We introduce a multi-scale metric on a space equipped with a diffusion semigroup. We prove, under some technical conditions, that the norm dual to the space of Lipschitz functions with respect to this metric is equivalent to two other norms, one of which is a weighted sum of the averages at each scale, and one of which is a weighted sum of the difference of averages across scales. The notion of 'scale' is defined by the semigroup. For both equivalent norms, bigger scales have greater contribution. When the function is a difference of two probability distributions, the dual norm is equal to the Earth Mover's Distance with ground distance equal to the multi-scale metric induced by the semigroup.

Earth Mover's Distance and Equivalent Metrics
for Spaces with Semigroups

R.R. Coifman† and W.E. Leeb†
Technical Report YALEU/DCS/TR-1481
July 22, 2013

† Dept. of Mathematics, Yale University, New Haven CT 06511

Approved for public release: distribution is unlimited.
Keywords: *earth mover's distance, lipschitz class, operator semigroup.*

1 Introduction

Given a measure space Ω, we consider a family of operators defined on a suitable space of functions on Ω satisfying the following properties:

1. A_0 is the identity
2. For all $t, s \geq 0$, $A_t A_s = A_{t+s}$ (the semigroup property)
3. Each A_t is self-adjoint
4. If $\mathbf{1}$ is the constant function 1 on Ω, then for all $t \geq 0$, $A_t \mathbf{1} = \mathbf{1}$
5. Each A_t is positive, that is, if $f \geq 0$, then $A_t f \geq 0$
6. $A_\infty f = \int_\Omega f(x)dx$.

We will assume in addition that A_t is given by a (symmetric) kernel $a_t(x, y) = a_t(y, x)$; that is, for functions f on Ω in the domain of A_t,

$$A_t f(x) = \int_\Omega a_t(x, y) f(y) dy$$

where dy is the measure on Ω.

Operator semigroups are ubiquitous in pure mathematics; see, for instance, [1]. In machine learning, the theory of *diffusion maps* [2] builds operators of this kind on arbitrary data sets equipped with a kernel $k(x, y)$ measuring the similarity of two points x, y. The semigroup is described by powers of an affinity matrix built by normalizing the matrix given by $k(x, y)$. We note that although in the original construction there is not in general a single matrix that is both stochastic and symmetric, one can easily define a measure on the data with respect to which the symmetric operator defined in [2] is also stochastic; see [3].

The *diffusion distance* at time t between $x, y \in \Omega$ can be defined as the L^2 distance between the probability distributions $a_t(x, \cdot)$ and $a_t(y, \cdot)$. As shown in [2], the eigenvectors of the operator A_t, weighted by their eigenvalues, give coordinates on Ω whose Euclidean distance is equal to $||a_t(x, \cdot) - a_t(y, \cdot)||_2$, the time t diffusion distance. The time t diffusion distance measures the connectivity of points at the given scale t.

The metric on Ω we consider in this paper replaces the L^2 diffusion distance with L^1 diffusion distance, and sums these over dyadic t, with weights depending on t. This new distance incorporates all scales at once, giving higher weight to the larger scales. We will define this distance in Section 2, and give alternate characterizations of functions that are Lipschitz with respect to this distance in Section 3. In Section we will use these characterizations to prove that the norm of a function viewed as a linear functional on the space of Lipschitz functions is equivalent to two norms, one of which is a weighted sum of the averages (by A_t) at dyadic scales, and one of which is a weighted sum of the difference of averages across scales.

2 Multi-scale Diffusion Metric

Suppose A_t is a symmetric diffusion semigroup on Ω with kernel $a_t(x,y) = a_t(y,x)$. We will consider discrete times $t = 2^{-k}$, and define

$$\begin{aligned} P_k &= A_{2^{-k}} \\ p_k(x,y) &= a_{2^{-k}}(x,y). \end{aligned}$$

We define a metric on Ω. First, for each k we can consider the L^1 diffusion distance at time $t = 2^{-k}$, defined by

$$d_k(x,y) = ||p_k(x,\cdot) - p_k(y,\cdot)||_1 = \int_\Omega |p_k(x,u) - p_k(y,u)|du.$$

When it makes sense, we consider the weighted sum of diffusion distances at all scales $t = 2^{-k}$ by

$$d_\alpha(x,y) = \sum_{k=-\infty}^{\infty} d_k(x,y) 2^{-k\alpha}. \tag{1}$$

We note that a sufficient condition for this series to be finite is if there is a K for which all the operators P_k, $k \le K$ are all equal to $P_{-\infty}$, namely $P_k f = \int f$; if we view the operators A_t as describing a diffusion process over Ω, we can think of the time $T = 2^{-K}$ as the diffusion time of the process, that is, the time it takes for any initial distribution of mass to become spread out uniformly over Ω. Since $p_k \equiv 1$ for all $k \le K$, $d_k(x,y) = 0$, and so

$$d_\alpha(x,y) = \sum_{k>K} d_k(x,y) 2^{-k\alpha} \le 2 \sum_{k>K} 2^{-k\alpha} < \infty$$

where we have used that $d_k(x,y) \le 2$ for all k, and the finiteness of geometric series. In fact we only need to assume that P_k converges sufficiently fast to the operator $f \mapsto \int f$.

We can gain some intuition as to the behavior of this distance by considering the special case of convolutions in $\mathbb{R}^n$. It will be easier for these computations to view (1) as a Riemann sum approximation to the following continuous version:

$$\tilde{d}_\alpha(x,y) = \int_0^\infty ||a_t(x,\cdot) - a_t(y,\cdot)||_1 t^{\alpha-1} dt.$$

Suppose $\Omega = \mathbb{R}^n$ and the kernel $a_t(x,y) = a_t(x-y)$ (abusing notation) for a function a_t of one variable defined by

$$a_t(x) = t^{-n\beta} a_1(t^{-\beta} x)$$

for some constant β and some radial function a_1. For example, if a_t is the heat kernel in $\mathbb{R}^n$ defined by

$$a_t(x) = c_n t^{-n/2} \exp\left(-\frac{||x||_2^2}{4t}\right),$$

then $\beta = 1/2$; and if a_t is the Poisson kernel

$$a_t(x) = c_n \frac{t}{(t^2 + ||x||_2^2)^{(n+1)/2}}$$

then $\beta = 1$.

For such kernels with sufficiently fast decay, we can show $\tilde{d}_\alpha(x, y)$ is equivalent to $||x - y||_2^{\alpha/\beta}$, the usual Euclidean distance raised to the power α/β. To see this, we have

$$\begin{aligned}
\tilde{d}_\alpha(x, y) &= \int_0^\infty \int_{\mathbb{R}^n} |a_t(x - u) - a_t(y - u)| du \ t^{\alpha-1} dt \\
&= \int_0^\infty t^{-n\beta} \int_{\mathbb{R}^n} |a_1(t^{-\beta}(x - u)) - a_1(t^{-\beta}(y - u))| du \ t^{\alpha-1} dt \\
&= \int_0^\infty \int_{\mathbb{R}^n} |a_1(v) - a_1(t^{-\beta})(y - x) + v| dv \ t^{\alpha-1} dt
\end{aligned}$$

where we have made the change of variable $v = t^{-\beta}(x - u)$. It is easy to see that the assumptions a_1 is radial implies that the function $\phi(w)$ defined by

$$\phi(w) = \int_{\mathbb{R}^n} |a_1(v) - a_1(w + v)| dv$$

depends only on $||w||_2$; hence we can denote it by $\phi(||w||_2)$ (again abusing notation). We then have

$$\tilde{d}_\alpha(x, y) = \int_0^\infty \phi(t^{-\beta} ||x - y||_2) t^{\alpha-1} dt$$

and making the substitution $s = t^\beta / ||x - y||_2$ gives

$$\tilde{d}_\alpha(x, y) = ||x - y||_2^{\alpha/\beta} \frac{1}{\beta} \int_0^\infty \phi(1/s) s^{\alpha+1/\beta-2} ds.$$

3 Alternate Characterizations of Lipschitz Functions

Define the operator Q_k by

$$Q_k = P_k - P_{k+1}.$$

Suppose from that $d_\alpha(x, y) < \infty$ for all $x, y \in \Omega$.

Theorem 1. *Suppose* $||P_k f - P_{k+1} f||_\infty \leq 2^{-k\alpha}$ *for every* k. *Then* $||f - P_k f||_\infty \leq C 2^{-k\alpha}$, *where* C *is a constant depending only on* α. *Conversely, if* $||f - P_k f||_\infty \leq 2^{-k\alpha}$ *for every* k, *then* $||P_k f - P_{k+1} f||_\infty \leq C' 2^{-k\alpha}$ *for some other constant* C'.

Proof. First, suppose $||f - P_k f||_\infty \leq 2^{-k\alpha}$ for every k. Then $||P_k f - P_{k+1} f||_\infty \leq ||f - P_k f||_\infty + ||f - P_{k+1} f||_\infty \leq 2 \cdot 2^{-k\alpha}$, proving the second part of the theorem with $C' = 2$.

For the first part, write f as a telescopic series:

$$f(x) - \int_\Omega f = \sum_{l=-\infty}^{\infty} (P_l f(x) - P_{l+1} f(x))$$

which holds since $P_\infty(f) = A_0(f) = f$, and $P_{-\infty}(f) = A_\infty(f) = \int f$. Similarly we can write $P_k f$ as a telescopic series

$$P_k f(x) - \int_\Omega f = \sum_{l=-\infty}^{k} (P_l f(x) - P_{l+1} f(x)).$$

Subtracting the two series gives:

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

which is the desired result, with $C = \frac{2^{-\alpha}}{1-2^{-\alpha}}$. □

We have the following theorem:

Theorem 2. *If* $||P_k f - P_{k+1} f||_\infty \leq 2^{-k\alpha}$ *for every* k*, then* f *is Lipschitz with respect to the metric* d_α*.*

Proof. Expand f in a telescopic series:

$$f(x) - \int_\Omega f = \sum_{k=-\infty}^{\infty} (P_k f(x) - P_{k+1} f(x)) = \sum_{k=-\infty}^{\infty} Q_k f(x) = \sum_{k=-\infty}^{\infty} (P_k + P_{k+1}) Q_k f(x)$$

where we have used Lemma (1). Take any $x, y \in \Omega$. Then

$$\begin{aligned} |P_k Q_k f(x) - P_k Q_k f(y)| &= \left| \int p_k(x,u)(Q_k f)(u) du - \int p_k(y,u)(Q_k f)(u) du \right| \\ &= \left| \int (p_k(x,u) - p_k(y,u))(Q_k f)(u) du \right| \\ &\leq 2^{-k\alpha} d_k(x,y). \end{aligned}$$

Similarly,

$$|P_{k+1} Q_k f(x) - P_{k+1} Q_k f(y)| \leq 2^{-k\alpha} d_{k+1}(x,y).$$

Consequently:

$$\begin{aligned}|f(x)-f(y)| &= \left|\sum_{k=-\infty}^{\infty}(P_k+P_{k+1})Q_kf(x)-\sum_{k=-\infty}^{\infty}(P_k+P_{k+1})Q_kf(y)\right| \\ &= \left|\sum_{k=-\infty}^{\infty}(P_kQ_kf(x)-P_kQ_kf(y))+\sum_{k=-\infty}^{\infty}(P_{k+1}Q_kf(x)-P_{k+1}Q_kf(y))\right| \\ &\leq \sum_{k=-\infty}^{\infty}2^{-k\alpha}d_k(x,y)+\sum_{k=-\infty}^{\infty}2^{-k\alpha}d_{k+1}(x,y) \\ &= (1+2^{\alpha})d_\alpha(x,y).\end{aligned}$$

□

We can prove a converse to this result under the following assumption on P_k, namely we assume that

$$\int_\Omega p_k(x,y)d_\alpha(x,y)dy \leq C2^{-k\alpha}. \tag{2}$$

where $C > 0$ is a constant.

Before proving the converse, let us gain some intuition as to why condition (2) is reasonable. Let us return to the case of continuous time. For a class of radial kernels in $\mathbb{R}^n$, we sketched the proof that $\tilde{d}_\alpha(x,y) \sim ||x-y||_2^{\alpha/\beta}$, where $\tilde{d}_\alpha$ is the distance given by (2) and β defines the homogeneity of the kernel. For such kernels, we then have, for all t,

$$\begin{aligned}\int_{\mathbb{R}^n} a_t(x,y)\tilde{d}_\alpha(x,y)dy &\sim \int_{\mathbb{R}^n} t^{-n\beta}a_1(t^{-\beta}(x-y))||x-y||_2^{\alpha/\beta}dy \\ &= \int_{\mathbb{R}^n} a_1(u)||t^\beta u||_2^{\alpha/\beta}du \\ &= t^\alpha\int_{\mathbb{R}^n} a_1(u)||u||_2^{\alpha/\beta}du \\ &\sim t^\alpha.\end{aligned}$$

Theorem 3. *Suppose (2) holds. Then for any f that has Lipschitz constant 1 with respect to d_α, $||f-P_kf||_\infty \leq C2^{-k\alpha}$ for all k.*

Proof. Since $p_k(x,\cdot)$ has integral 1 for every x, we have

$$\begin{aligned}|f(x)-P_kf(x)| = \left|f(x)-\int p_k(x,y)f(y)dy\right| &= \left|\int p_k(x,y)(f(x)-f(y))dy\right| \\ &\leq \int p_k(x,y)d_\alpha(x,y)dy \text{ (since } f \text{ has Lipschitz constant 1)} \\ &\leq C2^{-k\alpha} \text{ (from (2))}\end{aligned}$$

and consequently $||f-P_kf||_\infty \leq C2^{-k\alpha}$, as claimed. □

We note that condition (2) can be expressed in an equivalent form:

Theorem 4. *P_k satisfies* (2) *if and only if for all k, x, y*

$$\int_\Omega p_k(x,y) \sum_{l=-\infty}^{k} d_l(x,y) 2^{-l\alpha} \leq C' 2^{-k\alpha} \tag{3}$$

Proof. That (2) implies (3) is obvious, since

$$\int_\Omega p_k(x,u) \sum_{l=-\infty}^{k} d_l(x,y) 2^{-l\alpha} \leq \int_\Omega p_k(x,y) d_\alpha(x,y) dy.$$

For the other direction, observe that

$$\sum_{l=k+1}^{\infty} d_l(x,y) 2^{-l\alpha} \leq \sum_{l=k+1}^{\infty} 2^{-l\alpha} = \frac{2^{-(k+1)\alpha}}{1-2^{-\alpha}}$$

and consequently

$$\begin{aligned}
\int_\Omega p_k(x,y) d_\alpha(x,y) dy &= \int_\Omega p_k(x,y) \sum_{l=-\infty}^{k} d_l(x,y) 2^{-l\alpha} dy + \int_\Omega p_k(x,y) \sum_{l=k+1}^{\infty} d_l(x,y) 2^{-l\alpha} dy \\
&\leq \int_\Omega p_k(x,y) \sum_{l=-\infty}^{k} d_l(x,y) 2^{-l\alpha} dy + \frac{2^{-(k+1)\alpha}}{1-2^{-\alpha}} \\
&\leq \left(C' + \frac{2^{-\alpha}}{1-2^{-\alpha}} \right) 2^{-k\alpha}
\end{aligned}$$

where we have used (3). □

4 Three Equivalent Norms

We are now ready to introduce the three norms defined on the space of integrable functions on Ω with integral zero. The first is simply the dual norm to Lipschitz class

$$||f||_L = \sup \left\{ \int_\Omega f(x) g(x) dx : g \text{ with Lipschitz norm } 1 \right\}$$

where 'Lipschitz' is with respect to the metric d_α.

Now we define two equivalent norms by

$$||f||_P = \sum_{k=-\infty}^{\infty} 2^{-k\alpha} ||P_k f||_1 = \sum_{k=-\infty}^{\infty} 2^{-k\alpha} \int_\Omega |P_k f(x)| dx$$

and

$$||f||_Q = \sum_{k=-\infty}^{\infty} 2^{-k\alpha} ||Q_k f||_1 = \sum_{k=-\infty}^{\infty} 2^{-k\alpha} \int_\Omega |Q_k f(x)| dx.$$

Theorem 5. *$||\cdot||_Q$ and $||\cdot||_P$ are equivalent norms on the space of mean zero functions.*

Proof. We first show $||f||_Q \lesssim ||f||_P$:

$$\begin{aligned}||f||_Q &= \sum_{k=-\infty}^{\infty} 2^{-k\alpha} \int_\Omega |Q_k f(x)| dx = \sum_{k=-\infty}^{\infty} 2^{-k\alpha} \int_\Omega |P_k f(x) - P_{k+1} f(x)| dx \\ &\leq \sum_{k=-\infty}^{\infty} 2^{-k\alpha} \int_\Omega |P_k f(x)| dx + \sum_{k=-\infty}^{\infty} 2^{-k\alpha} \int_\Omega |P_{k+1} f(x)| dx \\ &= \sum_{k=-\infty}^{\infty} 2^{-k\alpha} \int_\Omega |P_k f(x)| dx + 2^\alpha \sum_{k=-\infty}^{\infty} 2^{-(k+1)\alpha} \int_\Omega |P_{k+1} f(x)| dx \\ &= (1+2^\alpha)||f||_P\end{aligned}$$

For the other direction, since f has mean zero, $P_{-\infty} f = A_\infty f = \int f = 0$. Then we can write P_k as the telescopic series

$$P_k f(x) = \sum_{l \leq k} (P_l f(x) - P_{l-1} f(x)) = -\sum_{l \leq k} Q_{l-1} f(x).$$

Then $||P_k f||_1 = ||\sum_{l \leq k} Q_{l-1} f||_1 \leq \sum_{l \leq k} ||Q_{l-1} f||_1$, and consequently

$$\begin{aligned}||f||_P = \sum_{k=-\infty}^{\infty} 2^{-k\alpha} ||P_k f||_1 &\leq \sum_{k=-\infty}^{\infty} \sum_{l \leq k} ||Q_{l-1} f||_1 \\ &= \sum_{l=-\infty}^{\infty} ||Q_{l-1} f||_1 \sum_{k \geq l} 2^{-k\alpha} \text{ (by Fubini's theorem)} \\ &= \sum_{l=-\infty}^{\infty} ||Q_{l-1} f||_1 \frac{2^{-l\alpha}}{1-2^{-\alpha}} = \frac{2^{-\alpha}}{1-2^{-\alpha}} ||f||_Q.\end{aligned}$$

Hence, we have shown

$$(1+2^\alpha)^{-1} ||f||_Q \leq ||f||_P \leq \frac{2^{-\alpha}}{1-2^{-\alpha}} ||f||_Q$$

for mean zero f. □

We next relate the norm $||\cdot||_L$ to the norms $||\cdot||_P$ and $||\cdot||_Q$. It is convenient to introduce a fourth norm whose definition is similar to that of $||\cdot||_L$; define

$$||f||'_L = \sup\left\{ \int_\Omega f(x) g(x) dx : ||Q_k g||_\infty \lesssim 2^{-k\alpha} \text{ for all } k \right\}.$$

From the equivalence (up to constants) of the condition $||Q_k g||_\infty \lesssim 2^{-k\alpha}$ to the Lipschitz condition under condition (2) on the semigroup, it follows easily that $||\cdot||_L$ and $||\cdot||'_L$ are equivalent norms. We have

Theorem 6. *For functions g, $||g||'_L \lesssim ||g||_P$.*

It is convenient to formulate the following lemma:

Lemma 1. $(P_k + P_{k+1})Q_k = Q_{k-1}$.

Proof of Lemma 1. This is a simple algebraic computation

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

□

Proof of Theorem 6. Suppose f satisfies $||Q_kf||_\infty \leq 2^{-k\alpha}$. Using Lemma 1 and the self-adjointness of P_k gives:

$$\begin{aligned}\int f(x)g(x)dx &= \int \sum_{k=-\infty}^{\infty} (Q_{k-1}f(x))g(x)dx = \sum_{k=-\infty}^{\infty} \int ((P_k + P_{k+1})Q_kf(x))g(x)dx \\ &= \sum_{k=-\infty}^{\infty} \int (Q_kf(x))(P_k + P_{k+1})g(x)dx \\ &\leq \sum_{k=-\infty}^{\infty} 2^{-k\alpha}\Big(\int |P_kg(x)|dx + \int |P_{k+1}g(x)|dx \Big) \\ &= (1 + 2^\alpha)||g||_P.\end{aligned}$$

So taking the supremum over all f with $||Q_kf||_\infty \leq 2^{-k\alpha}$ yields $||g||'_L \leq (1+2^\alpha)||g||_P$. □

Note that this also proves $||g||'_L \lesssim ||g||_Q$, from Theorem 5. Since $|| \cdot ||_L$ and $|| \cdot ||'_L$ are equivalent, this proves $|| \cdot ||_L \lesssim || \cdot ||_Q$ and $|| \cdot ||_L \lesssim || \cdot ||_P$.

For the other direction, we write $|| \cdot ||_L$ in an entirely different form, using the Kantorovich-Rubinstein Theorem for metric spaces:

$$\sup_g \Big\{ \int f(x)g(x)dx \Big\} = \inf_\pi \Big\{ \int \int d(x, y)d\pi(x, y) \Big\}$$

where the supremum on the left is over Lipschitz f, and the infimum on the right is over all non-negative measures π on $\Omega \times \Omega$ satisfying the difference of marginals condition

$$\pi(A \times \Omega) - \pi(\Omega \times A) = \int_A f(x)dx \tag{4}$$

for all measurable $A \subset \Omega$. (We assume as always that f has integral zero.) The theorem holds for a large class of metric spaces; see [4] for a proof. In applications where Ω is a finite set, the theorem is a result of duality from linear programming.

We can use this alternate description of $||f||_L$ to prove that it bounds above $||f||_P$ and $||f||_Q$. Take any measure π on $\Omega \times \Omega$ satisfying (4). This condition says that the measure $\mu(A) = \pi(A \times \Omega) - \pi(\Omega \times A)$ on Ω has Radon-Nikodym derivative $f(x)$; consequently, for any function g on Ω,

$$\int_\Omega g(x)f(x)dx = \int_\Omega g(x)d\mu(x) = \int_\Omega \int_\Omega g(x)d\pi(x,y) - \int_\Omega \int_\Omega g(x)d\pi(y,x).$$

Using this identity gives

$$\begin{aligned}
||A_t f||_1 &= \int_\Omega \left| \int_\Omega a_t(x,y)f(y)dy \right| dx \\
&= \int_\Omega \left| \int_\Omega \int_\Omega a_t(x,y)d\pi(y,u) - \int_\Omega \int_\Omega a_t(x,y)d\pi(u,y) \right| dx \\
&= \int_\Omega \left| \int_\Omega \int_\Omega (a_t(x,u) - a_t(x,y))d\pi(u,y) \right| dx \\
&\leq \int_\Omega \int_\Omega \int_\Omega |a_t(x,u) - a_t(x,y)|d\pi(u,y)dx \\
&= \int_\Omega \int_\Omega \int_\Omega |a_t(u,x) - a_t(y,x)|dxd\pi(u,y) \\
&= \int_\Omega \int_\Omega ||a_t(u,\cdot) - a_t(y,\cdot)||_1 d\pi(u,y).
\end{aligned}$$

When $t = 2^{-k}$ this shows

$$||P_k f||_1 \leq \int_\Omega \int_\Omega d_k(u,y)d\pi(u,y)$$

Using the definition of $d_\alpha = \sum_k 2^{-k\alpha} d_k$, we then have

$$\begin{aligned}
\int_\Omega \int_\Omega d_\alpha(x,y)d\pi(x,y) &= \sum_k 2^{-k\alpha} \int_\Omega \int_\Omega d_k(x,y)d\pi(x,y) \\
&\geq \sum_k 2^{-k\alpha} ||P_k f||_1 \\
&= ||f||_P.
\end{aligned}$$

Taking infimums over all suitable π gives the desired result, completing the proof that all three norms are equivalent.

References

[1] Stein, E.M. *Topics in Harmonic Analysis Related to the Littlewood-Paley Theory.* Princeton University Press, 1970.

[2] Coifman, R.R., Lafon, S. *Diffusion maps.* Applied and Computational Harmonic Analysis 2006; 21(1):5-30.

[3] Coifman, R.R., Hirn, M.J. *Bi-stochastic kernels via asymmetric affinity functions.* Applied and Computational Harmonic Analysis 2013; 35(1): 177-180.

[4] Villani, Cédric. *Topics in Optimal Transportation.* American Mathematical Society, 2003.