



This is the accepted manuscript made available via CHORUS. The article has been published as:

Nonuniform mixing

Jean-Luc Thiffeault

Phys. Rev. Fluids **6**, 090501 — Published 13 September 2021

DOI: [10.1103/PhysRevFluids.6.090501](https://doi.org/10.1103/PhysRevFluids.6.090501)

Nonuniform mixing

Jean-Luc Thiffeault

*Department of Mathematics, University of Wisconsin – Madison,
480 Lincoln Dr., Madison, WI 53706, USA*

Fluid mixing usually involves the interplay between advection and diffusion, which together cause any initial distribution of passive scalar to homogenize and ultimately reach a uniform state. However, this scenario only holds when the velocity field is non-divergent and has no normal component to the boundary. If either condition is unmet, such as for active particles in a bounded region, floating particles, or for filters, the ultimate state after a long time is not uniform, and may be time dependent. We show that in those cases of nonuniform mixing it is preferable to characterize the degree of mixing in terms of an f -divergence, which is a generalization of relative entropy, or to use the L^1 norm. Unlike concentration variance (L^2 norm), the f -divergence and L^1 norm always decay monotonically, even for nonuniform mixing, which facilitates measuring the rate of mixing. We show by an example that flows that mix well for the nonuniform case can be drastically different from efficient uniformly mixing flows.

I. INTRODUCTION

A. Uniform mixing

The standard paradigm for mixing in fluids is as follows [1–5]. Initially, some passive scalar (such as red dye or virus particles) is inhomogeneously distributed in a fluid. Given enough time, the dye would diffuse and spread uniformly throughout the domain; stirring the fluid greatly enhances the speed of this homogenization process.

The ultimate steady state is a fluid with uniform concentration of dye throughout the domain.

The mathematical underpinning for this process is straightforward. The dye concentration $\theta(\mathbf{x}, t)$ obeys the advection-diffusion equation

$$\partial_t \theta + \mathbf{u} \cdot \nabla \theta = D \nabla^2 \theta \quad (\text{I.1})$$

where the velocity field $\mathbf{u}(\mathbf{x}, t)$ is non-divergent ($\nabla \cdot \mathbf{u} = 0$), and $D > 0$ is the dye diffusivity. Since a constant θ solves (I.1), we can assume without loss of generality that $\int_{\Omega} \theta \, dV = 0$, that is, θ has zero mean over the bounded domain Ω . In that case we find after a few integrations by parts

$$\frac{d}{dt} \int_{\Omega} \theta^2 \, dV = -2D \int_{\Omega} |\nabla \theta|^2 \, dV \leq 0 \quad (\text{I.2})$$

where boundary terms vanish, assuming no-flux boundary conditions on θ . Equation (I.2) gives the evolution of the *concentration variance* or L^2 norm of θ , and the nonpositivity of the right-hand side shows that variance will decrease until θ is a constant throughout the whole domain Ω . This constant vanishes because of the zero-mean assumption, so the ultimate steady state is $\theta \equiv 0$ everywhere. We then declare the dye to be mixed. This argument makes no reference to the velocity $\mathbf{u}(\mathbf{x}, t)$, since the terms involving it have integrated away. Equation (I.2) thus cannot be used to get a useful estimate of the rate of mixing. Nevertheless, simply having an equation such as (I.2) is essential in mathematical analysis since it guarantees mixing for long enough times, no matter what the form of \mathbf{u} . It also validates the common use of variance as a measure of the degree of mixing. The right-hand side of Eq. (I.2) is called the *variance dissipation*, and the magnitude of its integrand is a useful proxy for regions where mixing is most active.

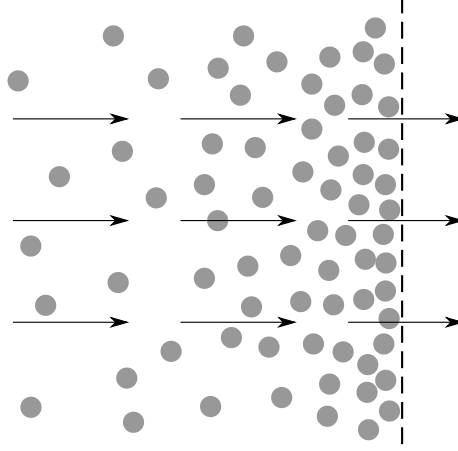


FIG. 1. In the presence of a suction flow, passive particles such as viruses accumulate near a filter. The filter is permeable to fluid but not to particles.

B. Compressibility

When the fluid is compressible, the fluid density $\rho(\mathbf{x}, t) > 0$ is solved for along with the concentration $\theta(\mathbf{x}, t)$, and instead of (I.1) we have the coupled equations

$$\partial_t \rho + \nabla \cdot (\mathbf{u} \rho) = 0, \quad \partial_t (\rho \theta) + \nabla \cdot (\mathbf{u} \rho \theta) = D \nabla^2 \theta. \quad (\text{I.3})$$

Notice that $\theta = \text{const.}$ is still a solution of (I.3), so the ultimate steady state remains uniform. The concentration variance equation (I.2) becomes

$$\frac{d}{dt} \int_{\Omega} \rho \theta^2 dV = -2D \int_{\Omega} |\nabla \theta|^2 dV \leq 0, \quad (\text{I.4})$$

again assuming no-flux boundary conditions on θ . The variance will relax to zero over time, implying that $\theta(\mathbf{x}, t)$ reaches the uniform mixed state. In that sense compressible mixing is also an instance of a uniform mixing scenario.

Note that setting $\rho = \text{const.}$ in (I.3) necessarily implies that $\nabla \cdot \mathbf{u} = 0$. Starting in the next section we shall allow for cases where $\nabla \cdot \mathbf{u} \neq 0$, but where fluid density

does not enter the problem. These cases are not the same as compressible mixing; we shall usually refer to them as *divergent flows* to avoid confusion.

C. Nonuniform mixing

There was an implicit assumption when we stated that $\theta = \text{const.}$ is a steady solution of (I.1): we required $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$ at the boundary $\partial\Omega$. Furthermore, when $\nabla \cdot \mathbf{u} \neq 0$ we must modify (I.1) to read

$$\partial_t \theta + \nabla \cdot (\mathbf{u} \theta) = D \nabla^2 \theta \quad (\text{I.5})$$

to ensure that $\int_{\Omega} \theta \, dV$ is conserved under no-flux boundary conditions. In order for $\theta = \text{const.}$ to be a steady solution of Eq. (I.1) or (I.5), we require both $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$ at the boundary $\partial\Omega$, as well as the non-divergence condition $\nabla \cdot \mathbf{u} = 0$. If either of these conditions is not satisfied, then the steady state is not uniform in space. In fact there may even be no steady state at all, in which case we instead refer to an *ultimate state*, which is reached after a long time. We will define this ultimate state more precisely later.

The no-penetration condition $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$ is usually quite reasonable: it says that fluid doesn't go through the walls. But in many relevant applications the fluid can go through boundaries, even if the passive scalar cannot. We give two examples of such a situation. (Note that we will use the term 'passive scalar' and 'particles' somewhat interchangeably. We usually denote by θ a passive scalar that can have either sign, and by p or n a particle density that cannot be negative.)

Particle filters. If the fluid is air and the passive scalar consists of virus particles, then a filter is a membrane that allows the passage of air but not of viruses (hopefully). This is shown schematically in Fig. 1: the virus particles naturally accumulate at the filter where $\mathbf{u} \cdot \hat{\mathbf{n}} > 0$ due to the suction effect. In this type of situation

the ‘mixed state’ is no longer uniform because of this accumulation.

Active particles. A popular model for 2D self-propelled active particles (so-called Janus particles [6]) assumes that the particles move at a constant speed U , in a swimming direction given by an angle ϕ that changes randomly with time [7, 8]. The probability density of particles $p(\mathbf{x}, \phi, t)$ obeys a Fokker–Planck (or Smoluchowski) equation

$$\partial_t p + (\mathbf{u} + U\hat{\mathbf{q}}) \cdot \nabla p = D \nabla^2 p + D_{\text{rot}} \partial_\phi^2 p \quad (\text{I.6})$$

with $\mathbf{q} = (\cos \phi, \sin \phi)$ and rotational diffusion D_{rot} . Equation (I.6) is exactly analogous to (I.1), except that the domain Ω involves spatial coordinates \mathbf{x} and the angle ϕ . The fluid velocity $\mathbf{u}(\mathbf{x}, t)$ obeys $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$ at boundaries, but the swimming velocity $U\hat{\mathbf{q}}$ does not: a particle may keep pushing against a boundary even after it makes contact. (It is prevented from entering the wall by the no-flux boundary condition on p .) Hence, the steady solution to Eq. (I.6) is not uniform: particles tend to accumulate near boundaries, in a manner similar to the filter example above [9–11].

There are two other effects that can lead to nonuniform ultimate states: divergence of the velocity ($\nabla \cdot \mathbf{u} \neq 0$) and the presence of sources and sinks. We give examples for each case.

Divergent velocity. Floating particles at the surface of the ocean are subjected to the fluid velocity field $\mathbf{u}(\mathbf{x}, t)$ evaluated at $z = 0$. Even though the three-dimensional velocity satisfies $\nabla \cdot \mathbf{u} = 0$, the two-dimensional velocity at the surface is in general divergent. In the long-time limit, particles will tend to congregate at downwellings, where the divergence is negative. The ultimate state is thus nonuniform [12].

The same type of model applies to surfactants, which are concentration scalar fields defined at the surface of a fluid. The equation for a surfactant concentra-

tion $\theta(\mathbf{x}, t)$ evaluated at a free surface is [13, 14]

$$\partial_t \theta + \nabla_s \cdot (\mathbf{u}_s \theta) = D \nabla_s^2 \theta - \theta (\nabla_s \cdot \hat{\mathbf{n}}_s) \mathbf{u} \cdot \hat{\mathbf{n}}_s \quad (\text{I.7})$$

where ∇_s is a gradient along the surface, \mathbf{u}_s is the component of \mathbf{u} parallel to the surface, and $\hat{\mathbf{n}}_s$ is the unit normal to the surface. The source-sink term on the right vanishes if the surface is flat ($\nabla_s \cdot \hat{\mathbf{n}}_s = 0$) or if it is not moving ($\mathbf{u} \cdot \hat{\mathbf{n}}_s = 0$). Even though the three-dimensional velocity is non-divergent, the surface divergence $\nabla_s \cdot \mathbf{u}_s$ is generally nonzero. Equation (I.7) thus has the form of Eq. (I.5), and the surfactant concentration can achieve a nonuniform ultimate state.

Heating a room. In the winter, a closed room may be heated by a space heater, which is a localized source of heat. A closed window somewhere else in the room may act as a sink of heat. The equilibrium state is nonuniform: after a long time, we still expect the temperature to be warmer near the heater, and cooler near the window.

Whenever Eq. (I.5) fails to have a uniform steady state, we are dealing with *nonuniform mixing*: any initial condition $\theta(\mathbf{x}, 0)$ still tends towards an ultimate state, and stirring can accelerate this convergence. However, mixing must be defined with respect to this ultimate state, not the uniform state. Note that this ultimate state may be time-dependent, which challenges our natural notion of mixing even further.

Note that even if $\mathbf{u} \cdot \hat{\mathbf{n}} \neq 0$ on the boundary $\partial\Omega$, it is still typically the case that

$$\int_{\Omega} \rho \mathbf{u} \cdot \hat{\mathbf{n}} \, dS = 0 \quad (\text{I.8})$$

where ρ is the fluid density. Equation (I.8) is a consequence of fluid mass conservation inside Ω . However, we shall not assume that Eq. (I.8) is satisfied in our development, since it is unnecessary, and there are cases where fluid mass might not be conserved (for instance, if there is some external source of fluid, such as rain).

D. Convergence to the ultimate state

When the variance equation (I.2) is modified to allow for a nonuniform ultimate state, as described in Section IC, we will see that it no longer implies monotonic convergence to that state, because of a nonvanishing term that is sign-indefinite. Concentration variance becomes an unreliable measure of mixing, as least from a mathematical viewpoint.

We will show that, in all cases where the ultimate state is nonuniform, the degree of mixing is better captured by a kind of entropy function, related to the relative entropy of information theory and statistical physics. This entropy function has a time evolution that is always nonincreasing, no matter the subtleties of the system, and therefore always predicts convergence to an ultimate state.

We also show that the L^1 norm of θ satisfies

$$\frac{d}{dt} \int_{\Omega} |\theta| dV = -2D \int_{\{\theta=0\}} |\nabla \theta| dS \leq 0 \quad (\text{I.9})$$

where the integral is taken over the zero level set of $\theta(\cdot, t)$. Equation (I.9) holds in the general case, unlike the variance equation (I.2) which depends on the non-divergence $\nabla \cdot \mathbf{u} = 0$ and $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$ at the boundary $\partial\Omega$. Thus, in general the L^1 norm is preferable to the L^2 norm as a measure of mixing, as we will make evident by simple numerical examples. In fact we will show that L^1 is the only L^q norm (with $1 \leq q \leq \infty$) having this monotonic decay property.

The main point of this article is that the non-divergence condition $\nabla \cdot \mathbf{u} = 0$ and no-penetration condition $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$ lead to a very special situation in that the ultimate mixed state is uniform. This is not true if either of these conditions is violated; we categorize the resulting situations as nonuniform mixing. We must then revise what we mean by the rate of mixing: instead of defining it as the rate of approach to a uniform state, it is preferable to use the rate at which any two initial states converge

to each other.

Nonuniform mixing can be very different from traditional mixing. For example, we will show by an example that a constant flow can be an exceedingly good mixer in the presence of suction boundary conditions, whereas such a flow is essentially useless for traditional mixing. The reason is that with the suction conditions the flow presses particles against one wall, which leads to a rapid convergence of any two initial conditions towards each other.

The presentation in this paper is unapologetically mathematical: the aim is to present the precise underpinnings in (hopefully) an agreeable language, without the rigorous burden of function spaces. In addition, the approach presented here relies on some techniques common in the analysis of convergence in Fokker–Planck equations [15–21], which are standard in statistical physics but less well known in fluid dynamics, even though the mathematical framework is similar. One major difference is that in statistical physics one is typically less concerned with specific boundary conditions, since the independent variables are often quantities like momenta, which live in unbounded spaces. In contrast, here we shall pay particularly close attention to the role of boundary conditions. Another difference is that much of the Fokker–Planck literature involves cases where the steady state exists and is easily identified, which will not be the case here for our more complicated, time-dependent examples.

It is worth noting that entropies have been used by several authors to quantify fluid mixing; see for instance [22–30]. However, their approaches are usually based on measuring statistical properties, whereas here we focus directly on differential equations to get rigorous bounds. Entropy and mixing are also often studied in the context of non-equilibrium thermodynamics, but in that case there is usually an equilibrium state such as a Maxwell–Boltzmann distribution towards which the system is tending. Our description will be more general and adapted to the context of fluid mixing. Approaches based on topological entropy [31–33] are complementary

but not closely related to ours, since they focus on properties of trajectories of $\mathbf{u}(\mathbf{x}, t)$.

Our paper is organized as follows. In Section II we define the system and derive some basic results. We consider the simplest ‘traditional’ case of non-divergent flow with impermeable boundary conditions in Section III, and show that the time-evolution equation for variance in that case predicts convergence to a uniform state. In Section IV we relax both the non-divergence and impermeability conditions. Now the ultimate state is no longer uniform, and may not even be steady. The variance equation no longer implies convergence, due to the addition of a sign-indefinite term. We remedy this by introducing the f -divergence associated with two probability densities p_1 and p_2 , a quantity that arises in information theory. (A special case of the f -divergence is the relative entropy of p_1 and p_2 .) We show that the time evolution of the f -divergence is nondecreasing, and that it must eventually decrease to zero.

We give some simple examples for flows that can be fully solved in Section V. In particular, we show that a constant flow with suction boundary conditions can be surprisingly effective at mixing. In Section VI we incorporate the effect of sources and sinks. For those we need to slightly generalize the definition of f -divergence, and we can still show convergence to an ultimate state. We discuss the time evolution of the L^1 norm in Section VII. Finally, we offer some concluding remarks in Section VIII.

II. A PARTICLE IN A CLOSED DOMAIN

Consider a particle in a closed, connected domain Ω . The particle could represent a virus, or some molecule of a pollutant. The particle evolves according to a velocity field (or drift) $\mathbf{u}(\mathbf{x}, t)$ and a diffusion tensor $\mathbb{D}(\mathbf{x}, t)$. The probability of finding the particle in a small volume $dV_{\mathbf{x}}$ centered on \mathbf{x} is $p(\mathbf{x}, t) dV_{\mathbf{x}}$, where the probability

density obeys the Fokker–Planck equation

$$\partial_t p + \nabla \cdot \mathbf{F}(p) = 0, \quad \mathbf{x} \in \Omega, \quad (\text{II.1})$$

with the probability flux defined as

$$\mathbf{F}(p) := \mathbf{u}(\mathbf{x}, t) p - \mathbb{D}(\mathbf{x}, t) \cdot \nabla p. \quad (\text{II.2})$$

The probability flux consists of an advective part and a diffusive part. In the fluid-dynamical context Eq. (II.1) is called an advection-diffusion equation.

The probability density satisfies $p \geq 0$ and $\int_{\Omega} p \, dV \leq 1$. (If the total probability is less than one, then the particle might not be in the domain at all.) We can integrate Eq. (II.1) over Ω and use the divergence theorem to get

$$\frac{d}{dt} \int_{\Omega} p(\mathbf{x}, t) \, dV = - \int_{\partial\Omega} \mathbf{F}(p) \cdot d\mathbf{S}, \quad (\text{II.3})$$

where $d\mathbf{S} = \hat{\mathbf{n}} \, dS$, with $\hat{\mathbf{n}}$ the outward unit normal to the boundary $\partial\Omega$. Equation (II.3) makes it clear that we can conserve total probability by requiring the no-flux boundary condition

$$\mathbf{F}(p) \cdot \hat{\mathbf{n}} = 0, \quad \mathbf{x} \in \partial\Omega. \quad (\text{II.4})$$

It is important to note that we have not made any assumptions on $\mathbf{u}(\mathbf{x}, t)$, other than a bit of smoothness. In particular we did not assume $\nabla \cdot \mathbf{u} = 0$. In addition, we did not assume $\mathbf{u} \cdot \hat{\mathbf{n}} = 0$, so the boundary condition Eq. (II.4) is of mixed type (i.e., a linear combination of p and ∇p).

We spoke of one particle in this section, but the description works equally well for N non-interacting particles, with N fixed, or if p is a non-negative quantity such as heat, appropriately normalized. Later in Section VI, we will introduce sources and sinks, so that N will be allowed to vary.

III. NON-DIVERGENT FLOW WITH IMPERMEABLE BOUNDARY

We make the additional assumptions

$$\nabla \cdot \mathbf{u} = 0, \quad \mathbf{x} \in \Omega, \quad (\text{III.1a})$$

$$\mathbf{u} \cdot \hat{\mathbf{n}} = 0, \quad \mathbf{x} \in \partial\Omega. \quad (\text{III.1b})$$

Equation (III.1a) is the *non-divergence condition*, and Eq. (III.1b) is the *impermeability condition*. The no-flux boundary condition (II.4) reduces to $\hat{\mathbf{n}} \cdot \mathbb{D} \cdot \nabla p = 0$.

Observe that, under conditions (III.1), Eq. (II.1) with boundary conditions (II.4) has the steady solution $p = \varphi$, with

$$\varphi(\mathbf{x}) = |\Omega|^{-1}, \quad (\text{III.2})$$

where $|\Omega|$ is the volume of Ω , so that $\int_{\Omega} \varphi \, dV = 1$. The solution $\varphi(\mathbf{x}) = |\Omega|^{-1}$ is called the uniform density on Ω . Two important remarks are in order: (i) *both* conditions in (III.1) are necessary for Eq. (III.2) to be a steady solution; (ii) Eq. (III.2) is a steady solution even when $\mathbf{u}(\mathbf{x}, t)$ and $\mathbb{D}(\mathbf{x}, t)$ are explicit functions of time.

We define mixing as the tendency for any initial condition $p(\mathbf{x}, t_0)$ to converge to $\varphi(\mathbf{x})$ as $t \rightarrow \infty$. A traditional way of characterizing this convergence is to first define the anomaly

$$\theta(\mathbf{x}, t) := p(\mathbf{x}, t) - \varphi(\mathbf{x}) \quad (\text{III.3})$$

so that $\int_{\Omega} \theta \, dV = 0$. The *variance* is then $\int_{\Omega} \theta^2 \, dV$; after a few integrations by parts, we find that it evolves according to

$$\frac{d}{dt} \int_{\Omega} \theta^2 \, dV = \int_{\Omega} \mathbf{u} \cdot \nabla \theta^2 \, dV - 2 \int_{\Omega} \nabla \theta \cdot \mathbb{D} \cdot \nabla \theta \, dV. \quad (\text{III.4})$$

The first integral on the right vanishes: from (III.1a) $\mathbf{u} \cdot \nabla \theta^2 = \nabla \cdot (\mathbf{u} \theta^2)$, followed by the divergence theorem and then (III.1b). Next we require that there exists a

constant $\sigma > 0$ such that

$$\mathbf{v} \cdot \mathbb{D}(\mathbf{x}, t) \cdot \mathbf{v} \geq \sigma |\mathbf{v}|^2 > 0, \quad \text{for all vectors } \mathbf{v}, \quad (\text{III.5})$$

i.e., the operator $\nabla \cdot (\mathbb{D} \cdot \nabla p)$ is *uniformly elliptic*. With (III.5), Eq. (III.4) now gives

$$\frac{d}{dt} \int_{\Omega} \theta^2 dV \leq -2\sigma \int_{\Omega} |\nabla \theta|^2 dV \leq -2\sigma \lambda \int_{\Omega} \theta^2 dV \quad (\text{III.6})$$

where in the last step we used the Poincaré–Wirtinger inequality $\|\nabla \theta\|_2^2 \geq \lambda \|\theta\|_2^2$ for a mean-zero function θ [34]. The constant $\lambda > 0$ depends only on the domain Ω . Grönwall’s lemma then yields the bound

$$\int_{\Omega} \theta^2(\mathbf{x}, t) dV \leq e^{-2\sigma\lambda(t-t_0)} \int_{\Omega} \theta^2(\mathbf{x}, t_0) dV \quad (\text{III.7})$$

which goes to zero as $t \rightarrow \infty$. We conclude that $\theta \rightarrow 0$, or $p \rightarrow \varphi$. Thus the ultimate fate of any initial $p(\mathbf{x}, t_0)$ is to be homogenized until the probability of finding the particle anywhere in Ω is uniform. The rate at which this happens is of order $2\sigma\lambda$, though this is generally an underestimate. In practice, the action of $\mathbf{u}(\mathbf{x}, t)$, called *stirring*, amplifies gradients so that $\nabla \theta$ in Eq. (III.4) can be much larger than required by the Poincaré–Wirtinger inequality. Nevertheless, Eq. (III.7) is useful in that it proves that variance *must* converge to zero. What we have just described is the basic idea of what is traditionally meant by mixing in the fluids community.

What can happen if we violate the uniform ellipticity condition Eq. (III.5)? For example, consider the heat equation

$$\partial_t p = T'(t) \nabla^2 p \quad (\text{III.8})$$

with time-dependent diffusion coefficient $D(t) = T'(t)$. If $T'(t) \sim t^{-\alpha}$ for large time, then the uniform ellipticity condition is violated when $\alpha > 0$. We can rescale and use T as a time coordinate, in which case we expect a long-time exponential decay of the form

$$p(\mathbf{x}, t) - |\Omega|^{-1} \sim e^{-\gamma T(t)}, \quad (\text{III.9})$$

where γ is the asymptotic decay rate for $T'(t) = 1$. Since $T(t) \sim t^{-\alpha+1}$, we see that p will fail to converge to the uniform density for $\alpha > 1$. Thus, the condition (III.5) is only sufficient: there may still be convergence to equilibrium even if it is not satisfied.

IV. DIVERGENT FLOW OR PERMEABLE BOUNDARY

The situation described in Section III is straightforward: for any velocity field $\mathbf{u}(\mathbf{x}, t)$ and diffusion tensor $\mathbb{D}(\mathbf{x}, t)$, we can expect convergence to a uniform density as long as conditions (III.1) and (III.5) are satisfied. Now we investigate what happens when either the flow is divergent (Eq. (III.1a) not satisfied), or when there is suction of fluid through the boundary (Eq. (III.1b) not satisfied).

First consider the autonomous case where $\mathbf{u}(\mathbf{x}, t) \rightarrow \mathbf{u}(\mathbf{x})$, and $\mathbb{D}(\mathbf{x}, t) \rightarrow \mathbb{D}(\mathbf{x})$. Then there is an equilibrium density $\varphi(\mathbf{x}) > 0$ that satisfies

$$\nabla \cdot (\mathbf{u} \varphi - \mathbb{D} \cdot \nabla \varphi) = 0, \quad \mathbf{x} \in \Omega; \quad \mathbf{F}(\varphi) \cdot \hat{\mathbf{n}} = 0, \quad \mathbf{x} \in \partial\Omega \quad (\text{IV.1})$$

and is normalized: $\int_{\Omega} \varphi \, dV = 1$. We can then define the anomaly as we did in Eq. (III.3); the only difference is that the reference state $\varphi(\mathbf{x})$ is no longer uniform. The variance evolution equation (III.4) is still valid, but now the first integral term on the right now longer vanishes. This term is not sign-definite: this means that we can no longer conclude from this equation alone that variance must decay. In fact, variance *does* eventually decay, but it might not do so monotonically. Equation (III.4) alone is not enough to conclude that p converges to φ .

It would be convenient, then, to have a quantity other than variance that does decay monotonically in this general case. To that end, consider the f -divergence of two normalized probability densities $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$ [35, 36]:

$$H_f[p_1, p_2] := \int_{\Omega} p_2 f(p_1/p_2) \, dV. \quad (\text{IV.2})$$

Here $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ is an arbitrary convex function with $f(1) = 0$. The f -divergence is non-negative; indeed, since p_2 is a probability density, by Jensen's inequality for the convex function f we have

$$H_f[p_1, p_2] \geq f\left(\int_{\Omega} (p_1/p_2) p_2 \, dV\right) = f(1) = 0. \quad (\text{IV.3})$$

The f -divergence is zero if and only if $p_1 \equiv p_2$; $H_f[p_1, p_2]$ measures how different p_1 and p_2 are from each other—hence the name ‘divergence.’ The f -divergence is not generally a metric for probability densities, since $H_f[p_1, p_2] \neq H_f[p_2, p_1]$, though for certain choices of f it can be made symmetric (see below).

We now further assume that f is strictly convex and twice-differentiable. If each $p_i(\mathbf{x}, t)$ evolves according to Eq. (II.1), with no-flux boundary condition (II.4), we show in Appendix A that

$$\dot{H}_f[p_1, p_2] = - \int_{\Omega} p_2 f''(p_1/p_2) \nabla(p_1/p_2) \cdot \mathbb{D} \cdot \nabla(p_1/p_2) \, dV \leq 0 \quad (\text{IV.4})$$

since $f'' > 0$ for a strictly convex function. For \mathbb{D} satisfying (III.5), notice that the right-hand side of (IV.4) is zero if and only if $p_1 \equiv p_2$. Hence, any two solutions to $\partial_t p = -\nabla \cdot \mathbf{F}(p)$ converge to each other; in the autonomous case they converge to the fixed point $p = \varphi$.

We emphasize that Eq. (IV.4) holds for any divergent flow, possibly with suction boundary conditions, with time-dependent \mathbf{u} and \mathbb{D} . In that sense the f -divergence is a better descriptor of mixing than variance: it monotonically decreases for any flow. The evolution equation (IV.4) also suggests how to define mixing in the non-autonomous context: p_1 and p_2 converge to some ultimate state $\varphi(\mathbf{x}, t)$, which is ‘locked’ to the time-dependence of \mathbf{u} and \mathbb{D} . Thus, the main characteristic of mixing is not that it leads to a homogeneous state, but rather that it leads to a state that has completely forgotten the initial condition. This ultimate state must be unique (for connected Ω), otherwise (IV.4) leads to a contradiction. Unfortunately, extracting

an explicit bound on the decay rate from (IV.4) is much more challenging than it was in the case of variance in Eq. (III.7), and is still a topic of ongoing research [17–20].

The discussion of H_f so far did not depend on a choice of the convex function f in (IV.2), as long as it exists. A simple choice for f is

$$f(u) = u \log u, \quad f''(u) = 1/u, \quad (\text{IV.5})$$

which corresponds to the *relative entropy* or *Kullback–Leibler divergence* (KLD), denoted by $H_{\text{KL}}(p_1, p_2)$ [37] [38]:

$$H_{\text{KL}}(p_1, p_2) = \int_{\Omega} p_1 \log(p_1/p_2) \, dV. \quad (\text{IV.6})$$

The KLD can be interpreted as the amount of information lost when p_2 is used to approximate p_1 . The KLD bounds the L^1 norm by *Pinsker’s inequality*:

$$\|p_1 - p_2\|_1^2 \leq 2 \log 2 \, H_{\text{KL}}(p_1, p_2). \quad (\text{IV.7})$$

However, $H_{\text{KL}}(p_1, p_2)$ is not symmetric in p_1 and p_2 , and is unbounded when p_2 vanishes anywhere in Ω . (We will discuss the time evolution of $\|p_1 - p_2\|_1$ in Section VII.)

A slightly more involved choice for f is

$$f(u) = \frac{1}{2}u \log u - \frac{1}{2}(1+u) \log[\frac{1}{2}(1+u)] \quad f''(u) = (2u(1+u))^{-1}, \quad (\text{IV.8})$$

which leads to the *Jensen–Shannon divergence* (JSD), denoted by $H_{\text{JS}}(p_1, p_2)$ [39]:

$$\begin{aligned} H_{\text{JS}}(p_1, p_2) &= \frac{1}{2} \{ H_{\text{KL}}(p_1, p_{12}) + H_{\text{KL}}(p_2, p_{12}) \} \\ &= \frac{1}{2} \int_{\Omega} \{ p_1 \log(p_1/p_{12}) + p_2 \log(p_2/p_{12}) \} \, dV \end{aligned} \quad (\text{IV.9})$$

where $p_{12} := \frac{1}{2}(p_1 + p_2)$. The JSD is symmetric in p_1 and p_2 , and its square root is a metric. Moreover, it is bounded:

$$H_{\text{JS}}(p_1, p_2) \leq \frac{1}{2}(\log 2) \|p_1 - p_2\|_1 \leq \log 2. \quad (\text{IV.10})$$

An intuitive interpretation of the JSD is not so straightforward, and will not be needed here; see for instance [39].

One remark is in order: notice that in (IV.2) and (IV.4) there are several divisions by p_2 , which should rightfully worry the reader since potentially p_2 could vanish at some points, for instance at the initial time. However, for any positive time p_2 immediately becomes strictly positive, because of diffusion. See the discussion in Arnold *et al.* [17, p. 161] for more careful considerations.

V. ONE-DIMENSIONAL EXAMPLES

To summarize the previous sections: for a velocity field $\mathbf{u}(\mathbf{x}, t)$ and diffusion tensor $\mathbb{D}(\mathbf{x}, t)$, we seek solutions to the advection-diffusion equation (II.1) with no-flux boundary conditions (II.4). Then the possible scenarios, in increasing order of complexity, can be characterized as follows.

1. If both conditions in (III.1) hold, then the uniform density is $\varphi(\mathbf{x}) = 1/|\Omega|$. This is true whether or not \mathbf{u} and \mathbb{D} are explicitly time-dependent (i.e., autonomous or non-autonomous). In this case the variance evolution equation (III.4) is sufficient to directly show convergence to the uniform state.
2. If either condition in (III.1) is unsatisfied, then there are two cases:
 - (a) For \mathbf{u} and \mathbb{D} time-independent (autonomous), any initial $p(\mathbf{x}, t_0)$ converges to a nonuniform invariant density $\varphi(\mathbf{x})$.
 - (b) For \mathbf{u} and \mathbb{D} time-periodic with period τ ,

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{u}(\mathbf{x}, t + \tau), \quad \mathbb{D}(\mathbf{x}, t) = \mathbb{D}(\mathbf{x}, t + \tau), \quad (\text{V.1})$$

any initial condition $p(\mathbf{x}, t_0)$ converges to a periodic limiting invariant density $\varphi(\mathbf{x}, t)$, with $\varphi(\mathbf{x}, t) = \varphi(\mathbf{x}, t + \tau)$.

- (c) For \mathbf{u} and \mathbb{D} time-dependent (non-autonomous), any initial condition $p(\mathbf{x}, t_0)$ converges to a time-dependent limiting invariant density $\varphi(\mathbf{x}, t)$.

In case 2 the f -divergence evolution equation (IV.4) can be used directly to show convergence to $\varphi(\mathbf{x}, t)$.

Since case 1 is familiar from the traditional view of mixing, we will give explicit examples for the subcategories of case 2.

A. Example of case 2(a): Convergence to a nonuniform density

Consider a simple one-dimensional model where the domain $\Omega = [0, L]$, the velocity $\mathbf{u} = U \hat{\mathbf{x}}$, and $(\mathbb{D})_{ij} = D \delta_{ij}$, with U and D constants. Then (II.1) simplifies to

$$\partial_t p + U \partial_x p - D \partial_x^2 p = 0, \quad 0 < x < L \quad (\text{V.2})$$

with no-flux boundary conditions

$$Up - D \partial_x p = 0, \quad x = 0, L. \quad (\text{V.3})$$

This may be regarded as a simple model of a filter: the flow is non-divergent and can pass through the membranes at $x = 0$ and L , but particles cannot cross those membranes. Since the velocity and diffusivity are time-independent, Eq. (V.2) has the invariant density

$$\varphi(x) = \frac{U}{D} \frac{e^{Ux/D}}{e^{UL/D} - 1}. \quad (\text{V.4})$$

The flow pushes particle against the boundary at $x = L$ (for $U > 0$), creating a boundary layer of thickness D/U .

Now we solve the initial value problem for Eq. (V.2). This is most generally done in terms of the Green's function $p = P(x, t | x_0, t_0) = P(x, t - t_0 | x_0, 0)$, which

satisfies (V.2)–(V.3) with initial condition $P(x, t_0 | x_0, t_0) = \delta(x - x_0)$. The solution is not completely straightforward, since the PDE is not self-adjoint, but it can be obtained using Laplace transforms as

$$P(x, t | x_0, t_0) = \varphi(x) + \frac{D}{L^3} \sum_{n=1}^{\infty} \frac{e^{-\gamma_n(t-t_0)}}{2\gamma_n} \phi_n^U(x) \phi_n^{-U}(x_0), \quad (\text{V.5})$$

with

$$\phi_n^U(x) = e^{Ux/2D} \{2\pi n \cos(n\pi x/L) + (|U|L/D) \sin(n\pi x/L)\} \quad (\text{V.6})$$

and decay rates

$$\gamma_n = D(\pi n/L)^2 + U^2/4D. \quad (\text{V.7})$$

The Green's function is plotted in Fig. 2. The relaxation rate to the invariant density $\varphi(x)$ is given by γ_1 , which is considerably enhanced by the constant flow U : the second term $U^2/4D$ is dominant for $UL/D > 2\pi$. Thus, unlike in ‘traditional’ mixing problems, a constant velocity can accelerate mixing substantially (though for a large domain size there could be an initial transient before the concentration reaches the wall). This acceleration is due to the flow squashing the concentration field against the boundary. Superficially, this does not sound like mixing, but it is in the sense that it causes the scalar field $p(x, t)$ to quickly forget its initial condition and converge to the invariant density $\varphi(x)$.

B. Example of case 2(b): Convergence to a time-periodic density

To illustrate convergence to a time-periodic invariant density $\varphi(x, t)$, we use the same system (V.2)–(V.3) as in the previous example. We mimic a time-periodic flow by reversing the direction of $\mathbf{u} = \pm U \hat{\mathbf{x}}$ at every half-period $\tau/2$. (This could represent the air flow reversing direction as a mask-wearer inhales and exhales.)

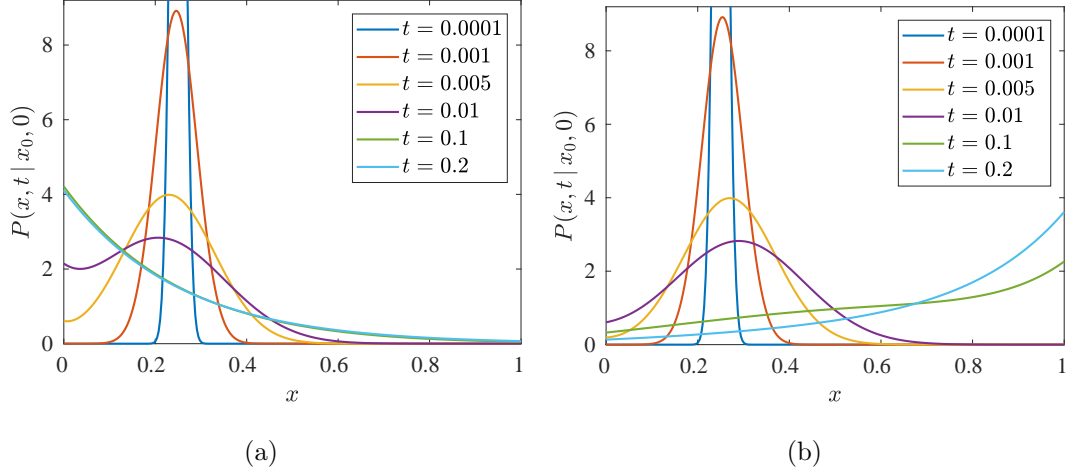


FIG. 2. The Green's function (V.5) with $x_0 = 1/4$, plotted at different times for $D = L = 1$ and (a) $U = -4$; (b) $U = 4$. The ultimate state is the invariant density Eq. (V.4).

Thus, the density $p(x, t)$ at time t is evolved to time $t + \frac{1}{2}\tau$ by

$$p(x, t + \frac{1}{2}\tau) = \int_0^L P_U(x, \frac{1}{2}\tau | x_0, 0) p(x_0, t) dx_0 \quad (\text{V.8})$$

where P_U is the Green's function (V.5); then, for the next half-period, we evolve the density with a flow $\mathbf{u} = -U\hat{\mathbf{x}}$ to the left:

$$\begin{aligned} p(x, t + \tau) &= \int_0^L P_{-U}(x, \frac{1}{2}\tau | x'_0, 0) p(x'_0, t + \frac{1}{2}\tau) dx'_0 \\ &= \int_0^L \mathcal{P}_\tau(x | x_0) p(x_0, t) dx_0 \end{aligned} \quad (\text{V.9})$$

where the period- τ kernel is

$$\mathcal{P}_\tau(x | x_0) := \int_0^L P_{-U}(x, \frac{1}{2}\tau | x'_0, 0) P_U(x'_0, \frac{1}{2}\tau | x_0, 0) dx'_0. \quad (\text{V.10})$$

Note that t in Eq. (V.9) is not arbitrary but is aligned with period boundaries: $t = t_k = k\tau$, for integer k . Equation (V.9) maps the density $p(x, t_k)$ to the beginning

of the next period at time $t_{k+1} = (k+1)\tau$. The invariant density $\varphi(x, t_k)$ may be found from

$$\varphi(x, t_k) = \int_0^L \mathcal{P}_\tau(x | x_0) \varphi(x, t_k) dx_0 \quad (\text{V.11})$$

which is a Fredholm integral equation of the second kind. Here $\varphi(x, t_k)$ is the periodic invariant density evaluated at the start of a period. Even for this simple time-periodic example it is not straightforward to compute $\varphi(x, t_k)$, or the rate of convergence to $\varphi(x, t_k)$.

In Fig. 3 we show a numerical solution of Eq. (V.9), for two different initial conditions: the first ($p_1(x, 0)$) has particles initially concentrated on the right side of the interval, and the second ($p_2(x, 0)$) on the left. The two solutions rapidly converge to each other after about 3 periods. The ultimate state $\varphi(x, t)$ may be considered ‘mixed’ even if it is not uniform. In Fig. 4 we compare the time evolution of variance $\int_\Omega |p_1 - p_2|^2 dV$ to the Jensen–Shannon divergence (IV.9). The variance is not at all monotonic: it oscillates about a decreasing trend. The JSD is nice and monotonic, which makes it much easier to assign a numerical value to the decay rate.

C. Example of case 2(c): Convergence to an aperiodic density

A simple way to produce an example that is neither steady nor time-periodic is to add some randomness [40]. Recall that in the periodic example of Section V B we imposed a flow $U\hat{\mathbf{x}}$ to the right for a time $\frac{1}{2}\tau$, followed by a flow $-U\hat{\mathbf{x}}$ to the left for a time $\frac{1}{2}\tau$, to obtain a period- τ map. One simple way to randomize this process is to select for every time interval $[t_k, t_k + \tau)$ a uniform independent random number $\alpha_k \in [0, 1]$, and impose a flow to the right for a time $\alpha_k\tau$, followed by a flow to the left for a time $(1 - \alpha_k)\tau$. The kernel Eq. (V.10) is then replaced by

$$\mathcal{P}_{\tau, \alpha_k}(x | x_0) := \int_0^L P_{-U}(x, (1 - \alpha_k)\tau | x'_0, 0) P_U(x'_0, \alpha_k\tau | x_0, 0) dx'_0 \quad (\text{V.12})$$

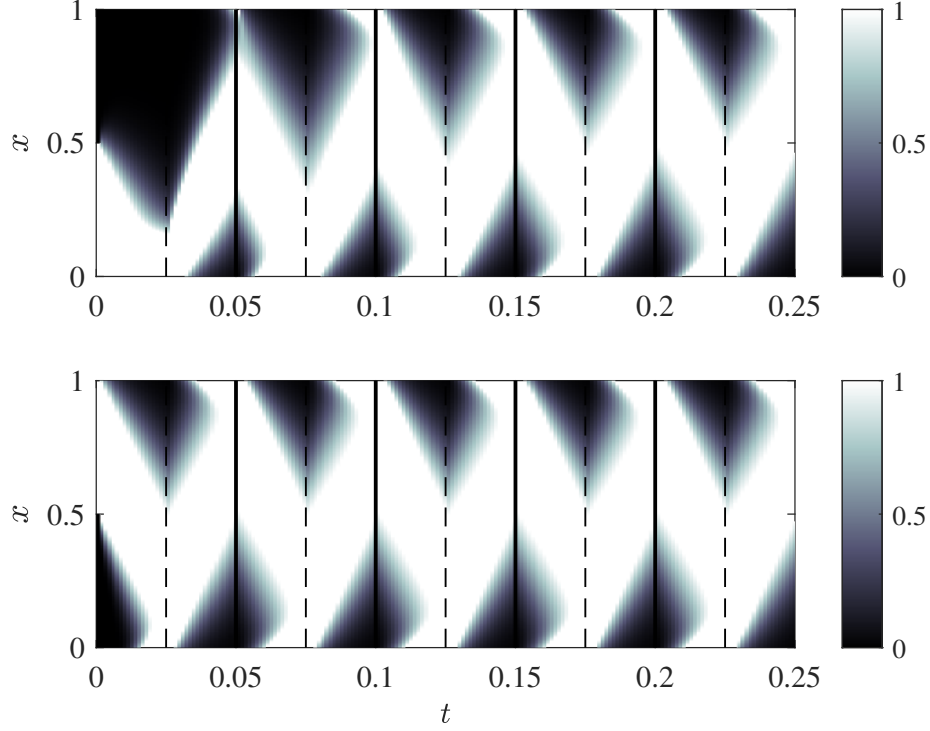


FIG. 3. Density $p(x, t)$ the periodic flow of Section V B, for two different initial conditions. The two initial conditions converge to the same periodic pattern $\varphi(x, t)$ after about 3 periods. Here the period $\tau = 0.05$ and drift $U = 20$, for a domain of width $L = 1$ and with diffusivity $D = 1$. The vertical lines indicate period boundaries, and the dashed lines are half-periods when the flow switches from right to left.

and the map from time t_k to $t_{k+1} = t_k + \tau$ is

$$p(x, t_k + \tau) = \int_0^L \mathcal{P}_{\tau, \alpha_k}(x | x_0) p(x_0, t_k) dx_0. \quad (\text{V.13})$$

In Fig. 5 we show a numerical solution of Eq. (V.13), for two different initial conditions, which rapidly converge to each other after about 3 periods. The ultimate state $\varphi(x, t)$ is ‘mixed’ even though it is neither uniform nor periodic. In Fig. 6 we compare the time evolution of variance $\int_{\Omega} |p_1 - p_2|^2 dV$ to the Jensen–Shannon

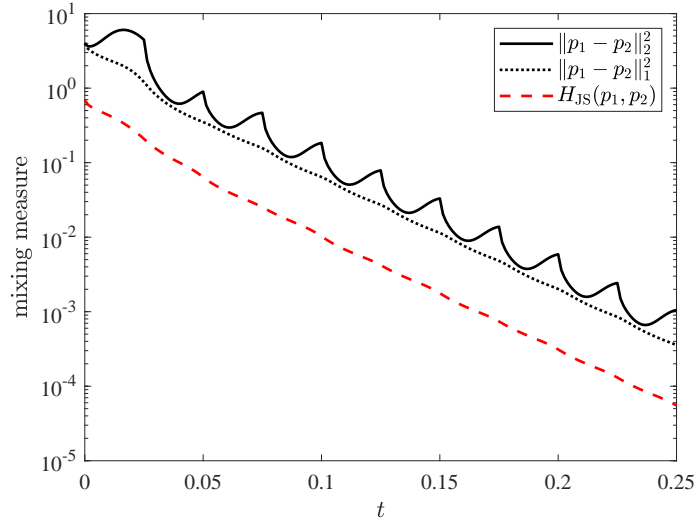


FIG. 4. Variance or L^2 norm (solid line), squared L^1 norm (dotted line), and Jensen–Shannon divergence (dashed line) between the two solutions in Fig. 3. The variance is nonmonotonic, whereas the L^1 norm and H_{JS} decreases monotonically.

divergence (IV.9). Much like the periodic case, the variance is not at all monotonic, whereas the JSD relentlessly decreases towards zero.

VI. SOURCES AND SINKS

A. Varying the number of particles

So far the number of particles was fixed. Now consider the particle density $n(\mathbf{x}, t) \geq 0$ (also sometimes called particle number or number density), which obeys the equation

$$\partial_t n + \nabla \cdot \mathbf{F}(n) = Q(\mathbf{x}, t; n), \quad \mathbf{x} \in \Omega, \quad (\text{VI.1})$$

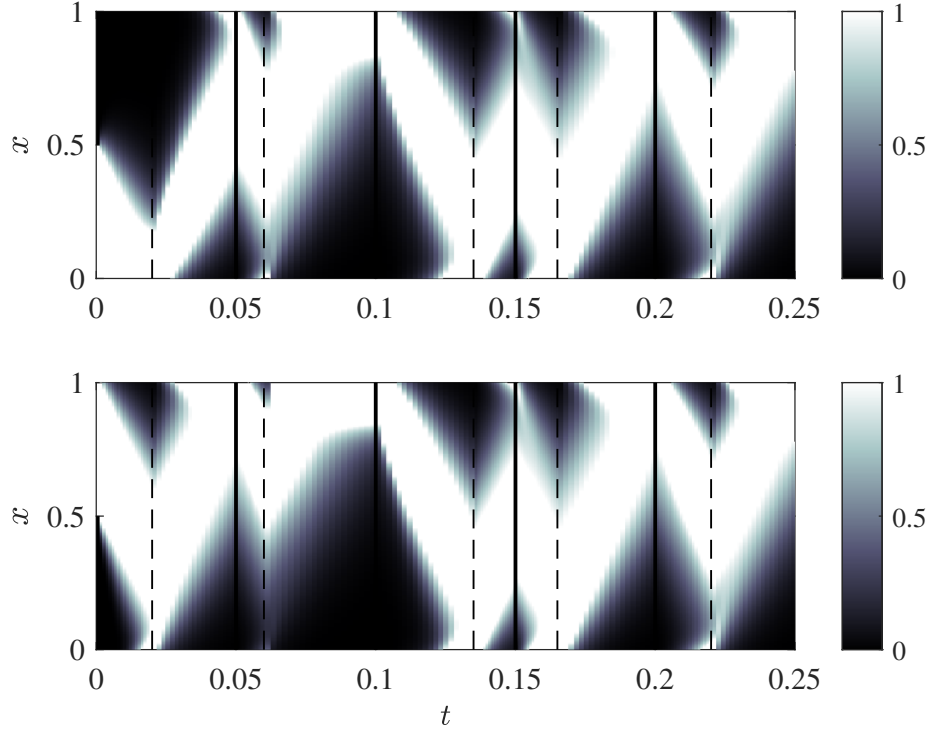


FIG. 5. Density $p(x, t)$ the random flow of Section V C, for two different initial conditions. The two initial conditions rapidly converge to the same random pattern $\varphi(x, t)$. The vertical lines indicate period boundaries t_k , and the dashed lines are the random times $t_k + \alpha_k \tau$ when the flow switches from right to left. Parameter values are as in Fig. 3.

with the particle flux $\mathbf{F}(n) = \mathbf{u}n - \mathbb{D} \cdot \nabla n$ defined as in (II.2). The particle density differs from the probability density $p(\mathbf{x}, t)$ in that the number of particles

$$N(t) = \int_{\Omega} n(\mathbf{x}, t) dV \quad (\text{VI.2})$$

is not 1, and can change with time. The number of particles $N(t)$ is not in general an integer. This can either be interpreted as a small error when N is very large, or n/N can be interpreted as a probability.

The source-sink function $Q(\mathbf{x}, t; n)$ is not completely arbitrary: it must preserve

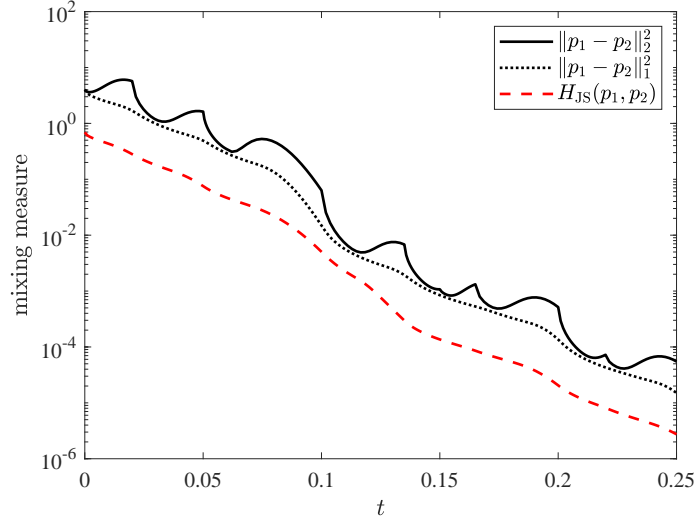


FIG. 6. Variance or L^2 norm (solid line), squared L^1 norm (dotted line), and Jensen–Shannon divergence (dashed line) between the two solutions in Fig. 5. The variance is nonmonotonic, whereas the L^1 norm and H_{JS} decreases monotonically.

the positivity of n . There is an asymmetry between adding and removing particles: we can always add particles, but we can only remove particles if there are particles present. One common form of Q that naturally enforces this is

$$Q(\mathbf{x}, t; n) = S(\mathbf{x}, t) - K(\mathbf{x}, t) n, \quad S \geq 0, K \geq 0 \quad (\text{VI.3})$$

for given non-negative functions S and K . The source S creates particles indiscriminately, but the sink $-Kn$ vanishes as $n \rightarrow 0$. Of course, more general forms than (VI.3) are possible.

The same considerations for the interior source-sink apply to the flux at the boundary: we should not remove particles if there are none present. Thus, we write for the boundary flux

$$q(\mathbf{x}, t; n) = -\mathbf{F}(n) \cdot \hat{\mathbf{n}} = s(\mathbf{x}, t) - k(\mathbf{x}, t) n, \quad s \geq 0, k \geq 0, \quad \mathbf{x} \in \partial\Omega \quad (\text{VI.4})$$

for given non-negative boundary functions s and k . The minus sign in front of $\mathbf{F} \cdot \hat{\mathbf{n}}$ in (VI.4) is because $\hat{\mathbf{n}}$ is an outward normal, so $\mathbf{F}(n) \cdot \hat{\mathbf{n}} > 0$ corresponds to particles leaving the domain Ω .

Using Eqs. (VI.1) and (VI.2) and the definition of q in (VI.4), we see that the time evolution of N satisfies

$$\dot{N} = \int_{\Omega} Q \, dV + \int_{\partial\Omega} q \, dS \quad (\text{VI.5})$$

where the first term is the ‘bulk’ source of particles, and the second is the flux of particles across the boundary of Ω .

B. Convergence to asymptotic state

In Section IV we showed that, for one particle (or equivalently a fixed number of noninteracting particles) we can use the f -divergence to prove that any two initial conditions will converge to the same ultimate state $\varphi(\mathbf{x}, t)$. The ultimate state may be nonuniform and time-dependent, but what characterizes it is that it is independent of the initial condition: it is an asymptotic state.

Having now allowed for sources and sinks in Section VIA, we can ask about defining $\varphi(\mathbf{x}, t)$ in that case. After all, adding and removing particles should not prevent two arbitrary initial conditions from converging to each other, as long as they are subjected to the same sources and sinks.

We define the difference $\theta = n_1 - n_2$ between any two solutions of Eq. (VI.1). The squared-integral of θ obeys an equation analogous to the variance evolution Eq. (III.4):

$$\begin{aligned} \frac{d}{dt} \int_{\Omega} \theta^2 \, dV = & \int_{\Omega} \mathbf{u} \cdot \nabla \theta^2 \, dV - 2 \int_{\Omega} \nabla \theta \cdot \mathbb{D} \cdot \nabla \theta \, dV \\ & - 2 \int_{\Omega} K \theta^2 \, dV - 2 \int_{\partial\Omega} k \theta^2 \, dV. \end{aligned} \quad (\text{VI.6})$$

The source S does not enter the equation; the last two terms are new but they are nonpositive, so they promote convergence to an equilibrium. However, the same sign-indefinite term involving the integral of $\mathbf{u} \cdot \nabla \theta^2$ appears on the right. This term does go away under the non-divergence and impermeability assumptions (III.1), in which case (VI.6) is enough to conclude convergence to an ultimate state $\varphi(\mathbf{x}, t)$, independent of initial condition.

However, in the divergent or permeable case, we have the same problem as before: the presence of a sign-indefinite term prevents us from guaranteeing convergence. A generalization of the f -divergence (IV.2) is needed, with a time evolution that allows us to conclude convergence. We tentatively define

$$H_f[n_1, n_2] = \int_{\Omega} n_2 f(n_1/n_2) dV. \quad (\text{VI.7})$$

This is not strictly speaking an f -divergence, since n_1 and n_2 are not normalized probability densities. The proof from Eq. (IV.3) that H_f is positive now reads

$$H_f[n_1, n_2] = N_2 \int_{\Omega} f(n_1/n_2) p_2 dV \geq N_2 f\left(\int_{\Omega} (n_1/n_2) p_2 dV\right) = N_2 f(N_1/N_2)$$

where $N_i = \int_{\Omega} n_i dV$, and $p_i = n_i/N_i$ are normalized probability densities. Hence, to guarantee $H_f[n_1, n_2] \geq 0$ we must add the additional requirement that $f \geq 0$, which is satisfied by the Jensen–Shannon choice (IV.8) for f . (To satisfy $f \geq 0$, the Kullback–Leibler choice (IV.5) can simply be modified to read $f(u) = u \log u - u + 1$, in which case H_f is sometimes called the *physical relative entropy*.) With this additional constraint on f , we have $H_f[n_1, n_2] = 0$ if and only if $n_1 \equiv n_2$.

With the same approach as in Appendix A we can show that the time evolution

of $H_f[n_1, n_2]$ is given by

$$\begin{aligned} \dot{H}_f[n_1, n_2] = & - \int_{\Omega} n_2 f''(n_1/n_2) \nabla(n_1/n_2) \cdot \mathbb{D} \cdot \nabla(n_1/n_2) dV \\ & - \int_{\Omega} (K n_2 f(n_1/n_2) + S g_f(n_1/n_2)) dV \\ & - \int_{\partial\Omega} (k n_2 f(n_1/n_2) + s g_f(n_1/n_2)) dS \leq 0, \quad (\text{VI.8}) \end{aligned}$$

where

$$g_f(u) := (u - 1)f'(u) - f(u) \geq 0, \quad g_f(1) = 0. \quad (\text{VI.9})$$

The inequality in (VI.8) follows from the positivity of n_i , the strict convexity of f ($f'' > 0$), the positive-definiteness of \mathbb{D} , the non-negativity of f , K , S , k , s , and the inequality (VI.9). (The latter is easy to prove: a differentiable convex function satisfies $f(x) \geq f(y) + (x - y)f'(y)$ for all x, y , since its graph is above all its tangents; set $x = 1$ and $y = u$ and use $f(1) = 0$.) The right-hand side of (VI.8) vanishes if and only if $n_1 = n_2$.

VII. TOTAL VARIATION DISTANCE AND L^1 NORM

As an alternative to the f -divergence, another measure of convergence of two densities is the *total variation distance* (or *variational distance*), which is equivalent to $\frac{1}{2}\|p_1 - p_2\|_1$ [37], where $\|\cdot\|_1$ is the L^1 norm on Ω . Compare the evolution of $\|p_1 - p_2\|_1^2$ to the concentration variance $\|p_1 - p_2\|_2^2$ in Figs. 4 and 6. Notice that the L^1 norm, much like H_f , decays monotonically, exhibiting none of the troublesome oscillations of the L^2 norm (variance). In this section we will show that the L^1 norm does indeed always decrease monotonically, so that it is a more reliable measure of mixing than the L^2 norm for nonuniform mixing.

We shall prove this for two general number densities n_1 and n_2 obeying Eq. (VI.1) with the source-sink (VI.3), and with boundary conditions (VI.4). Let $\theta = n_1 - n_2$,

which satisfies $\partial_t \theta = -\nabla \cdot \mathbf{F}(\theta) - K\theta$ and $\mathbf{F}(\theta) \cdot \hat{\mathbf{n}} = k\theta$ on $\partial\Omega$. For any function $G(\theta)$, we have

$$\begin{aligned} \frac{d}{dt} \int_{\Omega} G(\theta) dV &= \int_{\Omega} G''(\theta) \theta \mathbf{u} \cdot \nabla \theta dV - \int_{\Omega} G''(\theta) \nabla \theta \cdot \mathbb{D} \cdot \nabla \theta dV \\ &\quad - \int_{\Omega} G'(\theta) K \theta dV - \int_{\partial\Omega} G'(\theta) k \theta dS \end{aligned} \quad (\text{VII.1})$$

which is a generalization of Eq. (VI.6). Now let $G(\theta) = |\theta|$, so that $G'(\theta) = \text{sgn}(\theta)$ and $G''(\theta) = 2\delta(\theta)$. With that choice, Eq. (VII.1) becomes

$$\frac{d}{dt} \|\theta\|_1 = 2 \int_{\Omega} \delta(\theta) \theta \mathbf{u} \cdot \nabla \theta dV - 2 \int_{\Omega} \delta(\theta) \nabla \theta \cdot \mathbb{D} \cdot \nabla \theta dV - \int_{\Omega} K |\theta| dV - \int_{\partial\Omega} k |\theta| dS.$$

The first term on the right vanishes since the delta function forces $\theta = 0$; for the second term, we can turn the volume integral into a surface integral [41, Theorem 6.1.5]:

$$\frac{d}{dt} \|\theta\|_1 = -2 \int_{\{\theta=0\}} \nabla \theta \cdot \mathbb{D} \cdot \nabla \theta \frac{dS}{|\nabla \theta|} - \int_{\Omega} K |\theta| dV - \int_{\partial\Omega} k |\theta| dS \leq 0, \quad (\text{VII.2})$$

where the first integral is over the zero level set of $\theta(\cdot, t)$. We conclude that the total variation distance $\frac{1}{2} \|n_1 - n_2\|$ does indeed decrease monotonically, as was apparent from the earlier numerical simulations. (The level-set integral in Eq. (VII.2) appears in approaches based on tracer coordinates [42].)

The ‘proof’ presented here relies on the apparently strong assumption that $|\nabla \theta| \neq 0$ on the zero level set of θ . However, the uniform ellipticity bound (III.5) implies that $-\nabla \theta \cdot \mathbb{D} \cdot \nabla \theta / |\nabla \theta| \leq -\sigma |\nabla \theta|$, so singular points limit nicely to zero in the integrand. In Appendix B we show that the L^1 norm is the only L^q norm that decays monotonically in the nonuniform mixing case.

One possible advantage Eq. (VII.2) has over the corresponding equation (VI.8) for the f -divergence is that it shows convergence even when the source $S(\mathbf{x}, t)$ is negative, since the source has dropped out of (VII.2) completely. However Eq. (VI.8)

suggests that a positive source can actually improve the rate of convergence. Another weakness of Eq. (VII.2) compared to (VI.8) is that its right-hand side is difficult to compute: it requires tracking of the zero level set, which is a challenging problem in practice because of resolution and changes in topology. By comparison, the right-hand side of Eq. (VI.8) is readily computed and regions of large entropy production can be identified from the magnitude of the integrands.

VIII. DISCUSSION

The traditional view of mixing in non-divergent flow is that a stirred passive scalar will ultimately be homogenized to a uniform concentration. As we discussed, this requires both non-divergence of the velocity field and no-penetration boundary conditions. If either condition is violated, the ultimate state of the mixing process is no longer uniform, and may in fact be time-dependent for non-autonomous systems, where \mathbf{u} or \mathbb{D} are explicit functions of time. We refer to these systems as nonuniform mixing, because the passive scalar may be mixed even though its concentration is not uniform. Such nonuniform situations will arise in the presence of filters, which are membranes that permits the passage of fluid but not of particles (passive scalar).

Using the standard concentration variance as a proxy for mixing is less useful for nonuniform mixing, since the variance is not necessarily a monotonically-decreasing function of time. Of course, variance will eventually decrease to zero even in nonuniform mixing (as long as it is defined appropriately), but the excursions it undertakes can make it hard to ascribe a rate of mixing to the system (see Figs. 4 and 6). Instead of concentration variance, a more reliable proxy for mixing is the *f-divergence*, which is related to relative entropy. Instead of relying on an initial condition to become uniform, we define the rate of mixing in terms of the rate at which two arbitrary densities $p_1(\mathbf{x}, t)$ and $p_2(\mathbf{x}, t)$ approach each other. They will eventually both con-

verge to an ultimate density $\varphi(\mathbf{x}, t)$, which is independent of the initial condition. The f -divergence picture is easily adapted to cases with sources and sinks.

The connection between the f -divergence and mix-norms [4, 43, 44] is not completely clear. Mix-norms are used as a diagnostic for mixing, and are not guaranteed to decay monotonically for the types of examples presented here. Their behavior is thus probably more closely related to that of concentration variance than to f -divergence, though they have the advantage that they decay even when the diffusivity is set to zero, which renders them more useful for optimization [45–49]. Perhaps there is a hybrid approach that could marry the advantages of both.

Finally, note that nonuniform mixing suggests a different type of mixing optimization problem, where the goal is to decrease spatial or temporal variations of $\varphi(\mathbf{x}, t)$ itself rather than the rate of approach to $\varphi(\mathbf{x}, t)$. This was investigated previously for source-sink systems [50–53], but it could be effected in any problem involving nonuniform mixing. For example, a flow could be designed to minimize the concentration of viruses near a filter, in order to minimize the effect of inevitable imperfections in the membrane.

ACKNOWLEDGMENTS

The author thanks Yu Feng, Albion Lawrence, Noboru Nakamura, Bryan Oakley, Greg Pavliotis, Jim Thomas, Jeffrey Weiss, Bill Young, and an anonymous referee for insightful comments and discussions. Some of this research was completed at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611; travel there was supported by the Brandeis University Provost’s Re-

search Grant “Nonequilibrium Statistical Mechanics of the Ocean and Atmosphere.”

- [1] J.-L. Thiffeault, Scalar decay in chaotic mixing, in *Transport and Mixing in Geophysical Flows*, Lecture Notes in Physics, Vol. 744, edited by J. B. Weiss and A. Provenzale (Springer, Berlin, 2008) pp. 3–35, arXiv:nlin/0502011.
- [2] H. Aref, J. R. Blake, M. Budišić, S. S. Cardoso, J. H. Cartwright, H. J. Clercx, K. El Omari, U. Feudel, R. Golestanian, E. Guillard, G. F. van Heijst, T. S. Krasnopolskaya, Y. Le Guer, R. S. MacKay, V. V. Meleshko, G. Metcalfe, I. Mezić, A. P. de Moura, O. Piro, M. F. M. Speetjens, R. Sturman, J.-L. Thiffeault, and I. Tuval, Frontiers of chaotic advection, *Rev. Mod. Phys.* **89**, 025007 (2017).
- [3] W. R. Young, Stirring and mixing, in *Proceedings of the 1999 Summer Program in Geophysical Fluid Dynamics*, edited by J.-L. Thiffeault and C. Pasquero (Woods Hole Oceanographic Institution, Woods Hole, MA, 1999) <http://gfd.whoi.edu/proceedings/1999/PDFvol1999.html>.
- [4] J.-L. Thiffeault, Using multiscale norms to quantify mixing and transport, *Nonlinearity* **25**, R1 (2012), arXiv:1105.1101.
- [5] C. R. Doering and C. Nobili, Lectures on stirring, mixing and transport, in *Transport, Fluids, and Mixing* (De Gruyter Open Poland, 2020) pp. 8–34.
- [6] R. Golestanian, T. B. Liverpool, and A. Ajdari, Designing phoretic micro- and nano-swimmers, *New J. Phys.* **9**, 126 (2007).
- [7] S. van Teeffelen and H. Löwen, Dynamics of a Brownian circle swimmer, *Phys. Rev. E* **78**, 020101 (2008).
- [8] C. Kurtzhaler, S. Leitmann, and T. Franosch, Intermediate scattering function of an anisotropic active brownian particle, *Sci. Rep.* **6**, 36702 (2016).
- [9] C. F. Lee, Active particles under confinement: aggregation at the wall and gradient

- formation inside a channel, *New J. Phys.* **15**, 055007 (2013).
- [10] B. Ezhilan and D. Saintillan, Transport of a dilute active suspension in pressure-driven channel flow, *J. Fluid Mech.* **777**, 482 (2015).
 - [11] H. Chen and J.-L. Thiffeault, Shape matters: A Brownian microswimmer in a channel, *J. Fluid Mech.* **916**, A15 (2021).
 - [12] E. A. D’Asaro, A. Y. Shcherbina, J. M. Klymak, J. Molemaker, G. Novelli, C. M. Guigand, A. C. Haza, B. K. Haus, E. H. Ryan, G. A. Jacobs, and et al., Ocean convergence and the dispersion of flotsam, *Proceedings of the National Academy of Sciences* **115**, 1162 (2018).
 - [13] R. Aris, *Vectors, Tensors, and the Basic Equations of Fluid Mechanics* (Dover, New York, 1989).
 - [14] H. A. Stone, A simple derivation of the time-dependent convective-diffusion equation for surfactant transport along a deforming interface, *Physics of Fluids A: Fluid Dynamics* **2**, 111112 (1990).
 - [15] G. A. Pavliotis, *Stochastic Processes and Applications* (Springer, Berlin, 2014).
 - [16] H. Risken, *The Fokker–Planck Equation: Methods of Solution and Applications*, 2nd ed. (Springer, Berlin, 1996).
 - [17] A. Arnold, E. Carlen, and Q. Ju, Large-time behavior of non-symmetric Fokker–Planck type equations, *Communications on Stochastic Analysis* **2**, 10.31390/cosa.2.1.11 (2008).
 - [18] F. Achleitner, A. Arnold, and D. Stürzer, Large-time behavior in non-symmetric Fokker–Planck equations, *Riv. Mat. Univ. Parma* **6**, 1 (2015).
 - [19] A. Arnold, A. Einav, and T. Wöhrer, On the rates of decay to equilibrium in degenerate and defective Fokker–Planck equations, *Journal of Differential Equations* **264**, 6843 (2018).
 - [20] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter, On convex Sobolev inequal-

- ities and the rate of convergence to equilibrium for Fokker–Planck type equations, *Communications in Partial Differential Equations* **26**, 43 (2001).
- [21] T. Lelièvre, F. Nier, and G. A. Pavliotis, Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion, *Journal of Statistical Physics* **152**, 237 (2013).
 - [22] D. D’Alessandro, M. Dahleh, and I. Mezić, Control of mixing in fluid flow: A maximum entropy approach, *IEEE Transactions on Automatic Control* **44**, 1852 (1999).
 - [23] M. A. Stremler and B. A. Cola, A maximum entropy approach to optimal mixing in a pulsed source-sink flow, *Phys. Fluids* **18**, 011701 (2006).
 - [24] M. Camesasca, M. Kaufman, and I. Manas-Zloczower, Quantifying fluid mixing with the Shannon entropy, *Macromolecular Theory and Simulations* **15**, 595 (2006).
 - [25] N. C. Krützmann, A. J. McDonald, and S. E. George, Identification of mixing barriers in chemistry-climate model simulations using rényi entropy, *Geophysical Research Letters* **35**, 10.1029/2007gl032829 (2008).
 - [26] P. S. Fodor and M. Kaufman, Time evolution of mixing in the staggered herringbone microchannel, *Modern Physics Letters B* **25**, 1111 (2011).
 - [27] P. H. Lauritzen and J. Thuburn, Evaluating advection/transport schemes using inter-related tracers, scatter plots and numerical mixing diagnostics, *Quarterly Journal of the Royal Meteorological Society* **138**, 906 (2011).
 - [28] J. Grahm, *Rényi entropy and finite Lyapunov exponents as metrics of transport and mixing in an idealised stratosphere*, Master’s thesis, Chalmers University of Technology, Gothenburg, Sweden (2012).
 - [29] G. B. Brandani, M. Schor, C. E. MacPhee, H. Grubmiller, U. Zachariae, and D. Marenduzzo, Quantifying disorder through conditional entropy: An application to fluid mixing, *PLoS ONE* **8**, e65617 (2013).
 - [30] D. Perugini, C. P. De Campos, M. Petrelli, D. Morgavi, F. P. Vetere, and D. B. Dingwell, Quantifying magma mixing with the Shannon entropy: Application to sim-

- ulations and experiments, *Lithos* **236-237**, 299 (2015).
- [31] P. L. Boyland, H. Aref, and M. A. Stremler, Topological fluid mechanics of stirring, *J. Fluid Mech.* **403**, 277 (2000).
 - [32] J.-L. Thiffeault and M. D. Finn, Topology, braids, and mixing in fluids, *Philos. Trans. Royal Soc. Lond. A* **364**, 3251 (2006).
 - [33] E. Gouillart, M. D. Finn, and J.-L. Thiffeault, Topological mixing with ghost rods, *Phys. Rev. E* **73**, 036311 (2006).
 - [34] The L^q norm $\|f\|_q$ is defined by $\|f\|_q = \left(\int_{\Omega} |f|^q dV\right)^{1/q}$ for $1 \leq q \leq \infty$.
 - [35] F. Österreicher and I. Vajda, A new class of metric divergences on probability spaces and its applicability in statistics, *Annals of the Institute of Statistical Mathematics* **55**, 639 (2003).
 - [36] F. Liese and I. Vajda, On divergences and informations in statistics and information theory, *IEEE Transactions on Information Theory* **52**, 43944412 (2006).
 - [37] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. (Wiley, Hoboken, New Jersey, 2005).
 - [38] These entropies tend to decrease to zero with time, which is the opposite definition to that used in physics.
 - [39] D. M. Endres and J. E. Schindelin, A new metric for probability distributions, *IEEE Transactions on Information Theory* **49**, 1858 (2003).
 - [40] R. T. Pierrehumbert, Tracer microstructure in the large-eddy dominated regime, *Chaos Solitons Fractals* **4**, 1091 (1994).
 - [41] L. Hörmander, *The Analysis of Linear Partial Differential Operators*, 2nd ed., Vol. 1 (Springer, Berlin, 1990).
 - [42] N. Nakamura, Two-dimensional mixing, edge formation, and permeability diagnosed in an area coordinate, *J. Atmos. Sci.* **53**, 1524 (1996).
 - [43] G. Mathew, I. Mezić, and L. Petzold, A multiscale measure for mixing and its appli-

- cations, in *Proc. Conf. on Decision and Control, Maui, HI*, IEEE (IEEE, 2003).
- [44] G. Mathew, I. Mezić, and L. Petzold, A multiscale measure for mixing, *Physica D* **211**, 23 (2005).
 - [45] G. Mathew, I. Mezić, S. Grivopoulos, U. Vaidya, and L. Petzold, Optimal control of mixing in Stokes fluid flows, *J. Fluid Mech.* **580**, 261 (2007).
 - [46] Z. Lin, C. R. Doering, and J.-L. Thiffeault, Optimal stirring strategies for passive scalar mixing, *J. Fluid Mech.* **675**, 465 (2011).
 - [47] D. P. G. Foures, C. P. Caulfield, and P. J. Schmid, Optimal mixing in two-dimensional plane Poiseuille flow at finite Péclet number, *J. Fluid Mech.* **748**, 241 (2014).
 - [48] L. Vermach and C. P. Caulfield, Optimal mixing in three-dimensional plane Poiseuille flow at high Péclet number, *J. Fluid Mech.* **850**, 875 (2018).
 - [49] F. Marcotte and C. P. Caulfield, Optimal mixing in 2D stratified plane Poiseuille flow at finite Péclet and Richardson numbers, *J. Fluid Mech.* **853**, 359 (2018).
 - [50] J.-L. Thiffeault, C. R. Doering, and J. D. Gibbon, A bound on mixing efficiency for the advection–diffusion equation, *J. Fluid Mech.* **521**, 105 (2004).
 - [51] C. R. Doering and J.-L. Thiffeault, Multiscale mixing efficiencies for steady sources, *Phys. Rev. E* **74**, 025301(R) (2006).
 - [52] T. A. Shaw, J.-L. Thiffeault, and C. R. Doering, Stirring up trouble: Multiscale mixing measures for steady scalar sources, *Physica D* **231**, 143 (2007), [arXiv:physics/0607270](https://arxiv.org/abs/physics/0607270).
 - [53] J.-L. Thiffeault and G. A. Pavliotis, Optimizing the source distribution in fluid mixing, *Physica D* **237**, 918 (2008), [arXiv:physics/0703135](https://arxiv.org/abs/physics/0703135).

Appendix A: Derivation of Eq. (IV.4)

Write $\partial_t p_i = -\nabla \cdot \mathbf{F}(p_i)$, with $\mathbf{F}(p) := \mathbf{u} p - \mathbb{D} \cdot \nabla p$, and consider a domain $\Omega \subset \mathbb{R}^n$ with $\mathbf{F}(p_i) \cdot \hat{\mathbf{n}} = 0$ on $\partial\Omega$. Then by direct differentiation of Eq. (IV.2):

$$\begin{aligned} \dot{H}_f[p_1, p_2] &= \int_{\Omega} (\partial_t p_2 f(p_1/p_2) + p_2 f'(p_1/p_2) (\partial_t p_1/p_2 - p_1 \partial_t p_2/p_2^2)) dV \\ &= - \int_{\Omega} (\nabla \cdot \mathbf{F}(p_2) f(p_1/p_2) + f'(p_1/p_2) (\nabla \cdot \mathbf{F}(p_1) - (p_1/p_2) \nabla \cdot \mathbf{F}(p_2))) dV. \end{aligned}$$

We integrate by parts, and two terms containing $f'(p_1/p_2) \mathbf{F}(p_2) \cdot \nabla(p_1/p_2)$ cancel.

We are left with

$$\dot{H}_f[p_1, p_2] = \text{BT}[p_1, p_2] + \int_{\Omega} p_2^{-1} f''(p_1/p_2) \nabla(p_1/p_2) \cdot (p_2 \mathbf{F}(p_1) - p_1 \mathbf{F}(p_2)) dV \quad (\text{A.1})$$

with the boundary terms

$$\text{BT}[p_1, p_2] = - \int_{\partial\Omega} f(p_1/p_2) \mathbf{F}(p_2) \cdot \hat{\mathbf{n}} + p_2^{-1} f'(p_1/p_2) (p_2 \mathbf{F}(p_1) - p_1 \mathbf{F}(p_2)) \cdot d\mathbf{S}.$$

The boundary terms vanish when $\mathbf{F}(p_i) \cdot \hat{\mathbf{n}} = 0$ on $\partial\Omega$. Also,

$$\begin{aligned} p_2 \mathbf{F}(p_1) - p_1 \mathbf{F}(p_2) &= p_2 (\mathbf{u} p_1 - \mathbb{D} \cdot \nabla p_1) - p_1 (\mathbf{u} p_2 - \mathbb{D} \cdot \nabla p_2) \\ &= -p_2 \mathbb{D} \cdot \nabla p_1 + p_1 \mathbb{D} \cdot \nabla p_2 \\ &= -p_2^2 \mathbb{D} \cdot \nabla(p_1/p_2). \end{aligned} \quad (\text{A.2})$$

Inserting (A.2) into (A.1) recovers Eq. (IV.4).

Appendix B: Decay of L^q norms

For $1 \leq q \leq \infty$, does any L^q norm other than $q = 1$ decay monotonically? If we put $G(\theta) = |\theta|^q$ in Eq. (VII.1), then for q even we have $G(\theta) = \theta^q$, $G'(\theta) = q\theta^{q-1}$, and $G''(\theta) = q(q-1)\theta^{q-2}$, and find

$$\frac{d}{dt} \|\theta\|_q^q = q(q-1) \int_{\Omega} \theta^{q-1} \mathbf{u} \cdot \nabla \theta dV - q(q-1) \int_{\Omega} \theta^{q-2} \nabla \theta \cdot \mathbb{D} \cdot \nabla \theta dV \quad (\text{B.1})$$

where we set $K = k = 0$ for convenience. The first term on the right is not sign definite for any even q , so none of these norms will necessarily decay monotonically. For q odd, we have that $G'(\theta) = q\theta^{q-1}\text{sgn}(\theta)$ and $G''(\theta) = 2q\theta^{q-1}\delta(\theta) + q(q-1)\theta^{q-2}\text{sgn}(\theta)$, and Eq. (VII.1) becomes

$$\begin{aligned} \frac{d}{dt}\|\theta\|_q^q &= q(q-1) \int_{\Omega} \theta^{q-2}|\theta| \mathbf{u} \cdot \nabla \theta \, dV \\ &\quad - 2q \int_{\Omega} \theta^{q-1}\delta(\theta) \nabla \theta \cdot \mathbb{D} \cdot \nabla \theta \, dV - q(q-1) \int_{\Omega} |\theta|^{q-2} \nabla \theta \cdot \mathbb{D} \cdot \nabla \theta \, dV. \end{aligned} \quad (\text{B.2})$$

For $q = 1$ this reduces to Eq. (VII.2); for $q > 1$ we have

$$\frac{d}{dt}\|\theta\|_q^q = q(q-1) \int_{\Omega} \theta^{q-2}|\theta| \mathbf{u} \cdot \nabla \theta \, dV - q(q-1) \int_{\Omega} |\theta|^{q-2} \nabla \theta \cdot \mathbb{D} \cdot \nabla \theta \, dV. \quad (\text{B.3})$$

and again the first term is not sign-definite. We conclude that the L^1 norm is the only such norm that exhibits a monotonic decay to zero in the nonuniform mixing case.