QUALITATIVELY ACCURATE SPECTRAL SCHEMES FOR ADVECTION AND TRANSPORT*

HENRY O. JACOBS† AND RAM VASUDEVAN‡

Abstract. The transport and continuum equations exhibit a number of conservation laws. For example, scalar multiplication is conserved by the transport equation, while positivity of probabilities is conserved by the continuum equation. Certain discretization techniques, such as particle based methods, conserve these properties, but converge slower than spectral discretization methods on smooth data. Standard spectral discretization methods, on the other hand, do not conserve the invariants of the transport equation and the continuum equation. This article constructs a novel spectral discretization technique that conserves these important invariants while simultaneously preserving spectral convergence rates. The performance of this proposed method is illustrated on several numerical experiments.

1. Introduction. A topological space is, by the traditional definition, a set of points glued together by a topology. Starting from these notions, one builds the space of continuous functions, then the differentiable functions, and so on. In summary, points come first, then functions come second. The commutative version of the Gelfand-Naimark theorem turns the "points come first" perspective on its head. Specifically, the commutative Gelfand-Naimark theorem states that from any commutative C^*-algebra, $\mathcal{A}$, one could derive a topological space, X, for which $\mathcal{A} \cong C^0(X)$. In other words, it's perfectly feasible to consider the ring of functions first, and the points to be derived later.

While this functions-first perspective has flourished within the world of pure mathematics and mathematical physics, it has not penetrated far into the world of applied mathematics and numerical methods. It is the thesis of this paper, that certain problems in numerical mathematics stand to benefit from this distinct perspective. In particularly, the class of spectral schemes for advection and transport type equations can be improved by fully embracing the spirit of spectral discretization, viewing functions as foundational, and the domain as incidental. This should not be shocking, as spectral schemes take functions as the computational building block, in contrast to grid/particle based methods which take points as the central objects.

Let M be a manifold and let X be a C^2 vector-field on M. This paper is concerned with the following pair of partial differential equations (PDEs) given in local coordinates by:

$$\partial_t f + \sum_{i=1}^{n} X^i \partial_i f = 0 \tag{1}$$

$$\partial_t \rho + \sum_{i=1}^{n} \partial_i (\rho X^i) = 0 \tag{2}$$

for time dependent functions $f(t,\cdot) \in C^0(M)$ and a time dependent density $\rho(t,\cdot) \in L^1(M)$. Equation (1), which is sometimes called the "transport equation," describes how a scalar quantity is transported by the flow of X [34]. Equation (2), which is

*This research was supported by the University of Michigan, Ann Arbor.

†Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48104, USA (hojacobs@umich.edu)

‡Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48104, USA (ramv@umich.edu)

sometimes called the "continuum equation" or "Liouville's equation" describes how a density (e.g. a probability distribution) is transported by the flow of X. Such PDEs arise in a variety of contexts, ranging from mechanics [34] to control theory [20].

The solution to (1) takes the form $f(x;t) = f((\Phi_X^t)^{-1}(x);0) \equiv (\Phi_X^t)_* f(\cdot;0)$ where $\Phi_X^t : M \to M$ is the flow map of X at time t [24, Chapter 18]. This solution implies that (1) exhibits a variety of conservation laws. For example, if f and g are solutions to (1), then $f \cdot g$ and $f + g$ are also solutions. Similarly, the solution to (2) takes the form $\rho(x;t) = \det(D(\Phi_X^t)^{-1}(x))\rho((\Phi_X^t)^{-1}(x);0) := (\Phi_X^t)_*\rho(\cdot;0)$. One can deduce that the L^1-norm of $\rho(x;t)$ is conserved in time [24, Theorem 16.42]. Finally, (2) is the adjoint evolution equation to (1) in the sense that the integral $\langle f, \rho\rangle := \int f\rho$ is constant in time. To see this compute $\frac{d}{dt}\langle f, \rho\rangle = \int(\partial_t f)\rho + f(\partial_t \rho)$. One finds that the final integral vanishes upon substitution of (1) and (2) and applying integration by parts. This conservation law motivates the following definition of qualitative accuracy:

Definition 1. *A numerical method for* (1) *and* (2) *is* qualitatively accurate *if it conserves discrete analogs of scalar multiplication/addition, the L^1-norm and the total mass for densities and the sup-norm for functions.*

1.1. Our primary goal and motivation. Now that we have defined qualitative accuracy for (1) and (2), let us motivate the pursuit of qualitative accuracy. Naively, conservation laws for (1) and (2) are useful in numerics because they reduce the space of feasible solutions. However, more can be said in this case. Consider the following corollary to the Gelfand Naimark theorems [14].

Proposition 2 (follows from Corollary 1.7 of [17]). *Let M be a manifold. If $T : C^\infty(M) \to C^\infty(M)$ is a bounded automorphism preserving products/sums/complex conjugation then there is a unique homeomorphism $\Phi_T : M \to M$ such that $T[a](x) = a(\Phi_T(x))$. That is,* $\mathrm{Diff}(M) \equiv \mathrm{Aut}(C^\infty(M))$ *as a topological groups. Moreover, a linear evolution equation on $C^\infty(M)$ given by $\partial_t f + D[f] = 0$ for some differential operator, D, preserves the algebra of $C^\infty(M)$ if and only if $D[f] = X^i \frac{\partial f}{\partial x^i}$ for some vector-field X. The dual operator is then necessarily of the form $D^*[\rho] = \frac{\partial}{\partial x^i}(\rho X^i)$.*

In summary, (1) and (2) are the generators of all algebra-preserving automorphisms of $C^\infty(M)$. Thus, conservation of sums, products, and norms is more than just a random collection of properties exhibited by (1) and (2). *Conservation of sums, products, and norms is what defines* (1) *and* (2). As a result, it is natural for this to be reflected in a discretization.

A good example of a qualitatively accurate numerical scheme is a particle based method. For a continuous initial condition, f_0, for example, the method of characteristics [11] describes a solution to (1) as a time-dependent function $f(x_t;t) = f_0(x_0)$ where x_t is the solution to $\dot{x} = X(x)$. This suggests using a particle method to solve for f at a discrete set of points [25]. In fact, a particle method would inherit many discrete analogs of the conservation laws of (1), and would therefore be *qualitatively accurate*. For example, given the input $h_0 = f_0 \cdot g_0$, the output of a particle method is identical to the (componentwise) product of the outputs obtained from inputing f_0 and g_0 separately. However, particle methods converge much slower than their spectral counterparts when the function f is highly differentiable [15].

In the case where M is the unit circle, S^1, a spectral method can be obtained by converting (2) to the Fourier domain where it takes the form of an Ordinary Differential Equation (ODE): $\frac{d\hat{\rho}_k}{dt} + 2\pi i k \widehat{X}_{k-\ell}\hat{f}_\ell$ where $\hat{\rho}_k$ and $\widehat{X}_k$ denote the Fourier transforms of ρ and X [33]. In particular, this transformation converts (2) into an ODE on the space of Fourier coefficients. A standard spectral Galerkin discretization

is obtained by series truncation.

Such a numerical method is good for C^k-data, in the sense that the convergence rate, over a fixed finite time $T > 0$, is faster than $\mathcal{O}(N^{-k})$ where N is the order of truncation [15, 16]. In particular, spectral schemes converge faster than particle methods when the initial conditions have some degree of regularity. Unfortunately the spectral algorithm given above is not *qualitatively accurate*, as is demonstrated by several examples in Section 8.

In summary, the current status quo is one wherein we must sacrifice either rapid convergence or qualitative accuracy. The goal of this paper is to change the status quo and achieve both simultaneously. In particular, our goal is *to find a numerical algorithm for* (1) *and* (2) *which is simultaneously stable, spectrally convergent, and qualitatively accurate.*

1.2. Previous work. Within mechanics, spectral methods for the continuum and transport equation are common-place where they are viewed as special cases of first order hyperbolic PDEs [15]. Various Galerkin discretizations of the Koopman operator have been successfully used for generic dynamic systems [6, 28], most notably fluid systems [30] where such discretizations serve as a generalization of dynamic mode decomposition [32]. Dually, Ulam-type discretizations of the Frobenius-Perron operator [23, 35] have been used to find invariant manifolds of systems with uniform Gaussian noise [12, 13]. In continuous time, Petrov-Galerkin discretization of the infinitesimal generator of the Frobenius Perron operator converge in the presence of noise [4] and preserve positivity in a Haar basis [22].

1.3. Main contributions. In this paper we develop numerical schemes for (1) and (2). Our schemes exhibit spectral convergence rates and qualitative accuracy (Section 7). We end the paper by demonstrating these findings in numerical experiments in (Section 8). In experiments we observe our algorithm for (2) to be superior to a standard spectral Galerkin scheme, both in terms of numerical accuracy and qualitative accuracy.

1.4. Notation. Throughout the paper M denotes a smooth compact n-manifold without boundary. The space of continuous functions is denoted $C(M)$ and has a topology induced by the sup-norm, $\|\cdot\|_\infty$ (see [33, 9]). Given a separable Hilbert space $\mathcal{H}$ we denote the Banach-algebra of bounded operators by $B(\mathcal{H})$ and topological group of unitary operators by $U(\mathcal{H})$. The adjoint of an operator $L : \mathcal{H} \to \mathcal{H}$ is denoted by $L^\dagger$. The trace of a trace class operator, L, is denoted by $\mathrm{Tr}(L)$. The commutator bracket for operators A, B on $\mathcal{H}$ is denoted by $[A, B] := A \cdot B - B \cdot A$ (see [9]). Finally, we will use the term "canonical" with respect to the category of smooth manifolds. For example, we could not say that a vector-field is divergence free, because divergence is only well-defined with respect to a reference volume form and there is no canonical volume form associated to a smooth manifold.

2. The canonical Hilbert space of a manifold. For the moment, assume M is orientable and let μ be a volume form on M. Given μ we may define the L^2_μ-inner product

$$\langle f, g\rangle_{L^2_\mu} := \int_M (f \cdot g) \cdot \mu$$

for functions $f, g \in C^0(M)$. If we complete the resulting inner-product space we obtain the Hilbert space, $L^2_\mu(M)$. In the case where M is non-orientable we can perform virtually the same construction by replacing μ with a positive density [24].

It's notable that these constructions are not canonically associated to M, but only to the pair (M, μ). If we change μ, then the inner-products and norms are transformed as well. However, there is a canonical Hilbert space associated to M, independent of an arbitrarily chosen μ. We obtain this space by dividing by the following equivalence relation. We we can consider pairs of functions and volume forms, (f, μ), where $f \in L^2_\mu(M)$. We say that $(f_1, \mu_1) \sim (f_2, \mu_2)$ if there exists a positive function, ϕ, such that $f_1 = \phi^{1/2} \cdot f_2$ and $\phi \cdot \mu_1 = \mu_2$. In fact, $\sim$, is an equivalence relation and so we may consider the space of equivalence classes. In particular, we define the *canonical Hilbert space of* M to be

$$L^2(M) := \{L^2_\mu(M) \times \{\mu\} \mid \mu \in \mathrm{Vol}(M)\}/\sim .$$

By construction, $L^2(M)$ is a vector-space with the addition operator defined independently of any reference to a volume form (or positive density). As topological vector-space $L^2(M)$ is isomorphic to $L^2_\mu(M)$ for a fixed μ. Thus, for any $\psi \in L^2(M)$ and any volume form μ, there is a unique $f \in L^2_\mu(M)$ such that $\psi = (f, \mu)/\sim$. Therefore, we can define a bijection $i_\mu : L^2(M) \to L^2_\mu(M)$.

For any two elements $\psi_1, \psi_2 \in L^2(M)$ we can define the canonical inner-product

$$\langle \psi_1 \mid \psi_2 \rangle := \int_M (i_\mu(\psi_1) \cdot i_\mu(\psi_2)) \cdot \mu \tag{3}$$

for an arbitrary $\mu \in \mathrm{Vol}(M)$.

PROPOSITION 3. *The right hand side of* (3) *is defined independently of* μ.

Proof. Let μ' be another volume form. Then there exists a positive function ϕ such that $\mu' = \phi\mu$. Moreover, Let $f'_1 = i_{\mu'}(\psi_1)$ (that is to say, $f'_1 \in L^2_{\mu'}(M)$ is the unique element such that $\psi_1 = (f'_1, \mu')/\sim$). Then it must be the case that $f'_1 = \phi^{-1/2} \cdot f_1$ because $\sim$ is an equivalence relation and $\psi_1 = (f_1, \mu)/\sim$. Similarly, $\psi_2 = (f'_2, \mu')/\sim$ for $f'_2 = \phi^{-1/2} \cdot f_2$. The right hand side of (3) can be written in terms of the primed variables as

$$\begin{aligned}\int_M (f_1 \cdot f_2) \cdot \mu &= \int_M ((\phi^{-1/2} \cdot f_1) \cdot (\phi^{-1/2} \cdot f_2)) \cdot (\phi \cdot \mu) \\ &= \int_M (f'_1 \cdot f'_2) \cdot \mu'.\end{aligned}$$

However, this is nothing but $\langle \psi_1 \mid \psi_2 \rangle$ written in terms of the volume form μ'. ∎

Locally, we may understand elements of $L^2(M)$ by how they transform under a change of coordinates. The choice of a local coordinate system induces a choice of a local volume form, given by the flat metric on $\mathbb{R}^n$, and so an element $\psi \in L^2(M)$ can, locally, be pictured as a square integrable function $\psi(x)$. Under a local change of coordinations, Φ, the element ψ transforms as

$$\Phi_* \psi(x) = |\det(D\Phi^{-1}(x)|^{1/2}\, \psi(\Phi^{-1}(x)). \tag{4}$$

From the perspective of transformation theory, it is abundantly clear that elements of $L^2(M)$ are not functions, as the law for transforming functions would send $f(x)$ to $f(\Phi^{-1}(x))$. The reason for the multiplication by $|\det(D\Phi^{-1}(x)|^{1/2}$ is because Φ warps the volume of space, and the local volume form is being used in order to depict ψ as a function. Said differently, elements of $L^2(M)$ are a distinct geometric

objects, known as a *half-densities* [3, 18]. The origin of this terminology comes from the following observation. If $\rho = \psi^2$, then the change of coordinate formula for ρ is

$$\Phi_* \rho(x) = |\det(D\Phi^{-1}(x))|\, \rho(\Phi^{-1}(x)).$$

This is nothing but the change of coordinate formula for a density [24]. Therefore, half-densities are objects that yield densities when you square them. In fact, for any two half-densities $\psi_1, \psi_2 \in L^2(M)$, the product obtained from point-wise multiplication, $\psi_1 \cdot \psi_2$, is a density. As a density can be integrated, we can write the inner-product on $L^2(M)$ more simply as

$$\langle \psi_1 \mid \psi_2 \rangle = \int_M \psi_1 \cdot \psi_2.$$

As the integration is a canonical operation on M we observe the following.

PROPOSITION 4. *Diffeomorphisms of M act unitarily on $L^2(M)$.*

Proof. Locally, the action of a diffeomorphism, $\Phi : M \to M$, is given by the change of coordinates formula (4). For an arbitrary half-density ψ we observe

$$\begin{aligned} \|\Phi_* \psi\|_{L^2}^2 &= \langle \Phi_* \psi \mid \Phi_* \psi \rangle \\ &= \int_M |\det(D\Phi^{-1}(x)\, \psi(\Phi^{-1}(x))^2 dx. \end{aligned}$$

By the change of variables formula with the change of variable $x' = \Phi(x)$ we find

$$= \int_M \psi(x)^2 dx = \|\psi\|_{L^2}^2$$

As ψ was chosen arbitrarily, we have observe that Φ preserves the L^2-norm. By polarization, Φ also preserves the inner-product $\langle \cdot \mid \cdot \rangle$, and therefore Φ acts unitarily on $L^2(M)$.[1] ∎

For the purposes of this paper, it is very important that $L^2(M)$ consist of entities distinct from functions. Without such a distinction, we will run into ambiguities and contradictions. For example, in later sections we will canonically embed $C^0(M)$ into the space of Hermitian operators on $L^2(M)$. We will find that deformations of M act on $L^2(M)$, and thus induce an actions on operators of $L^2(M)$ which descends to the standard action of a deformation on $C^0(M)$. However, if $C^0(M)$, also embedded in $L^2(M)$ we would have two group actions to consider on $C^0(M)$ with no discernible reason to prefer one over the other. Thankfully, no such ambiguity arises since $C^0(M)$ does not embed into $L^2(M)$ canonically.

3. Sobolev spaces. Just as there is a canonical space $L^2(M)$, there is also a canonical space $H^s(M)$. We define $H^s(M)$ as a subspace of $L^2(M)$ with certain regularity properties. In particular, the Lie derivative of $\psi \in L^2(M)$ with respect to a vector field, $X \in \mathfrak{X}(M)$, is given by

$$\pounds_X[\psi] := -\frac{d}{dt}(\Phi_X^t)_* \psi$$

[1] Strictly speaking, the integrals in this proof should be computed with respect to a partition of unity. However, the notational complexity incurred was hard to justify in light of the clarity that was lost when we did this.

Where Φ_X^t is the time-t flow of X, and $(\Phi_X^t)_*\psi$ is locally represented by (4). In local coordinates where $X(x) = \sum_{i=1}^n X^i(x)\frac{\partial}{\partial x^i}$ the Lie derivative is

$$\pounds_X[\psi] = \frac{1}{2}\left(\sum_{i=1}^n X^i\partial_i\psi + \partial_i(\psi X^i)\right) \tag{5}$$

As the Lie derivative operator is nothing but the action of an infinitesimal unitary transformation (this is Proposition 4), we obtain the following corollary.

COROLLARY 5. *The Lie-derivative operator is anti-symmetric with respect to the L^2-inner product on the subspace in which it is bounded.*

As $\pounds_X$ is an (unbounded) anti-symmetric operator, we may consider raising it to fractional powers in order to define the space $H^s(M)$ as

$$H^s(M) := \{\psi \in L^2(M) \mid \|(\pounds_X)^s \cdot \psi\| < \infty, \forall X \in \mathfrak{X}(M)\}.$$

This definition of $H^s(M)$ in terms of semi-norms is a generalization of the standard definition on $\mathbb{R}^n$ using the seminorms $\|\partial_k^s f\|_{L^2}$. However, for the purposes of analysis, it will be convenient to refer to a single H^s norm, rather than an infinite family of semi-norms. There is no canonical choice, so we are forced to non-canonically choose one. The clearest way to do this is to arbitrarily choose a Riemmanian metric, g.[2] The choice in metric induces a Riemannian density, μ_g, and a Laplace-Beltrami operator, Δ_g. We then define the H^s-norm with respect to g as

$$\|\psi\|^2_{H^s_g} := \int_M f \cdot [(1-\Delta_g)^s \cdot f] \cdot \mu_g$$

where $f = i_{\mu_g}(\psi)$. It is fairly clear from merely counting derivatives in a local coordinate chart that $\psi \in H^s(M)$ if and only if $\|\psi\|_{H^s_g} < \infty$ for an arbitrary g. We will not concern ourselves with how to choose a metric g. When we write $\|\cdot\|_{H^s}$ we will imply that an arbitrary metric has been chosen.

4. Quantization. In physics, "quantization" refers to the process of substituting certain physically relevant functions with operators on a Hilbert space, while attempting to preserve the symmetries and conservation laws of the classical theory [3, 18]. In this section, we quantize (1) and (2) by replacing functions and densities with bounded and trace-class operators on $L^2(M)$. This is useful in Section 5 when we discretize.

To begin, let us quantize the space of continuous real-valued functions $C(M)$. For each $f \in C(M)$, there is a unique bounded Hermitian operator, $H_f : L^2(M) \to L^2(M)$ given by scalar multiplication. That is to say $(H_f \cdot \psi)(x) = f(x)\psi(x)$ for any $\psi \in L^2(M)$. By inspection one can observe that the map "$f \mapsto H_f$" is a monomorphism from the algebra $C(M)$ into $\mathrm{B}(L^2(M))$ because $H_{f\cdot g+h} = H_f \cdot H_g + H_h$ and $\|H_f\|_{op} = \|f\|_\infty$.

Dually, for any trace class operator A there is a unique $\rho_A \in C(M)'$ such that:

$$\int_M f\,\rho_A d\mathbf{x} = \mathrm{Tr}(\mathbf{H}_\mathbf{f}^\dagger \cdot \mathbf{A})$$

for any $f \in C(M)$. Said differently, the map "$A \mapsto \rho_A$" is the adjoint of the map "$f \mapsto H_f$". Therefore "$A \mapsto \rho_A$" is an epimorphism (i.e. surjective).

[2] We know such a metric exists as long as M is paracompact.

By Stone's theorem, $\mathcal{L}_X$ is the infinitesimal generator of a one-parameter semigroup on the topological Lie group $\mathrm{U}(L^2(M))$. We can relate this transformation to the flow of the vector field $\Phi_X^t : M \to M$ in a precise sense. In particular

$$(e^{\mathcal{L}_X} \cdot \psi)(x) = \det(D[(\Phi_X^t)^{-1}](x))^{1/2} \psi((\Phi_X^t)^{-1}(x)).$$

We can now convert the evolution PDEs (1) and (2) into ODEs of operators on $L^2(M)$.

THEOREM 6. *Let $X(t) \in \mathfrak{X}(M)$ be a time-dependent vector-field. Then f satisfies (1) if and only if H_f satisfies*

$$\frac{dH_f}{dt} + [H_f, \mathcal{L}_X] = 0. \tag{6}$$

If A is trace-class and satisfies

$$\frac{dA}{dt} + [A, \mathcal{L}_X] = 0, \tag{7}$$

then ρ_A satisfies (2). Finally, if ψ satisfies

$$\partial_t \psi + \mathcal{L}_X[\psi] = 0 \tag{8}$$

then $\rho_A = \psi^2$ satisfies (2) and $\psi \otimes \psi^\dagger$ satisfies (7).

Proof. Let f satisfy (1). For an arbitrary $\psi \in L^2(M)$ we observe that $[H_f, \mathcal{L}_X] \cdot \psi$ is given by:

$$\begin{aligned}([H_f, \mathcal{L}_X] \cdot \psi)(x) = f(x) \sum_{j=1}^n \left(\frac{1}{2} X^j \frac{\partial \psi}{\partial x^j} + \frac{1}{2} \frac{\partial}{\partial x^j}(\psi X^j) \right)(x) \\ - \frac{1}{2} X^j \left.\frac{\partial}{\partial x^j}\right|_x (f\psi) + \frac{1}{2} \left.\frac{\partial}{\partial x^j}\right|_x (f\psi X^j)\end{aligned}$$

where we have used (5). Application of the product rule to each of these terms yields a number of cancellations and we find:

$$[H_f, \mathcal{L}_X] \cdot \psi = -X^j \frac{\partial f}{\partial x^j} \psi = (\partial_t f)\psi = \frac{dH_f}{dt} \cdot \psi.$$

As ψ is arbitrary, we have shown that H_f satisfies (6). Each line of reasoning is reversible, and so we have proven the converse as well. This proves the first claim.

For the second claim let f be an arbitrary time-dependent continuous function. Then we find

$$\begin{aligned}\int (\partial_t \rho_A) f dx = \mathrm{Tr}(H_f^\dagger \partial_t A) = \mathrm{Tr}(H_f^\dagger \cdot [A, \mathcal{L}_X]) \\ = \mathrm{Tr}([\mathcal{L}_X, H_f]^\dagger A)\end{aligned}$$

By the previous claim the above expression is equivalent to

$$\begin{aligned}= \int_M \left(\sum_{i=1}^n X^i \partial_i f \right) \rho_A dx \\ = - \int_M f \left(\sum_{i=1}^n \partial_i (\rho_A X^i) \right) dx\end{aligned}$$

where the boundary terms vanish because X is tangential to M. Lastly, if ψ satisfies (8) and $\rho = \psi^2$ then we see

$$\begin{aligned}\partial_t \rho &= \partial_t(\psi^2) = 2(\partial_t \psi)\,\psi \\ &= -2\left(-\frac{1}{2}\partial_i(\psi X^i) - \frac{1}{2}X^i\partial_i\psi\right)\psi = \left(-X^i\partial_i\psi - \psi\partial_i X^i\right)\psi \\ &= -2\psi(\partial_i\psi)X^i - (\partial_i X^i)\psi^2 = -\partial_i(\rho)X^i - (\partial_i X^i)\rho = -\partial_i(\rho X^i). \qquad \square\end{aligned}$$

The benefit of using (6) and (7) to represent the PDEs of concern is that (6) and (7) may be discretized using a standard least squares projections on $L^2(M)$ without sacrificing qualitative accuracy.

5. Discretization. This section presents the numerical algorithms for solving (1) and (2). The basic ingredient for all the algorithms in this section are a Hilbert basis and an ODE solver. Denote a Hilbert basis by $\{e_0, e_1, \dots\}$ for $L^2(M)$. For example, we may use the eigen-functions of the Laplace operator. This is known as the Fourier basis. To ensure convergence, we assume:

ASSUMPTION 7. *Our basis $\{e_k\}$ is such that there exists a metric g for which the unitary transformation which sends the basis $\{e_k\}$ to the Fourier basis is bounded with respect to the $\|\cdot\|_{s,2}$-norm for some $s > 1$.*

In this section we provide a spectral discretization of space in order to convert the PDEs (1) and (2) into ODEs. Sometimes this is referred to as a semi-discretization [16]. In particular, we will assume access to a finite dimensional ODE solver, denoted "OdeSolve." whose time-step discretization error is dwarfed by the spatial discretization error of our semi-discretization. In practice a high-order Runge-Kutta method or a well tested software library such as [5] could be used. Most notably, the method of [7] is specialized to isospectral flows such as (6) and (7) by using discrete-time isospectral flows. More explicitly, let OdeSolve(F, x_0, t) denote the numerically computed solution $x(t)$ to the ODE "$\dot{x} = F(x)$" at time $t \in \mathbb{R}$, with initial condition $x_0 \in M$. Before constructing an algorithm to spectrally discretize (1) and (2) in a qualitatively accurate manner, we first solve (8) using a spectral Galerkin discretization in Algorithm 1.

Algorithm 1 A spectral discretization to solve (8) for half densities.

Require: $\psi(0) \in L^2(M)$, $t \in \mathbb{R}$, $N \in \mathbb{N}$.
  initialize $z(0) = \vec{0} \in \mathbb{C}^N$.
  initialize $X_N = [0] \in \mathbb{C}^{N\times N}$
  for $i = 1, \dots, N$ **do**
    $[z(0)]_i = \int_M \bar{e}_i(x)\psi(0,x)$
    for $j = 1, \dots, N$ **do**
      $[X_N]^i{}_j = \frac{1}{2}\int_M \bar{e}_i(x)(X^\alpha\partial_\alpha e_j + \partial_\alpha(X^\alpha e_j))(x)$
    end for
  end for
  initialize the function $F : \mathbb{C}^N \to \mathbb{C}^N$ given by $F(z) = X_N \cdot z$.
  $z(t) = \text{OdeSolve}(F, z(0), t)$
  return $\psi_N(t) = \sum_{i=1}^n [z(t)]_i e_i$.

To summarize, Algorithm 1 produces a half-density $\psi_N(t_k) \in V_N = \text{span}(e_1, \dots, e_N)$ by projecting (8) to V_N. This projection is done by constructing the operator $X_N =$

$\pi_N \circ \pounds_X|_{V_N} : V_N \to V_N$. In Section 6 we prove that $\psi_N(t_k)$ converges to the solution of (8) as $N \to \infty$. We see that $\psi_N(t)$ evolves by unitary transformations, just as the exact solution to (8) does. This correspondence is key in providing the qualitative accuracy of algorithms that follow, so we formally state it here.

PROPOSITION 8. *The output of Algorithm 1 is given by $U_N(t_k) \cdot \psi_N(0)$ when $\psi(0) \in L^2(M)$ is the input to Algorithm 1 where $\psi_N(0) = \pi_N(\psi(0))$ and $U_N(t)$ is the unitary operator generated by X_N.*

Proof. The operator X_N in Algorithm 1 is anti-Hermitian on V_N. It therefore generates a unitary action on $V_N \subset L^2(M)$ when inserted into OdeSolve. ∎

Before continuing, we briefly state a sparsity result that aides in selecting a basis. We say an operator $A : L^2(M) \to L^2(M)$ is *sparse banded diagonal* with respect to a Hilbert basis $\{e_0, e_1, \dots\}$ if there exists an integer $W \in \mathbb{N}$ such that $A(e_i)$ is a finite sum elements of the form $e_{i+\delta_j}$ for fewer than W offsets δ_j for $i = 0, 1, 2, \dots$.

THEOREM 9. *If $\pounds_{\frac{\partial}{\partial x^j}}$ and H_{f_k} are sparse banded diagonal, and if the vector-field X is given in local coordinates by $X^i = \sum_k c_k^i f_k$ with fewer than $W > 0$ of c_k^i's being non-zero for each $i = 1, \dots, n$, then the matrix X_N in Algorithm 1 is sparse banded diagonal and the sparsity of X_N is $\mathcal{O}(W/N)$.*

Proof. The result follows directly from counting. ∎

Theorem 9 suggests selecting a basis where W is small, or at least finite. For example, if M were a torus, and the vector-field was made up of a finite number of sinusoids, then a Fourier basis would yield a W equal to the maximum number of terms along all dimensions.

By Theorem 6, the square of the result of Algorithm 1 is a numerical solution to (2). We can use this to produce a numerical scheme to (2) by finding the square root of a density. Given a $\rho \in L^1(M)$ which admits a square root, let ρ^+ denote the positive part and ρ^- denote the negative part so that $\rho = \rho^+ - \rho^-$, then $\psi = \sqrt{\rho^+} - i\sqrt{\rho^-}$ is a square root of ρ since $\rho = \psi^2$. This yields Algorithm 2 to spectrally discretize (2) in a qualitatively accurate manner for densities which admit a square root.

Algorithm 2 A spectral discretization to solve (2) for densities

Require: $\rho(0) \in L^1(M), t \in \mathbb{R}, N \in \mathbb{N}$.
  Initialize $\psi(0) = \sqrt{\rho^+(0)} - i\sqrt{\rho^-(0)}$
  Set $\psi_N(t) = \text{Algorithm\_1}(\psi(0), t, N)$
  return $\rho_N(t, x) = \psi_N(t, x)^2$.

Alternatively, we could have considered the trace-class operator $A_N(t_k) = \psi_N(t_k) \otimes \psi_N(t_k)^\dagger$ as an output. This would be an numerical solution to (7), and would be related to our original output in that $\rho_N(t_k) = \rho_{A_N(t_k)}$. Finally, we present an algorithm to solve (6) (in lieu of solving (1)). This algorithm is presented for theoretical interest at the moment.

Algorithm 3 A spectral discretization to solve (6) for functions

Require: $f(0) \in C(M), t \in \mathbb{R}, N \in \mathbb{N}$.
initialize $F_N(0), X_n \in \mathbb{C}^{N\times N}$.
initialize the linear map $B : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times N}$ given by $B(H) = -[H, X_N]$.
for $i, j = 1, \ldots, N$ **do**
$[F_N(0)]^i{}_j = \int_M \bar{e}_i(x) f(x) e_j(x)$
$[X_N]^i{}_j = \frac{1}{2}\int_M \bar{e}_i(x)(X^\alpha \partial_\alpha e_j + \partial_\alpha(X^\alpha e_j))(x)$
end for
$F_N(t) = \text{OdeSolve}(B, F_N(0), t)$
return The (compact) operator $H_{f,N}(t_k) = \sum_{i,j=1}^N [F_N(t_k)]^i{}_j e^i \otimes e^\dagger_j$.

We find that the ouput of Algorithm 3 bears algebraic similarities similarities to the exact solution to the infinite dimensional ODE, (6) (which is isomorphic to (1) by Theorem 6). This is stated in a proposition analogous to Proposition 8.

PROPOSITION 10. *$H_{f,N}(t) = U_N(t) \cdot H_{f,N}(0) \cdot U_N(t)^\dagger$ for any $t \in \mathbb{R}$. Moreover, $U_N(t)$ is identical to the unitary transformation of Proposition 8. Lastly, the exact solution of (6) is of the form $H_f(t) = U(t) \cdot H_f(0) \cdot U(t)^\dagger$ as well.*

Proof. This follows from the fact that algorithm outputs the solution to an isospectral flow "$\dot{F}_N + [F_N, X_N]$" where X_N is anti-Hermitian and that the H_f satisfies the isospectral flow (6). ∎

6. Error analysis. In this sections we derive convergence rates. We find that the error bound for Algorithm 1 induces error bounds for the Algorithms 2 and 3. Therefore, we first derive a useful error bound for Algorithm 1. Our proof is a generalization of the convergence proof in [29], where (8) is studied (modulo a factor of two time rescaling) on the torus. We begin by proving an approximation bound. In all that follows, let $\pi_N : L^2(M) \to V_N$ denote the orthogonal projection.

PROPOSITION 11. *If $\psi \in H^{\bar{s}}(M)$ and $\bar{s} > s \geq 0$, then*

$$\|\psi - \pi_N(\psi)\|_{s,2} < \frac{n\, C_{\bar{s},s}}{\bar{s} - s} \|\psi\|_{\bar{s},2}\, N^{-2(\bar{s}-s)/n}$$

for some constant $C_{\bar{s},s}$ and $n = \dim(M)$.

Proof. We can assume that $e_1, e_2, \ldots$ is a Fourier basis. The results are unchanged upon applying Assumption 7 and converting to the Fourier basis. Any $\psi \in H^s(M; g)$ can expanded as $\psi = \hat{\psi}_k e_k$ where $\hat{\psi}_k = \langle e_k \mid \psi \rangle$. As $\psi \in H^s(M; g)$ it follows that

$$\|\psi\|^2_{\bar{s},2} = \sum_{k=0}^{\infty} \left|\hat{\psi}_k\right|^2 (1 + \lambda_k)^{\bar{s}} < \infty. \tag{9}$$

A corollary of Weyl's asymptotic formula is that λ_k is $\mathcal{O}(k^{2/n})$ for large k [8, page 155]. After substitution of this asymptotic result into (9) for large k, we see that $|\hat{\psi}_k|^2$ is asymptotically dominated by $Ck^{-1-2\bar{s}/n}$ for some constant C. For sufficiently large N we find

$$\|\psi - \pi_N(\psi)\|_{s,2} = \sum_{k>N} (1 + \lambda_k)^s |\hat{\psi}_k|^2 \leq C \sum_{k>N} (1 + \lambda_k)^s k^{-1-2\bar{s}/n}$$

and by another application of the Weyl formula

$$\|\psi - \pi_N(\psi)\|_{s,2} \leq \tilde{C} \sum_{k>N} \frac{1}{k^{1+2(\bar{s}-s)/n}} \leq C_{s,\bar{s}} \frac{d}{\bar{s}-s} N^{-2(\bar{s}-s)/n}.$$

Where the last inequality is derived by bounding the infinite sum with an integral. ∎

With this error bound for the approximation error we can derive an error bound for Algorithm 1:

THEOREM 12. *Let $\psi(0) \in H^{\bar{s}}(M)$ for $\bar{s} > s > 1$. Let $T > 0$ and $t \in [0, T]$. Let $\psi(t)$ be denote the solution to (8) with initial condition $\psi(0)$. Finally, let $\psi_N(t)$ be the output of Algorithm 1 with respect to the inputs $\psi(0), t, N$ for some $N \in \mathbb{N}$. Then the error $\varepsilon_N(t) := \|\psi(t) - \psi_N(t)\|_{s,2}$ satisfies:*

$$\varepsilon_N(t) \leq \|\psi(0)\|_{\bar{s},2} K_T \left(N^{-2(s-1)} t + \frac{n}{\bar{s}-s} N^{-2(\bar{s}-s)/n} \right) e^{C_T t}$$

where K_T and C_T are positive and constant with respect to N,s, and $\bar{s}$. In particular for $s = (\bar{s}+1)/2$:

$$\varepsilon_N(t) \leq \|\psi(0)\|_{\bar{s},2} K_T \left(N^{1-\bar{s}} t + \frac{n}{\bar{s}-1} N^{(1-\bar{s})/n} \right) e^{C_T t}.$$

To prove Theorem 12, we need a perturbed version of Gronwall's inequality:

LEMMA 13. *If $\frac{du}{dt} \leq Ku + \epsilon$ for some $K > 0$ then $u(t) \leq (\epsilon t + u(0))e^{Kt}$.*

Proof. Let $w(t) = u(t)e^{-Kt}$. Then for $t \geq 0$ we find

$$\frac{dw}{dt} = \frac{du}{dt} e^{-Kt} - Kw \leq (Ku + \epsilon)e^{-Kt} - Kw = \epsilon e^{-Kt} \leq \epsilon$$

Thus $w(t) \leq \epsilon t + w(0) = \epsilon t + u(0)$. ∎

Now we can prove Theorem 12:

Proof (Proof of Theorem 12). Note that $\frac{d\varepsilon_N}{dt} = \frac{1}{2\varepsilon_N} \langle \psi - \tilde{\psi} \mid (1+\Delta)^s \frac{d}{dt}(\psi - \tilde{\psi}) \rangle$ By the Cauchy-Schwarz inequality

$$\begin{aligned} \frac{d\varepsilon_N}{dt} &\leq \frac{1}{2} \|\mathcal{L}_X[\psi] - \pi_N(\mathcal{L}_X[\psi_N])\|_{s,2} \\ &= \frac{1}{2} \|\mathcal{L}_X[\psi] - \pi_N(\mathcal{L}_X[\psi - (\psi - \psi_N)])\|_{s,2}. \end{aligned}$$

By the triangle inequality and the definition of the operator norm:

$$\frac{d\varepsilon_N}{dt} \leq \|(1 - \pi_N)\|_{H^{s-1},op} \|X\|_{H^s,op} \|\psi\|_{s,2} + \|\pi_N\|_{op} \|X\|_{H^s,op} \varepsilon_N$$

From direct observation, we can verify that $\psi(t)$ is related to ψ_0 through the flow of X, Φ_X^t via the equation

$$\psi(t, x) = \det(D[(\Phi_X^t)^{-1}](x)) \psi_0((\Phi_X^t)^{-1}(x))$$

As Φ_X^t is smooth we can observe that $\|\psi(t)\|_{s,2}$ is bounded by a scalar multiple of $\|\psi_0\|_{s,2}$. Thus we may write the above bound in the form

$$\frac{d\varepsilon_N}{dt} \leq K' \|1 - \pi_N\|_{H^{s-1},op} \|\psi_0\|_{s,2} + C_T \varepsilon_N$$

for constants C_T and K'. As $s > 1$, for sufficiently large N we can compute that $\|1 - \pi_N\|_{H^{s-1},op} \leq (1 + \lambda_{N+1})^{-(s-1)}$ where λ_N denotes the Nth eigenvalue of the Laplace operator. This is accomplished by observing the operator $1 - \pi_N$ in a Fourier basis and applying to appropriate norms. By Weyl's asymptotic formula [8, Theorem B.2], λ_N asymptotically behaves like $N^{2/n}$. Therefore by Lemma 13 with $\epsilon = C_T n^{-2(s-1)/d} \|\psi_0\|_{s,2}$:

$$\varepsilon_N(t) \leq (K' N^{-2(s-1)/n} \|\psi_0\|_{s,2} t + \varepsilon_N(0)) e^{C_T t}.$$

That $\varepsilon_N(0)$ behaves as $K'' \|\psi_0\|_{\bar{s},2} n^{-2(\bar{s}-s)/d}$ is a re-statement of Proposition 11. We then set $K_T = \max(K', K'')$. ∎

Having derived an error bound for Algorithm 1, we can derive an error bound for Algorithm 2.

THEOREM 14. *Let $\rho(0) \in W^{\bar{s},1}(M)$ for $\bar{s} > s > 1$. Let $T > 0$ and $t \in [0, T]$ be fixed. Let $\rho(t)$ be the solution of (2) at time t. Finally, let $\rho_N(t)$ be the output of Algorithm 2 with respect to the input $(\rho(0), t, N)$ for some $N \in \mathbb{N}$. Then:*

$$\|\rho(t) - \rho_N(t)\|_1 \leq \|\rho(0)\|_{\bar{s},1} K \left(N^{-2(s-1)} t + \frac{d}{\bar{s} - s} N^{-2(\bar{s}-s)/n} \right) e^{C_T t}$$

where K is constant with respect to N, and C_T is the same constant as in Theorem 12.

Proof. Without loss of generality, assume that ρ is non-negative (otherwise split it into its non-negative and non-positive components). Let $\psi \in L^2(M)$ be such that $\rho = \psi^2$, as described in Algorithm 2. It follows that $\psi \in H^s(M)$ and we compute

$$\|\rho(t) - \rho_N(t)\|_1 = \int_M |\rho(t) - \rho_N(t)| dx = \int_M |\psi^2 - \psi_N^2| dx$$

If we let $\phi_N = \psi - \psi_N$ then we can re-write the above as

$$\begin{aligned} \|\rho(t) - \rho_N(t)\|_1 &= \int_M |\psi^2 - (\psi - \phi_N)^2| dx = \int_M |2\psi\phi_N - \phi_N^2| dx \\ &\leq 2\|\psi\|_2 \|\phi_N\|_2 + \|\phi_N\|_2^2 = 2\|\rho\|_1^{1/2} \cdot \|\phi_N\|_2 + \|\phi_N\|_2^2 \end{aligned}$$

Above we have applied Holder's inequality to $L^2(M)$, which still holds upon using the isometry generated by $\mathcal{L}_X$. Theorem 12 provides a bound for $\|\phi_N\|$. Substitution of this bound into the above inequality yields the theorem. ∎

Finally, we prove that Algorithm 3 converges to a solution of (6), which is equivalent to a solution of (1) courtesy of Theorem 6:

PROPOSITION 15. *Let $f \in C^k(M)$ and let $H_{f,N} = \pi_N \circ H_f \circ \pi_N$. Then*

$$\|H_f - H_{f,N}\|_{H^s,op} \leq D \frac{n}{s} N^{-2s/n} \|\hat{f}\|_{op}$$

where $s > k \geq 1$, and D is constant.

Proof. Let $\pi_N^\perp = 1 - \pi_N$. By Proposition 11 we know that

$$\|\pi_N^\perp(\psi)\|_2 \leq \frac{n}{s} N^{-2s/n} \|\psi\|_{s,2} \tag{10}$$

for $s > 0$, then:

$$\begin{aligned}
\|H_f - H_{f,N}\|_{H^s,op} &= \sup_{\|\psi\|_{s,2}=1} \langle \psi \mid H_f - H_{f,N} \mid \psi \rangle \\
&= \sup_{\|\psi\|_{s,2}=1} \left(\langle \psi \mid H_f \mid \psi \rangle - \langle \psi - \pi_N^{\perp}(\psi) \mid H_f \mid \psi - \pi_N^{\perp}(\psi) \rangle \right) \\
&= \sup_{\|\psi\|_{s,2}=1} \left(2\Re \langle \pi_N^{\perp}(\psi) \mid H_f \mid \psi \rangle - \langle \pi_N^{\perp}(\psi) \mid H_f \mid \pi_N^{\perp}(\psi) \rangle \right) \\
&\leq \sup_{\|\psi\|_{s,2}=1} (\|\pi_N^{\perp}(\psi)\|_2 - \|\pi_N^{\perp}(\psi)\|_2^2) \|H_f\|_{op}
\end{aligned}$$

By (10) the result follows. ∎

THEOREM 16. *Let $T > 0$ and $t \in [0, T]$ be fixed. Let $f(t)$ denote the solution to (1) at time t with initial condition $f(0) \in C^k(M)$. Let $H_{f,N}(t)$ denote the output of Algorithm 3 with respect to the inputs $(f(0), t, N)$ for some $N \in \mathbb{N}$. Then:*

$$\begin{aligned}
\|H_{f(t)} - H_{f,N}(t)\|_{H^s,op} \leq D \frac{n}{s} N^{-2s/n} \|H_{f(t)}\|_{op} \\
+ K_T \|H_{f,N}(t)\|_{op} \left(N^{1-s} + \frac{2n}{s-1} N^{(1-s)/n} \right) e^{C_T t}
\end{aligned}$$

for the same constant D as in Proposition 15 and the same constants C_T, K_T as in Theorem 12.

Proof. We find

$$\|H_{f(t)} - H_{f,N}(t)\|_{H^s,op} = \sup_{\|\psi\|_{s,2}=1} \langle \psi \mid H_{f(t)} - H_{f,N}(t) \mid \psi \rangle$$

In light of Proposition 10 we find

$$= \sup_{\|\psi\|_{s,2}=1} \langle \psi \mid U(t) \cdot H_{f(0)} \cdot U(t)^{\dagger} - U_N(t) \cdot H_{f,N}(0) \cdot U_N(t)^{\dagger} \mid \psi \rangle.$$

The output of Algorithm 3 indicates that $H_{f,N}(0) = \pi_N^{\perp} \circ H_{f(0)} \circ \pi_N^{\perp}$. Therefore, the above inline equation becomes

$$= \sup_{\|\psi\|_{s,2}=1} \langle U(t)^{\dagger}\psi \mid H_{f(0)} \mid U(t)^{\dagger}\psi \rangle - \langle U_N(t)^{\dagger}\psi \mid \pi_N^{\perp} \circ H_{f(0)} \circ \pi_N^{\perp} \mid U_N(t)^{\dagger}\psi \rangle.$$

and finally

$$= \sup_{\|\psi\|_{s,2}=1} \langle U(t)^{\dagger}\psi \mid H_{f(0)} - H_{f,N}(0) \mid U(t)^{\dagger}\psi \rangle - \langle \phi(t) \mid H_{f,N}(0) \mid \phi(t) \rangle \tag{11}$$

where $\phi(t) = U_N(t)^{\dagger}\psi - U(t)^{\dagger}\psi$.

The first term is bounded by Proposition 15. To bound the second term we must bound ϕ. As $U_N(t)^{\dagger}\psi$ is the backwards time numerical solution to (8) and $U(t)^{\dagger}\psi$ is the exact backward time solution to (8), Theorem 12 prescribes the existence of constants K and C such that:

$$\|\phi\|_{\underline{s},2} = \|U_N(t)^{\dagger}\psi - U(t)^{\dagger}\psi\|_{\underline{s},2} \leq K \|\psi\|_{s,2} \left(N^{-2(\underline{s}-1)} + \frac{n}{s-\underline{s}} N^{-2(s-\underline{s})/n} \right) e^{Ct}$$

for any $\underline{s} < s$. This expression can be simplified by noting that $\|\psi\|_{s,2} = 1$, setting $\underline{s} = (1+s)/2$, and noting that the $H^{\underline{s}}$ norm is stronger than the L^2-norm to get:

$$\|\phi\|_2 \leq K\left(N^{1-s} + \frac{2n}{s-1}N^{(1-s)/n}\right)e^{Ct}.$$

By applying the Cauchy-Schwarz inequality to (11) and our derived bound on ϕ:

$$\begin{aligned}\|H_{f(t)} - H_{f,N}(t)\|_{H^s,op} \leq \|H_{f(0)} - H_{f,N}(0)\|_{H^s,op} \\ + K\|H_{f,N}\|_{op}\left(N^{1-s} + \frac{2n}{s-1}N^{(1-s)/n}\right)e^{Ct}\end{aligned}$$

Upon invoking Proposition 15 we get the desired result. ∎

7. Qualitative Accuracy. In this section, we prove that our numerical schemes are qualitatively accurate. We begin by illustrating the preservation of appropriate norms. Throughout this section let $\psi_N(t)$, $\rho_N(t)$, and $H_{f,N}(t)$ denote the sequence of outputs of Algorithms 1, 2, and 3 with respect to initial conditions $\psi(0) \in H^s(M;g), \rho(0) \in W^{s,1}$ and $f(0) \in C^s(M)$ for $N = 1, 2, \ldots$.

THEOREM 17. *Let ψ, ρ, f denote solutions to (8), (2), and (1) respectively. Let $\psi_N(t), \rho_N(t)$, and $H_{f,N}(t)$, denote outputs from algorithms 1, 2, and 3 respectively for a time $t < \infty$. Then $\|\psi_N(t)\|_2, \|\rho_N(t)\|_1$, and $\|H_{f,N}(t)\|_{op}$ are constant with respect to t for arbitrary $N \in \mathbb{N}$. Moreover,*

$$\begin{aligned}\lim_{N\to\infty} \|\psi_N(t)\|_2 &= \|\psi(\cdot;t)\|_2, \\ \lim_{N\to\infty} \|\rho_N(t)\|_{nuc} &= \|\rho(\cdot;t)\|_1, \\ \lim_{N\to\infty} \|H_{f,N}(t)\|_{op} &= \|f(\cdot;t)\|_{\sup}.\end{aligned}$$

Proof. To prove $\|H_{f,N}\|_{op}$ is conserved note that the evolution is isospectral [7]. We have already shown that $H_{f,N}(t)$ converges to $H_f(t)$ in the operator norm. Convergence of the norms follows from the fact that $\|H_f(t)\|_{op} = \|f\|_{\sup}$. An identical approach is able to prove the desired properties for $\rho_N(t)$ and $\psi_N(t)$ as well. ∎

Theorem 17 is valuable because each of the norms is naturally associated to the entity which it bounds, and these quantities are conserved for the PDEs that this paper approximates. For example, $\|H_f\|_{op} = \|f\|_{\sup}$ for a function f, and this is constant in time when f is a solution to (1). A discretization constructed according to Algorithm 3 according to Theorem 17 is constant for any N, no matter how small.

The full Banach algebra $C(M)$ is conserved by advection too. This property is encoded in our discretization as well.

THEOREM 18. *Let $f(x;t), g(x;t)$, and $h(x;t)$ be solutions of (1) and let $k = f \cdot g + h$. Let $H_{f,N}, H_{g,N}$ and $H_{h,N}$ be numerical solutions constructed by Algorithm 3, then $H_{k,N}(t) = H_{f,N} \cdot H_{g,N} + H_{h,N}$ satisfies*

$$\frac{d}{dt}H_{k,N} = [X_N, H_{k,N}].$$

Moreover, $H_{k,N}(t)$ strongly converges to H_k as $N \to \infty$ in the operator norm on $H^s(M)$ when $f, g, h \in C^s(M)$ for $s > 1$.

Proof. By construction, the output of Algorithm 3 is the result of an isospectral flow, and is therefore of the form

$$\begin{aligned} H_{f,N}(t) &= U_N(t) H_{f,N}(0) U_N(t)^\dagger \\ H_{g,N}(t) &= U_N(t) H_{g,N}(0) U_N(t)^\dagger \\ H_{h,N}(t) &= U_N(t) H_{h,N}(0) U_N(t)^\dagger. \end{aligned}$$

We then observe

$$\begin{aligned} H_{k,N}(t) &= U_N(t) H_{k,N}(0) U_N(t)^\dagger = U(t) \left(H_{f,N}(0) H_{g,N}(0) + H_{h,N}(0) \right) U(t)^\dagger \\ &= U(t) H_{f,N}(0) U(t)^\dagger U(t) H_{g,N}(0) U(t)^\dagger + U(t) H_{h,N}(0) U(t)^\dagger \\ &= H_{f,N}(t) H_{g,N}(t) + H_{h,N}(t). \end{aligned}$$

Differentiation in time implies the desired result. Convergence follows from Theorem 16. ∎

Finally, the duality between functions and densities is preserved by advection. If f satisfies (1) and ρ satisfies (2) then $\int f\rho$ is conserved in time. Algorithms 2 and 3 satisfy this same equality:

THEOREM 19. *For each $N \in \mathbb{N}$, $\mathrm{Tr}(H_{f,N} A_{\rho,N}(t))$ is constant in time where $A_{\rho,N}(t) = \psi_N(t) \otimes \psi_N(t)^\dagger$. Moreover, $\mathrm{Tr}(H_{f,N} A_{\rho,N})$ converges to the constant $\int f\rho$ as $N \to \infty$.*

Proof. As $H_{f,N}(t) = U_N(t) H_{f,N}(0) U_N(t)^\dagger$ and $\psi_N(t) = U_N \cdot \psi_N(0)$ we observe that

$$\begin{aligned} &\mathrm{Tr}(H_{f,N}(t)(\psi_N(t) \otimes \psi_N^\dagger(t))) \\ &\quad = \mathrm{Tr}(U_N(t) H_{f,N}(0) U_N(t)^\dagger U)_N(t)(\psi_N(0) \otimes \psi_N(0)^\dagger) U_N(t)^\dagger) \\ &\quad = \mathrm{Tr}(H_{f,N}(0)(\psi_N(0) \otimes \psi_N(0)^\dagger)) \end{aligned}$$

Convergence follows from Theorems 16 and 14. ∎

8. Numerical Experiments. This section describes two numerical experiments. First, a benchmark computation to illustrate the spectral convergence of our method and the conservation properties in the case of a known solution is considered.

8.1. Benchmark computation. Consider the vector field $\dot{x} = -\sin(2x)$ for $x \in S^1$. The flow of this system is given by:

$$\Phi_X^t(x) = \mathrm{atan}\left(e^{2t} \tan(x) \right).$$

If the initial density is a uniform distribution, ρ_0, then the the exact solution of (2) is:

$$\rho(x;t) = \left(e^{2t} \sin^2(x) + e^{-2t} \cos^2(x) \right)^{-1} \tag{12}$$

Figure 1 depicts the evolution of $\rho(x;t)$ at $t = 1.5$ with an initial condition. Figure 1a depicts the exact solution, given by (12), Figure 1b depicts the numerical solution computed from a standard Fourier discretization of (2) with 32 modes, and Figure 1c (page 17) depicts the numerical solution solution computed using Algorithm 2 with 32 modes.

Here we witness how Algorithm 2 has greater qualitative accuracy than a standard spectral discretization, in the "soft" sense of qualitative accuracy. For example, standard spectral discretization exhibits negative mass, which is not achievable in the exact system. Moreover, the L^1-norm is not conserved in standard spectral discretization. In contrast, Theorem 17 proves that the L^1-norm is conserved by Algorithm 2. A plot of the L^1-norm is given in Figure 2 (page 17). Finally, a convergence plot is depicted in Figure 3 (page 17). Note the spectral convergence of Algorithm 2. In terms of numerical accuracy, Algorithm 2 appears to have a lower coefficient of convergence.

In general, Algorithm 3 is very difficult to work with, as it outputs an operator rather than a classical function. However, Algorithm 3 is of theoretical value, in that it may inspire new ways of discretization (in particular, if one is only interested in a few level sets). We do not investigate this potentiality here in the interest of focusing on the qualitative aspects of this discretization. For example, under the initial conditions $g_0(x) = \cos(x)$ and $f_0 = \sin(x)$ the exact solutions to (6) are:

$$g(x,t) = \cos(x)\left(e^{4t}\sin^2(x) + \cos^2(x)\right)^{-1/2}$$
$$f(x,t) = \sin(x)\left(\sin^2(x) + e^{-4t}\cos^2(x)\right)^{-1/2}$$

Under the initial condition $h_0 = f_0 \cdot g_0 = \sin(x)\cos(x)$ the exact solution to (6) is:

$$h(x,t) = f(x,t)g(x,t) = \cos(x)\sin(x)\left(\cos^2(x) + e^{4t}\sin^2(x)\right)^{-1}.$$

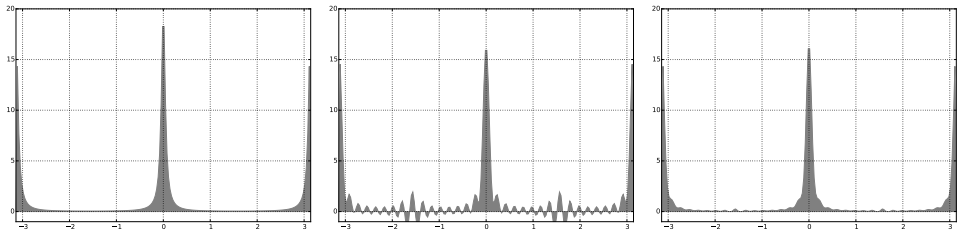
One can compute h by first multiplying the initial conditions and then using Algorithm 3 to evolve in time, or we may evolve each initial condition in time first, and multiply the outputs. If one uses Algorithm 3, then both options, as a result of Theorem 18, yield the same result up to time discretization error (which is obtained with error tolerance $1e{-}8$ in our code). In contrast, if one uses a standard spectral discretization, then these options yield different results with a discrepancy. This discrepancy between the order of operations for both discretization methods is depicted in Figure 4 (page 20).

Finally, the sup-norm is preserved by the solution of (1). As shown in Theorem 6, the sup-norm is equivalent to the operator norm when the functions are represented as operators on $L^2(M)$. As proven by Theorem 17, the operator-norm is conserved by Algorithm 3. In contrast, the sup-norm drifts over time under a standard discretization. This is depicted in Figure 5 (page 20).

8.2. A modified ABC flow. Consider the system

$$\dot{x} = A\sin(2\pi z) + C\cos(2\pi y) + D\cos(2\pi x)$$
$$\dot{y} = B\sin(2\pi z) + A\cos(2\pi y) + D\cos(2\pi y)$$
$$\dot{z} = A\sin(2\pi z) + B\cos(2\pi y) + D\cos(2\pi z)$$

on the three-torus for constants $A, B, C, D \in \mathbb{R}$. When $D = 0$ this system is the well studied volume conserving system known as an Arnold-Beltrami-Childress flow [1]. When $A > B > C > 0$, $D = 0$, and $C << 1$, then the solutions to this ODE are chaotic, with a uniform steady state distribution [26]. When $D = 0$ the operator $\pounds_X$ of (8) is identical to the operator $\partial_\alpha(\rho X^\alpha)$ that appears in (2), and Algorithm 1 do not differ from a standard spectral discretization. Therefore we consider the case where $D > 0$ to see how our discretization are differs from the standard one. When $D > 0$ volume is no longer conserved and there is a non-uniform steady-state distribution.



(a) Exact (b) Standard spectral (c) Algorithm 2

Fig. 1: A benchmark illustration of Algorithm 2 on the example described in Section 8.1.

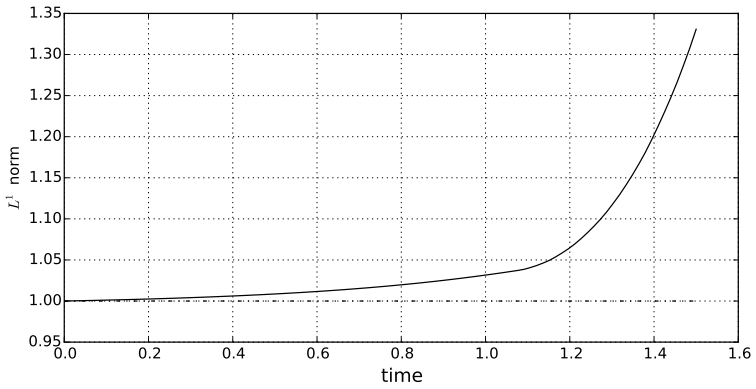


Fig. 2: A plot of the L^1-norm vs time of a standard spectral discretization (solid) and the result of Algorithm 2 (dotted) on the example described in Section 8.1.

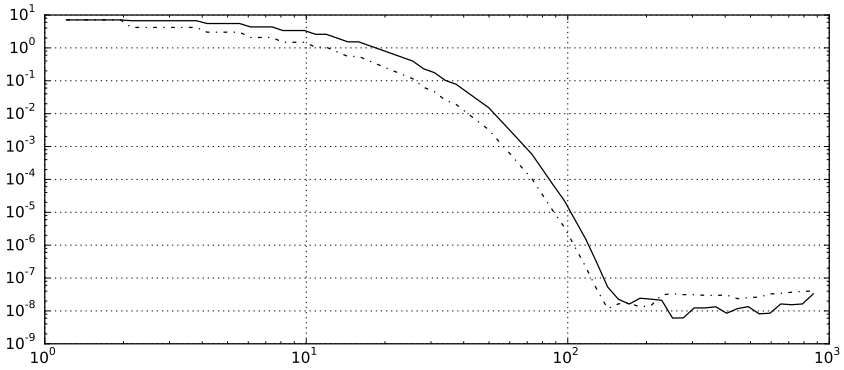


Fig. 3: Convergence plot for Algorithm 2 (dotted) and a standard spectral method (solid) in the L^1-norm.

For the following numerical experiment let $A = 1.0, B = 0.5, C = 0.2$, and $D = 0.5$. As an initial condition consider a wrapped Gaussian distribution with anisotropic variance $\sigma = (0.2, 0.3, 0.3)$ centered at $(0, 0, 0)$. Equation (2) is approximately solved using Algorithm 2, Monte-Carlo, and a standard spectral method. The results of the z-marginal of these densities are illustrated in Figure 6 on page 21. The top row depicts the results from using Algorithm 2 using 33 modes along each dimension. The middle row depicts the results from using a Monte-Carlo method with $15^3 = 3375$ particles as a benchmark computation. Finally, the bottom row depicts the results from using a standard Fourier based discretization of (2) using 33 modes along each dimension. Notice that Algorithm 2 performs well when compared to the standard discretization approach.

9. Conclusion. In this paper we constructed a numerical scheme for (1) and (2) that is spectrally convergent and qualitatively accurate, in the sense that natural invariants are preserved. The result of obeying such conservation laws is a robustly well-behaved numerical scheme at a variety of resolutions where legacy spectral methods fail. This claim was verified in a series of numerical experiments which directly compared our algorithms with standard Fourier spectral algorithms. The importance of these conservation laws was addressed in a short discussion on the Gelfand Transform. We found that conservation laws completely characterize (1) and (2), and this explains the benefits of using qualitatively accurate scheme at a more fundamental level.

9.1. Acknowledgements. This paper developed over the course of years from discussions with many people whom we would like to thank: Jaap Eldering, Gary Froyland, Darryl Holm, Peter Koltai, Stephen Marsland, Igor Mezic, Peter Michor, Dmitry Pavlov, Tilak Ratnanather, and Stefan Sommer. This research was made possible by funding from the University of Michigan.

REFERENCES

[1] V. I. Arnold and B. A. Khesin, *Topological Methods in Hydrodynamics*, vol. 24 of Applied Mathematical Sciences, Springer Verlag, 1992.

[2] N. Balci, B. Thomases, M. Renardy, and C. R. Doering, *Symmetric factorization of the conformation tensor in viscoelastic fluid models*, Journal of Non-Newtonian Fluid Mechanics, 166 (2011), pp. 546–553.

[3] S. Bates and A. Weinstein, *Lectures on the geometry of quantization*, vol. 8 of Berkeley Mathematics Lecture Notes, American Mathematical Society, Providence, RI, 1997.

[4] A. Bittracher, P. Koltai, and O. Junge, *Pseudogenerators of spatial transfer operators*, SIAM Journal on Applied Dynamical Systems, 14 (2015), pp. 1478–1517.

[5] P. N. Brown, G. D. Byrne, and A. C. Hindmarsh, *VODE: a variable-coefficient ODE solver*, SIAM Journal on Scientific and Statistical Computing, 10 (1989), pp. 1038–1051.

[6] M. Budišić, R. Mohr, and I. Mezić, *Applied Koopmanism*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 22 (2012).

[7] M. Calvo, A. Iserles, and A. Zanna, *Numerical solution of isospectral flows*, Mathematics of Computation of the American Mathematical Society, 66 (1997), pp. 1461–1486.

[8] I. Chavel, *Eigenvalues in Riemannian geometry*, vol. 115 of Pure and Applied Mathematics, Academic Press, Inc., Orlando, FL, 1984.

[9] J. B. Conway, *A course in functional analysis*, vol. 96 of Graduate Texts in Mathematics, Springer-Verlag, New York, second ed., 1990.

[10] K. Crane, U. Pinkall, and P. Schröder, *Robust fairing via conformal curvature flow*, ACM Transactions on Graphics, 32 (2013).

[11] L. C. Evans, *Partial differential equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, second ed., 2010.

[12] G. Froyland, O. Junge, and P. Koltai, *Estimating long-term behavior of flows without trajectory integration: the infinitesimal generator approach*, SIAM J. Numer. Anal., 51

(2013), pp. 223–247.
[13] G. Froyland and K. Padberg, *Almost-invariant sets and invariant manifolds—connecting probabilistic and geometric descriptions of coherent structures in flows*, Phys. D, 238 (2009), pp. 1507–1523.
[14] I. Gelfand and M. Neumark, *On the imbedding of normed rings into the ring of operators in Hilbert space*, Rec. Math. [Mat. Sbornik] N.S., 12(54) (1943), pp. 197–213.
[15] D. Gottlieb and J. Hesthaven, *Spectral methods for hyperbolic problems*, Journal of Computational and Applied Mathematics, 128 (2001), pp. 83 – 131.
[16] D. Gottlieb and S. A. Orszag, *Numerical analysis of spectral methods: theory and applications*, vol. 26, SIAM, 1977.
[17] J. M. Gracia-Bondía, J. C. Várilly, and H. Figueroa, *Elements of noncommutative geometry*, Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks], Birkhäuser Boston, Inc., Boston, MA, 2001.
[18] V. Guillemin and S. Sternberg, *Geometric Asymptotics*, vol. 14 of Mathematical Surveys and Monographs, American Mathematical Society, 1970.
[19] E. Hebey, *Nonlinear analysis on manifolds: Sobolev spaces and inequalities*, vol. 5 of Courant Lecture Notes in Mathematics, New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 1999.
[20] D. Henrion and M. Korda, *Convex computation of the region of attraction of polynomial control systems*, IEEE Transactions on Automatic Control, 59 (2014), pp. 297–312.
[21] R. S. Ismagilov, *The unitary representations of the group of diffeomorphisms of the space R^n, $n \geq 2$*, Functional Analysis and its applications, 9(2) (1975), pp. 154–155.
[22] P. Koltai, *Efficient approximation methods for the global long-term behavior of dynamical systems: theory, algorithms and examples*, Logos Verlag Berlin GmbH, 2011.
[23] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise*, Applied Mathematical Sciences, Springer Verlag, 1994.
[24] J. M. Lee, *Introduction to smooth manifolds*, vol. 218 of Graduate Texts in Mathematics, Springer-Verlag, 2nd ed., 2006.
[25] R. J. LeVeque, *Numerical methods for conservation laws*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 1992.
[26] A. J. Majda and A. L. Bertozzi, *Vorticity and incompressible flow*, vol. 27 of Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2002.
[27] P. A. Meyer, *Quantum probability for probabilists*, vol. 1538 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1993.
[28] I. Mezić, *Spectral properties of dynamical systems, model reduction and decompositions*, Nonlinear Dynamics, 41 (2005), pp. 309–325.
[29] J. E. Pasciak, *Spectral and pseudospectral methods for advection equations*, Mathematics of Computation, 35 (1980), pp. 1081–1092.
[30] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, *Spectral analysis of nonlinear flows*, Journal of Fluid Mechanics, 641 (2009), pp. 115–127.
[31] T. Sakai, *Riemannian geometry*, vol. 149 of Translations of Mathematical Monographs, American Mathematical Society, Providence, RI, 1996.
[32] P. J. Schmid, *Dynamic mode decomposition of numerical and experimental data*, Journal of Fluid Mechanics, 656 (2010), pp. 5–28.
[33] M. Taylor, *Pseudo differential operators*, Lecture Notes in Mathematics, Vol. 416, Springer-Verlag, Berlin-New York, 1974.
[34] C. Truesdell, *A First Course in Rational Continuum Mechanics: General Concepts*, Academic Press, 1991.
[35] S. M. Ulam and J. Von Neumann, *On combination of stochastic and deterministic processes-preliminary report*, Bulletin of the American Mathematical Society, 53 (1947), pp. 1120–1120.
[36] A. M. Vershik, I. M. Gelfand, and M. I. Graev, *Representations of the group of diffeomorphisms*, Uspehi Mat. Nauk, 30 (1975), pp. 1–50.

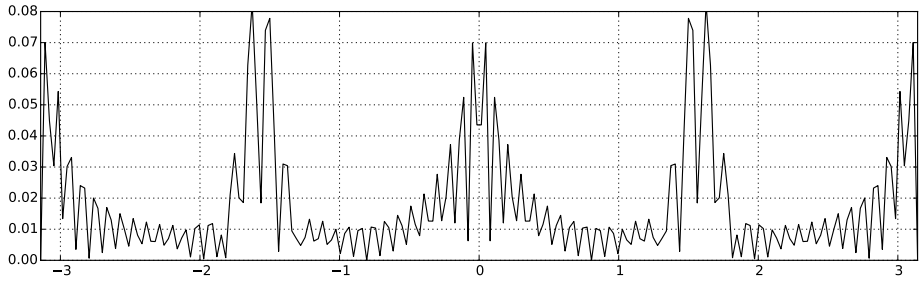


Fig. 4: The discrepancy due to non-preservation of scalar products under a standard spectral Galerkin discretization. The discrepancy of Algorithm 3 (not plotted) is attributable to our time-discretization scheme where we only tolerated error of 10^{-8} in this instance.

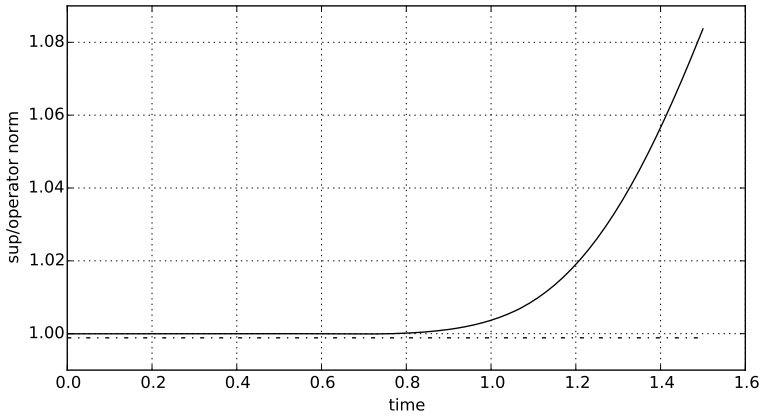


Fig. 5: A plot of the sup-norm vs time of a standard spectral discretization (blue) and the result of Algorithm 3 (red) on the example described in Section 8.1.

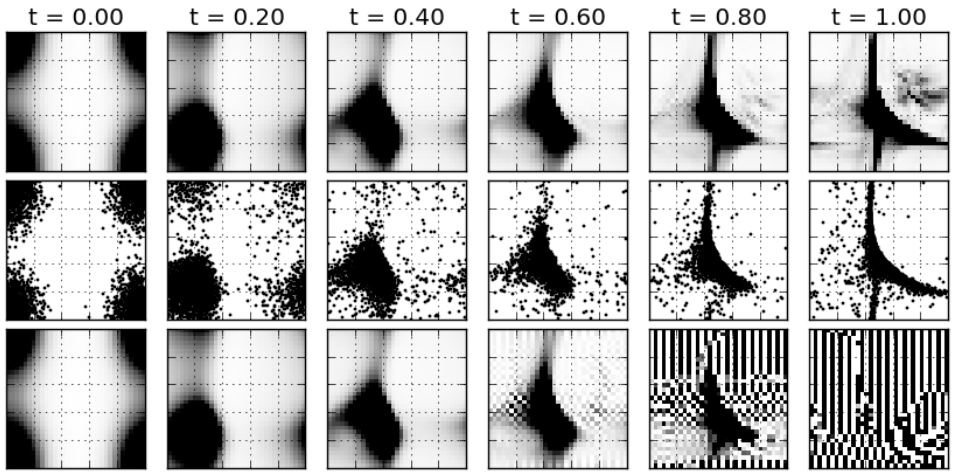


Fig. 6: An illustration of the performance of Algorithm 2 (top row), Monte Carlo (middle row) and a standard spectral Galerkin (bottom row) on the example described in Section 8.2. The domain is the 2-torus. Here we've consider an initial probability density given by a wrapped Gaussian. Darker regions represent areas of higher-density.