|_{\Lambda_{\alpha,\beta}} \simeq 1$  and whose inner product with  $T$  achieves the norm  $\|T\|_{\Lambda_{\alpha,\beta}^*}^{(2)}$ . It is easy to check that the function  $f$  defined by

$$\begin{aligned} f = & \sum_{k,l \geq 0} 2^{-k\alpha} 2^{-l\beta} d_{P,l} d_{Q,l} \operatorname{sgn}(d_{P,k}^* d_{Q,l}^* T) + \sum_{k \geq 0} 2^{-k\alpha} d_{P,k} Q_0 \operatorname{sgn}(d_{P,k}^* Q_0^* T) \\ & + \sum_{l \geq 0} 2^{-l\beta} d_{Q,l} P_0 \operatorname{sgn}(d_{Q,l}^* P_0^* T) + P_0 Q_0 \operatorname{sgn}(P_0^* Q_0^* T). \end{aligned}$$

satisfies the necessary conditions; in fact, each of the four terms defining  $f$  have mixed Hölder–Lipschitz norm bounded independently of  $T$ , and  $\langle f, T \rangle = \|T\|_{\Lambda_{\alpha,\beta}^*}^{(2)}$ . We have therefore shown:

**Theorem 7** *The norms  $\|T\|_{\Lambda_{\alpha,\beta}^*}$ ,  $\|T\|_{\Lambda_{\alpha,\beta}^*}^{(1)}$ , and  $\|T\|_{\Lambda_{\alpha,\beta}^*}^{(2)}$  are equivalent on  $\Lambda_{\alpha,\beta}^*$ .*

**Acknowledgments** The authors thank Maxim Goldberg for many insightful and stimulating discussions throughout the development of this work. We also thank the anonymous reviewers for their helpful comments. Ronald Coifman was supported by NSF Grant No. 1309858.

## References

1. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15**(6), 1373–1396 (2003)
2. Bui, H.Q., Duong, X.T., Yan, L.: Calderón reproducing formulas and new Besov spaces associated with operators. *Adv. Math.* **229**(4), 2449–2502 (2012)
3. Butzer, P.L., Berens, H.: Semi-groups of Operators and Approximation. Springer, Berlin (1967)
4. Charikar, M.S.: Similarity estimation techniques from rounding algorithms. In: Proceedings of the 34th Annual ACM Symposium on Theory of Computing, pp. 380–388 (2002)
5. Chavel, I.: Eigenvalues in Riemannian Geometry. Academic Press, Orlando (1984)
6. Coifman, R.R., Lafon, S.: Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**(1), 5–30 (2006)
7. Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Nadler, B., Warner, F., Zucker, S.W.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc. Natl. Acad. Sci. USA* **102**(21), 7426–7431 (2005)
8. Coifman, R.R., Maggioni, M.: Diffusion wavelets. *Appl. Comput. Harmon. Anal.* **21**(1), 53–94 (2006)
9. Dachkovski, S.: Anisotropic function spaces and related semi-linear hypoelliptic equations. *Math. Nachr.* **248**(1), 40–61 (2003)
10. Do Carmo, M.P.: Riemannian Geometry. Birkhäuser, Boston (1992)
11. Donoho, D.L., Johnstone, I.M.: Neo-classical minimax problems, thresholding and adaptive function estimation. *Bernoulli* **2**(1), 39–62 (1996)
12. Dudley, R.M.: Real Analysis and Probability. Cambridge University Press, Cambridge (2002)
13. Farkas, W.: Atomic and subatomic decompositions in anisotropic function spaces. *Math. Nachr.* **209**, 83–113 (2000)
14. Folland, G.B.: Real Analysis: Modern Techniques and Their Applications, 2nd edn. Wiley, New York (1999)
15. Goldberg, M.J., Kim, S.: Some remarks on diffusion distances. *J. Appl. Math.* **2010**, 464815 (2010)
16. Goldberg, M.J., Kim, S.: An efficient tree-based computation of a metric comparable to a natural diffusion distance. *Appl. Comput. Harmon. Anal.* **33**(2), 261–281 (2012)
17. Grigor'yan, A., Liu, L.: Heat kernel and Lipschitz-Besov spaces. *Forum Math.* (2014)
18. Heinonen, J.: Lectures on Analysis on Metric Spaces. Springer, New York (2001)
19. Indyk, P., Thaper, N.: Fast image retrieval via embeddings. In: 3rd International Workshop on Statistical and Computational Theories of Vision (at ICCV) (2003)

20. Li, P., Yau, S.T.: On the parabolic kernel of the Schrödinger operator. *Acta Math.* **156**(1), 153–201 (1986)
21. Little, A.V., Jung, Y.M., Maggioni, M.: Multiscale estimation of intrinsic dimensionality of data sets. In: AAAI fall symposium: manifold learning and its applications (2009)
22. Marinai, S., Miotti, B., Soda, G.: Using earth mover's distance in the bag-of- visual-words model for mathematical symbol retrieval. In: 2011 International Conference on Document Analysis and Recognition, pp. 1309–1313 (2011)
23. Meyer, Y.: Wavelets and Operators. Cambridge University Press, Cambridge (1992)
24. Neumann, M.H.: Multivariate wavelet thresholding in anisotropic function spaces. *Stat. Sin.* **10**(2), 399–432 (2000)
25. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.* **40**(2), 99–121 (2000)
26. Sandler, R., Lindenbaum, M.: Nonnegative matrix factorization with earth mover's distance metric for image analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8), 1590–1602 (2011)
27. Schmeisser, H., Triebel, H.: Topics in Fourier Analysis and Function Spaces. Wiley, Chichester (1987)
28. Shirdhonkar, S., Jacobs, D.W.: Approximate earth mover's distance in linear time. In: IEEE conference on computer vision and pattern recognition, pp. 1–8 (2008)
29. Stein, E.M.: Topics in Harmonic Analysis, Related to the Littlewood-Paley Theory. Princeton University Press, Princeton (1970)
30. Temlyakov, V.N.: Approximation of Periodic Functions. Computational Mathematics and Analysis Series. Nova Science Publishers Inc, New York (1993)
31. Triebel, H.: Theory of Function Spaces. Birkhäuser, Basel (1983)
32. Wan, X.: A novel document similarity measure based on earth mover's distance. *Inf. Sci.* **177**(18), 3718–3730 (2007)
33. Zolotarev, V.M.: One-Dimensional Stable Distributions. American Mathematical Society, Providence (1986)