An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems

Gary Froyland *

School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia

ARTICLE INFO

Article history:
Received 17 October 2012
Received in revised form 9 January 2013
Accepted 12 January 2013
Available online 30 January 2013
Communicated by T. Wanner

Keywords:
Transfer operator
Coherent set
Finite-time dynamics

ABSTRACT

The study of transport and mixing processes in dynamical systems is particularly important for the analysis of mathematical models of physical systems. Barriers to transport, which mitigate mixing, are currently the subject of intense study. In the autonomous setting, the use of transfer operators (Perron–Frobenius operators) to identify invariant and almost-invariant sets has been particularly successful. In the nonautonomous (time-dependent) setting, coherent sets, a time-parameterised family of minimally dispersive sets, are a natural extension of almost-invariant sets. The present work introduces a new analytic transfer operator construction that enables the calculation of finite-time coherent sets (sets that minimally disperse over a finite time interval). This new construction also elucidates the role of diffusion in the calculation and we show how properties such as the spectral gap and the regularity of singular vectors scale with noise amplitude. The construction can also be applied to general Markov processes on continuous state space.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Transport and mixing processes [1–3] play an important role in many natural phenomena and their mathematical analysis has received considerable interest in the last two decades. Transport is a key descriptor of the impact of the system’s dynamics in the physical world, and coherent structures and metastability provide information on the multiple timescales of the system, which is crucial for efficient modelling approaches. Persistent barriers to transport (or often more correctly pseudo-barriers, which allow very low transport) play fundamental roles in geophysical systems by organising fluid flow and obstructing transport. For example, ocean eddy boundaries strongly influence the horizontal distribution of heat in the ocean, and atmospheric vortices can trap chemicals and pollutants. These time-dependent persistent transport barriers, or Lagrangian coherent structures, are often difficult to detect and track by measurement (for example, by satellite observations), and many coherent structures present in the flow are very difficult to detect, map, and track to high precision by analytical means.

Coherent structures have been the subject of intense research for over a decade, primarily by geometric approaches. The notion that geometric structures such as invariant manifolds play a key role in dynamical transport and mixing for fluid-like flow dates back at least two decades. In autonomous settings, invariant cylinders and tori form impenetrable dynamical barriers. This follows directly from the uniqueness of trajectories of the underlying ordinary differential equation. Slow mixing and transport in periodically driven maps and flows can sometimes be explained by lobe dynamics of invariant manifolds [4,5]. In non-periodic time-dependent settings, finite-time hyperbolic material lines and surfaces, and more generally Lagrangian coherent structures (LCSs) [6–9] have been put forward as a geometric approach to identifying barriers to mixing. LCSs are material (i.e. following the flow) co-dimension 1 objects that are locally the strongest repelling or attracting objects. In practice these objects are often numerically estimated by a finite-time Lyapunov exponent (FTLE) field, with additional requirements [9]. Recent work [10] considers minimal stretching material lines.

Probabilistic approaches to studying transport, based around the transfer operator (or Perron–Frobenius operator) have been developed for autonomous systems, including [11–15]. These approaches use the transfer operator (a linear operator providing a global description of the action of the flow on densities) to answer questions about global transport properties. The transfer operator is particularly suited to identifying global metastable or almost-invariant structures in phase space [11,16–18]. Related methods, which attempt to decompose phase space into ergodic components include [19,20]. Transfer operator-based methods have been very successful in resolving almost-invariant objects in a variety of applied settings, including molecular dynamics (to identify molecular conformations) [16], ocean dynamics...
Furthermore, the maximising $x$ are the $u_l$, $l = 1, \ldots, M$.

From this, we obtain:

\[ \lambda_l = \min_{\forall \text{codim} V \leq l - 1 < M} \max_{0 \neq x \in V} \frac{\langle A x, x \rangle_X}{\langle x, x \rangle_X}, \quad l = 1, \ldots, M. \]
Proposition 1.

\[ \sigma_i := (\lambda_i)^{1/2} = \min_{V: \text{codim} V = \{\cdot\} \leq 1} \max_{0 \neq \varepsilon \in V, 0 \neq y \in Y} \frac{\langle Lx, y \rangle_Y}{\|x\|_X \|y\|_Y}, \quad l = 1, \ldots, M. \]

The maximising \( x \) is \( u_l \) and the maximising \( y \) is \( Lu_l/\|Lu_l\|_Y \), \( l = 1, \ldots, M \).

**Proof.** Taking the square root of (2) one has

\[ \frac{\langle Ax, x \rangle_X^{1/2}}{\|x\|_X} = \frac{\langle Lx, x \rangle_Y^{1/2}}{\|x\|_X} = \max_{0 \neq y \neq 0} \frac{\langle Lx, y \rangle_Y/\|y\|_Y}{\|x\|_X}, \]

where the maximising \( y/\|y\|_Y \) is \( Lx/\|Lx\|_Y \). \( \square \)

We will call the maximising unit \( x \) and \( y \) in (3) the left and right singular vectors of \( L \), respectively. The \( \sigma_i \) are the singular values of \( L \).

3. Transfer operators

We now begin to be more specific about the objects in the previous section. We introduce measure spaces \((X, \mathcal{B}_X, \mu)\) and \((Y, \mathcal{B}_Y, \nu)\), where we imagine \( X \) as our domain at an initial time, with \( \mu \) being a reference measure describing the mass distribution of the object of interest. We now transform \( \mu \) forward by some finite-time dynamics (including advection and diffusion) to arrive at a probability measure \( \nu \) describing the distribution at this later time. The set \( Y \) is the support of this transformed measure \( \nu \), and may be thought of as the “reachable set” for initial points in \( X \). In applications, \( \mu \) may describe the distribution of air or water particles, for instance, and \( \nu \) the distribution at a later time.

We set \( \mathcal{X} = L^2(X, \mu) \) and \( \mathcal{Y} = L^2(Y, \nu) \) and denote the standard inner product on \( L^2(X, \mu) \) by \( \langle \cdot, \cdot \rangle_{\mu} \) and the inner product on \( L^2(Y, \nu) \) by \( \langle \cdot, \cdot \rangle_{\nu} \). We begin to place some assumptions on our transfer operator \( L : \mathcal{X} \to \mathcal{Y} \) and its dual.

**Assumption 1.**

1. \((\mathcal{A}f)(y) = \int k(x, y)f(x) \, dx\), where \( k \in L^2(\mathbb{R} \times \mathbb{R}) \) is non-negative and \( \mathcal{A}1_x = 1_y \), equivalently, \( \int k(x, y) \, dy = 1 \) for \( \mathcal{A}f \) and \( \mathcal{A}^*1_Y = 1_X \), equivalently, \( \int k(y, x) \, dx = 1 \) for \( \mathcal{A}^*f \).

Non-negativity of the stochastic kernel \( k \) in **Assumption 1(1)** is a consistency requirement that says if \( f \) represents some distribution of mass with respect to \( \mu \), then \( \mathcal{A}f \) also represents some mass distribution. Square integrability of \( k \) will provide compactness of \( \mathcal{A} \); we discuss this shortly. **Assumption 1(2)** says that the function \( 1 \) (the density function for the measure \( \mu \) in \( L^2(X, \mu) \)) is mapped to \( 1 \) (the density function for the measure \( \nu \) in \( L^2(Y, \nu) \)). This is a normalisation condition on \( \mathcal{A} \). **Assumption 1(3)** says that \( \mathcal{A} \) preserves integrals. That is, \( \int Y \mathcal{A}f \, d\nu = \int X f \, d\mu \). This follows since \( \int Y \mathcal{A}f \, d\nu = (\mathcal{A}(\mathcal{A}^*1_Y), f) = \int X f \, d\mu \).

**Lemma 1.** Under **Assumption 1(1)**, \( \mathcal{A} : \mathcal{L}(X, \mu) \to \mathcal{L}(Y, \nu) \) and \( \mathcal{A}^* : \mathcal{L}(Y, \nu) \to \mathcal{L}(X, \mu) \) are both compact operators.

**Proof.** See Appendix. \( \square \)

Compactness ensures that the spectrum of \( \mathcal{A}^* \mathcal{A} =: \mathcal{A} : \mathcal{L}(X, \mu) \to \mathcal{L}(Y, \nu) \) is discrete; in particular, the eigenvalue 1 is isolated and of finite multiplicity. Further, self-adjointness and positivity of \( \mathcal{A} \) implies the spectrum is non-negative and real, so the only eigenvalue of magnitude 1 is the eigenvalue 1 itself.

**Assumption 2.** The leading singular value \( \sigma_1 \) of \( \mathcal{A} \) is simple.

We will show in Section 5 that the “small random perturbation of a deterministic process” kernel studied in Sections 4–6 satisfies **Assumption 2**.

**Proposition 2.** Under **Assumption 1**,

1. The largest singular value of \( \mathcal{L} \) is \( \sigma_1 = 1 \) and the corresponding left and right singular vectors are \( 1_X \) and \( 1_Y \), and
2. Under **Assumption 2**,

\[ \sigma_2 = \max_{f \in L^2(X, \mu), g \in L^2(Y, \nu)} \left\{ \frac{\langle \mathcal{A}f, g \rangle_{\nu}}{\|f\|_X \|g\|_Y} : (f, 1)_\mu = (g, 1)_\nu = 0 \right\} < 1, \]

where the maximising \( f \) and \( g \) are \( u_2 \) (the second largest singular vector for \( \mathcal{L} \)) and \( Lu_2/\|Lu_2\| \), respectively.

**Proof.**

1. By **Lemma 8** (see Appendix) using **Assumption 1(2)** and (3), one has \( \|\mathcal{L}\| = \|\mathcal{L}^*\| \leq 1 \). But \( \mathcal{A}1_X = 1_Y \), so \( \|\mathcal{A}\| = 1 \).

2. By compactness \( \sigma_1 = 1 \) is isolated; non-negativity of the spectrum of \( \mathcal{A} \) and simplicity of \( \sigma_1 \) trivially implies \( \sigma_2 < 1 \). To show that (4) = \( \sigma_2 \), we note that by **Assumption 2**, when \( l = 2 \) **Proposition 1** becomes

\[ \sigma_2^+ = \min_{V : \text{codim} V \leq 1} \max_{f \in L^2(X, \mu), g \in L^2(Y, \nu)} \frac{\langle \mathcal{A}f, g \rangle_{\nu}}{\|f\|_X \|g\|_Y} \]

The minimising codimension 1 subspace \( V \) is \( \text{sp} \{1\} \) (orthogonal to the leading singular vector \( 1 \)). Thus (5) becomes

\[ \max_{f \in L^2(X, \mu), g \in L^2(Y, \nu)} \left\{ \frac{\langle \mathcal{A}f, g \rangle_{\nu}}{\|f\|_X \|g\|_Y} : (f, 1)_\mu = 0 \right\}. \]

Arguing as in the proof of **Proposition 1** the maximising \( g/\|g\| \) above is \( \mathcal{L}f/\|\mathcal{L}f\| \), and as \( \langle \mathcal{A}f, 1 \rangle_{\nu} = (f, 1)_\mu \) for any \( f \in L^2(X, \mu) \) one may include the restriction \( (g, 1)_\nu = 0 \) without effect. \( \square \)
3.1. Partitioning X and Y

We wish to measurably partition X and Y as \( X_1 \cup X_2 \) and \( Y_1 \cup Y_2 \) respectively, where

**Goals 1.**

1. \( \mathcal{L} 1_{X_k} \approx 1_{Y_k}, \ k = 1, 2. \)
2. \( \mu(X_k) = \nu(Y_k), \ k = 1, 2. \)

The dynamical reasoning for these goals is that if \( X_1 \) is a coherent set, then one should be able to find a set \( Y_1 \subset Y \) that is approximately the image of \( X_1 \) under the (advective and diffusive) dynamics. Similarly for \( X_2 \), the complement of \( X_1 \) in \( X \). While \( Y_1 \) is possibly only an approximate image of \( X_1 \), we will insist that there is no loss of mass under the dynamics (the action of \( \mathcal{L} \)). Thus \( \mu(X_1) = \nu(Y_1) \) and similarly for \( X_2, Y_2 \). We use the shorthand \( [X_1, X_2] \circ X \) to mean that \( X_1 \) and \( X_2 \) are a measurable partition of \( X \).

Consider the Set-based problem:

\[
\begin{align*}
\text{(S)} \quad \max_{(X_1, X_2) \in X \times Y} & \left\{ \mathcal{L} \left( \frac{\mu(X_2)}{\mu(X_1)} 1_{X_1} - \frac{\mu(X_1)}{\mu(X_2)} 1_{X_2} \right) \cdot \left( \frac{\nu(Y_1)}{\nu(Y_2)} 1_{Y_1} - \frac{\nu(Y_2)}{\nu(Y_1)} 1_{Y_2} \right) : \mu(X_1) = \nu(Y_1), k = 1, 2 \right\}. \\
\end{align*}
\]

(6)

Using the shorthand \( \psi_{X_1, X_2} = \sqrt{\frac{\mu(X_2)}{\mu(X_1)}} 1_{X_1} - \sqrt{\frac{\mu(X_1)}{\mu(X_2)}} 1_{X_2} \) and \( \psi_{Y_1, Y_2} = \sqrt{\frac{\nu(Y_1)}{\nu(Y_2)}} 1_{Y_1} - \sqrt{\frac{\nu(Y_2)}{\nu(Y_1)}} 1_{Y_2} \), one can easily check that \( \| \psi_{Y_1, Y_2} \|_\mu = 1 \) and \( \langle \psi_{X_1, X_2} \rangle = \langle \psi_{Y_1, Y_2} \rangle = 0 \). We have chosen \( \psi_{X_1, X_2} \) and \( \psi_{Y_1, Y_2} \) to be differences of characteristic functions of our partition sets, and to satisfy the normalisation and orthonormalisation conditions above; these will become useful shortly. We now justify the expression (S) in terms of our goals. The constraints in (6) directly capture \( \mu(X_1) = \nu(Y_1), k = 1, 2 \) from Goals 1(2) above. Noting that \( \mu(X_1) = \nu(Y_1), k = 1, 2 \), the objective may be rewritten as:

\[
\begin{align*}
&\left( \frac{\mu(X_2)}{\mu(X_1)} \right) \int \mathcal{L} 1_{X_1} \cdot 1_{Y_1} \, dv + \left( \frac{\mu(X_1)}{\mu(X_2)} \right) \int \mathcal{L} 1_{X_2} \cdot 1_{Y_2} \, dv - \left( \int \mathcal{L} 1_{X_1} \cdot 1_{Y_2} \, dv + \int \mathcal{L} 1_{X_2} \cdot 1_{Y_1} \, dv \right) \\
&= \left( \frac{\mu(X_2)}{\mu(X_1)} \right) \langle \mathcal{L} 1_{X_1}, 1_{Y_1} \rangle + \left( \frac{\mu(X_1)}{\mu(X_2)} \right) \langle \mathcal{L} 1_{X_2}, 1_{Y_2} \rangle - \langle \mathcal{L} 1_{X_1}, 1_{Y_2} \rangle - \langle \mathcal{L} 1_{X_2}, 1_{Y_1} \rangle \\
&= \left( \frac{\mu(X_2)}{\mu(X_1)} \right) \langle \mathcal{L} 1_{X_1}, 1_{Y_1} \rangle + \left( \frac{\mu(X_1)}{\mu(X_2)} \right) \langle \mathcal{L} 1_{X_2}, 1_{Y_2} \rangle - \langle \mathcal{L} 1_{X_1}, 1_{Y_2} \rangle - \langle \mathcal{L} 1_{X_2}, 1_{Y_1} \rangle \\
&= \frac{\langle \mathcal{L} 1_{X_1}, 1_{Y_1} \rangle + \langle \mathcal{L} 1_{X_2}, 1_{Y_2} \rangle}{\mu(X_1)} = \frac{\langle \mathcal{L} 1_{X_2}, 1_{Y_2} \rangle}{\mu(X_2)}.
\end{align*}
\]

(7)

Thus, the problem (S) is a natural way to achieve Goals 1(1) and (2) above.

The set-based problem (6) is difficult to solve because the functions \( \psi_{X_1, X_2} \) and \( \psi_{Y_1, Y_2} \) are restricted to particular forms (differences of two characteristic functions). We therefore relax this condition:

\[
\begin{align*}
\text{(S)} \quad &\max_{(X_1, X_2) \in X \times Y} \left\{ \langle \mathcal{L} \psi_{X_1, X_2}, \psi_{Y_1, Y_2} \rangle \right\} \\
&\leq \max_{f \in L^2((\mu, g), \mathbb{R})} \left\{ \frac{\langle \mathcal{L} f, g \rangle}{\|f\|_\mu \cdot \|g\|_\nu} : \langle f, 1 \rangle_\mu = \langle g, 1 \rangle_\nu = 0 \right\} := \langle \mathcal{R} \rangle,
\end{align*}
\]

where we call (R) the Relax problem as it is a relaxation of the Set-based problem (S). Combining this with Proposition 2 we have

**Theorem 2.** **Under Assumptions 1 and 2,**

\[
\max_{(X_1, X_2) \in X \times Y} \left\{ \frac{\langle \mathcal{L} 1_{X_1}, 1_{Y_1} \rangle}{\mu(X_1)} + \frac{\langle \mathcal{L} 1_{X_2}, 1_{Y_2} \rangle}{\mu(X_2)} : \mu(X_k) = \nu(Y_k), k = 1, 2 \right\} \leq 1 + \sigma_2.
\]

(8)

This result is related to the upper bound in Theorem 2 [37]. The main difference is that [37] is concerned with the dynamics of a reversible Markov operator at equilibrium and seeks fixed metastable sets, while here we have no assumption on reversibility of the dynamics, no restriction on the choice of \( \mu \) (we need not, for example, be invariant under the dynamics), and seek pairs of coherent sets \( X_k, Y_k, k = 1, 2 \), where the \( X_k \) are not fixed, but map approximately onto \( Y_k \) (which need not even intersect \( X_k \)).

Throughout this work, we have, for simplicity considered partitions of \( X \) and \( Y \) into two sets each. Our ideas can be extended to multiple partitions using multiple singular vector pairs, as has been done in the autonomous setting with eigenvectors of transfer operators to find multiple almost-invariant sets [17,38,39]. As one progresses down the spectrum of sub-unit singular values, the corresponding singular vector pairs provide independent information on coherent separations of the phase space, which can in principle be combined using techniques drawn from the autonomous setting.

**Remark 1.** In practice, solving (R) is straightforward once a suitable numerical approximation of \( \mathcal{L} \) has been constructed. Following [27] we will take the optimal \( f \) and \( g \) from (4) and use them heuristically to create partitions \( [X_1, X_2] \) and \( [Y_1, Y_2] \) via \( X_1 = \{ f > b \}, X_2 = \{ f < b \} \), \( Y_1 = \{ g > c \}, Y_2 = \{ g < c \} \), where \( b \) and \( c \) are chosen so that \( \mu(X_k) = \nu(Y_k), k = 1, 2 \). This can be achieved, for example, by a line search on the value of \( b \), where for each choice of \( b \), there is a corresponding choice of \( c \) that will match \( \mu(X_k) = \nu(Y_k) \), \( k = 1, 2 \); we refer the reader to [27,34] for implementation details.
4. Perron–Frobenius operators on smooth manifolds

We now consider the situation where the operator \( \mathcal{L} \) arises from a Perron–Frobenius operator of a deterministic dynamical system. Our dynamical system \( T : M \to M \) acts on a compact subset \( M \subset \mathbb{R}^d \). The action of \( T \) might be derived from either a discrete time or continuous time dynamical system. In the latter case, \( T \) will be a time-\( t \) flow of some differential equation. There are no assumptions about stationarity, in fact, our machinery is specifically designed to handle nonautonomous, random, and non-stationary systems. We seek to analyse the one-step, finite-time action of \( T \).

Our domain of interest will be \( X \subset M \). On \( X \) we have a reference probability measure \( \mu \) describing the quantity we are interested in tracking the transport of. We assume that \( \mu \) has an \( L^2(X, \ell) \) density \( h_\mu \) with respect to Lebesgue measure \( \ell \). We will discuss two situations: purely deterministic dynamics, and deterministic dynamics preceded and followed by a small random perturbation. In the former case we denote the set by \( Y = T(X) \) and the measure \( \nu \) by \( \nu_0 \) (with density \( h_{\nu_0} = d\nu_0/d\ell \), while in the latter case \( Y \) is denoted \( Y \supset T(X) \) and \( \nu \) is denoted \( \nu \) (with density \( h_{\nu} = d\nu/d\ell \)); this terminology emphasises whether or not a perturbation is present. We define these objects formally in the coming subsections. The same subscript notation will also shortly be extended to \( L^2 \) and \( A \) with the obvious meanings.

Two notable aspects of our approach are (i) \( \mu \) may be supported on a small subset \( X \subset M \), enabling one to neglect the remainder of the phase space \( M \) for computations, and (ii) the possibility that \( X \cap Y = \emptyset \); thus \( X \) need not be close to invariant, but in fact, all points may leave \( X \) under the action of \( T \).

4.1. The deterministic setting

Let \( \ell \) denote the Lebesgue measure on \( M \) and suppose that \( T : X \to Y_0 \) is non-singular w.r.t. Lebesgue measure (i.e. \( \ell(A) = 0 \Rightarrow T^{-1}(A) = 0 \)) for all measurable \( A \subset Y_0 \). The evolution of a density \( f \in L^1(X, \ell) \) under \( T \) is described by the Perron–Frobenius operator \( \mathcal{P} : L^1(X, \ell) \to L^1(Y_0, \ell) \) defined by \( \mathcal{P} \mathcal{P} f \, d\ell = f \, d\ell \) for all measurable \( A \subset Y_0 \). If \( T \) is differentiable, one may write \( \mathcal{P} f = \sum_{T \in T^{-1}(X)} \mathcal{P} \mathcal{P} f \, \ell(\cdot, \mu) \nu \), however, in the remainder of this section we do not require differentiability of \( T \).

If we define \( \mathcal{L}_0 \) by \( \mathcal{L}_0 f = \mathcal{P}(f \cdot h_\mu)/h_{\nu_0}, \) then \( \mathcal{L}_0^\alpha : L^\infty(Y_0, \nu_0) \to L^\infty(X, \mu) \) is given by \( \mathcal{L}_0^\alpha g = g \circ T \) since,

\[
\langle \mathcal{L}_0^\alpha, g \rangle_{\nu_0} = \int \mathcal{P}(f \cdot h_\mu)_{h_{\nu_0}} \cdot g \, d\nu_0 = \int \mathcal{P}(f \cdot h_\mu) \cdot g \, d\ell = \int f \cdot h_\mu \cdot g \circ T \, d\ell = \int f \cdot g \circ T \, d\mu = \langle f, \mathcal{L}_0^\alpha g \rangle_\mu,
\]

using standard \( L^1/L^\infty \) duality of \( \mathcal{P} \) (see e.g. p. 48 [40]).

We now briefly argue that it is not instructive to use \( \mathcal{L} = \mathcal{L}_0 \) created directly from \( \mathcal{P} \). Firstly, if \( T \) is invertible (e.g. arising as a time-\( t \) map of a flow), then a simple computation shows that \( \mathcal{A}_0 = \mathcal{L}_0^\alpha \mathcal{L}_0 \) is the identity operator and so there are no “second largest” eigenvalues of \( \mathcal{A}_0 \), only the eigenvalue 1. Secondly, without the invertibility assumption, one can informally (due to non-compactness of \( \mathcal{L}_0 \)) connect this dynamically with Theorem 2. We note that the LHS of (8) (substituting \( \mathcal{L} = \mathcal{L}_0 \)) can be equivalently written as

\[
\max \left\{ \frac{\{1_{X_1}, L_{Y_1}^\alpha 1_{Y_2}\}_\mu}{\mu(X_1)} \right\} = \max \left\{ \frac{(1_{X_1}, 1_{T^{-1} Y_1})_\mu}{\mu(X_1)} \right\} = \max \left\{ \frac{\mu(X_1 \cap T^{-1} Y_1)}{\mu(X_1)} \right\}.
\]

By choosing \( \{Y_1, Y_2\} \) to be any nontrivial measurable partition and \( X_0 = T^{-1}(Y_k), k = 1, 2 \), the value of the above expression becomes 2, forcing \( \sigma_2 = 1 \). Thus, we see that defining \( \mathcal{L} = \mathcal{L}_0 \) using the deterministic Perron–Frobenius operator \( \mathcal{P} \) does not allow us to find a “distinguished” coherent partition; all partitions of this form are equally coherent. In practice, one is typically interested in coherent sets that are robust in the presence of noise of a certain amplitude, or in the presence of diffusion inherent in a dynamical or physical model. In the following section we find such distinguished coherent partitions by adding small random perturbations to the deterministic dynamics.

4.2. Small random perturbations

To create “distinguished” coherent sets in purely advective dynamics, indicated by isolated singular values close to 1, we construct \( \mathcal{L} = \mathcal{L}_0 \) from the Perron–Frobenius operator and diffusion operators. We will apply diffusion before and after the application of the Perron–Frobenius operator. For \( X \subset C \subset M \), we define a local diffusion operator \( \mathcal{D}_{X, \epsilon} : L^1(X, \ell) \to L^1(X, \ell) \) via a bounded stochastic kernel: \( \mathcal{D}_{X, \epsilon} g = \int_{Y \in \mathbb{R}^+} \alpha_{X, \epsilon}(y) (g(x) - g(y)) \, dx \), where \( \alpha_{X, \epsilon} : M \to \mathbb{R}^+ \) satisfies \( \int_{Y \in \mathbb{R}^+} \alpha_{X, \epsilon}(y - x) \, dy = 1 \) for \( x \in X \). Similarly, for \( X_0 \subset Y_0 \subset M \) we define a local diffusion operator \( \mathcal{D}_{Y_0} : L^1(Y_0, \ell) \to L^1(Y_0, \ell) \) by \( \mathcal{D}_{Y_0} g = \int_{Y_0} \alpha_{Y_0}(y - x) g(x) \, dx \), where \( \alpha_{Y_0} \) is bounded and satisfies \( \int_{Y_0} \alpha_{Y_0}(y) \, dy = 1 \) for \( y \in Y_0 \). We have in mind that \( \alpha_{X, \epsilon} \) and \( \alpha_{Y_0} \) are supported in an \( \epsilon \)-neighbourhood of the origin, and that \( x = \text{supp}(\Delta_{X, 1} 1_X), y = T(X), \) and \( y = \text{supp}(\Delta_{X, 1} 1_Y) \). Thus, we have

\[
L^1(X, \ell) \xrightarrow{\mathcal{D}_{X, \epsilon}} L^1(X, \ell) \xrightarrow{\mathcal{P}} L^1(Y_0, \ell) \xrightarrow{\mathcal{D}_{Y_0}} L^1(Y_0, \ell).
\]
We now define \( \mathcal{P}_e : L^1(X, \ell) \to L^1(Y_e, \ell) \) by \( \mathcal{P}_e f(y) = \mathcal{D}_{X_e, e} \mathcal{P} \mathcal{D}_{X_e} f(y) \) and note that

\[
\mathcal{D}_{X_e, e} \mathcal{P} \mathcal{D}_{X_e} f(y) = \int_{Y_e} \alpha_{X_e, e}(y - x) (\mathcal{P} \mathcal{D}_{X_e} f)(x) \, dx
\]

\[
= \int_{X_e} \alpha_{X_e, e}(y - x) \mathcal{P} \left( \int_{X} \alpha_{X, e}(x - z) f(z) \, dz \right) \, dx
\]

\[
= \int_{X_e} \alpha_{X_e, e}(y - T x) \left( \int_{X} \alpha_{X, e}(x - z) f(z) \, dz \right) \, dx \quad \text{by duality of } \mathcal{P}
\]

\[
= \int_{X_e} \left( \int_{X} \alpha_{X_e, e}(y - T x) \alpha_{X, e}(x) \, dx \right) f(z) \, dz.
\]

We note that to define \( \mathcal{P}_e \) via the final displayed expression above, \( T \) need only be measurable; in particular, \( T \) need not be nonsingular (nor invertible).

Finally, we define an operator \( \mathcal{L}_e \) (which, under usable conditions on \( \alpha_{X_e, e} \) and \( \alpha_{Y_e, e} \) discussed shortly, will act on functions in \( L^2(X, \mu) \)) by

\[
\mathcal{L}_e f(y) = \int_{X_e} \alpha_{X_e, e}(y - T z) \alpha_{X_e, e}(z - x) f(y) \, dx = \int_{X} k_e(x, y) f(x) \, d\mu(x),
\]

(10)

where \( k_e(x, y) = \int_{X_e} \alpha_{X_e, e}(y - T z) \alpha_{X_e, e}(z - x) \, dz \). We denote the normalising term in the denominator \( \int_{X_e} \alpha_{X_e, e}(y - T z) \alpha_{X_e, e}(z - x) \, dz \) by \( \mathcal{P}_e h_{\mu} = h_{\mu}(y) \) and define \( v_{\gamma} = d\mu / d\ell \). By Lemma 1, if \( k_e(x, y) \in L^2(X \times Y_e, \mu \times v_{\gamma}) \) then \( \mathcal{L}_e : L^2(X, \mu) \to L^2(Y_e, v_{\gamma}) \) and is compact.

We note that the dual operators \( \mathcal{D}_{X_e, e}^* : L^\infty(X_e, \ell) \to L^\infty(X, \ell) \) and \( \mathcal{D}_{Y_e, e}^* : L^\infty(Y_e, \ell) \to L^\infty(Y_e, \ell) \) are given by \( \mathcal{D}_{X_e, e}^* f(x) = \int_{X_e} \alpha_{X_e, e}(x - y) f(y) \, dy \) and \( \mathcal{D}_{Y_e, e}^* g(x) = \int_{Y_e} \alpha_{Y_e, e}(x - y) g(y) \, dy \), respectively. Explicitly, for this choice of \( \alpha_e \), \( \mathcal{L}_e \) is analogous to \( \mathcal{L}_e^* \) (see Section 4.1), with the addition of pre-and post-diffusion.

We now verify that \( \mathcal{L}_e \) satisfies Assumption 1. For Assumption 1(2), we set \( f = 1_k \) in (10) and for Assumption 1(3), we set \( g = 1_{Y_e} \), i.e.,

\[
\mathcal{L}_e f(y) = \int_{X_e} \alpha_{X_e, e}(y - T z) \alpha_{X_e, e}(z - x) f(x) \, dx = \int_{X_e} \alpha_{X_e, e}(y - z) \alpha_{X_e, e}(z - x) f(x) \, dx
\]

\[
= \int_{X_e} \alpha_{X_e, e}(z - x) \alpha_{X_e, e}(z - y) f(x) \, dx
\]

\[
= \mathcal{D}_{X_e, e} \mathcal{K} \circ \mathcal{D}_{X_e}^* g(x)
\]

(11)

(12)

(13)

(14)

where \( \mathcal{K} g = g \circ T \) is the Koopman operator. Thus, the expression for \( \mathcal{L}_e^* \) is analogous to \( \mathcal{L}_e^* \) (see Section 4.1), with the addition of pre- and post-diffusion.

We now verify that \( \mathcal{L}_e \) satisfies Assumption 1. For Assumption 1(2), we set \( f = 1_k \) in (10) and for Assumption 1(3), we set \( g = 1_{Y_e} \), i.e.,

\[
\mathcal{L}_e f(y) = \int_{X_e} \alpha_{X_e, e}(y - T z) \alpha_{X_e, e}(z - x) f(x) \, dx
\]

\[
= \mathcal{D}_{X_e, e} \mathcal{K} \circ \mathcal{D}_{X_e}^* g(x)
\]

(15)

\[
= \mathcal{D}_{X_e, e} \mathcal{K} \circ \mathcal{D}_{X_e}^* g(z) \, dz
\]

(16)

3 In the case where \( M \) has boundaries so that \( B_\ell(Tx) \supset M \) for some \( x \in X \), we assume that the kernel \( \alpha_e \) is suitably modified at these boundaries to remain stochastic.
Thus, \( \mathcal{L}_\epsilon f \) at \( y \in Y_\epsilon \) is a double-average over the pull-back by \( T \) of an \( \epsilon \)-neighbourhood of \( y \). Clearly, \( \mathcal{L}_\epsilon 1_X = 1_{Y_\epsilon} \). Also, by (13):

\[
\mathcal{L}_\epsilon^*g(x) = \frac{1}{\ell(B_\epsilon(0)\epsilon^2)} \int_{Y_\epsilon} \int_{X_\epsilon} 1_{B_\epsilon(0)}(y - Tz)1_{B_\epsilon(0)}(z - x) \, dz \, g(y) \, dy
= \frac{1}{\ell(B_\epsilon(0))} \int_{X_\epsilon} 1_{B_\epsilon^*(x)}(x) \left( \frac{1}{\ell(B_\epsilon(0))} \int_{Y_\epsilon} 1_{B_\epsilon(Tz)}(y) \, dy \right) \, dz
= \frac{1}{\ell(B_\epsilon(0))} \int_{X_\epsilon} 1_{B_\epsilon^*(x)}(x) \left( \frac{1}{\ell(B_\epsilon(0))} \int_{X_\epsilon} g(y) \, dy \right) \, dz
\]

noting that \( B_\epsilon(Tz) \subset Y_\epsilon \) for \( z \in X_\epsilon \) and \( B_\epsilon(x) \subset X_\epsilon \) for \( x \in X \). Using these facts, one may compute that \( \mathcal{L}_\epsilon^* 1_{Y_\epsilon}(x) = 1_X \).

**Lemma 2.** If \( \ell(Y_\epsilon) < \infty \) and \( \alpha_{Y,\epsilon}(x) = \alpha_{Y,\epsilon}(x) = 1_{B_\epsilon(0)}(x)/\ell(B_\epsilon(0)) \) then \( \kappa_\epsilon \in L^2(X \times Y_\epsilon, \mu \times \nu_\epsilon) \) and thus \( \mathcal{L}_\epsilon \) is compact.

**Proof.** We show that \( k_\epsilon(x, y) \in L^2(X \times Y_\epsilon, \mu \times \nu_\epsilon) \); the result will then follow from **Lemma 1**. We use the fact that \( \int_{X_\epsilon} \alpha_{Y,\epsilon}(y - Tz)\alpha_{X,\epsilon}(z - x) \, dz = \ell(B_\epsilon(x) \cap T^{-1}B_\epsilon(y)) / (\ell(B_\epsilon))^2 \) as shown in (16).

\[
\|k\|_2^2 = \int_{Y_\epsilon} \int_{X_\epsilon} \left( \int_{X_\epsilon} \alpha_{Y,\epsilon}(y - Tz)\alpha_{X,\epsilon}(z - x) \, dz \right)^2 \, d\mu(x) \, d\nu_\epsilon(y)
= \int_{Y_\epsilon} \int_{X_\epsilon} \left( \int_{X_\epsilon} \alpha_{Y,\epsilon}(y - Tz)\alpha_{X,\epsilon}(z - x) \, dz \right)^2 \, d\mu(x) \, d\nu_\epsilon(y)
= \frac{1}{\ell(B_\epsilon(0))^2} \int_{X_\epsilon} \int_{X_\epsilon} \ell(T^{-1}B_\epsilon(y) \cap B_\epsilon(x))^2 \, d\mu(x) \, d\nu_\epsilon(y)
= \frac{1}{\ell(B_\epsilon(0))^2} \int_{X_\epsilon} \int_{X_\epsilon} \ell(T^{-1}B_\epsilon(y) \cap B_\epsilon(x))^2 \, d\mu(x) \, d\nu_\epsilon(y)
\leq \frac{\ell(Y_\epsilon)}{\ell(B_\epsilon(0))} < \infty.
\]

By **Lemma 1** we have that \( \mathcal{L}_\epsilon \) and \( \mathcal{L}_\epsilon^* \) are both compact. \( \square \)

**Corollary 1.** If \( \ell(Y_\epsilon) < \infty \) and \( \alpha_{Y,\epsilon}(x) = \alpha_{Y,\epsilon}(x) = 1_{B_\epsilon(0)}(x)/\ell(B_\epsilon(0)) \), the operator \( \mathcal{L}_\epsilon \) satisfies **Assumption 1**.

In Section 5 we show that the above choice of \( \alpha_{Y,\epsilon}, \alpha_{Y,\epsilon} \) satisfies **Assumption 2**.

### 4.3. Objectivity

We demonstrate that our analytic framework for identifying finite-time coherent sets is **objective or frame-invariant**, meaning that the method produces the same features when subjected to time-dependent “proper orthogonal + translational” transformations; see [41,42].

In continuous time, to test for objectivity, one makes a time-dependent coordinate change \( x \mapsto Q(t)x + b(t) \) where \( Q(t) \) is a proper orthogonal linear transformation and \( b(t) \) is a translation vector, for \( t \in \{t_0, t_1\} \). The discrete time analogue is to consider our initial domain \( X \) transformed to \( \hat{X} = \{Q(t_0)x + b(t_0) : x \in X\} \) and our final domain \( Y_\epsilon \) transformed to \( \hat{Y}_\epsilon = \{Q(t_1)y + b(t_1) : y \in Y_\epsilon\} \). For shorthand, we use the notation \( \Phi_{t_0} \) and \( \Phi_{t_1} \) for these transformations, so that \( \hat{X}_\epsilon = \Phi_{t_0}(X) \) and \( \hat{Y}_\epsilon = \Phi_{t_1}(Y_\epsilon) \). The deterministic transformation \( \hat{T} : \Phi_{t_0}(M) \rightarrow \Phi_{t_1}(M) \), which we wish to analyse using our transfer operator framework, is given by \( \hat{T} = \Phi_{t_1} \circ T \circ \Phi_{t_0}^{-1} \). The change of frames is summarised in the commutative diagram below.

\[
\begin{array}{ccc}
M & \xrightarrow{T} & M \\
\Phi_{t_0}(M) & \xrightarrow{T} & \Phi_{t_1}(M)
\end{array}
\]

We define \( \hat{\mu} = \mu \circ \Phi_{t_0}^{-1} \) and \( \hat{\nu}_\epsilon = \nu_\epsilon \circ \Phi_{t_1}^{-1} \) as the transformed versions of \( \mu \) and \( \nu_\epsilon \); \( \hat{\mu} \) and \( \hat{\nu}_\epsilon \) are probability measures on \( \hat{X} \) and \( \hat{Y}_\epsilon \), respectively. We further define the Perron–Frobenius operators for \( \Phi_{t_0} \) and \( \Phi_{t_1} \), namely \( \mathcal{D}_{\Phi_{t_0}} : L^2(X, \mu) \rightarrow L^2(\hat{X}, \hat{\mu}) \) and \( \mathcal{D}_{\Phi_{t_1}} : L^2(Y_\epsilon, \nu_\epsilon) \rightarrow L^2(\hat{Y}_\epsilon, \hat{\nu}_\epsilon) \) by \( \mathcal{D}_{\Phi_{t_0}} f = f \circ \Phi_{t_0}^{-1} \) and \( \mathcal{D}_{\Phi_{t_1}} f = f \circ \Phi_{t_1}^{-1} \).

The operators \( \mathcal{D}_{\Phi_{t_0}} \) and \( \mathcal{D}_{\Phi_{t_1}} \) are defined on \( \hat{X} \) and \( \hat{Y}_\epsilon \) analogously to the definitions of \( \mathcal{D}_{X,\epsilon} \) and \( \mathcal{D}_{Y,\epsilon} \); that is, \( \mathcal{D}_{\Phi_{t_0}}(y) = \int_{\hat{X}} \alpha_{\Phi_{t_0}}(y - x)g(x) \, dx \) and \( \mathcal{D}_{\Phi_{t_1}}(y) = \int_{\hat{Y}_\epsilon} \alpha_{\Phi_{t_1}}(y - x)g(x) \, dx \), where \( \alpha_{\Phi_{t_0}} = \alpha_{X,\epsilon} = 1_{B_\epsilon(0)}(x)/\ell(B_\epsilon(0)) \). Note in particular, that these operators are created independently of the operators \( \mathcal{D}_{X,\epsilon} \) and \( \mathcal{D}_{Y,\epsilon} \), and that \( \mathcal{D}_{\Phi_{t_0}} \) and \( \mathcal{D}_{\Phi_{t_1}} \) are not simply defined by direct transformation of
\(D_{\kappa, \epsilon} \text{ and } D_{\kappa', \epsilon}, \text{ with } P_{\kappa_0}, P_{\kappa_1}. \) While the transformed deterministic dynamics \(\hat{T}\) must of course be defined by conjugation with \(\Phi_{\kappa_0}, \Phi_{\kappa_1},\) our intention with the “small random perturbation” model is to use the same diffusion operators \(D_{\kappa, \epsilon}, D_{\kappa', \epsilon}, \text{ without any knowledge of coordinate changes, to construct an } \hat{T}. \) Note that because \(\Phi_{\kappa_0} \text{ and } \Phi_{\kappa_1}\) are proper orthogonal affine transformations, one has
\[
P_{\kappa_0} \circ D_{\kappa, \epsilon} = D_{\kappa, \epsilon} \circ P_{\kappa_0}, \text{ and}
P_{\kappa_1} \circ D_{\kappa', \epsilon} = D_{\kappa', \epsilon} \circ P_{\kappa_1}
\]
geometrically, the LHSs are an averaging over \(\epsilon\)-balls followed by the transformations \(\Phi_{\kappa_0} \text{ (resp. } \Phi_{\kappa_1})\) first and then average. The result is the same because \(B_{\kappa_0, \epsilon}^{-1}(B_{\kappa_1, \epsilon}) = B_{\kappa_1, \epsilon}^{-1}(B_{\kappa_0, \epsilon});\) similarly, for \(\Phi_{\kappa_1}^{-1}\).

The Perron–Frobenius operator for \(\hat{T}\), denoted \(P\) is \(P = P_{\kappa_1} \circ P \circ P^{-1}_{\kappa_0}.\) Denote \(\hat{h}_\mu = P_{\kappa_0} h_\mu.\) Then, \(\hat{L}_\epsilon : L^2(\hat{X}, \hat{v}) \to L^2(\hat{Y}, \hat{v})\) is defined by \(\hat{L}_\epsilon f = D_{\kappa', \epsilon} \circ P_{\kappa_1} \circ D_{\kappa, \epsilon} (f \cdot h_\mu)/D_{\kappa, \epsilon} \circ P_{\kappa_1} \circ D_{\kappa', \epsilon} (h_\mu)\) is defined by \(\hat{L}_\epsilon f = D_{\kappa', \epsilon} \circ P_{\kappa_1} \circ D_{\kappa, \epsilon} (f \cdot h_\mu)/D_{\kappa, \epsilon} \circ P_{\kappa_1} \circ D_{\kappa', \epsilon} (h_\mu)\).

**Theorem 3.** The generator \(L\) in the original frame and the operator \(\hat{L}_\epsilon\) in the transformed frame satisfy the commutative diagram:

\[
\begin{align*}
L^2(X, \mu) &\xrightarrow{\ell} L^2(X, \nu) \\
\downarrow P_{\kappa_0} &\xrightarrow{\ell} \downarrow P_{\kappa_1} \\
L^2(\hat{X}, \hat{\mu}) &\xrightarrow{\hat{L}_\epsilon \ell} L^2(\hat{Y}, \hat{v})
\end{align*}
\]

**Proof.** Note that because \(P_{\kappa_0}\) is a composition operator, one has \(P_{\kappa_0} (f \cdot g) = P_{\kappa_0} f \cdot P_{\kappa_0} g\) (and similarly for \(P_{\kappa_1}\); we use this fact below. We also make use of (18).

\[\hat{L}_\epsilon f = D_{\kappa', \epsilon} \circ P_{\kappa_1} \circ D_{\kappa, \epsilon} (f \cdot h_\mu)/D_{\kappa, \epsilon} \circ P_{\kappa_1} \circ D_{\kappa', \epsilon} (h_\mu) = D_{\kappa', \epsilon} \circ P_{\kappa_1} \circ P \circ P_{\kappa_0}^{-1} \circ D_{\kappa, \epsilon} (f \cdot h_\mu)/P_{\kappa_1} \circ D_{\kappa', \epsilon} \circ P \circ P_{\kappa_0}^{-1} \circ D_{\kappa, \epsilon} (h_\mu) = P_{\kappa_1} \circ P \circ P_{\kappa_0}^{-1} (f \cdot h_\mu)/P_{\kappa_1} \circ P \circ P_{\kappa_0}^{-1} (h_\mu) = P_{\kappa_1} \circ P (P_{\kappa_0}^{-1} f \cdot h_\mu)/P_{\kappa_1} \circ P (P_{\kappa_0}^{-1} h_\mu) = P_{\kappa_1} \left( P (P_{\kappa_0}^{-1} f \cdot h_\mu) \right) = P_{\kappa_1} \left( P (P_{\kappa_0}^{-1} h_\mu) \right)\]

as required. \(\square\)

**Corollary 2.** If \(L g = \lambda g\) where \(f \text{ and } g \text{ are left and right singular vectors of } L, \text{ respectively, then } \hat{L}_\epsilon P_{\kappa_0} f = \lambda P_{\kappa_1} g \text{ where } P_{\kappa_0} f \text{ and } P_{\kappa_0} g \text{ are left and right singular vectors of } \hat{L}_\epsilon, \text{ respectively.}\)

It follows from Corollary 2 that the coherent sets extracted on \(X \text{ and } Y\) as eg. level sets from the singular vectors of \(L\) will be transformed versions (under \(\Phi_{\kappa_0} \text{ and } \Phi_{\kappa_1}\)) of those extracted from \(L\) on \(X \text{ and } Y,\) as required for objectivity.

More generally, if \(D_{\kappa, \epsilon} \text{ and } D_{\kappa', \epsilon}\) are compact operators representing diffusion, then \(P_{\kappa_0} \circ D_{\kappa, \epsilon} = D_{\kappa, \epsilon} \circ P_{\kappa_0} \text{ and } P_{\kappa_1} \circ D_{\kappa', \epsilon} = D_{\kappa', \epsilon} \circ P_{\kappa_1}\) is a sufficient condition for the method to be objective.

**5. The case of T a diffeomorphism**

In this section we specialise to the case where \(M = X = Y \subset \mathbb{R}^d\) is compact, \(T: M \to M\) is a diffeomorphism, and \(|\det DT|\) and \(h_\mu\) are bounded uniformly above and below.

**5.1. Simplicity of the leading singular value of \(L_\epsilon\)**

Our main result of this section states that the leading singular value of \(L_\epsilon\) is simple when \(\alpha_{\kappa, \epsilon} = \alpha_{\kappa', \epsilon} = \mathbf{1}_{B_{\kappa_0}}(0)/\ell(B_{\kappa_0}(0));\) thus Assumption 2 is satisfied. To set notation, note that one has \(\hat{L}_\epsilon^* g(x) = \int_X k(x, y) g(y) \, d\nu(y),\) so \(A_\mu f (y) = \int_Y k(x, y) f(x) \, d\mu(x),\) where \(k(x, y) = \int k(x, y, z) k(x, z) \, dv(z).\) Let \(A_\mu f (y) := \int_Y k(x, y) \, d\mu(x).\) The following technical lemma provides sufficient conditions for simplicity of the leading singular value of \(L_\epsilon\).

**Lemma 3.** If there exists an integer \(q > 0, \text{ a } G \in L^1(X, \mu) \text{ such that } k_\mu(x, y) \leq G(y), \text{ and a set } A \subset X \text{ with } \mu(A) > 0 \text{ such that } k_\mu(x, y) > 0 \text{ for all } x \in X \text{ and } y \in A, \text{ then the leading eigenvalue value of } A_\mu, \text{ namely } \sigma_1 = 1, \text{ is simple.}\)

**Proof.** By Theorem 5.7.4 [40] under the hypotheses of the lemma, the Markov operator \(A\) is “asymptotically stable” in \(L^1(X, \mu),\) meaning that there exists a unique \(h \in L^1 \text{ (scaled so that } \int_X h \, d\mu = 1) \text{ such that } Ah = h \text{ and } \lim_{\epsilon \to 0} \|A^\epsilon f - h\|_1 = 0 \text{ for all } f \in L^1 \text{ scaled}
Lemma 5. Let $f$ be a Lipschitz function on $X$ and $\epsilon > 0$. Then $D_{\epsilon}f$ is globally Hölder on $X_{\epsilon}$ with Hölder constant $C$.

\[ |D_{\epsilon}x f(x) - D_{\epsilon}y f(y)| \leq \|f\|_{L^1(X, \mu)} \frac{C(d)}{\epsilon^{(1+d)/2}} \|x - y\|^{1/2}, \quad \text{for all } x, y \in X_{\epsilon}, \]

where $C(d) = \sqrt{2/(d+2)}$, and $d = \dim X$.

5.2. Regularity of singular vectors of $L_e$

A standard heuristic for obtaining partitions from functions is to threshold on level sets; such an approach has been used in many previous applications of transfer operator methods to determine almost-invariant and metastable sets for autonomous or time-independent dynamical systems [17,18]. By employing this heuristic, forming, e.g. $X_1 = \{ x \in M : f(x) < c \}$ where $f$ is a sub-dominant singular vector of $L_e$, and $c \in \mathbb{R}$ is a threshold, if $f$ has some regularity, this places some limitations on the geometrical form of $X_1$.

We derive explicit expressions for the Lipschitz and Hölder constants for the eigenfunctions of $L_e$, $L_{\epsilon}$, and $L_{\epsilon}^*$, showing how these constants vary as a function of the perturbation parameter $\epsilon$. We begin with two lemmas that demonstrate the regularity of $D_{\epsilon}f$, for $f \in L^2(X, \mu)$. Completely analogous results hold for the regularity of $D_{\epsilon}g$, $g \in L^2(Y, \nu)$. In order to obtain a Lipschitz bound on $D_{\epsilon}f$, one requires $f$ to be bounded; on the other hand, the Hölder bound is in terms of the $L^2$-norm of $f$, which is better suited to our setup. In one direction (right singular vectors of $L_e$) we can combine the Hölder bound and the Lipschitz bound to provide a better estimate than a direct Hölder bound.

We remark that our diffusion kernels $\alpha_{\epsilon, x}$, $\alpha_{\epsilon, y}$ need not be Lipschitz nor Hölder. Some prior work has considered the regularity of the range of $\mathcal{P}$ followed by a smoothing operator, denoted by $D_{\epsilon}$, in either a $C^0$ or $L^1$ setting: this includes Zeeman, Lemma 5 [43], who discusses the equicontinuity of the image of the unit sphere in $C^0$ and $\alpha_{\epsilon}(x) = n \exp(-||x||^2)/2\pi$ is smooth, and Junge, Proposition 3.1 [44], who bounds the Lipschitz constant of $D_{\epsilon}\mathcal{P}$ in terms of the $L^1$-norm of $f$ using a Lipschitz kernel $\alpha_{\epsilon}$ (the bound is $L_{\epsilon\alpha} f \leq L_{\epsilon} f$).

As most of the proofs of the following results are technical, we have deferred them to the Appendix.

Lemma 4. Let $f \in L^\infty(X, \mu)$, where $X \subset \mathbb{R}^d$, $1 \leq d \leq 3$. Let $\alpha_{\epsilon} = 1_{B_\epsilon}/\mu(B_\epsilon)$. Then $D_{\epsilon}f$ is globally Hölder on $X_{\epsilon}$ with Hölder constant bounded above by $C_\epsilon/d/d$, where $C_\epsilon = 1, 4/\pi, 3/2$, for dimensions $1, 2, 3$, respectively.

Lemma 5. Let $f \in L^2(X, \mu)$, where $X \subset \mathbb{R}^d$, $1 \leq d \leq 3$. Let $\alpha_{\epsilon} = 1_{B_\epsilon}/\mu(B_\epsilon)$. Then $D_{\epsilon}f$ is globally Hölder on $X_{\epsilon}$:

\[ |D_{\epsilon}x f(x) - D_{\epsilon}y f(y)| \leq \|f\|_{L^2(X, \mu)} \frac{C(d)}{\epsilon^{(1+d)/2}} \|x - y\|^{1/2}, \quad \text{for all } x, y \in X_{\epsilon}, \]

where $C(d) = 1/\sqrt{2}, \sqrt{2/\pi}, 3/(2\sqrt{2})$, for dimensions $1, 2, 3$, respectively.

Proposition 4. If $f \in L^2(X, \mu)$ with $\int f \, d\mu = 0$, then $L_e f$ is Hölder and

\[ |L_e f(x) - L_e f(y)| \leq \|f\|_{L^2(X, \mu)} C(d) \cdot C_{\mu} \cdot (1/\epsilon^{d+1}) \|x - y\|^{1/2}, \quad \text{for all } x, y \in X_e, \]

in dimensions $d = 1, 2, 3$, where $0 < C_{\mu} < \infty$ is a constant depending on properties of $T$ and $h_\mu$. 

Proof. By Lemma 9 (see Appendix), one has that $k_\epsilon$ is bounded. Note that

\[ k(x, y) = \int_X k(y, z)k(x, z) \, d\nu(z) \leq \|k\|_\infty \int_X k(y, z) \, d\nu(z) = \|k\|_\infty. \]

Thus, in the hypotheses of Lemma 3 we may take $G(y) \equiv \|k\|_\infty^{-1} \in L^1(X, \mu)$. By Lemma 10 (see Appendix) the covering hypothesis of Lemma 3 is satisfied; the result follows by Lemma 3. \qed
2. If \( f \) is bounded with \( \int f \, d\mu = 0 \), then \( \mathcal{L}f \) is Lipschitz and
\[
|\mathcal{L}_\epsilon f(x) - \mathcal{L}_\epsilon f(y)| \leq \|f\|_\infty C(\epsilon) C'_\epsilon (1/\epsilon^2) \|x - y\|, \quad \text{for all } x, y \in Y_{\epsilon},
\]
in dimensions \( d = 1, 2, 3 \), where \( 0 < C'_\epsilon < \infty \) is a constant depending on properties of \( T, h, \) and \( M \).

The explicit forms of \( C'_\epsilon \) and \( C_\epsilon \) are given in the proofs in the Appendix.

**Proposition 5.** If \( g \in L^2(Y, \nu_\epsilon) \) then \( \mathcal{L}_\epsilon^* g \) is Hölder and
\[
|\mathcal{L}_\epsilon^* g(x) - \mathcal{L}_\epsilon^* g(y)| \leq \|g\|_{L^2(\nu_\epsilon)} 1/\sqrt{A} \|x - y\|^{1/2}, \quad \text{for all } x, y \in Y_{\epsilon},
\]
in dimensions \( d = 1, 2, 3 \), where \( A = \min_{x \in X_{\epsilon}} |\det DT(x)| \).

As the random perturbations or noise of amplitude \( \epsilon \) is increased, Propositions 4 and 5 show that the images of \( L^2 \) functions under \( \mathcal{L}_\epsilon \) and \( \mathcal{L}_\epsilon^* \) become more regular in a Hölder (or Lipschitz) sense. This information can be used to imply similar regularity results for the left and right singular vectors of \( \mathcal{L}_\epsilon \). As we anticipate that the optimal \( \psi_{x_1, x_2} \), \( \psi_{y_1, y_2} \) in the set-based problem (5) (Section 3.1) will have coefficients of \( O(1) \), we are interested in the minimum spatial distance in phase space that can be traversed by an \( O(1) \) difference in value of the singular vectors of \( \mathcal{L}_\epsilon \). Lower bounds on the \( \epsilon \)-scaling of these distances are the content of the following corollary. In both the statement and proof, for brevity we drop the explicit constants and make statements only about the scaling behaviour; the explicit constants can be constructed using the proofs in the Appendix.

**Corollary 3.** 1. Let \( f \in L^2(\mu) \) be a subdominant left singular vector of \( \mathcal{L}_\epsilon \) (an eigenfunction of \( \mathcal{L}_\epsilon^* \mathcal{L}_\epsilon \) corresponding to an eigenvalue less than 1), normalised so that \( \|f\|_{L^2(\mu)} = 1 \). An “\( O(\epsilon) \)” feature has width of at least order \( \epsilon^{d+1} \).

2. Let \( g \in L^2(\nu_\epsilon) \) be a subdominant right singular vector of \( \mathcal{L}_\epsilon \) (an eigenfunction of \( \mathcal{L}_\epsilon^* \mathcal{L}_\epsilon \) corresponding to an eigenvalue less than 1), normalised so that \( \|g\|_{L^2(\nu_\epsilon)} = 1 \). An “\( O(1) \)” feature has width of at least order \( \epsilon^{(d+1)/2} \).

**Proof.** 1. A subdominant normalised singular vector \( f \in L^2(X, \mu) \) arises as an eigenvector of \( \mathcal{L}_\epsilon^* \mathcal{L}_\epsilon \), and in particular \( f = \mathcal{L}_\epsilon^* g \) for some \( g \in L^2(Y, \nu_\epsilon) \) with \( \int g \, d\nu_\epsilon = 0 \) and \( \|g\|_{L^2(\nu_\epsilon)} = 1 \). By Proposition 5 we see that the Hölder constant of \( f = \mathcal{L}_\epsilon^* g \) is \( O(1/\epsilon^{(d+1)/2}) \). Thus, if along a given direction, \( g \) increases from zero to \( O(1) \) and decreases to zero again, the minimal distance required is \( O(\epsilon^{(d+1)/2}) = O(\epsilon^{d+1}) \). In detail, if \( 1 \leq H_{1/2}(f) : |x - y|^{1/2} \) then
\[
|f(x) - f(y)| \geq 1/H_{1/2}(f)^2 \geq A/C(d)^2 \epsilon^{d+1}.
\]

2. A subdominant normalised singular vector \( g \in L^2(Y, \nu_\epsilon) \) arises as an eigenvector of \( \mathcal{L}_\epsilon^* \mathcal{L}_\epsilon \). We begin with a \( g \in L^2(\nu_\epsilon) \) and apply Proposition 5 to obtain \( \|\mathcal{L}_\epsilon^* g\|_\infty \leq \|g\|_{L^2(\nu_\epsilon)} \leq C(\epsilon) C'_\epsilon \|g\|_{L^2(\nu_\epsilon)} \) using the fact that \( \int g \, d\nu_\epsilon = 0 \). We now apply Proposition 4(2) to obtain
\[
L(f) \leq \frac{UB\text{diam}(M)^{1/2}C(\nu)C'_\epsilon(d)}{A^{3/2}Lk^{(d+3)/2}} + \frac{UB^2\text{diam}(M)^{3/2}C(\nu)C'_\epsilon(d)^2}{A^{5/2}Lk^{(5d+3)/2}}.
\]
Now, in order to have an \( O(1) \) feature we require \( 1 \leq L(f) : |x - y| \) or \( |x - y| \geq 1/L(f) \). Since we have an upper bound for \( L(f) \), the width of an \( O(1) \) feature must be at least \( O(\epsilon^{(d+1)/2}) \). 

**Remark 2.** Note that if we use only Proposition 4(1) in the proof of Corollary 3(2), we would obtain \( O(\epsilon^{d+2}) \), which is worse than the estimate in the corollary.

5.3. Scaling of \( \sigma_{2,\epsilon} \) with \( \epsilon \)

We now demonstrate a lower bound on \( \sigma_{2,\epsilon} \), for small \( \epsilon \), where \( \sigma_{2,\epsilon} \) is the second singular value of \( \mathcal{L}_\epsilon \). A bound similar is spirit to Corollary 4 was developed in the autonomous two-dimensional area-preserving setting [45]. Related numerical studies of advection–diffusion partial differential equations include [46,47]. We begin by choosing some \( \epsilon^* > 0 \) and a fixed partition \( \{Y_{k,\epsilon}\}_{k=1}^{k_{\epsilon}} \) of \( Y_{\epsilon} \). The partition \( \{Y_{k,\epsilon}\}_{k=1}^{k_{\epsilon}} \) induces compatible partitions for \( Y_0 \), \( 0 < \epsilon < \epsilon^* \), namely, \( \{Y_{k,\epsilon}\}_{k=1}^{k_{\epsilon}} \), where \( Y_{k,\epsilon} := Y_{k,\epsilon} \cap Y_\epsilon \), the restriction of \( Y_{k,\epsilon} \) to \( Y_\epsilon \). In what follows, it will be useful to consider the “\( \epsilon \)-interior” of a set \( A \), denoted \( \bar{A} := \{x \in A : B_\epsilon(x) \subset A\} \).

**Lemma 6.** Let \( T : X \to Y_0 \) be non-singular and suppose that \( \mu \) is supported on \( X \) and absolutely continuous. One has
\[
\frac{\mu(T^{-1}Y_{1,\epsilon})}{\mu(T^{-1}Y_{2,\epsilon})} + \frac{\mu(T^{-1}Y_{2,\epsilon})}{\mu(T^{-1}Y_{1,\epsilon})} \leq \frac{\langle \mathcal{L}_\epsilon 1_{X_{1,\epsilon}}, 1_{Y_{1,\epsilon}} \rangle_{\mu}}{\mu(X_1)} + \frac{\langle \mathcal{L}_\epsilon 1_{X_{2,\epsilon}}, 1_{Y_{2,\epsilon}} \rangle_{\mu}}{\mu(X_2)} \leq 1 + \frac{\sigma_2}{\epsilon^*},
\]
for all \( 0 < \epsilon < \epsilon^* \).

**Proof.** First note that by Theorem 2, one has
\[
\sigma_2 \geq \frac{\langle \mathcal{L}_\epsilon 1_{X_{1,\epsilon}}, 1_{Y_{1,\epsilon}} \rangle_{\mu}}{\mu(X_{1,\epsilon})} + \frac{\langle \mathcal{L}_\epsilon 1_{X_{2,\epsilon}}, 1_{Y_{2,\epsilon}} \rangle_{\mu}}{\mu(X_{2,\epsilon})} - 1,
\]
for any partition \( \{X_1, X_2\} \) of \( X \) and \( \{Y_{1,\epsilon}, Y_{2,\epsilon}\} \) of \( Y_\epsilon \). To partition \( X \) we choose \( X_{k,\epsilon} := T^{-1}Y_{k,\epsilon} \cap X \); one may check that in fact \( X_k := X_{k,0} = X_{k,\epsilon} \) for all \( 0 \leq \epsilon \leq \epsilon^* \) and that \( \{X_k\}_{k=1}^{k_{\epsilon}} \) partitions \( X \).
In what follows, we will use the following two inequalities: For \( W \subset Y \),
\[
\mathcal{D}_{W(x)}^* 1_W(x) = \frac{1}{\ell(B_x(0))} \int_{Y_x} 1_{B_x(\epsilon)}(y) 1_W(y) \, dy
\]
\[
= \frac{\ell(Y_x \cap B_x(\epsilon) \cap W)}{\ell(B_x(\epsilon))} \geq 1_{W}(x),
\]
and for \( V \subset X \),
\[
\mathcal{D}_{V(x)}^* 1_V(x) = \frac{1}{\ell(B_x(0))} \int_{X_x} 1_{B_x(\epsilon)}(y) 1_V(y) \, dy
\]
\[
= \frac{\ell(X_x \cap B_x(\epsilon) \cap V)}{\ell(B_x(\epsilon))} \geq 1_{V}(x).
\]
Now, for \( k = 1 \) (and identically for \( k = 2 \)) and \( 0 < \epsilon \leq \epsilon^* \) we have
\[
(L_x 1_{Y_1,\epsilon}, 1_{Y_1,\epsilon})_{\nu} = (L_x 1_{T^{-1}Y_1,\epsilon} \cdot 1_{Y_1,\epsilon})_{\mu}
\]
\[
\geq (1_{T^{-1}Y_1,\epsilon}, D_{X_1,\epsilon} \cdot D_{X_1,\epsilon}^* 1_{Y_1,\epsilon})_{\mu}
\]
\[
\geq (1_{T^{-1}Y_1,\epsilon}, D_{X_1,\epsilon}^* 1_{T^{-1}Y_1,\epsilon})_{\mu}
\]
\[
= \mu(T^{-1}Y_1,\epsilon).
\]
Since \( \mu(X_1) \leq \mu(T^{-1}Y_1,\epsilon) \) we have
\[
\frac{(L_x 1_{X_1}, 1_{Y_1,\epsilon})_{\nu}}{\mu(X_1)} \geq \frac{\mu(T^{-1}Y_1,\epsilon)}{\mu(T^{-1}Y_1,\epsilon)},
\]
and similarly for \( X_2, Y_2,\epsilon \), and the result follows. \( \square \)

**Lemma 7.** Using the notation of Lemma 6, suppose that \( Y_0 \) is a smooth manifold with smooth boundary, \( T^{-1} \) is smooth, and \( \mu \) is absolutely continuous with density bounded above and below. Given some \( \epsilon^* > 0 \) one may choose \( Y_{1,\epsilon^*}, Y_{2,\epsilon^*} \) so that there exists \( 0 < C < \infty \) with
\[
2 - C \epsilon \leq \frac{\mu(T^{-1}Y_{1,\epsilon})}{\mu(T^{-1}Y_{1,\epsilon})} + \frac{\mu(T^{-1}Y_{2,\epsilon})}{\mu(T^{-1}Y_{2,\epsilon})}
\]
for all \( 0 < \epsilon < \epsilon^* \).

**Proof.** In order to achieve a tight lower bound for \( \sigma_{2,\epsilon} \), we should choose \( Y_{1,\epsilon^*}, Y_{2,\epsilon^*} \) judiciously. Under the hypotheses, \( Y_{\epsilon} \) has smooth boundary. Choose a partition \([Y_{k,\epsilon^*}]_{k=1,2}\) of two simply connected sets with non-empty interior so that each element has smooth boundary, and so that \( \mu(T^{-1}Y_{k,\epsilon^*}) = 1/2 \). Now for all \( 0 < \epsilon < \epsilon^* \), for small enough \( \epsilon \), for each \( k = 1, 2 \), one has \( \mu(T^{-1}Y_{k,\epsilon^*})/\mu(T^{-1}Y_{k,\epsilon^*}) \geq 1 - C \epsilon \), where the constant \( C \) depends on \( \mu \), \( T \), and the curvature of the boundaries of the \( Y_{k,\epsilon^*} \). \( \square \)

**Remark 3.** Note that in order to optimise (minimise) the constant in the \( C \) above, one chooses \( Y_{k,\epsilon^*}, k = 1, 2 \) so that the co-dimension 1 volume of the shared boundary of \( Y_{1,\epsilon^*} \) and \( Y_{2,\epsilon^*} \) is small and the co-dimension 1 volume of the shared boundary of \( T^{-1}Y_{1,\epsilon^*} \) and \( T^{-1}Y_{2,\epsilon^*} \) is small. We see here the core of the reason for the “double” diffusion (pre- and post- \( T \)-dynamics) in the definition of \( L_\epsilon \). If we defined \( L_\epsilon \) as \( D_{\epsilon}^* \cdot D_{\epsilon} \) (resp. \( D_{\epsilon}^* \cdot D_{\epsilon}^* \)), then we would choose \( Y_{k,\epsilon^*} \) with small shared boundary (resp. choose \( T^{-1}Y_{k,\epsilon^*} \) with small shared boundary), and not care about the boundaries of \( T^{-1}Y_{k,\epsilon^*} \) (resp. \( Y_{k,\epsilon^*} \)). By defining \( L_\epsilon = D_{\epsilon}^* \cdot D_{\epsilon}^* \), we require small shared boundaries for both initial and final time partitions, and guide the 2nd singular vectors (and the subsequently constructed coherent sets) towards having smooth boundaries at initial and final times.

**Corollary 4.** In the setting of Lemma 6 and under the hypotheses of Lemma 7, one has \( 1 - \sigma_{2,\epsilon} \leq C \epsilon \) for \( 0 < \epsilon < \epsilon^* \).

### 6. Numerical example

We now numerically investigate the constructions of the previous sections. Our numerical example is a quasi-periodically forced flow system representing an idealised stratospheric flow in the northern or southern hemisphere (see [48]), defined by
\[
\begin{align*}
\frac{dx}{dt} &= -\frac{\partial \psi}{\partial y} \\
\frac{dy}{dt} &= \frac{\partial \psi}{\partial x} \\
\frac{dz}{dt} &= \text{some function of } x, y
\end{align*}
\]
with streamfunction

\[ \Psi(x, y, t) = c_1 y - U_0 L \tanh(y/L) + A_1 U_0 \text{sech}^2(y/L) \cos(k_1 x) + A_2 U_0 \text{sech}^2(y/L) \cos(k_2 x - s_2 t) \\
+ A_3 U_0 \text{sech}^3(y/L) \cos(k_3 x - s_1 t). \]

We use the parameter values as in [49], i.e. \( U_0 = 5.41, A_1 = 0.075, A_2 = 0.4, A_3 = 0.2, L = 1.770, c_2/U_0 = 0.205, c_3/U_0 = 0.7, k_1 = 2/r_e, k_2 = 4/r_e, k_3 = 6/r_e \), where \( r_e = 6.371 \) as well as \( s_2 = k_2 (L - c_2) \), \( s_1 = s_2 (1 + \sqrt{5})/2 \), where we have dropped the physical units for brevity. We seek coherent sets for the finite-time duration from \( t = 10 \) to \( t = 20 \). As the flow is area-preserving, we set \( h_\mu \) to be constant. Rypina et al. [48] show that there is a time-varying jet core oscillating in a band around \( y = 0 \). The parameters studied in [48] were chosen so that the jet core formed a complete transport barrier between the two Rossby wave regimes above and below it. In [27] some of the parameters were modified to remove the jet core band, allowing a small amount of transport between the two Rossby wave regimes. Thus we expect that there are two coherent sets; one above the removed jet core, and one below it. We demonstrate that we can numerically find these coherent sets, and illustrate the effect of diffusion.

In order to carry out a numerical investigation we require a finite-rank approximation of \( \mathcal{L}_\epsilon \). To construct such an approximation, we use the numerical method of [27], whereby the domains \( X \) and \( Y_\epsilon \) are partitioned into boxes and an estimate of \( \mathcal{P}_\epsilon \), and then \( \mathcal{L}_\epsilon \) is obtained. We carry out two experiments at differing \( \epsilon \) values.
6.1. Pure advection with implicit numerical diffusion

Firstly, we directly use the technique from [27]. We partition \([0, 20] \times [-2.5, 2.5]\) into \(2^{15}\) boxes, leading to boxes of radius \(0.0391 \times 0.0195\). We obtain finite rank estimates of \(\mathcal{P}\) and \(\mathcal{L}_0\), which are \(32\,768 \times 39\,465\) sparse non-negative matrices \(P\) and \(L\), respectively; these matrices were produced using 400 test points per box (see [27] for details). The leading singular value of \(L\) is 1, and the second singular value is \(\sigma_2 = 0.9969\). The discretisation procedure used to estimate \(\mathcal{L}_0\) leads to a "numerical diffusion" so that one effectively estimates \(\mathcal{L}_e\) with \(\varepsilon \approx 0.0391\); this is the reason for the spectral gap appearing in the numerical estimate of \(\mathcal{L}_0\). An estimate of \(h_{\varepsilon, \mu}\), produced as \(\mathcal{P}_\varepsilon\), is shown in Fig. 1 (left). The second left (resp. right) singular vector is shown in the left image of Fig. 2 (resp. Fig. 3). Note that the value of the singular vectors in Figs. 2 and 3 is predominantly in the vicinity of \(\pm \varepsilon\), and that there is a clear separation into two coherent sets, consistent with the known facts about transport in this system. An optimal level set thresholding of the second left (resp. right) singular vector is shown in the left image of Fig. 4 (resp. Fig. 5); see Remark 1 for details. The maximal value in (7) obtained from this level set thresholding procedure was computed to be 1.9854, consistent with (and close to) the upper bound of 1.9969 given by (8). The \(\mu\)-measures of the red (resp. blue) sets are approximately 0.5064 (resp. 0.4936).

6.2. Advection with explicit diffusion

Secondly, we explicitly apply a diffusion of radius \(\epsilon = 0.1\) before and after the action of \(\mathcal{P}_\varepsilon\). Numerically, this is achieved by applying a mask of 37 points in an \(\varepsilon\)-ball about each of the 36 test points per box, before and after the action of the deterministic dynamics. This results in a \(20\,480 \times 29\,071\) matrices \(P_{\epsilon}\) and \(L_{\epsilon}\) with the latter having a leading singular value of 1 and second singular value \(\sigma_2 = 0.9793\). An estimate of \(h_{\varepsilon, \mu}\), produced as \(\mathcal{P}_{\epsilon, \mu}\), is shown in Fig. 1 (left). The second left (resp. right) singular vector is shown in the right image of Fig. 2 (resp. Fig. 3). An optimal level set thresholding of the second left (resp. right) singular vector is shown in the right image of Fig. 4 (resp. Fig. 5); see Remark 1 for details. The maximal value in (7) obtained from this level set thresholding procedure was computed to be 1.9544, consistent with the upper bound of 1.9793 given by (8). The \(\mu\)-measures of the red (resp. blue) sets are approximately 0.5031 (resp. 0.4969).

Furthermore, the increasing regularity of the left and right singular vectors with increasing \(\epsilon\), as predicted by Propositions 4 and 5 is evident in Figs. 2 and 3.

7. Conclusion

In this paper we created a formal transfer operator framework for identifying finite-time maximally coherent sets in time-dependent dynamical systems. Maximally coherent sets (one at an initial time and one at a final time) are those sets for which there is minimal "mass" leakage over the finite-time duration. These structures arise naturally in (for example) geophysical flows, and provide valuable information for transport and mixing processes. Coherent sets are good transporters of mass precisely because they move about with minimal dispersion (for example, oceanic eddies are good transporters of water that is warmer/cooler/saltier than the surrounding water), however these important features are frequently invisible to observations.

The new operator framework described here extended the matrix-based numerical approach of [27], enabling a formal posing and solving of a natural transport minimisation problem in an \(L^2\) functional setting. The analysis proceeded from general transfer or Markov operator constructions through to small random perturbations of smooth deterministic dynamics. The small random perturbations, explicitly effected by diffusion operators, were key to finding maximally coherent sets that are distinguished by their robustness to noise or diffusion.

At a general level, we proved the frame-invariance or objectivity of the method. At the level of diffeomorphisms, we studied the scaling of an upper bound for the spectral gap to the second singular value (an indicator of the strength of coherence) with diffusion amplitude, and we developed upper bounds for the regularity of the singular vectors (which control the geometric regularity of the coherent sets) as functions of diffusion amplitude. A numerical study of the dependence of the coherent sets on the finite-time duration and the diffusion radius has recently been undertaken in [50].

Acknowledgements

The author is grateful to Kathrin Padberg-Gehle for feedback on an earlier draft, and to George Haller for posing the question of frame-invariance. The author received partial support from the Australian Research Council (DP110100068).

Appendix. Proofs

A.1. Boundedness and compactness of \(\mathcal{L}_e\) and \(\mathcal{L}_e^\ast\)

The following lemma is a straightforward modification of Proposition II.1.6 in Conway [35], which we include for completeness.

**Lemma 8.** Let \((X, \mu)\) and \((Y, \nu)\) be measure spaces and suppose \(k : X \times Y \rightarrow \mathbb{R}\) is measurable, and that

\[
\int_X |k(x, y)| \, d\mu(x) \leq c_1 \quad \text{for } \nu\text{-almost all } y \in Y, \tag{23}
\]

\[
\int_Y |k(x, y)| \, d\nu(y) \leq c_2 \quad \text{for } \mu\text{-almost all } x \in X. \tag{24}
\]

If \(K : L^2(X, \mu) \rightarrow L^2(Y, \nu)\) is defined by \(Kf(x) = \int_Y k(x, y)f(y) \, d\mu(y)\) then \(K\) is a bounded linear operator and \(\|K\| \leq (c_1c_2)^{1/2}\).
Proof.

\[ |Kf(y)| \leq \int_X |k(x, y) | |f(x)| \, d\mu(x) \]
\[ = \int_X |k(x, y)|^{1/2} |k(x, y)|^{1/2} |f(x)| \, d\mu(x) \]
\[ \leq \left( \int_X |k(x, y)| \, d\mu(x) \right)^{1/2} \left( \int_X |f(x)|^2 \, d\mu(x) \right)^{1/2} \]
\[ = c_1^{1/2} \left( \int_X |k(x, y)| \, d\mu(x) \right)^{1/2} . \]

Thus,

\[ \|Kf\|_{L^2(v)}^2 = \int_Y |Kf(y)|^2 \, dv(y) \]
\[ \leq \int_Y c_1 \int_X |k(x, y)| |f(x)|^2 \, d\mu(x) \, dv(y) \]
\[ = \int_X |f(x)|^2 c_1 \int_Y |k(x, y)| \, dv(y) \, d\mu(x) \]
\[ = \int_X c_1 c_2 |f(x)|^2 \, d\mu(x) = c_1 c_2 \|f\|_{L^2(\mu)}^2 . \]

Proof of Lemma 1. We follow the proofs of Proposition II.4.7 and Lemma II.4.8 [35], generalising from \(L^2(X \times X, \mu \times \mu)\) to \(L^2(X \times Y, \mu \times \nu)\). Inner products and norms on \(L^2(X, \mu)\) and \(L^2(Y, \nu)\) will have subscripts \(\mu\) and \(\nu\) respectively; no subscript means an inner product or norm on \(L^2(X \times Y, \mu \times \nu)\). The assumptions on the integrability of \(k\) guarantee that \(\mathcal{L} : L^2(X, \mu) \to L^2(Y, \nu)\) and its dual are bounded linear operators: \(\|\mathcal{L} f\|_{L^2(\nu)}^2 = \int \int k(x, y) f(x) \, d\mu(x) \, dv(y) \leq \int \int k(x, y)^2 \, d\mu(x) \cdot \int f(x)^2 \, d\mu(x) \, dv(y) = \|k\|^2 \|f\|_{L^2(\mu)}^2\); similarly for \(\mathcal{L}^*\).

Let \(\{e_i\}\) be an orthonormal basis of \(L^2(X, \mu)\) and \(\{f_j\}\) be an orthonormal basis of \(L^2(Y, \nu)\). Define \(\phi_{ij}(x, y) = \tilde{e}_i(x) \tilde{f}_j(y)\). It is easy to check that \(\{\phi_{ij}\}\) is an orthonormal set in \(L^2(X \times Y, \mu \times \nu)\):

\[ \langle \phi_{ij}, \phi_{kl} \rangle = \int \tilde{e}_i(x) \tilde{f}_j(y) \cdot e_k(x) \tilde{f}_l(y) \, d\mu(x) \, dv(y) \]
\[ = \int \left( \int \tilde{e}_i(x) e_k(x) \, d\mu(x) \right) \tilde{f}_j(y) \tilde{f}_l(y) \, dv(y) \]
\[ = \delta_{ik} \delta_{jl} . \]

One also has

\[ \langle k, \phi_{ij} \rangle = \int k(x, y) \tilde{\phi}_{ij}(x, y) \, d\mu(x) \, dv(y) \]
\[ = \int k(x, y) e_i(x) \tilde{f}_j(y) \, d\mu(x) \, dv(y) \]
\[ = \langle \mathcal{L} e_i, f_j \rangle . \]

Therefore \(\|k\|^2 \geq \sum_{i,j} |\langle k, \phi_{ij} \rangle|^2 = \sum_{i,j} |\langle \mathcal{L} e_i, f_j \rangle|^2\). Since \(k \in L^2(X \times X, \mu \times \nu)\) there are at most a countable number of \(i, j\) such that \(\langle k, \phi_{ij} \rangle \neq 0\); denote these by \(\{\psi_m\}\) and note that \(\langle k, \phi_{ij} \rangle = 0\) unless \(\phi_{ij} \in \{\psi_m\}\). Let \(\psi_m(x, y) = \tilde{e}_i(x) f_m(y)\), let \(P_n : L^2(X, \mu) \to L^2(X, \mu)\) be the orthogonal projection onto \(\{e_i : 1 \leq i \leq n\}\) and let \(Q_m : L^2(Y, \nu) \to L^2(Y, \nu)\) be the orthogonal projection onto \(\{f_m : 1 \leq m \leq n\}\). Define \(\mathcal{L}_n = \mathcal{L} P_n + Q_m \mathcal{L} - Q_m \mathcal{L} P_n\), a finite-rank operator. We will show that \(\|\mathcal{L} - \mathcal{L}_n\| \to 0\) as \(n \to \infty\), showing \(\mathcal{L}\) is compact.

Let \(g \in L^2(X, \mu)\) with \(\|g\|_{L^2(\mu)} \leq 1\), and write \(g = \sum_i a_i e_i\). One has

\[ \|\mathcal{L} g - \mathcal{L}_n g\|_{L^2(\nu)}^2 = \sum_j |\langle \mathcal{L} g - \mathcal{L}_n g, f_j \rangle|^2 \]
\[ = \sum_j \left| \sum_i a_i \langle \mathcal{L} - \mathcal{L}_n e_i, f_j \rangle \right|^2 \]
\[ = \sum_m \left| \sum_i a_i \langle \mathcal{L} - \mathcal{L}_n e_i, f_m \rangle \right|^2 \]
\[ \leq \sum_m \left( \sum_i |a_i|^2 \right) \left( \sum_i |\langle \mathcal{L} - \mathcal{L}_n e_i, f_m \rangle|^2 \right) \]
\[ = \|g\|_{L^2(\mu)}^2 \sum_m \sum_i |\langle \mathcal{L} e_i, f_m \rangle|^2 - \|\mathcal{L}_n g\|_{L^2(\nu)}^2 \]
The penultimate equality holds since when \( l \leq n \), \( P_n e_l = e_l \) and when \( m \leq n \), \( Q_n f_m = f_m \); thus if either \( l \leq n \) or \( m \leq n \), the entire expression is zero. Finally, since \( \sum_{m=n}^{\infty} |(k, \phi_m)|^2 \leq ||k|| < \infty \), given \( \epsilon > 0 \), choose \( n \) large enough so that this sum is less than \( \epsilon^2 \). Since \( \mathcal{L} \) is compact, so is \( \mathcal{L}^* \). □

A.2. Simplicity of the leading singular value of \( \mathcal{L}_\epsilon \)

For the two lemmas in this section we assume that \( X = Y = M \subset \mathbb{R}^d \) is compact, that \( T \) is a diffeomorphism, that \( \mu \) is absolutely continuous with positive density, and that \( \alpha_{X, \epsilon} = \alpha_{Y, \epsilon} = 1_{B_{R_0}(0)}/\ell(B_{r_0}(0)) \).

Lemma 9. \( k_e(x, y) \) is bounded.

Proof. Recall

\[
\kappa_e(x, y) = \frac{\ell(B_r(x) \cap T^{-1}B_r(y))}{\int_X \ell(B_r(x) \cap T^{-1}B_r(y)) \, d\mu(x)}.
\]

Clearly the numerator is uniformly bounded above by \( \ell(B_{r_0}(0)) \); we now show that the denominator is uniformly bounded below. One has \( \ell(T^{-1}B_r) \geq \ell(B_{r_0}(0))/\sup_{p \in M} |\det DT(x)| \). Define \( \delta^* = \sup(\delta : \text{for all } y \in M, 3\epsilon = z(y) \text{ such that } B_z(x) \subset T^{-1}B_r(y)) \). Since \( |\det DT| \) is uniformly bounded above, \( \delta^* > 0 \). Then \( \int \ell(B_r(x) \cap T^{-1}B_r(y)) \, d\mu(x) \geq \delta^* \cdot \ell(B_{r/2}(0)) \) for all \( y \in M \), where \( D^* = \inf_{p \in M} \mu([-T^{-1}B_{r/2}(0) : T^{-1}B_{r/2}(0)]) \). Since \( \mu \) is absolutely continuous with positive density, \( D^* > 0 \). □

Lemma 10. There exists a \( q > 0 \) and an \( A \subset X \) with \( \mu(A) > 0 \) such that \( \kappa_q(x, y) > 0 \) for \( \mu \)-a.a. \( x \in X \) and \( y \in A \).

Proof. Consider the action of \( A_e = \mathcal{L}_e^* \mathcal{L}_e \) on a distribution \( \delta \). As \( \mathcal{L}_e(f) = D\mathcal{U}_e \circ D\mathcal{X}_e \circ D\mathcal{Y}(f)/D\mathcal{U}_e \circ D\mathcal{X}_e \circ D\mathcal{Y}(f) \), the initial application of \( \mathcal{D}_e \) produces a function with support \( B_r(x) \) (in fact, \( 1_{B_{r_0}} \)). The application of \( \mathcal{P} \) now creates a function with support \( T(B_r(x)) \) (because \( |\det DT| \) is uniformly bounded above) and finally the application of \( D\mathcal{U}_e \) produces a support of \( B_{r_0}(T(B_r(x))) \). Now applying \( \mathcal{L}_e^* \) again applies \( \mathcal{D}_e \), then \( T^{-1} \), then \( D\mathcal{U}_e \), producing a function with support \( S = B_r(T^{-1}(B_{r_0}(T(B_r(x)))))) \). Clearly \( B_{r/2}(x) \subset S \). At each iteration of \( A_e \) the support expands by at least \( \epsilon/2 \). As \( X \) is bounded, eventually the support fills \( X \) after \( q \) iterations for some \( q \). Thus \( \kappa_q(x, y) > 0 \) for \( \mu \)-a.a. \( x \in X \) and \( y \in X \). □

A.3. Regularity of singular vectors of \( \mathcal{L}_e \)

Proof of Lemma 4. One has

\[
\left| D_{X_e} f(y + \gamma) - D_{X_e} f(y) \right| / ||y|| = \left| \int_X (\alpha_{X, \epsilon}(y + \gamma - x) - \alpha_{X, \epsilon}(y - x)) f(x) \, dx \right| / ||y|| \\
\leq \left| f \right|_\infty \cdot \int_X |\alpha_{X, \epsilon}(y + \gamma - x) - \alpha_{X, \epsilon}(y - x)| \, dx / ||y|| \\
= \left| f \right|_\infty \cdot \int_X \frac{1_{B_{r_0}(y + \gamma)} - 1_{B_{r_0}(y)}}{\ell(B_{r_0}(y)) ||y||} \, dx / ||y|| \\
\leq \left| f \right|_\infty / ||y|| \cdot \frac{\ell(B_{r_0}(y + \gamma) \Delta B_{r_0}(y))}{\ell(B_{r_0}(y)) ||y||}.
\]

We show that \( \limsup_{||y|| \to 0} \left| D_{X_e} f(y + \gamma) - D_{X_e} f(y) \right| / ||y|| \leq C \left| f \right|_\infty / \epsilon \), for all \( x \in X \). From this, global Lipschitzness follows by Lemma 16. We now detail the computations for dimensions 1, 2, and 3. The main estimate is for \( \ell(B_{r_0}(y + \gamma) \Delta B_{r_0}(y)) \) in (25), which only depends on \( \gamma \) and \( \epsilon \).

Dimension 1: \( B_r \) is an interval of length \( 2\epsilon \) and clearly \( \ell(B_{r_0}(y + \gamma) \Delta B_{r_0}(y)) = 2|\gamma| \) for \( 0 \leq |\gamma| \leq 2\epsilon \). Thus for \( 0 \leq |\gamma| \leq 2\epsilon \), (25) = \( \left| f \right|_\infty / ||y|| \leq 2\epsilon \), symmetric difference area \( \ell(B_{r_0}(y + \gamma) \Delta B_{r_0}(y)) = A = 2(\pi \epsilon^2 - (2\epsilon^2 \cos^{-1}(||y||/2\epsilon) - (1/2)[1/\sqrt{4\epsilon^2 - ||y||^2}]) \) [51]. To first order in \( ||y|| \), \( \cos^{-1}(||y||/2\epsilon) = \pi/2 - ||y||/2\epsilon + O(||y||^2) \). Thus \( A = 2(\pi \epsilon^2 - \epsilon^2 + \epsilon||y|| + (1/2)\sqrt{4\epsilon^2 - ||y||^2}) + O(||y||^2) \). Thus \( \lim_{||y|| \to 0} \left| f \right|_\infty / ||y|| \leq 2\epsilon \), \( V = 2(4/3 \pi \epsilon^3 - (\pi/12)(4\epsilon^3 - (12\epsilon^3 - 12\epsilon^3 ||y||^2))) \). To first order in \( ||y|| \), \( V = 2(4/3 \pi \epsilon^3 - (\pi/12)(16\epsilon^3 - 12\epsilon^3 ||y||^2)) + O(||y||^2) \). Thus \( \lim_{||y|| \to 0} \left| f \right|_\infty / ||y|| \leq 4/3 \pi \epsilon^3 / \epsilon \).

Dimension 3: \( B_r \) is a circle of radius \( \epsilon \). For \( 0 \leq ||y|| \leq 2\epsilon \), the symmetric difference area \( \ell(B_{r_0}(y + \gamma) \Delta B_{r_0}(y)) = V = 2(4\pi \epsilon^3 - (\pi/12)(4\epsilon^3 + (2\epsilon - ||y||^2))) \) [52]. To first order in \( ||y|| \), \( V = 2(4\pi \epsilon^3 - (\pi/12)(16\epsilon^3 - 12\epsilon^3 ||y||^2)) + O(||y||^2) \). Thus \( \lim_{||y|| \to 0} \left| f \right|_\infty / ||y|| \leq 2\pi \epsilon^3 / \epsilon \).

By Lemma 16 the constants found above are also global Lipschitz constants. □
Proof of Lemma 5. Let $x = y + \gamma$. One has

$$
\left| \mathcal{D}_{x,e} f (y + \gamma) - \mathcal{D}_{x,e} f (y) \right| = \left| \int_X (\alpha_{x,e} (y + \gamma - x) - \alpha_{x,e} (y - x)) f (x) \, dx \right|
$$

$$
\leq \| f \|_{L^2 (\mathcal{O})} \left( \int_X (\alpha_{x,e} (y + \gamma - x) - \alpha_{x,e} (y - x))^2 \, dx \right)^{1/2}
$$

$$
= \| f \|_{L^2 (\mathcal{O})} \left( \int_X (1_{B_\gamma (y + \gamma)} - 1_{B_\gamma (y)})^2 \, dx \right)^{1/2} / \ell (B_\gamma (y))
$$

$$
\leq \| f \|_{L^2 (\mathcal{O})} \left( \ell (B_\gamma (y + \gamma) \Delta B_\gamma (y)) \right)^{1/2} / \ell (B_\gamma (y)).
$$

(26)

We now detail the computations for dimensions 1, 2, and 3. The main estimate is for $\ell (B_\gamma (y + \gamma) \Delta B_\gamma (y))$ in (26), which only depends on $\epsilon$ and $\gamma$.

**Dimension 1:** $B_\gamma$ is an interval of length $2\epsilon$ and clearly

$$
\ell (B_\gamma (y + \gamma) \Delta B_\gamma (y)) = \begin{cases} 
2 \| y \|, & \| y \| \leq 2\epsilon; \\
2\epsilon, & \| y \| > 2\epsilon.
\end{cases}
$$

Thus

$$
\left( \ell (B_\gamma (y + \gamma) \Delta B_\gamma (y)) \right)^{1/2} / \ell (B_\gamma (y)) = \begin{cases} 
\sqrt{2}\| y \| / 2\epsilon, & \| y \| \leq \epsilon; \\
2\epsilon / 2\epsilon = 1, & \| y \| > \epsilon,
\end{cases}
$$

and one has $|\mathcal{D}_{x,e} f (x) - \mathcal{D}_{x,e} f (y)| \leq \| f \|_{L^2 (\mathcal{O})} \cdot \sqrt{1 / 2} \| x - y \|^{1/2}$.

**Dimension 2:** $B_\gamma$ is a disc of radius $\epsilon$. For $0 \leq \| y \| \leq 2\epsilon$, the symmetric difference area $\ell (B_\gamma (y + \gamma) \Delta B_\gamma (y))$ is $A = 2(\pi \epsilon^2 - (2\epsilon^2 \cos^{-1}(\| y \| / 2\epsilon)) - (1/2)\| y \| \sqrt{4\epsilon^2 - \| y \|^2})$ [51]. Thus,

$$
\ell (B_\gamma (y + \gamma) \Delta B_\gamma (y)) = \begin{cases} 
2(\pi \epsilon^2 - (2\epsilon^2 \cos^{-1}(\| y \| / 2\epsilon)) - (1/2)\| y \| \sqrt{4\epsilon^2 - \| y \|^2}), & \| y \| \leq 2\epsilon; \\
2\pi \epsilon^2, & \| y \| > 2\epsilon.
\end{cases}
$$

Then

$$
\left( \ell (B_\gamma (y + \gamma) \Delta B_\gamma (y)) \right)^{1/2} / \ell (B_\gamma (y)) = \begin{cases} 
\sqrt{2}(\pi \epsilon^2 - (2\epsilon^2 \cos^{-1}(\| y \| / 2\epsilon)) - (1/2)\| y \| \sqrt{4\epsilon^2 - \| y \|^2}) / (\pi \epsilon^2), & \| y \| \leq 2\epsilon; \\
\sqrt{2}\pi \epsilon^2 / (\pi \epsilon^2), & \| y \| > 2\epsilon.
\end{cases}
$$

and one has

$$
|\mathcal{D}_{x,e} f (x) - \mathcal{D}_{x,e} f (y)| \leq \| f \|_{L^2 (\mathcal{O})} \cdot \sqrt{2/\pi} \pi / \epsilon^{1/2} \| x - y \|^{1/2}
$$

for all $x, y \in X$.

**Dimension 3:** $B_\gamma$ is a disc of radius $\epsilon$. For $0 \leq \| y \| \leq 2\epsilon$, the symmetric difference volume $\ell (B_\gamma (y + \gamma) \Delta B_\gamma (y))$ is [52]

$$
V = 2(4/3\pi \epsilon^3 - (\pi / 12(4\epsilon + \| y \|))(2\epsilon - \| y \|^2))
$$

$$
= 8\pi \epsilon^3 / 3 - (\pi / 6)(4\epsilon + \| y \|)(4\epsilon^2 - 4\| y \| \epsilon + \| y \|^2))
$$

$$
= 8\pi \epsilon^3 / 3 - (\pi / 6)(16\epsilon^2 - 12\epsilon \| y \|^2 + \| y \|^3)
$$

$$
= 2\pi \| y \| \epsilon^2 - \pi \| y \|^3 / 6.
$$

Thus

$$
\ell (B_\gamma (y + \gamma) \Delta B_\gamma (y)) = \begin{cases} 
(2\pi \| y \| \epsilon^2 - \pi \| y \|^3 / 6), & \| y \| \leq 2\epsilon; \\
8\pi \epsilon^3 / 3, & \| y \| > 2\epsilon.
\end{cases}
$$

Thus

$$
\left( \ell (B_\gamma (y + \gamma) \Delta B_\gamma (y)) \right)^{1/2} / \ell (B_\gamma (y)) = \begin{cases} 
\sqrt{(2\pi \| y \| \epsilon^2 - \pi \| y \|^3 / 6)} / (4\pi \epsilon^2), & \| y \| \leq 2\epsilon; \\
\sqrt{3/2\pi} \sqrt{\epsilon / \epsilon^2}, & \| y \| > 2\epsilon.
\end{cases}
$$

Then

$$
\left( \ell (B_\gamma (y + \gamma) \Delta B_\gamma (y)) \right)^{1/2} / \ell (B_\gamma (y)) = \begin{cases} 
\sqrt{1/\pi \sqrt{9\| y \|^2 / 8 - 3\| y \|^3 / 32\epsilon^2}} \cdot 1 / \epsilon^2, & \| y \| \leq 2\epsilon; \\
\sqrt{3/2\pi} \sqrt{\epsilon / \epsilon^2}, & \| y \| > 2\epsilon.
\end{cases}
$$

and one has

$$
|\mathcal{D}_{x,e} f (x) - \mathcal{D}_{x,e} f (y)| \leq \| f \|_{L^2 (\mathcal{O})} \cdot (3/2) \sqrt{1/2\pi} \cdot 1 / \epsilon^2 \| x - y \|^{1/2}
$$

for all $x, y \in X$. □
Setting notation for the following lemmas, $T : M \to M$ is a diffeomorphism with $0 < A \leq |\det DT| \leq B < \infty$ and $0 < L \leq h_\mu \leq U < \infty$. Moreover $X = Y = Y_\epsilon = M$.

**Lemma 11.** $\|f \cdot h_\mu\|_{L^2(\epsilon)} \leq U^{1/2}\|f\|_{L^2(\mu)}$.

**Proof.**

\[\|f \cdot h_\mu\|_{L^2(\epsilon)}^2 = \int f^2 \cdot h_\mu^2 \, d\epsilon = \int f^2 \cdot h_\mu \, d\mu \leq U \int f^2 \, d\mu = U\|f\|_{L^2(\mu)}^2. \]

**Lemma 12.** If $f \in L^1(\epsilon)$ satisfies $c \leq f \leq d$ then $c \leq D_X f \leq d$.

**Proof.** This follows since $D_X \cdot f$ is an averaging operator.

**Lemma 13.** $\|D_X \cdot f\|_{L^2(\epsilon)} \leq \|f\|_{L^2(\epsilon)}$.

**Proof.** This follows directly from Lemma 8, putting $k(x, y) = 1_{B_\epsilon}(y) / \epsilon(B_\epsilon(0))$ one has $\int k(x, y) \, dx = 1$ for all $y \in X_\epsilon$ and $\int_{X_\epsilon} k(x, y) \, dy \leq 1$ for all $x \in X$.

**Lemma 14.** $\|Pf\|_{L^2(\epsilon)} \leq 1/A^{1/2}\|f\|_{L^2(\epsilon)}$.

**Proof.**

\[
\int (Pf)^2 \, d\epsilon = \int \left( \frac{f \circ T^{-1}}{|\det DT \circ T^{-1}|} \right)^2 \, d\epsilon \\
= \int \left( \frac{f}{|\det DT|} \right)^2 \cdot \frac{1}{|\det DT^{-1} \circ T|} \, d\epsilon \quad \text{by change of variables under } T \\
= \int \frac{f^2}{|\det DT|} \, d\epsilon \\
\leq 1/A \int f^2 \, d\epsilon. \]

**Proof of Proposition 4.** 1. Let $H_{1/2}(f)$ denote the $1/2$-Hölder exponent for $f$. Since $\mathcal{L}_f = D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} (f \cdot h_\mu)$, we have

\[H_{1/2}(\mathcal{L}_f) \leq H_{1/2}(D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} (f \cdot h_\mu)) : \frac{1}{D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} h_\mu} + H_{1/2} \left( \frac{1}{D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} h_\mu} \right) : |D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} (f \cdot h_\mu)|_{\infty}.\]

By Lemmas 11, 13 and 14 we have

\[\|Pf\|_{L^2(\epsilon)} \leq \|f\|_{L^2(\epsilon)}\|f\|_{L^2(\mu)} \leq (U/A)^{1/2} \|f\|_{L^2(\mu)} \]

and so applying Lemma 5 we have

\[H_{1/2}(D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} (f \cdot h_\mu)) \leq (U/A)^{1/2} \|f\|_{L^2(\mu)} C(d)/\epsilon^{(d+1)/2} \]

where $C(d)$ is from Lemma 5. Moreover, since $h_\mu = D_X h_\mu$ is bounded below and above by $L$ and $U$, respectively (by Lemma 12) we have $h_\mu = P h_\mu$ is bounded below and above by $L/B$ and $U/A$ respectively. Finally, applying Lemma 12 again, we have $A/U \leq 1/(D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} h_\mu) \leq B/L$.

For the second term, note

\[H_{1/2} \left( \frac{1}{D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} h_\mu} \right) \leq \frac{H_{1/2}(D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon})}{(\min h_{Y_\epsilon})^2} \leq \frac{C(d)}{\epsilon^{(d+1)/2}} \]

by Lemmas 5 and 12

\[\leq \frac{(\max h_{Y_\epsilon})^{1/2}}{(\min h_{Y_\epsilon})^{1/2}} \frac{C(d)}{\epsilon^{(d+1)/2}} \]

as $\|h_{Y_\epsilon}\|_{L^1(\epsilon)} = 1$ we have $\|h_{Y_\epsilon}\|_{L^2(\epsilon)} \leq (\max h_{Y_\epsilon})^{1/2}$

\[\leq \frac{(U/A)^{1/2}}{(L/B)^{1/2}} \cdot C(d)/\epsilon^{(d+1)/2}. \]

Finally, to bound $|D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} (f \cdot h_\mu)|_{\infty}$ we note that since $\int f \, d\mu = 0$ we have $\int D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} (f \cdot h_\mu) \, d\epsilon = 0$ since $D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon}$ and $P$ preserve $\epsilon$-integrals. Thus, $|D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} (f \cdot h_\mu)|_{\infty} \leq H_{1/2}(D_{Y_\epsilon} \cdot P \cdot D_{X_\epsilon} (f \cdot h_\mu)|_{\text{diam}(M)})^{1/2}$. Putting this all together we have

\[H_{1/2}(\mathcal{L}_f) \leq (U/A)^{1/2} \|f\|_{L^2(\epsilon)} \|f\|_{L^2(\mu)} C(d)/\epsilon^{(d+1)/2} + (U/A)^{1/2} \cdot C(d)/\epsilon^{(d+1)/2} \cdot (U/A)^{1/2} \|f\|_{L^2(\mu)} C(d)/\epsilon^{(d+1)/2} |\text{diam}(M)|^{1/2}. \]
2. Let \( L(f) \) denote the Lipschitz exponent for \( f \). Since \( L(f) \leq L(D_{\nu_1}^\ast \mathcal{P} D_{\nu_2} f (\cdot, h_\mu) / D_{\nu_1}^\ast \mathcal{P} D_{\nu_2} h_\mu \) we have

\[
L(L(f)) \leq L(D_{\nu_1}^\ast \mathcal{P} D_{\nu_2} (f (\cdot, h_\mu)) \cdot \left| \frac{1}{D_{\nu_1}^\ast \mathcal{P} D_{\nu_2} (f (\cdot, h_\mu))} \right| + L \left( \frac{1}{D_{\nu_1}^\ast \mathcal{P} D_{\nu_2} h_\mu} \right) \cdot |D_{\nu_1}^\ast \mathcal{P} D_{\nu_2} (f (\cdot, h_\mu))|_\infty.
\]

We have that \( |D_{\nu_1}^\ast \mathcal{P} D_{\nu_2} (f (\cdot, h_\mu))| \leq (U/A)\|f\|_\infty \) and applying Lemma 4 we have \( L(D_{\nu_1}^\ast \mathcal{P} D_{\nu_2} (f (\cdot, h_\mu)) \leq (U/A)\|f\|_\infty C_4 (d) / \epsilon \) where \( C_4 (d) \) is from Lemma 4. Similarly, we have \( A/U \leq |1/D_{\nu_1}^\ast \mathcal{P} D_{\nu_2} h_\mu| \leq B/L.\)

For the second term, note

\[
\frac{1}{D_{\nu_1}^\ast \mathcal{P} D_{\nu_2} h_\mu} \leq \frac{1}{(U/A)\|f\|_\infty C_4 (d) / \epsilon ^2} \leq \frac{(U/A)^2}{(L/B)^2} \|f\|_\infty (\text{diam}(M)) C_4 (d) / \epsilon ^2. \quad \square
\]

**Lemma 15.** Consider \( \mathcal{K} : L^2(Y_\epsilon, \epsilon) \rightarrow L^2(X_\epsilon, \epsilon) \). One has \( \|\mathcal{K}g\|_{L^2(X_\epsilon, \epsilon)} \leq \|g\|_{L^2(Y_\epsilon, \epsilon)} / \sqrt{\lambda}. \)

**Proof.**

\[
\|\mathcal{K}g\|_{L^2(X_\epsilon, \epsilon)} = \left( \int_{X_\epsilon} (g(\Delta x))^2 \, dx \right)^{1/2}
\]

\[
= \left( \int_{Y_\epsilon} (g(\Delta x))^2 / \text{det}(\Delta^{-1} x) \, dx \right)^{1/2}
\]

\[
\leq (1/\sqrt{\lambda}) \left( \int_{Y_\epsilon} (g(\Delta x))^2 \, dx \right)^{1/2} = \|g\|_{L^2(Y_\epsilon, \epsilon)} / \sqrt{\lambda}. \quad \square
\]

**Proof of Proposition 5.** Note that \( L^\ast_\epsilon = D^\ast_\epsilon \mathcal{K} D^\ast_\epsilon \mathcal{K} \). We first claim the \( L^2 \)-norm of \( \mathcal{K} D^\ast_\epsilon \mathcal{K} \) is \( 1/\sqrt{\lambda}. \) By Lemma 8, putting \( X = Y_\epsilon, \epsilon = Y_\epsilon, \) and \( k(x, y) = \mathbf{1}_{B(y)}(x) / \text{det}(B(y)) \), we have that \( \|D^\ast_\epsilon \mathcal{K} \|_{L^2(Y_\epsilon, \epsilon)} \leq \|g\|_{L^2(Y_\epsilon, \epsilon)}. \) This follows since \( \int_{Y_\epsilon} \mathbf{1}_{B(y)}(x) / \text{det}(B(y)) \, dy \leq 1 \) for \( y \in Y_\epsilon, \) and \( \int_{Y_\epsilon} \mathbf{1}_{B(y)}(x) \, dy = 1 \) for \( x \in Y_\epsilon. \) By Lemma 15 the claim follows.

We now consider \( |D^\ast_\epsilon \mathcal{K} (f (x) - D^\ast_\epsilon \mathcal{K} (f (y)) | \) for \( f \in L^2(Y_\epsilon, \epsilon) \). Because of the symmetry of the kernel \( K(x, y) = \mathbf{1}_{B(y)}(x) / \text{det}(B(y)) \), the only difference between \( D^\ast_\epsilon \) and \( D^\ast_\epsilon \) is the domain of integration (respectively \( X \) and \( X_\epsilon)). \) The bound of Lemma 5 thus also applies to \( D^\ast_\epsilon \).

**Setting:** \( f = K^\ast_\epsilon \mathcal{K} D^\ast_\epsilon \mathcal{K} \) we have

\[
|D^\ast_\epsilon \mathcal{K} D^\ast_\epsilon \mathcal{K} (x) - D^\ast_\epsilon \mathcal{K} D^\ast_\epsilon \mathcal{K} (y) | \leq (C/\sqrt{\lambda}) (1/\sqrt{\lambda}) \|g\|_{L^2(Y_\epsilon, \epsilon)} \cdot \|x - y\|^{1/2}. \quad \square
\]

**Lemma 16.** Let \( X \subset \mathbb{R}^d \) compact and \( F : X \rightarrow \mathbb{R} \). If

\[
\limsup_{|x| \rightarrow 0} \frac{|F(x + y) - F(x)|}{\|y\|} \leq C \quad \text{for all } x \in X,
\]

then \( F \) is globally Lipschitz with Lipschitz constant \( C \).

**Proof.** Note \( \limsup_{|y| \rightarrow 0} \frac{|F(x + y) - F(x)|}{\|y\|} = \lim_{|y| \rightarrow 0} \sup \{\|F(x + y) - F(x)\| / \|y\| : \|y\| < \|y\|, \|y\| \neq 0\}. \) Thus given \( \epsilon > 0 \) there is \( \Delta = \Delta(\epsilon, x) > 0 \) such that \( |F(x + y) - F(x)| / \|y\| < C + \epsilon \) for all \( \|y\| \neq 0 \) with \( \|y\| < \Delta. \) Form an open cover of \( X \) as \( \{B(\Delta(x), x) : x \in X\}. \) By compactness of \( X \) we can find a finite subcover.

Consider arbitrary \( x, y \in X \) and write \( y = x + \gamma y \) (note \( \gamma \) is arbitrary from now on and need not satisfy \( \|\gamma\| < \Delta \)). Draw a line segment from \( x \) to \( x + y; \) this line segment \( [x + \gamma y : 0 \leq \gamma \leq 1] \) passes through a subcollection of open balls in our finite subcover. Denote by \( X \) the finite set of centres of the open balls in our subcollection.

We now trace out the line segment again, identifying a finite sequence of points \( y_k \) and ball centres \( X_k \) as we go. Begin at \( x, \) set \( y_0 = x \) and choose an \( X_0 \) so that \( x \in B(x_0, X_0) \). Now increase \( \theta \) until \( x_0 - (x + \theta y) \| = 0.9\Delta (x, x_0) \). If \( x + \theta y \) lies in some \( B(\Delta(x), x) \), \( x \neq x_0 \), then set \( y_1 = x + \theta y \); otherwise, increase \( \theta \) until \( x + \theta y \) lies in \( B(\Delta(x_0), x) \cap B(\Delta(x_1), x) \). Repeat the procedure: in general, we have a \( y_k \in B(\Delta(x_k), x) \) and we increase \( \gamma \) until \( x_k - (y_k + \gamma y) \| = 0.9\Delta(x, x_0). \) If \( y_k + \gamma y \) lies in some \( B(\Delta(x), x_k) \), \( x_k \neq x_0 \), then set \( y_{k+1} = y_k + \gamma y \); otherwise, increase \( \gamma \) until \( y_k + \gamma y \) lies in \( B(\Delta(x_0) \cap B(\Delta(x_1), x) \). We set \( y_{k+1} = y_k + \gamma y \) and again set \( y_{k+1} = y_k + \gamma y \). Finally we make a step from \( y_{k-1} \) to \( y_k = x + y \) where \( y_{k-1}, x + y \in B(\Delta(x_0), x). \)
By construction, \( y_0 \in B_{\Delta(t,x_0)} \), \( y_1 \in B_{\Delta(t,x_{t-1})} \cap B_{\delta(t,x_{t-1})} \) for \( 1 \leq t \leq L - 1 \), and \( y_L \in B_{\Delta(t,x_L)} \) for \( \ell = 0, \ldots, L - 1 \). We may estimate:

\[
F(x) - F(x + y) = \sum_{\ell=0}^{L-1} F(y_\ell) - F(y_{\ell+1})
\]

\[
\leq \sum_{\ell=0}^{L-1} \|F(y_\ell) - F(y_{\ell+1})\| \leq (C + \varepsilon) \sum_{\ell=0}^{L-1} \|y_\ell - y_{\ell+1}\| = (C + \varepsilon) \|y_0 - y_L\| = (C + \varepsilon) \|x - y\|,
\]

where the final equality follows since the \( y_\ell \) are collinear. As \( x, y \in X \) were arbitrary, the result follows.

References


