

Transport and mixing in control volumes through the lens of probability

John Craske¹  and Paul M. Mannix¹ 

¹Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK

Corresponding author: John Craske, john.craske07@imperial.ac.uk

(Received 30 September 2024; revised 12 May 2025; accepted 16 August 2025)

A partial differential equation governing the evolution of the joint probability distribution of multicomponent flow observations, drawn randomly from one or more control volumes, is derived and applied to examples involving irreversible mixing. Unlike local probability density methods, this work adopts an integral perspective by regarding a control volume as a sample space with an associated probability distribution. A natural and general definition for the boundary of such control volumes comes from the magnitude of the gradient of the sample space distribution, which can accommodate Eulerian or Lagrangian frames of reference as particular cases. The formulation exposes contributions made by uncertain or stochastic boundary fluxes and internal cross-gradient mixing in the equation governing the observables' joint probability distribution. Advection and diffusion over a control volume's boundary result in source and drift terms, respectively, whereas internal mixing, in general, corresponds to the sign-indefinite diffusion of probability density. Several typical circumstances for which the corresponding diffusion coefficient is negative semidefinite are identified and discussed in detail. The framework is a natural setting for examining available potential energy, the incorporation of uncertainty into bulk models, and establishing a link with the Feynman–Kac formula and Kolmogorov equations that are used to analyse stochastic processes.

Key words: turbulent mixing, coupled diffusion and flow, general fluid mechanics

1. Introduction

1.1. Bulk models and uncertainty

Bulk, integral, lumped or coarse-grained models in fluid mechanics involve integrating equations that govern the local (pointwise in space) evolution of a system over a control

volume. Such models are useful for providing a macroscopic picture at scales that are directly relevant to a given application. A prevalent example comes from the concept of entrainment across the boundary of a control volume, which is typically defined to demarcate the turbulent parts of a flow from a non-turbulent surrounding environment (Ellison & Turner 1959; van Reeuwijk, Vassilicos & Craske 2021). However, over a century of intensive research into fluid turbulence has shown that nature does not always yield to coarse representations, demanding, in return, case-dependent closures and probabilistic approaches.

Probabilistic approaches are particularly appropriate for bulk models because the contents of control volumes is typically heterogeneous. A relevant example comes from the field of building ventilation, where it is often assumed that the air in each room or ‘zone’ of a building is well mixed and, therefore, of uniform temperature. However, it is now acknowledged, not least due to concerns raised during the COVID-19 pandemic, that the secondary flows and temperature structures within rooms play an important role in determining the fate of contaminants, energy demands and thermal comfort (see, for example, Vouriot *et al.* 2023; Bhagat *et al.* 2024). An additional complication, which renders the underlying challenge probabilistic, is that boundary and internal forcing conditions are almost never known precisely and typically fluctuate stochastically in time (see, for example, Andrian & Craske 2023). Rare events arising from the mentioned considerations can have a significant effect in a wide range of natural and artificial environments (Villiermaux 2019). On this basis, the present work forsakes detailed local deterministic information in physical space for a limited amount of probabilistic information about a control volume; we focus on what is inside a control volume at the expense of knowing precisely where it is occurring.

Although probabilistic approaches to turbulence are common (see, for example, Monin & Yaglom 2013), the complete statistical (sometimes referred to as ‘functional’, due to the fact that the sought-after probability distributions are characterised by functionals of infinite dimensional spaces) formulation (Hopf 1952; Lewis & Kraichnan 1962) is less common. The functional formulation has been explored, in a variety of guises, by mathematicians, physicists and engineers in parallel and somewhat independently. It is therefore appropriate to summarise the similarities and limitations of the various perspectives before highlighting their connection with the concept of available potential energy and stating the specific aims of the present work.

1.2. Background

In the full functional formulation of the Navier–Stokes equations (Hopf 1952; Lewis & Kraichnan 1962; Lundgren 1967; McComb 1990; Monin & Yaglom 2013), one seeks to understand the evolution of a flow’s probability distribution in phase space, which, consisting of infinite dimensions, produces a problem that is typically intractable. However, a happy consequence of ‘lifting’ the problem to an infinite-dimensional space is that the functional formulation renders the problem linear. The resulting Frobenius–Perron and Liouville operators, depending on whether time is regarded as discrete or continuous, respectively, underpin statistical mechanics (Gaspard 1998; Lasota & Mackey 1998; Ruelle 2004).

An alternative perspective is afforded by observables of a flow, whose expected evolution is determined by Koopman and Lie operators for discrete and continuous time, respectively (Koopman 1931; Mezić 2005, 2013). Although the two (Frobenius–Perron and Koopman) approaches are dual in certain settings, they typically rely on different approximations (Brunton *et al.* 2022). Approximations of the Frobenius–Perron operator

commonly involve the evolution of a large number of initial conditions for a relatively short time using Ulam's method (Li 1976), whereas approximations of the Koopman operator involve evolution from a relatively small number of initial conditions over a long time (Klus, Koltai & Schütte 2016).

Frobenius–Perron and Koopman operators naturally accommodate stochastic differential equations, the latter by considering expectations of observables (Lasota & Mackey 1998; Črnjarić-Žić *et al.* 2020). In this regard, the two perspectives described in the previous paragraphs correspond to forward and backward Kolmogorov equations for continuous Markov processes, respectively (Pavliotis 2014).

Independently, and particularly in an engineering context, probability density function (p.d.f.) methods were developed towards the latter half of the twentieth century, primarily in the fields of combustion (Pope 1981, 1985; Kollmann 1990; Pope 1994, 2000; Fox 2003) and those involving turbulent dispersion (Fischer *et al.* 1979; Hunt 1985). The equations used in p.d.f. methods typically focus on conditional probability distributions from single points in space and can therefore be regarded as a subset or projection of the full functional formulations described previously. Therefore, in a prognostic capacity, they contain expectations of gradients that require closure (see, for example, Minier & Pozorski 1997; Pope 2000; Fox 2003). Diagnostically, however, they have been used recently to study vorticity dynamics (Li, Qian & Zhou 2022) and to analyse Rayleigh–Bénard convection (Lülff *et al.* 2011, 2015). Although well suited to the task, p.d.f. models have not been used extensively to describe systems with external stochastic forcing. In this regard, their link with the more formal derivations of forward and backward Kolmogorov equations in stochastic processes, or with Frobenius–Perron/Koopman formalisms, is not well established.

Although an attractive feature of p.d.f. methods is that advection and state-dependent forcing terms, such as buoyancy, appear in closed form (Pope 2000), their potential as either a diagnostic or prognostic tool for buoyancy-driven flows and stratified turbulence has not been explored. This is surprising, because, as highlighted by Tseng & Ferziger (2001), the bulk quantities that are routinely used to diagnose the energetics behind mixing in such fields are expressed naturally as functionals of p.d.f.s. In particular, the reference state that is used to define global available potential energy (Margules 1903; Lorenz 1955; Winters *et al.* 1995; Tailleux 2013), which quantifies the maximum amount of potential energy that can be released during a volume-preserving and adiabatic rearrangement of fluid parcels, is a functional of the joint p.d.f. of buoyancy and geopotential height. While such constructions are difficult to wrestle with in physical space, due to the global nature of the rearrangement, their expression in terms of joint p.d.f.s is natural and therefore convenient. There is consequently an opportunity for advances made in p.d.f. methods to be applied to a broader class of problems and, conversely, for the knowledge about the available energetics behind transport and mixing to inform closures. Establishing a connection in this regard might provide a means of generating operational stochastic models or, at least, an alternative means of interpreting existing low-dimensional stochastic models of stratified turbulence (Kerstein 1999; Wunsch & Kerstein 2005).

The aim of the present work is to develop a general and extensible framework for deriving evolution equations for the joint probability distribution of a set of spatially heterogeneous and stochastic flow observables over arbitrary control volumes. To this end, we generalise and shed new light on classical p.d.f. methods in the following ways.

- (i) The present perspective is macroscopic (top down) in dealing with joint probability distributions for the heterogeneous contents of distributed regions in space (such as those discussed by Villiermaux (2019)), rather than point observations. The concept

is simple to implement, yet general in encompassing both Eulerian and Lagrangian control volumes as particular cases. In this context, a prevalent role is played by boundary fluxes of probability density into and out of zones, which, rather than possessing a sharp boundary, can overlap to form a so-called mixture distribution (Lindsay 1995) of a parent domain.

- (ii) We make use of the classical derivation of the (weaker) forward Kolmogorov equation from the (stronger) backward Kolmogorov equation from stochastic processes and the Feynman–Kac formula (Pavliotis 2014). An advantage in proceeding from this dual perspective is that one more readily obtains a versatile and complete framework, in which the forward and backward equations both have distinct physical significance. The dual perspective provides an alternative to using a ‘fine-grained’ microscopic p.d.f., corresponding to a delta distribution (Pope 1985), to derive the forward Kolmogorov equation. In doing so, we establish a closer connection with the field of stochastic processes, in the hope that its methods can be applied to the development of bulk stochastic models that faithfully account for the underlying fluid mechanics.

The core part of the framework described previously is presented in § 3, and is followed by a discussion of drift and diffusion coefficients in §§ 4 and 5, respectively. The decomposition of a parent domain into non-uniform overlapping conditional components (see point (i)) is explained in § 6. Example applications are discussed in § 7, before approaches to closure are summarised in § 8. First, we illustrate the basic ideas behind viewing the contents of a control volume probabilistically, with relatively simple introductory examples for one- and two-dimensional domains.

2. Introductory examples

This work addresses the following question: what is the probability that a given vector of field variables $\mathbf{Y}$, evaluated at a randomly selected point ω from a control volume Ω , parameterised by coordinates $\mathbf{X}$, will have values lying in a given range?

Here, the term ‘randomly’ refers to a prescribed probability distribution f_X for space, which determines the probability of selecting or sampling a given region from the control volume. An Eulerian point ω becomes an element of a sample space $\Omega \ni \omega$ parametrised by coordinate functions $\mathbf{X}$. The field variables $\mathbf{Y} = \varphi(\mathbf{X})$ are quantities such as velocity, temperature or scalar concentration. We are therefore interested in the distribution of $\mathbf{Y}$ with respect to the given sampling distribution in space.

The answer to the question is given by the probability density f_Y (or, more generally, distribution) associated with $\mathbf{Y}$, which does not contain information about the precise relationship between $\mathbf{Y}$ and $\mathbf{X}$. Indeed, the probability density corresponds to an infinite number of possible functions φ of $\mathbf{X}$ that produce the same distribution of $\mathbf{Y}$. Informally, the construction involves tipping the fields $\mathbf{Y}$ into a sack that does not store values of $\mathbf{X}$; the probability density ‘weighs’ the various values of $\mathbf{Y}$, without caring about from where they came.

Figure 1 illustrates an example in one dimension for which the vector ‘ $\mathbf{Y}$ ’ can therefore be replaced with the scalar ‘ Y ’. As explained more precisely in Appendix A, the construction of f_Y involves determining the proportion of Ω taken up by a given value of Y . Analytically, this involves considering intervals of Ω over which $Y = \varphi(X)$ is strictly monotone with respect to X and, therefore, invertible. Such intervals lie between the stationary points of φ . Each point within the interval contributes a density that is inversely proportional to the gradient of φ with respect to X at that point, because relatively large gradients in φ account for a relatively small proportion of Ω .

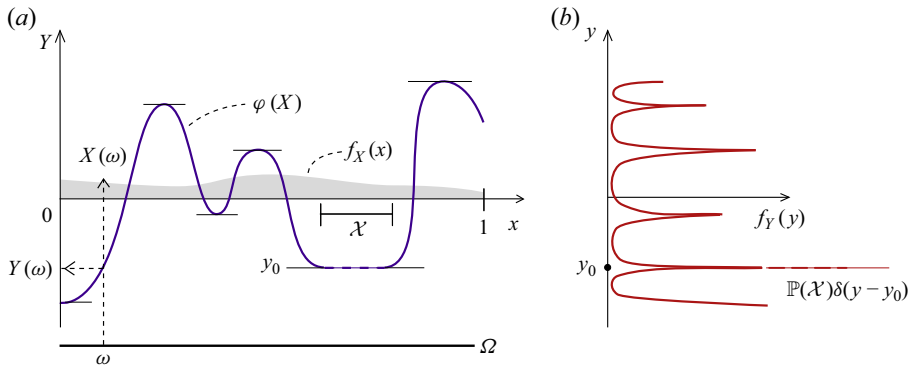


Figure 1. (b) Probability distribution f_Y of (a) the value of a function $Y = \varphi(X)$ parametrised by the coordinate X , which has an associated density f_X . Stationary points of the function Y (highlighted by horizontal lines) correspond to singularities in the distribution f_Y . To account for parts of Y that are constant (dashed), the distribution contains a Dirac measure δ , weighted by the probability $\mathbb{P}(\mathcal{X})$ associated with the selection of a point in the domain over which Y is equal to the given constant.

Intervals $\mathcal{X} \ni X$ over which φ is invertible contribute to the density f_Y in proportion to their probability $\mathbb{P}(\mathcal{X}) := \int_{\mathcal{X}} f_X(x) dx$, which gives the density the appearance of consisting of folds and caustics (see also figure 2b). The overall contribution that a given value of Y makes to the probability density therefore depends on the number of occurrences of that value in Ω as well as its gradients. For example, the probability density of $Y = \varphi(X) = \cos(\pi n X)$ for $n \in \mathbb{N}^+$ and uniformly distributed $X \in [0, 1]$ is $f_Y(y) = \pi^{-1}(1 - y^2)^{-1/2}$ for $y \in (-1, 1)$, and is therefore independent of n .

Points at which Y is stationary produce singularities in the distribution f_Y shown on the right of figure 1. In exceptional cases, intervals of finite size over which $Y = y$ is constant are not described by a density function but rather a Dirac measure or distribution, weighted by the proportion of the domain over which Y is constant, as indicated by the horizontal dashed part of $\varphi(X)$ and corresponding δ in figure 1 (see also chapter 1 of Chung (2001)).

We now consider a two-dimensional example using velocity and buoyancy fields from Lorenz's 1963 model for convection (Lorenz 1963) shown in figure 2(a) (details of the model and calculations required to construct the corresponding distribution f_Y can be found in Appendix B). In this example, $\mathbf{X} := (X^1, X^2)^\top \in [0, 2\pi/k] \times [0, 1]$ are the horizontal and vertical coordinates, k is a horizontal wavenumber, $f_X \equiv k/(2\pi)$ is a uniform distribution, and $\mathbf{Y}_t := (Y_t^1, Y_t^2)^\top = \varphi(\mathbf{X}) \in \mathbb{R}^2$ denotes the vertical velocity and buoyancy, relative to the static state of linear conduction, at time t , respectively. Figure 2(b) depicts the joint probability density $f_Y(-, t) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ corresponding to the fields shown in figure 2(a), such that the probability of finding a value of $\mathbf{Y}_t$ in any subset $\mathcal{Y} \subset \mathbb{R}^2$ of the function's codomain is

$$\mathbb{P}\{\mathbf{Y}_t \in \mathcal{Y}\} = \int_{\mathcal{Y}} f_Y(\mathbf{y}, t) d\mathbf{y}, \quad (2.1)$$

where $\mathbf{y} := (y^1, y^2)^\top$ denotes the argument to the probability density corresponding to the capitalised random variable $\mathbf{Y}_t$. The density f_Y determines the expectation of any observable g , which can also be computed as an integral over physical coordinates:

$$\mathbb{E}[g(\mathbf{Y}_t)] = \int_{\mathbb{R}^2} g(\mathbf{y}) f_Y(\mathbf{y}, t) d\mathbf{y} = \int_{\mathbb{R}^2} g(\varphi(\mathbf{x})) f_X(\mathbf{x}) d\mathbf{x}, \quad (2.2)$$

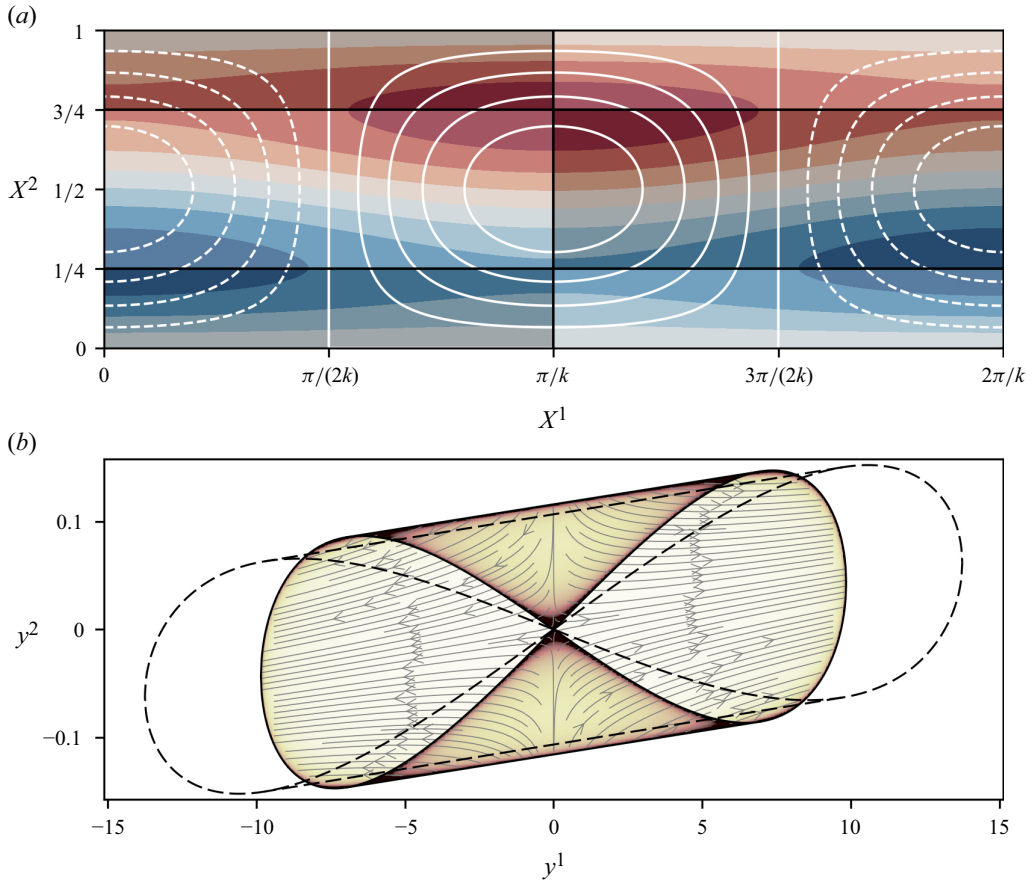


Figure 2. (a) Relative buoyancy (blue to red) and vertical velocity isolines (white and dashed for negative velocity) at time t of the two-dimensional fields corresponding to a point on the Lorenz attractor described in [Appendix B](#) ($r = 28$, $s = 10$, $b := 4\pi^2(k^2 + \pi^2)^{-1} = 8/3$ for horizontal wave number k). The shaded/unshaded rectangles correspond to regions over which the Jacobian $\partial\varphi/\partial X$ is non-singular. (b) Joint probability density (shaded colour) of vertical velocity (y^1) and relative buoyancy (y^2) corresponding to the field shown in panel (a). The solid black line marks singularities in the density and the dashed black line corresponds to the position of the singularities at $t + 0.04$. The light grey arrows are tangential to the probability flux induced by the Lorenz equations (B3) that is responsible for the time evolution of the density. <https://www.cambridge.org/S0022112025106319/JFM-Notebooks/files/fig02/fig02.ipynb>.

where, in general, the sample distribution f_X accounts for the physical extent and distributed weight assigned to the control volume, but is uniform for the example illustrated in [figure 2](#). Note that f_Y does not provide information about how Y_t is correlated with X . In particular, f_Y does not provide information about multipoint statistics or spatial gradients, unless they are included in Y_t .

To understand the p.d.f. shown in [figure 2\(b\)](#), it is helpful to consider the regions over which φ is invertible. Such regions are highlighted as shaded/unshaded rectangles in [figure 2\(a\)](#), in which the Jacobian $\partial\varphi/\partial X$ is non-singular. The solid black lines, separating the regions, denote points for which $|\partial\varphi/\partial X| = 0$, which account for the singularities in [figure 2\(b\)](#). In particular, $|\partial\varphi/\partial X| = 0$ over the sets $S_1 := \{X^1 = \pi/k, X^2 \in [0, 1]\}$ and $S_2 := \{X^1 \in [0, 2\pi/k], X^2 = 1/2 \pm 1/4\}$. The set S_1 corresponds to the solid black line that looks like ‘ ∞ ’ in [figure 2\(b\)](#), while S_2 corresponds to the nearly horizontal lines that define its convex hull. As explained for the previous one-dimensional example, the folded

appearance of f_Y shown in figure 2(b) is due to the fact that a given value of Y_t contributes to f_Y from more than one region in the domain.

As the fields shown in figure 2(a) evolve in time, the density is transported over the phase space shown in figure 2(b). Since the density integrates to unity, it is useful to regard probability, like mass, as a conserved quantity. From this perspective, the governing equations for Y_t over the entire control volume correspond to a two-dimensional velocity field that produces a flux of density in phase space. The direction of the density flux for this example is shown in figure 2(b) by grey lines. The density flux determines the subsequent evolution of f_Y , whose singularities a short time after t are depicted with dashed lines in figure 2(b). Understanding how the evolution of f_Y depends on the evolution of the field variables Y_t is the central topic of this article.

3. Governing equations

In this section, an evolution equation is derived for the joint probability density f_Y of a multicomponent random variable Y_t drawn from a domain according to the spatial distribution f_X . First, § 3.1 defines an infinitesimal generator, which gives the expected time rate of change of an observable evaluated along the trajectories of a stochastic differential equation. Here, the stochastic differential equation, defined in (3.1), is a generic transport equation that can represent the Navier–Stokes or advection–diffusion equations as special cases. Next, in § 3.2, the generator is decomposed to distinguish transport across a bounding region defined by $|\nabla f_X|$ and internal mixing over f_X . Then, § 3.3 uses the resulting operator to define a backward Kolmogorov equation, whose solution backwards in time gives the expected value of an observable, conditioned on an earlier state. Finally, the dual (forward Kolmogorov) equation describing the evolution of probability density is obtained in § 3.4.

3.1. Local governing equations

For the purposes of emphasising a probabilistic perspective that focuses on control volumes, we regard the spatial domain Ω as a sample space, $X : \Omega \rightarrow \mathbb{R}^d$ as coordinate functions and $Y_t : \Omega \rightarrow \mathbb{R}^n$ as random variables, as illustrated in figure 3. Let the Eulerian evolution of Y_t at a given point in the domain be determined by the stochastic differential equation

$$dY_t = (I_t - \nabla \cdot J_t) dt + \sigma dW_t, \quad (3.1)$$

where $I_t \in \mathbb{R}^n$ represents internal sources/sinks, $J_t \in \mathbb{R}^{d \times n}$ is a matrix of fluxes, $\nabla \cdot J_t \in \mathbb{R}^n$ is the divergence of the flux of each component of Y_t and $W_t \in \mathbb{R}^m$ is a vector of independent Wiener processes. In general, the matrix coefficient σ could depend on both X and Y_t , such that $\sigma(X, Y_t) \in \mathbb{R}^{n \times m}$ determines m modes of forcing from W_t for each of the n components of Y_t .

In what follows, it should be remembered that with Ω regarded as a sample space, the precise relationship $Y_t = \varphi_t(X)$ is not assumed to be known. In this context, the gradient ∇Y_t should be regarded as an unknown random variable, rather than a known gradient of $\varphi_t(X)$. It is unknown in the sense that (3.1) does not provide ∇Y_t with its own governing equation. Noting that J_t typically depends on ∇Y_t , the system (3.1) is therefore unclosed because it consists of more unknowns than equations. If spatial derivatives were included in Y_t , their evolution equations would include higher derivatives and one would be faced with the infinite hierarchy of equations that constitutes the closure problem of turbulence (see McComb 1990, and § 8 for a discussion of closures in the context of the present work).

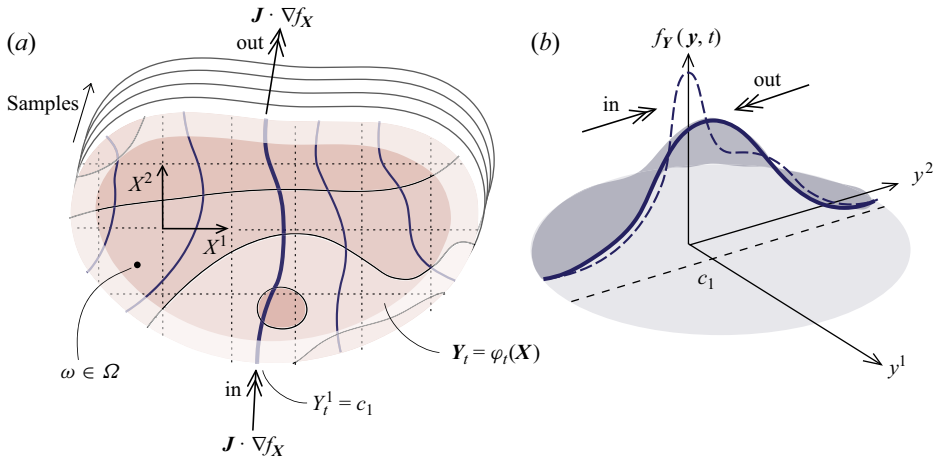


Figure 3. Sample space Ω as the domain of random variables corresponding to coordinates $\mathbf{X}$ and field variables $\mathbf{Y}_t = \varphi_t(\mathbf{X})$. (a) Two fields Y_t^1 (dark blue isolines) and Y_t^2 (red/pink filled isoregions) over the region in space defined by the density f_X . In panel (a), the arrows labelled ‘in’ and ‘out’ denote fluxes of $\mathbf{Y}_t$ at a fixed value of $Y_t^1 = c_1$ over a boundary region defined by ∇f_X (see § 3.2). (b) Section through the current and future probability density f_Y with thick and dashed blue lines, respectively. As demonstrated in § 3 and illustrated by the ‘in’ and ‘out’ arrows in panel (b), the boundary fluxes lead to drift in the equation governing the joint distribution f_Y (for $y^1 = c_1$, the drift is in the positive y^2 direction for those values of Y^2 for which $\mathbf{J} \cdot \nabla f_X > 0$ and in the negative y^2 direction for those values of Y^2 for which $\mathbf{J} \cdot \nabla f_X < 0$). At the same time, f_Y is made narrower due to irreversible mixing, which homogenises the fields.

To determine the evolution of the probability distribution of $\mathbf{Y}_t$, consider the infinitesimal generator $\mathcal{L}$ (Pavliotis 2014, § 2.3) acting on an observable $g : \mathbb{R}^n \rightarrow \mathbb{R}$:

$$\mathcal{L}g(\mathbf{y}, s) := \lim_{t \rightarrow s} \frac{\mathbb{E}[g(\mathbf{Y}_t) | \mathbf{Y}_s = \mathbf{y}] - g(\mathbf{y})}{t - s}, \quad (3.2)$$

where the conditional expectation accounts for the behaviour of all points within the domain where $\mathbf{Y}_s = \mathbf{y}$. In a deterministic context, without the expectation around $g(\mathbf{Y}_t)$, (3.2) corresponds to the infinitesimal generator of the Koopman operator (Brunton *et al.* 2022).

Using Itô’s formula to determine the time rate of change of the observable g , the limit in (3.2) can, noting that the time integral of the resulting Wiener process has an expected value of zero, be expressed as (see § 3.4 and Lemma 3.2, Pavliotis 2014):

$$\mathcal{L}g(\mathbf{y}, s) = \mathbb{E} \left[\left(I_s^i - \nabla \cdot \mathbf{J}_s^i \right) \frac{\partial g(\mathbf{Y}_s)}{\partial Y_s^i} \middle| \mathbf{Y}_s = \mathbf{y} \right] + \frac{1}{2} \Sigma_s^{ij} \frac{\partial^2 g(\mathbf{y})}{\partial y^i \partial y^j}, \quad (3.3)$$

where

$$\Sigma_s := \mathbb{E}[\sigma \sigma^\top | \mathbf{Y}_s = \mathbf{y}], \quad (3.4)$$

and Einstein summation notation has been used over components of $\mathbf{Y}_s$.

We will follow standard arguments to construct backward and forward Kolmogorov equations in terms of the generator $\mathcal{L}$ (see, e.g. Gardiner 1990; Pavliotis 2014). Along the way, integration by parts will be applied to produce boundary and irreversible interior mixing terms that will appear in the associated Kolmogorov equations as forcing and anti-diffusion terms, respectively.

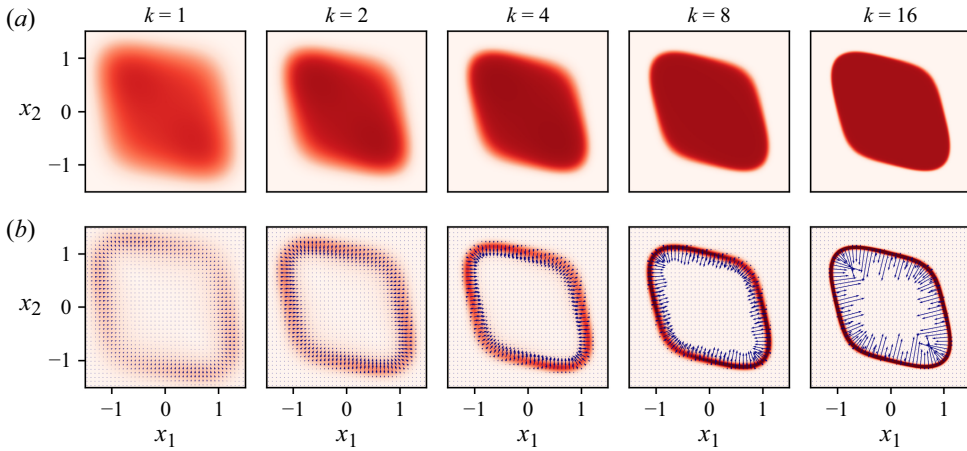


Figure 4. (a) the spatial sample distribution $f_X(\mathbf{x}) := N_k / (1 + \exp(kp(\mathbf{x})))$, where $p(\mathbf{x}) := x_1^4 + x_2^4 + x_1x_2 - 1$ and N_k is a normalisation constant. (b) the bounding region of f_X defined by the unnormalised density $|\nabla f_X|$ (red) and vector field ∇f_X (blue). As $k \rightarrow \infty$, the distribution describes well-defined subsets of $\mathbb{R}^2$ with a sharp boundary.

3.2. Transport and mixing

To understand the physics behind the evolution of the probability density of $\mathbf{Y}_t$ with respect to a prescribed spatial sample distribution f_X , it is useful to decompose the transport terms in (3.3) into those associated with internal irreversible mixing and transport across the sample distribution's ‘boundary’, which will now be defined.

If the sample distribution f_X is uniform over a subset of Ω , the boundary associated with f_X is simply the boundary of the associated subset. However, more generally, f_X could be non-uniform and would therefore not correspond to a classical set with a sharp bounding surface. As explained later and illustrated in figure 4, a natural definition of the bounding region in this regard proves to be the unnormalised density $|\nabla f_X|$, which defines a sharp surface as f_X becomes proportional to an indicator function. The approach therefore generalises the traditional notion of a bounding surface in a way that can be expressed in terms of probability distributions. The manipulations described later are relatively simple because of (rather than in spite of) the generality of this approach.

To establish a connection between expectations obtained via conditioning on $\mathbf{Y}$ compared with conditioning on $\mathbf{X}$, recall the elementary property

$$\mathbb{E}[\square] = \mathbb{E}[\mathbb{E}[\square | \mathbf{Y}]] = \mathbb{E}[\mathbb{E}[\square | \mathbf{X}]] = \int_{\mathbb{R}^d} \mathbb{E}[\square | \mathbf{X} = \mathbf{x}] f_X(\mathbf{x}) \, d\mathbf{x}, \quad (3.5)$$

which means that expectations can be computed from those conditioned on $\mathbf{Y}$ or those conditioned on $\mathbf{X}$. The latter are useful because they correspond to variables in physical space to which it is possible to apply integration by parts to distinguish boundary fluxes from interior mixing. In this regard, note that use of the product and chain rule implies

$$-(\nabla \cdot \mathbf{J})^i \frac{\partial g(\mathbf{Y})}{\partial Y^i} = -\nabla \cdot \left(\mathbf{J}^i \frac{\partial g(\mathbf{Y})}{\partial Y^i} \right) + \mathbf{J}^i \cdot \nabla Y^j \frac{\partial^2 g(\mathbf{Y})}{\partial Y^i \partial Y^j}. \quad (3.6)$$

It will be assumed that $f_X(\mathbf{x})$ either has compact support on $\mathbb{R}^d$ or tends to zero sufficiently fast as $|\mathbf{x}| \rightarrow \infty$. Then, making use of (3.5), (3.6) and applying integration by parts to the first term on the right-hand side of (3.6) yields

$$\begin{aligned}
 -\mathbb{E} \left[(\nabla \cdot \mathbf{J})^i \frac{\partial g(\mathbf{Y})}{\partial Y^i} \right] &= \int_{\mathbb{R}^d} \mathbb{E} \left[\mathbf{J}^i \frac{\partial g(\mathbf{Y})}{\partial Y^i} \middle| \mathbf{X} = \mathbf{x} \right] \cdot \nabla f_X(\mathbf{x}) \, d\mathbf{x} \\
 &\quad + \int_{\mathbb{R}^d} \mathbb{E} \left[\mathbf{J}^i \cdot \nabla Y^j \frac{\partial^2 g(\mathbf{Y})}{\partial Y^i \partial Y^j} \middle| \mathbf{X} = \mathbf{x} \right] f_X(\mathbf{x}) \, d\mathbf{x}, \quad (3.7)
 \end{aligned}$$

where the first and second term on the right-hand side correspond to boundary transport and internal mixing, respectively. The first can be expressed in terms of the original density by including the weight $\nabla f_X / f_X = \nabla \log f_X$ inside the expectation:

$$\int_{\mathbb{R}^d} \mathbb{E}[\square \middle| \mathbf{X} = \mathbf{x}] \cdot \nabla f_X(\mathbf{x}) \, d\mathbf{x} = \int_{\mathbb{R}^d} \mathbb{E}[\square \cdot \nabla \log f_X \middle| \mathbf{X} = \mathbf{x}] f_X(\mathbf{x}) \, d\mathbf{x}; \quad (3.8)$$

hence, using (3.5),

$$\begin{aligned}
 -\mathbb{E} \left[(\nabla \cdot \mathbf{J})^i \frac{\partial g(\mathbf{Y})}{\partial Y^i} \middle| \mathbf{Y} \right] &= \underbrace{\mathbb{E} \left[\mathbf{J}^i \frac{\partial g(\mathbf{Y})}{\partial Y^i} \cdot \nabla \log f_X \middle| \mathbf{Y} \right]}_{\text{boundary transport}} + \underbrace{\mathbb{E} \left[\mathbf{J}^i \cdot \nabla Y^j \frac{\partial^2 g(\mathbf{Y})}{\partial Y^i \partial Y^j} \middle| \mathbf{Y} \right]}_{\text{internal mixing}}. \quad (3.9)
 \end{aligned}$$

Physically, the factor $\nabla \log f_X$ conditions the flux against the normal direction of the boundary region. To see why, note that $-\nabla f_X / |\nabla f_X|$ defines an outward unit vector that is perpendicular to isosurfaces of f_X and that $|\nabla f_X|$ is the (unnormalised) density associated with the bounding region. Therefore, introducing the normalisation factor N and using $\tilde{\mathbb{E}}$ to denote the expectation of a random variable $\square$ over the boundary region:

$$\tilde{\mathbb{E}}[\square] := \frac{1}{N} \int_{\mathbb{R}^d} \square |\nabla f_X| \, d\mathbf{x}, \quad (3.10)$$

the formula for conditional expectations under a change in measure, removing a factor of N^{-1} from both sides, yields (see, for example Shreve 2004, § 5.2.1)

$$\mathbb{E}[\square \cdot \nabla \log f_X \middle| \mathbf{Y}] = \tilde{\mathbb{E}} \left[\square \cdot \frac{\nabla f_X}{|\nabla f_X|} \middle| \mathbf{Y} \right] \mathbb{E} \left[\frac{|\nabla f_X|}{f_X} \middle| \mathbf{Y} \right]. \quad (3.11)$$

The first term on the right-hand side of (3.11) is the expected flux $\square$ into the region defined by f_X . The second term is the expected value of the normalised density ratio $N^{-1} |\nabla f_X| / f_X$ multiplied by N and conditioned on $\mathbf{Y}$. Now consider the probability that $\mathbf{Y} \in \mathcal{Y}$ (where $\mathcal{Y} \subseteq \mathbb{R}^n$ is an arbitrary subset) over the boundary region using the indicator function $\mathbb{I}_{\mathcal{Y}}$:

$$\tilde{\mathbb{P}}\{\mathbf{Y} \in \mathcal{Y}\} := \tilde{\mathbb{E}}[\mathbb{I}_{\mathcal{Y}}] = \mathbb{E} \left[\mathbb{I}_{\mathcal{Y}} \frac{|\nabla f_X|}{N f_X} \right] = \int_{\mathcal{Y}} \underbrace{\mathbb{E} \left[\frac{|\nabla f_X|}{f_X} \middle| \mathbf{Y} = \mathbf{y} \right]}_{=: \tilde{f}_Y(\mathbf{y})} \frac{f_Y(\mathbf{y})}{N} \, d\mathbf{y}, \quad (3.12)$$

which defines the probability density $\tilde{f}_Y$ of $\mathbf{Y}$ when $\mathbf{X}$ is sampled from the boundary region and enables the final term in (3.11) to be expressed as $N \tilde{f}_Y / f_Y$. The normalisation constant N can be interpreted as the relative size of the boundary region. In particular, as f_X tends to the indicator function $\mathbb{I}_{\mathcal{X}}$, selecting a subset $\mathcal{X} \subseteq \mathbb{R}^d$, N tends to the ratio of the surface area of $\mathcal{X}$ to its volume. If, for example, $\mathcal{X}$ is a unit interval of $\mathbb{R}$, the unnormalised distribution $|\nabla f_X|$ would tend to a delta measure at either end of the interval and the required normalisation constant N would, therefore, tend to 2.

3.3. Backward Kolmogorov equation

For random variables $\square$ and $\mathbf{Y}$, and a function $g : \mathbb{R}^n \rightarrow \mathbb{R}$, the expected value of $g(\mathbf{Y})$, conditional on $\mathbf{Y} = \mathbf{y}$, is $g(\mathbf{y})$. Therefore, $\mathbb{E}[\square g(\mathbf{Y}) | \mathbf{Y} = \mathbf{y}] = \mathbb{E}[\square | \mathbf{Y} = \mathbf{y}] g(\mathbf{y})$ and (3.3) becomes

$$\mathcal{L}g(\mathbf{y}) = \mathbf{D}_1^i \frac{\partial g}{\partial y^i} + \mathbf{D}_2^{ij} \frac{\partial^2 g}{\partial y^i \partial y^j}. \quad (3.13)$$

Using the results from § 3.2, the so-called drift velocity is

$$\mathbf{D}_1 := \mathbb{E}[\mathbf{I}_s + \mathbf{J}_s \cdot \nabla \log f_X | \mathbf{Y}_s = \mathbf{y}] \quad (3.14)$$

and the symmetric diffusion coefficient is

$$\mathbf{D}_2 := \frac{1}{2} \mathbb{E}[\mathbf{J}_s^\top \nabla \mathbf{Y}_s + \nabla \mathbf{Y}_s^\top \mathbf{J}_s | \mathbf{Y}_s = \mathbf{y}] + \frac{1}{2} \boldsymbol{\Sigma}. \quad (3.15)$$

If $u(\mathbf{y}, s) := \mathbb{E}[g(\mathbf{Y}_t) | \mathbf{Y}_s = \mathbf{y}]$, then

$$-\partial_s u(\mathbf{y}, s) = \mathcal{L}u(\mathbf{y}, s) \quad (3.16)$$

is solved backwards in time from the end condition $u(\mathbf{y}, t) = g(\mathbf{y})$, so that $u(\mathbf{y}, s)$ is the expected value of $g(\mathbf{Y}_t)$ given $\mathbf{Y}_s = \mathbf{y}$ for $s \leq t$. The generator $\mathcal{L}$ therefore ‘pulls back’ the observation g along the dynamics specified by (3.1).

A local (in space) diffusivity with what looks like the same form as (3.15) is found in classical p.d.f. methods (Pope 1981). Here, however, the spatial conditioning is distributed over an arbitrary control volume, as defined by the sample distribution f_X (see item (i) at the end of § 1), which leads to the boundary fluxes in (3.14) and has been derived from the dual perspective of observables g (see item (ii) at the end of § 1). Consequently, (3.15) readily accounts for distributed stochastic forcing through $\boldsymbol{\Sigma}$. The following section explains how $\mathbf{D}_1$ and $\mathbf{D}_2$ determine the evolution of probability density using duality.

3.4. Forward Kolmogorov equation

The ‘forward’ equation corresponding to (3.16) is derived by expressing the time rate of change of the conditional expectation $\mathbb{E}[g(\mathbf{Y}_t) | \mathbf{Y}_s = \mathbf{y}]$ in terms of a conditional probability density $f_{Y_t | Y_s}$, as described by Pavliotis (2014, pp. 45–50). Here, we present an abridged version to highlight the main idea without introducing new notation.

If the density $f_Y(\mathbf{y}, t)$ is understood as being conditional on a specified density $f_Y(\mathbf{y}, 0)$ at $t = 0$, the expected value of g resulting from the initial density $f_Y(\mathbf{y}, 0)$ satisfies

$$\partial_t \mathbb{E}[g(\mathbf{Y}_t)] = \partial_t \int_{\mathbb{R}^n} g(\mathbf{y}) f_Y(\mathbf{y}, t) d\mathbf{y} = \int_{\mathbb{R}^n} g(\mathbf{y}) \partial_t f_Y(\mathbf{y}, t) d\mathbf{y}. \quad (3.17)$$

Alternatively, using $\mathcal{L}$, integrating by parts and assuming that either f_Y or g decay sufficiently fast as $|\mathbf{y}| \rightarrow \infty$,

$$\partial_t \mathbb{E}[g(\mathbf{Y}_t)] = \int_{\mathbb{R}^n} \mathcal{L}g(\mathbf{y}) f_Y(\mathbf{y}, t) d\mathbf{y} = \int_{\mathbb{R}^n} g(\mathbf{y}) \mathcal{L}^\dagger f_Y(\mathbf{y}, t) d\mathbf{y}. \quad (3.18)$$

Therefore, equating (3.17) and (3.18), and noting that they are valid for all suitable observables g ,

$$\partial_t f_Y(\mathbf{y}, t) = \mathcal{L}^\dagger f_Y(\mathbf{y}, t), \quad (3.19)$$

where

$$\mathcal{L}^\dagger f_Y(\mathbf{y}, t) := -\frac{\partial}{\partial y^i}(\mathbf{D}_1^i f_Y(\mathbf{y}, t)) + \frac{\partial^2}{\partial y^i \partial y^j}(\mathbf{D}_2^{ij} f_Y(\mathbf{y}, t)). \quad (3.20)$$

The forward equation evolves f_Y forwards in time from the initial density $f_Y(\mathbf{y}, 0)$.

3.5. Physical interpretation

Sources $\mathbf{I}$ and boundary fluxes $\mathbf{J} \cdot \nabla f_X / |\nabla f_X|$ into/out of the control volume contribute to positive/negative drift $\mathbf{D}_1$ in the evolution of the density f_Y . As discussed in §3.2 beneath (3.12), the boundary fluxes are scaled to account for the relative size of the boundary. Both processes contribute to drift because they determine a rate of gradual change in $\mathbf{Y}$, rather than an immediate addition or removal of fluid with a given value of $\mathbf{Y}$. An exception to this interpretation occurs when advection due to a velocity field is included in $\mathbf{J}$, which will be discussed in §4.

Expectations of the cross-gradient mixing terms $\mathbf{J}^\top \nabla \mathbf{Y}$ determine the diffusion coefficient $\mathbf{D}_2$ in (3.15). The sign of $\mathbf{D}_2$ will be discussed in detail in §5. In many practical applications, and according to Fick's law, each column of $\mathbf{J}$ points in the direction of each column of $-\nabla \mathbf{Y}$, making it reasonable to expect $\mathbf{D}_2$ to be negative semidefinite. This property corresponds to the fact that down-gradient molecular transport homogenises $\mathbf{Y}$, which leads to greater certainty in the value of $\mathbf{Y}_t$ and lower entropy of the distribution f_Y .

The following sections (§§4 and 5) consider fluxes that comprise advection and diffusion:

$$\mathbf{J} = \mathbf{U} \otimes \mathbf{Y} - \nabla \mathbf{Y} \boldsymbol{\alpha}, \quad (3.21)$$

where $\mathbf{U}$ is a solenoidal velocity field (i.e. $\nabla \cdot \mathbf{U} \equiv 0$) and $\boldsymbol{\alpha} \in \mathbb{R}^{n \times n}$ is a (typically diagonal) multicomponent diffusion matrix. More generally, $\boldsymbol{\alpha}$ could also account for anisotropic diffusion, which would make it a four-dimensional tensor.

4. Stirring due to advection

To understand the role that the advective component of the flux $\mathbf{J}$ in (3.21) plays it is useful to revisit the identity (3.6) in §3.2, recognising that if $\nabla \cdot \mathbf{U} \equiv 0$, the chain rule implies

$$-\nabla \cdot (\mathbf{U} Y^i) \frac{\partial g(\mathbf{Y})}{\partial Y^i} = -\nabla \cdot (\mathbf{U} g). \quad (4.1)$$

Consequently, by applying the steps described in §3.2 to (4.1), the contributions that the flux $\mathbf{U} \otimes \mathbf{Y}$ makes to the right-hand side of (3.20) through $\mathbf{D}_1$ and $\mathbf{D}_2$ can be replaced with a single sink term $-V f_Y$:

$$\partial_t f_Y = (\mathcal{L}^\dagger - V) f_Y, \quad (4.2)$$

where

$$V(\mathbf{y}, t) := -\mathbb{E}[\mathbf{U}_t \cdot \nabla \log f_X | \mathbf{Y}_t = \mathbf{y}]. \quad (4.3)$$

Unlike the drift and diffusion terms in (3.20), V modifies f_Y by accounting for the addition or removal of fluid, with a scalar field value of $\mathbf{y}$, over the sample distribution's boundary. The inclusion of V is the subject of the Feynman–Kac formula, which is discussed in Appendix C and provides an interpretation of V as modifying the expectation with respect to which $u(\mathbf{y}, s)$ in (3.16) is defined. As expected, if $\mathbf{U} \cdot \nabla f_X \equiv 0$, then the boundary fluxes are zero and the velocity field serves to stir the various scalar fields internally without affecting f_Y . The incorporation of time-dependent sample distributions

f_X , which produce further contributions to V , are discussed in [Appendix C.1](#). In particular, [Appendix C.1](#) demonstrates that if f_X is defined as a Lagrangian control distribution, advected by the fluid, then $V \equiv 0$.

For f_Y to evolve as a probability density, the integral of Vf_Y over $\mathbb{R}^n$ must vanish. Although it looks as though the sink term on the right-hand side of (4.2) might violate this requirement, it should be recalled, using (3.5), that

$$\int_{\mathbb{R}^n} Vf_Y(\mathbf{y}, t) d\mathbf{y} = \mathbb{E}[\mathbb{E}[U_t \cdot \nabla \log f_X | Y_t = \mathbf{y}]] = \mathbb{E}[\mathbb{E}[U_t \cdot \nabla \log f_X | X = \mathbf{x}]], \quad (4.4)$$

where, using integration by parts, the final expectation can be expressed as

$$\int_{\mathbb{R}^d} \mathbb{E}[U_t | X = \mathbf{x}] \cdot \nabla f_X d\mathbf{x} = -\mathbb{E}[\nabla \cdot U_t] = 0, \quad (4.5)$$

which demonstrates that the integral of f_Y is invariant to the sources and sinks arising from advection over the boundary region. A similar expression to (4.5) can be found from Pope (2000, § 12.6.2, (12.217)) when f_X is regarded as a particle position density.

5. Mixing due to diffusion

In this section, the focus will be on the effect that mixing has in the forward equation for probability density (3.19). It is therefore convenient to start by neglecting Σ , which, being positive semidefinite by construction, typically acts in (3.15) to oppose the effects of mixing.

Neglecting Σ , using (3.21) and applying the result from § 4, the diffusion coefficient (3.15) becomes

$$D_2 := -\frac{1}{2} \mathbb{E}[\alpha^\top \nabla Y^\top \nabla Y + \nabla Y^\top \nabla Y \alpha | Y = y]. \quad (5.1)$$

If, in particular, a sampling distribution f_X completely encompasses a scalar field, such that the boundary fluxes in D_1 are zero, then the forward equation for probability density (4.2) for a single scalar Y with diffusivity $\alpha > 0$ can be multiplied by y^2 and integrated to give the well-known evolution of variance (Zeldovich 1937):

$$\frac{1}{2} \frac{d}{dt} \mathbb{E}[Y_t^2] = -\alpha \mathbb{E}[|\nabla Y_t|^2], \quad (5.2)$$

which illustrates that the diffusivity α reduces the variance of Y_t due to the fact that $D_2 < 0$.

Indeed, for many practical applications, it is reasonable to expect D_2 to be negative semidefinite, corresponding to the fact that down-gradient molecular transport homogenises Y , which leads to greater certainty in the value of Y . For example, in physical space, the eventual steady state of a scalar subjected to diffusion in an insulated domain will be uniform, which corresponds to a Dirac measure in the probability distribution.

More generally, however, the properties of (5.1) are intriguing because they depend on both the relative diffusivities α of the observed quantities Y and the *a priori* unknown correlation between gradients ∇Y . It is therefore useful to understand when to expect negative semidefinite D_2 (which will be denoted $D_2 \leq 0$ or, equivalently, $-D_2 \geq 0$ when it is convenient to refer to positive semidefinite matrices).

Begin by noting that outer products $\mathbf{r} \otimes \mathbf{r}$ for $\mathbf{r} \in \mathbb{R}^n$ are extreme rays and, therefore, generators of the convex cone of positive semidefinite $n \times n$ matrices (Blekherman, Parrilo & Thomas 2012). In particular, the matrix product $\nabla Y^\top \nabla Y \geq 0$, because, noting

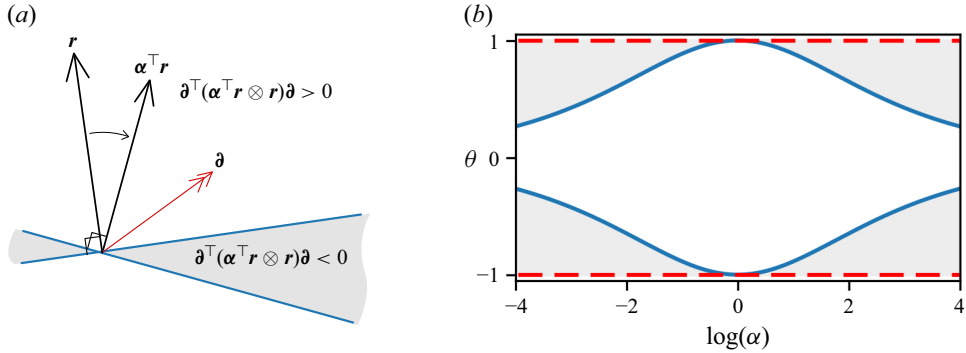


Figure 5. (a) Violation of positive semidefiniteness in the generator $\alpha^\top \mathbf{r} \otimes \mathbf{r}$ of $-\mathbf{D}_2$ when the quantities $\mathbf{Y}$ have different diffusivities α . (b) Relationship between the ratio of diffusivities α and the correlation coefficient θ that ensures that $\mathbf{D}_2$ in (5.4) is negative semidefinite (white region).

that $\nabla \mathbf{Y} \in \mathbb{R}^{d \times n}$, it can be represented as the sum of d outer products (one for gradients with respect to each spatial dimension $X^1, X^2, \dots, X^d$). Indeed, the expectation of such products also produces positive semidefinite matrices, because $\partial^\top \mathbb{E}[\nabla \mathbf{Y}^\top \nabla \mathbf{Y} | \mathbf{Y} = \mathbf{y}] \partial = \mathbb{E}[(\partial^\top \nabla \mathbf{Y}^\top)(\nabla \mathbf{Y} \partial) | \mathbf{Y} = \mathbf{y}] \geq 0$ for all non-zero vectors $\partial \in \mathbb{R}^n$ is a sum of squares, which implies that

$$\mathbb{E}[\nabla \mathbf{Y}^\top \nabla \mathbf{Y} | \mathbf{Y} = \mathbf{y}] \geq 0. \quad (5.3)$$

The (negative) diffusion coefficient $-\mathbf{D}_2$ in (3.15), however, is effectively generated by the outer product of the vectors $\alpha^\top \mathbf{r}$ and $\mathbf{r}$. If the diffusivities of each observable quantity $\mathbf{Y}$ are equal and non-negative, such that α can be replaced with a single scalar $\alpha \in \mathbb{R}_{\geq 0}$, then $\alpha^\top \mathbf{r}$ and $\mathbf{r}$ point in the same direction and, following the above-mentioned arguments, $\mathbf{D}_2$ is negative semidefinite ($-\mathbf{D}_2 \geq 0$). More generally, however, when α cannot be replaced with a scalar, the outer product $\alpha^\top \mathbf{r} \otimes \mathbf{r}$ creates the possibility of $\mathbf{D}_2$ being sign indefinite, as illustrated by the shaded region of figure 5(a). The determining factor in such cases is the correlation between the gradients of the various components of $\mathbf{Y}$, since it is possible for the sum of sign indefinite matrices to be positive semidefinite. Weaker correlation between the gradients provides stronger mitigation of the effects of unequal diffusivities in α . For example, if α is any non-negative diagonal matrix and the n columns of $\nabla \mathbf{Y}$ are uncorrelated, such that $\mathbb{E}_Y[\nabla Y^i \cdot \nabla Y^j] = 0$ for $i \neq j$, then $\mathbf{D}_2 \leq 0$ is a negative semidefinite diagonal matrix. More generally, it is also worth noting that the scalar diffusion coefficients associated with the marginal distributions of $Y^1, Y^2, \dots, Y^n$ correspond to the diagonal elements of α .

A simple two-dimensional example illustrates the combined effects of correlation in the gradients $\nabla \mathbf{Y}$ and unequal diffusivities in α . Let $\alpha_{11} = 1$, $\alpha_{22} = \alpha$ and $\alpha_{12} = \alpha_{21} = 0$, and assume, without loss of generality, that $\mathbb{E}[\nabla Y^1 | \mathbf{Y} = \mathbf{y}] = \mathbb{E}[\nabla Y^2 | \mathbf{Y} = \mathbf{y}] = 1$ to normalise the problem. Define the correlation coefficient $\theta := \mathbb{E}[\nabla Y^1 \cdot \nabla Y^2 | \mathbf{Y} = \mathbf{y}]$, such that

$$-\mathbf{D}_2 = \begin{bmatrix} 1 & \theta(1 + \alpha)/2 \\ \theta(1 + \alpha)/2 & \alpha \end{bmatrix} \quad (5.4)$$

from (3.15). The (negative) symmetric diffusion coefficient $-\mathbf{D}_2$ in (5.4) is positive semidefinite if its minimum eigenvalue

$$\lambda = \frac{(1 + \alpha) - \sqrt{(1 + \alpha)^2 + \theta^2(1 + \alpha)^2 - 4\alpha}}{2} \quad (5.5)$$

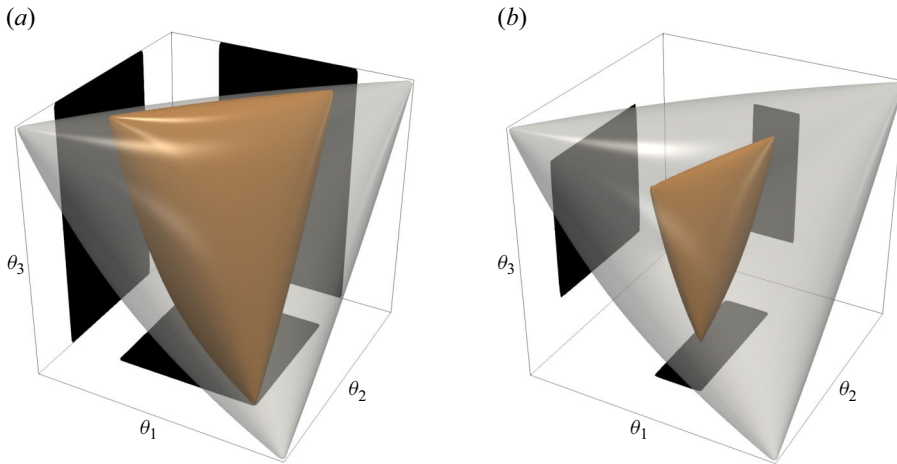


Figure 6. Spectrahedra corresponding to the condition that $-\mathbf{D}_2 \geq 0$ for correlation coefficients θ_1 , θ_2 and θ_3 , and different relative diffusivities α_2 and α_3 in (5.7). (a) $\alpha_2 = 1$ and $\alpha_3 = 10$, and (b) $\alpha_2 = 0.1$ and $\alpha_3 = 10$. The red regions correspond to $\mathbf{D}_2 \leq 0$, whilst the larger grey region corresponds to the feasibility requirement (5.3) for the correlations. The range of all axes is $[-1, 1]$.

is non-negative, which means that

$$|\theta| \leq \frac{2\sqrt{\alpha}}{1 + \alpha} \quad (5.6)$$

is the required relation between correlation and diffusivity that guarantees $-\mathbf{D}_2 \geq 0$ and is illustrated in figure 5(b). If $\theta \in \{-1, 1\}$, any difference in α from unity will lead to a sign indefinite diffusion coefficient $\mathbf{D}_2$, as motivated in the text following (5.3) and figure 5(a). However, for weaker correlations $|\theta| < 1$, the ratio of the diffusivities α has to be either large or small to produce a sign indefinite diffusion coefficient $\mathbf{D}_2$, as illustrated by the grey regions in figure 5(b). For reference, the ratio of thermal diffusivity to salt diffusivity in the oceans is approximately 100 and can be regarded as corresponding to α in the previous example. Therefore, according to (5.6), in that case, $|\theta|$ would need to be less than approximately 0.2 for $\mathbf{D}_2$ to be negative semidefinite.

When α is prescribed, it is useful to know the correlation coefficients $\theta \in \mathbb{R}^{n(n-1)/2}$ that produce negative semidefinite diffusion coefficients $\mathbf{D}_2$. In such cases, the condition that $-\mathbf{D}_2 \geq 0$ can be represented by spectrahedra, which are formally defined as the intersection of the convex cone of positive semidefinite matrices (5.3) in $\mathbb{R}^{n \times n}$ with an affine subspace (Blekherman *et al.* 2012). In three dimensions, setting $\alpha_{11} = 1$, $\alpha_{22} = \alpha_2$ and $\alpha_{33} = \alpha_3$,

$$-\mathbf{D}_2 = \begin{bmatrix} 1 & \theta_3(1 + \alpha_2)/2 & \theta_2(1 + \alpha_3)/2 \\ \theta_3(1 + \alpha_2)/2 & \alpha_2 & \theta_1(\alpha_2 + \alpha_3)/2 \\ \theta_2(1 + \alpha_3)/2 & \theta_1(\alpha_2 + \alpha_3)/2 & \alpha_3 \end{bmatrix}, \quad (5.7)$$

where θ_1 , θ_2 and θ_3 are correlation coefficients. Spectrahedra corresponding to $-\mathbf{D}_2 \geq 0$ are shown in red in figure 6. The larger spectrahedron shown in grey corresponds to (5.3), which constrains the possible values of θ . Values of θ that lie between the red and the grey regions therefore correspond to permissible but sign indefinite diffusion coefficients $\mathbf{D}_2$. In figure 6(a), $\alpha_2 = 1$ and $\alpha_3 = 10$, which leads to a narrower range of θ_1 and θ_2 values for which $-\mathbf{D}_2 \geq 0$. In figure 6(b), $\alpha_2 = 0.1$ and $\alpha_3 = 10$, which narrows the range further and increases the possibility that $\mathbf{D}_2$ is sign indefinite.

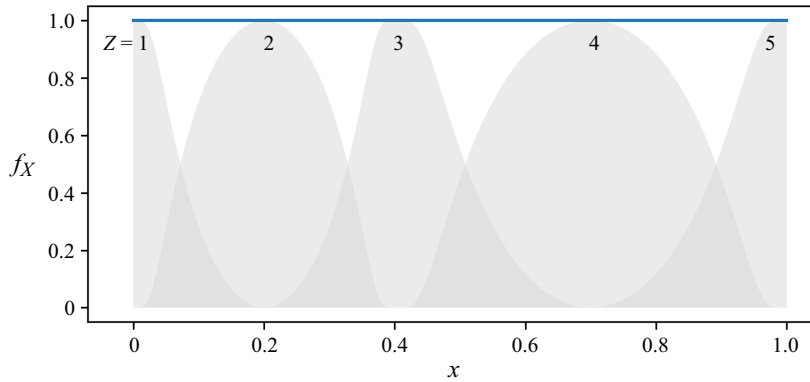


Figure 7. A uniform mixture distribution $f_X \equiv 1$ over the unit interval (blue line) obtained from the sum of $f_{X|Z}(x|z)\mathbb{P}\{Z = z\}$ according to (6.1) for $Z \in \{1, 2, 3, 4, 5\}$. In the regions where the distributions overlap, the probability that X is associated with a given component (i.e. $f_{Z|X}$) is $f_{X|Z}(x|z)\mathbb{P}\{Z = z\}$.

For situations involving stochastic forcing, diagnosing the sign of $\mathbf{D}_2$ is complicated by the presence of $\Sigma \geq 0$ in (3.15), which, in contrast to mixing, leads to greater uncertainty in the state of the system.

6. Partitions and mixture distributions

The construction in § 3 gives an evolution equation for f_Y that is conditional on the specification of the spatial probability density f_X . It therefore generalises immediately to collections of densities describing the distribution of random variables within different zones of a domain. In statistics, such a decomposition is called a mixture distribution. To paraphrase Lindsay (1995) giving a simple example, they would arise naturally when sampling from a population consisting of several homogeneous subpopulations, highlighting the potential of mixture distributions to decompose a heterogeneous flow into (approximately) homogeneous sub flows.

Using the ‘latent’ random variable $Z \in \mathbb{Z}$ to denote a given zone, defined by the conditional density $f_{X|Z}$, the original density can be recovered from Bayes’ theorem:

$$f_X(\mathbf{x}) = \sum_{z \in \mathbb{Z}} f_{X|Z}(\mathbf{x}|z)\mathbb{P}\{Z = z\}, \quad (6.1)$$

where $\mathbb{P}\{Z = z\}$ is the proportion of the original control volume attributed to zone z . An illustration of (6.1) is shown in figure 7, consisting of a decomposition of the uniform distribution over the interval $X \in [0, 1]$.

The distribution f_Y , or any distribution of Y conditional on Z belonging to a subset of $\mathbb{Z}$, can be decomposed in a similar way:

$$f_Y(\mathbf{y}) = \sum_{z \in \mathbb{Z}} f_{Y|Z}(\mathbf{y}|z)\mathbb{P}\{Z = z\}, \quad (6.2)$$

where the distribution $f_{Y|Z}$ associated with each zone satisfies the forward Kolmogorov equation (3.19), provided that f_X is replaced with $f_{X|Z}$.

For periodic domains with a uniformly distributed mixture distribution f_X (cf. figure 7), boundary fluxes arising from $\mathbf{J}$ in $\mathbf{D}_1$ vanish because $\nabla f_X \equiv \mathbf{0}$. Consequently, the sum of boundary fluxes for each component $f_{X|Z}$, representing exchange between the

components, also vanishes:

$$\sum_{z \in \mathbb{Z}} \mathbb{E} [\mathbf{J} \cdot \nabla \log f_{X|Z} | \mathbf{Y} = \mathbf{y}, Z = z] f_{Y|Z} \mathbb{P}\{Z = z\} = 0. \quad (6.3)$$

In this way, the decomposition of a mixture distribution generalises the classical partition of a domain into non-overlapping control volumes with sharp boundaries to components of f_X that can overlap, which will be used in § 7.4 to analyse Rayleigh–Bénard convection.

7. Example applications

The following subsections each illustrate particular aspects of the results derived in § 3 by providing simple example applications.

7.1. Advection and diffusion by the ABC flow

An Arnold–Beltrami–Childress (ABC) flow with coefficients equal to unity is the three-dimensional divergence-free velocity field

$$\mathbf{U} := (\sin(X^3) + \cos(X^2), \sin(X^1) + \cos(X^3), \sin(X^2) + \cos(X^1))^T. \quad (7.1)$$

It is an exact solution of the Euler equations in a periodic domain and is known to exhibit chaotic streamlines (Dombre *et al.* 1986), which makes it an ideal candidate to study mixing. Couched in the format of (3.1), the evolution of a passive scalar concentration Y_t due to the combined effects of advection by $\mathbf{U}$ and diffusion is

$$\frac{dY_t}{dt} = -\mathbf{U} \cdot \nabla Y_t + \alpha \Delta Y_t, \quad (7.2)$$

where α denotes the constant scalar diffusivity. The corresponding forward Kolmogorov equation governing the evolution of $f_Y(y, t)$ is given by (3.19):

$$\frac{\partial f_Y}{\partial t} = - \underbrace{\frac{\partial^2}{\partial y^2} (\alpha \mathbb{E}[|\nabla Y_t|^2 | Y_t = y])}_{=:-D_2} f_Y. \quad (7.3)$$

The coefficients V and D_1 from § 3.3 are both zero because the periodic domain does not have boundaries. Although (7.3) cannot be solved, as D_2 is unknown, we can examine estimates of D_2 by solving (7.2) numerically. Space is treated using a Fourier pseudo-spectral method and time using a Crank–Nicolson discretisation with a resolution of 64^3 modes and time step of $\Delta t = 5 \times 10^{-3}$, respectively. Choosing an initial condition consisting of a front separating two regions of different concentration $Y_0 = \tanh(10\mathbf{X}^T \mathbf{1})$ and setting $\alpha = 1/10$, we numerically integrate with respect to $t \in [0, 4]$. Then, by approximating the density $f_Y(y, t)$ and $\mathbb{E}[|\nabla Y_t|^2 | Y_t = y]$ with a histogram, constructed from a single snapshot of field data, we estimate the terms in (7.3).

A time evolution of a cross-section of the scalar field Y_t alongside its corresponding density $f_Y(y, t)$ and the conditional expectation $\mathbb{E}[|\nabla Y_t|^2 | Y_t = y]$ are shown in figure 8. At $t = 0.5$, the scalar field, consisting of two regions of approximately uniform concentration separated by a relatively sharp interface, is represented in probability space by spikes at $y = \pm 1$. As the concentration field is mixed irreversibly, the amplitude of these spikes reduces and the density associated with values of y in the vicinity of zero increases. Mixing subsequently homogenises the scalar field to the extent that no evidence of the initial distribution remains by $t = 2$. As there is no injection of concentration to balance the homogenising effects of diffusion, Y_t will ultimately tend to a constant uniformly over the domain and the density $f_Y(y, t)$ in the limit $t \rightarrow \infty$ would tend to a Dirac distribution.

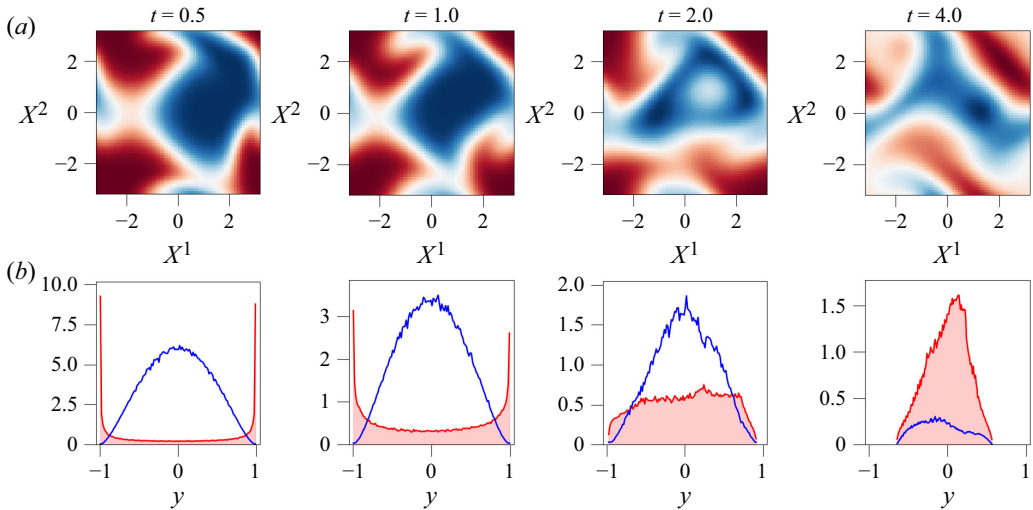


Figure 8. (a) Time evolution of a cross-section of the concentration field Y_t and (b) its corresponding probability density $f_Y(y, t)$ (shaded) and (negative) diffusion coefficient $\alpha \mathbb{E}[|\nabla Y_t|^2 | Y_t = y]$. <https://www.cambridge.org/S0022112025106319/JFM-Notebooks/files/fig08/fig08.ipynb>.

The expectation $\mathbb{E}[|\nabla Y_t|^2 | Y_t = y]$, which quantifies the rate of mixing and, therefore, negative diffusion in (7.3) is maximised by $y = 0$. For $t = 0.5$ and $t = 1.0$, scalar concentrations that are less probable are associated with the most mixing. This observation is not surprising because, as discussed in the examples in § 2 and illustrated in figure 1, large gradients are associated with small probability densities, unless they occur over sufficiently many branches of the function's inverse. In this regard, it should also be noted that $\mathbb{E}[|\nabla Y_t|^2 | Y_t = y]$ is zero for the minimum ($y = -1$) and maximum ($y = 1$) values of Y_t , at which $\nabla Y_t = \mathbf{0}$. In the absence of diffusion at these values of y , the density f_Y consequently maintains compact support (i.e. mixing interpolates and cannot produce values that lie outside the range of values that were there in the first place).

7.2. Stochastic boundary conditions

To sustain a finite variance of a diffusive scalar field, it is necessary to apply a forcing. Given that a wide variety of physical systems are sustained by a boundary forcing, often characterised by substantial fluctuations and uncertainty, it is natural to consider how these forces manifest in the evolution equation governing the system's probability density. A simple example of such a problem that is sufficient to highlight several of the terms discussed in § 3 is the diffusion of heat through a one-dimensional rod of unit length forced by Ornstein–Uhlenbeck processes (i.e. normally distributed thermal fluctuations) at the boundary. When cast in the form of (3.1), for unit thermal diffusivity ($\alpha = 1$), the temperature Y_t evolves according to

$$\frac{dY_t}{dt} = \partial_X^2 Y_t \quad \text{for } X \in [0, 1], \quad (7.4)$$

where values of Y_t for points on the boundary $\partial\Omega$ (corresponding to $X \in \{0, 1\}$) evolve according to

$$dY_t(\omega) = -aY_t(\omega) dt + \sigma dW_t(\omega) \quad \text{for } \omega \in \partial\Omega, \quad (7.5)$$

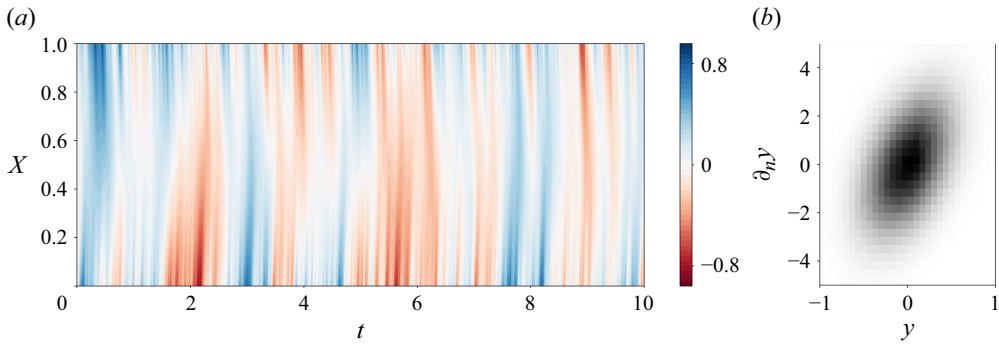


Figure 9. (a) Space–time plot of Y_t for the one-dimensional diffusion (7.4) forced by an Ornstein–Uhlenbeck process (7.5) with $a = 5$, $\sigma = 1$, and (b) joint density $\tilde{f}_{Y, \partial_n Y}$ conditioned on the boundaries. The positive covariance of this plot reflects the tendency of heat to flow down gradients at the boundaries. <https://www.cambridge.org/S0022112025106319/JFM-Notebooks/files/fig09-10/fig09-10.ipynb>.

where a, σ are real constants and $W_t(\omega)$ is a Wiener process. In the context of (3.1), $J_t = -\partial_X Y_t$ is the (diffusive) scalar flux. Since we are interested in the entire domain, the sampling distribution is $f_X \equiv 1$ and $\partial_X \log f_X$ in the calculation of $\mathbf{D}_1$ should be regarded as changing the distribution under which the expectation is calculated from f_X to point evaluation at the (two) boundaries (see discussion following (3.9) in § 3.2):

$$\mathbb{E}[J_t \partial_X \log f_X | Y_t] f_Y = 2 \mathbb{E}[\partial_n Y_t | Y_t] \tilde{f}_Y, \quad (7.6)$$

where $\partial_n Y_t$ is the boundary normal derivative of Y_t , $\tilde{f}_Y$ is the probability density of Y_t , conditional on evaluation at a boundary, and the factor 2 accounts for there being two boundaries separated by a distance of one unit.

Figure 9(a) shows a space–time plot of the system, for which the temperature at the boundaries fluctuates on the integral time scale ($1/a$) of the Ornstein–Uhlenbeck processes (7.5). The solution is obtained numerically using second-order finite differences in space with 32 grid points and an Euler–Maruyama discretisation in time with a time step of 2.5×10^{-3} . Towards the centre of the domain, the effects of diffusion reduce the amplitude of the fluctuations.

Using (7.6), the forward Kolmogorov equation, which describes how the probability density $f_Y(y, t)$ evolves in time, is given by (3.20):

$$\frac{\partial f_Y}{\partial t} = - \underbrace{\frac{\partial}{\partial y} \left(2 \mathbb{E}[\partial_n Y_t | Y_t = y] \frac{\tilde{f}_Y}{f_Y} f_Y \right)}_{=: \mathbf{D}_1} - \underbrace{\frac{\partial^2}{\partial y^2} \left(\mathbb{E}[|\partial_X Y_t|^2 | Y_t = y] f_Y \right)}_{=: -\mathbf{D}_2}, \quad (7.7)$$

According to (7.5), Y_t sampled at the boundary is an Ornstein–Uhlenbeck process and therefore has a stationary density given by $\tilde{f}_Y(y) = \sqrt{a/(\pi\sigma^2)} \exp(-ay^2/\sigma^2)$ (Gardiner 1990), which, using the approach described following (3.12), is used to express $\mathbf{D}_1$.

As there is no advection in (7.4), the coefficient V from § 4 is zero. However, due to conduction and, in turn, boundary fluxes of heat, (7.7) has non-zero drift and diffusion terms. Although (7.7) cannot be solved prognostically without a model for the unknown coefficients, the coefficients can be approximated numerically by constructing an ensemble satisfying (7.4) and (7.5).

Figure 10 shows the drift $\mathbf{D}_1$ and diffusion $\mathbf{D}_2$ coefficients, as well as the stationary density $f_Y(y)$ using time and ensemble averaging. An ensemble of 500 paths each

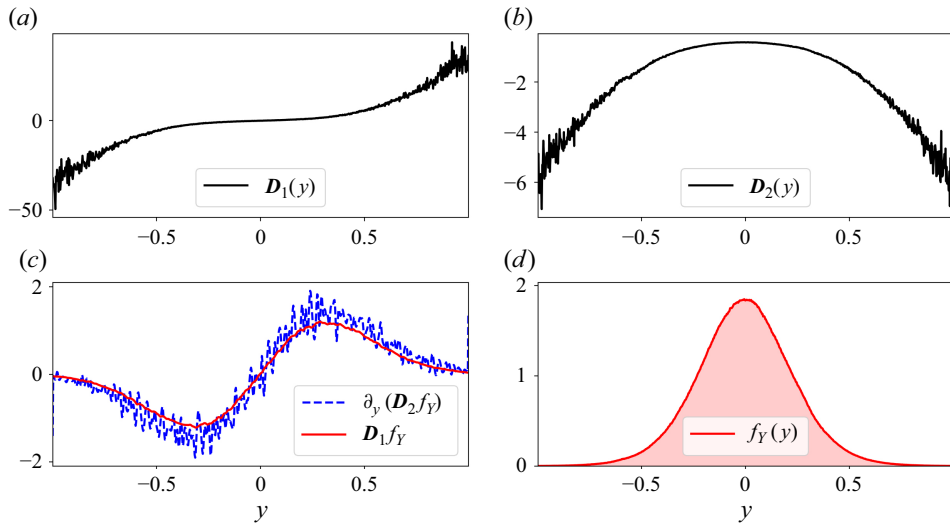


Figure 10. (a,b,d) Drift coefficient $\mathbf{D}_1$, diffusion coefficient $\mathbf{D}_2$ and probability density $f_Y(y)$, respectively, of the forward Kolmogorov equation corresponding to a stationary state of the system (7.4) with $a = 5$, $\sigma = 1$. (c) A numerical verification of the balance between boundary forcing and mixing. <https://www.cambridge.org/S0022112025106319/JFM-Notebooks/files/fig09-10/fig09-10.ipynb>.

containing 1950 snapshots in time was constructed numerically by simulating (7.4), (7.5) for $t \in [0, 5]$ and discarding the initial 50 snapshots. In § 7.1, molecular diffusion acted to homogenise the scalar field Y_t in the ABC flow and drive its corresponding density towards a Dirac measure. Were the value of Y_t at the boundaries in this example to be fixed at zero, the distribution would, due to the negative diffusion $\mathbf{D}_2$ (figure 10d), also tend towards a Dirac measure at zero, irrespective of the precise initial conditions. Instead, random forcing at the system's boundary creates variance that balances the destruction of variance described by $\mathbf{D}_2$. To illustrate how, figure 9(b) shows the conditional joint density $\tilde{f}_{Y, \partial_n Y_t}$ of Y_t and its boundary-normal gradient $\partial_n Y_t$. The conditional density reveals the expected positive correlation between Y_t and $\partial_n Y_t$, which means that heat flux into/out of the domain is typically accompanied by positive/negative temperatures at the boundary. Therefore, $\mathbf{D}_1 \leq 0$ for $y \leq 0$, as shown in figure 10(a), which implies that $\mathbf{D}_1$ corresponds to divergent transport of probability density away from the origin.

Computing the balance of the terms in the right-hand side of (7.7), as shown in figure 10(c), verifies (7.7). It is worth noting that while incorporating uncertainty into (7.4) required a Monte Carlo approach to be employed in order to estimate $f_Y(y)$, such boundary conditions pose no particular additional difficulty in (7.7), which accommodates stochastic forcing naturally.

7.3. Boussinesq equations

The following example demonstrates that coordinates, regarded here as functions $\mathbf{X}$ over the sample space Ω , can be included in the state vector $\mathbf{Y}_t$. For instance, if one wishes to understand probability distributions over horizontal slices of a domain, the vertical coordinate can be included in $\mathbf{Y}_t$. More generally, other functions, such as geopotential height, could be incorporated into $\mathbf{Y}_t$. The equations developed in § 3 are therefore very general in being able to generate equations that are specific to a particular problem and circumvent the need for case-specific derivations.

In a Boussinesq context, the pointwise deterministic equations governing the behaviour of the velocity field $\mathbf{U}_t$ and the buoyancy B_t are

$$\begin{aligned}\frac{d\mathbf{U}_t}{dt} &= B_t \mathbf{e}_3 - \nabla P_t - \mathbf{U}_t \cdot \nabla \mathbf{U}_t + \alpha_1 \Delta \mathbf{U}_t, \\ \nabla \cdot \mathbf{U}_t &= 0, \\ \frac{dB_t}{dt} &= -\mathbf{U}_t \cdot \nabla B_t + \alpha_2 \Delta B_t,\end{aligned}\quad (7.8)$$

where $\mathbf{e}_3$ is the unit vector in the vertical direction. Since the vertical direction plays a distinguished role in corresponding to the direction of gravity, let $\mathbf{Y}_t := (W_t, B_t, Z)^\top$, where $W_t := \mathbf{U}_t \cdot \mathbf{e}_3$ is the vertical velocity (not to be confused with the vector of Wiener processes $\mathbf{W}_t$ used in § 3) and $Z = X^3$ is the (time independent) vertical coordinate. Writing $dZ/dt = W_t - \mathbf{U}_t \cdot \nabla Z = 0$, and noting that $\alpha_{11} = \alpha_1$, $\alpha_{22} = \alpha_2$, $\alpha_{33} = 0$ and $\alpha_{ij} = 0$ for $i \neq j$ implies that V is given by (4.3),

$$\mathbf{D}_1 = \begin{pmatrix} b \\ 0 \\ w \end{pmatrix} - \mathbb{E} \left[\begin{array}{c} \partial_Z P_t + \alpha_1 \nabla W_t \cdot \nabla \log f_X \\ \alpha_2 \nabla B_t \cdot \nabla \log f_X \\ 0 \end{array} \middle| \mathbf{Y}_t = \mathbf{y} \right], \quad (7.9)$$

and

$$\mathbf{D}_2 = -\frac{1}{2} \mathbb{E} \left[\begin{array}{ccc} 2\alpha_1 |\nabla W_t|^2 & (\alpha_1 + \alpha_2) \nabla W_t \cdot \nabla B_t & \alpha_1 \partial_Z W_t \\ (\alpha_1 + \alpha_2) \nabla W_t \cdot \nabla B_t & 2\alpha_2 |\nabla B_t|^2 & \alpha_2 \partial_Z B_t \\ \alpha_1 \partial_Z W_t & \alpha_2 \partial_Z B_t & 0 \end{array} \middle| \mathbf{Y}_t = \mathbf{y} \right], \quad (7.10)$$

where $\mathbf{y} := (w, b, z)^\top$. In $\mathbf{D}_1$, b is responsible for drift in the w direction, because buoyancy increases W_t and, in turn, w is responsible for drift in the z direction, because vertical velocity increases Z . When interpreting the latter, it should be borne in mind that for the closed domain in this example $\mathbb{E}[W_t|Z] = 0$, which, therefore, does not affect the marginal distribution of Z (i.e. the domain does not change its shape over time). The unknown conditionally averaged vertical pressure gradient also affects the evolution of the joint density through $\mathbf{D}_1$, whose remaining terms account for boundary fluxes of W_t and B_t , scaled by the relative size of the boundary.

While the drift $\mathbf{D}_1$ is responsible for moving and stretching the joint density, the diffusion coefficient $\mathbf{D}_2$ accounts for the effects of irreversible mixing. As discussed in § 5, $\mathbf{D}_2$ is expected to be negative semidefinite ($\mathbf{D}_2 \leq 0$) in most applications, which means that it typically represents anti-diffusion. In particular, if $\alpha_1 = \alpha_2 = 1$, then $\mathbf{D}_2 \leq 0$ can be guaranteed. For other combinations of α_1 and α_2 , $\mathbf{D}_2 \leq 0$ depends on the correlations among ∇W_t , ∇B_t and ∇Z , as discussed in § 5. The gradients, ∇W_t and ∇B_t are *a priori* unknown and therefore require closure (see § 8), which would involve postulating their dependence on the joint density $f_{\mathbf{Y}}$ and/or the independent variables $\mathbf{y} := (w, b, z)^\top$.

7.4. Rayleigh–Bénard convection with a stochastic boundary condition

This section describes an example of the domain partition discussed in § 6 applied to Rayleigh–Bénard convection. Here, the zonal sample distributions are defined using the diffusive vertical buoyancy flux and therefore separate the bulk central part of the flow from the top and bottom boundary layers (see figure 11, which displays instantaneous fields alongside weighted sample distributions $f_{X|Z}$ for each zone). To illustrate how the partition yields useful information, we focus on the contributions to the equations governing $f_{\mathbf{Y}|Z}$ from V and $\mathbf{D}_1$ arising from transport between the zones.

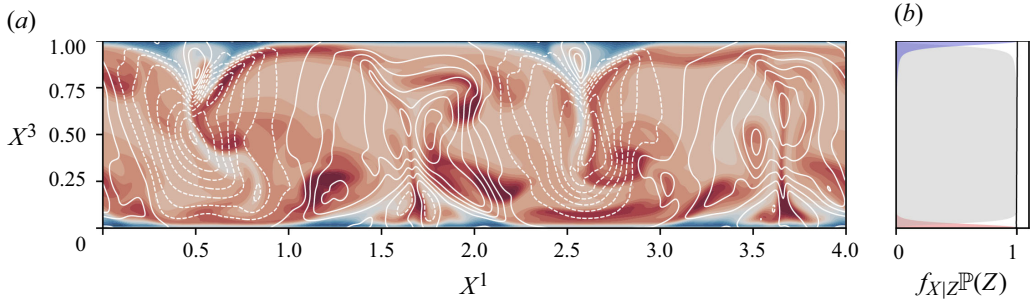


Figure 11. (a) Snapshot of the vertical velocity $Y_t^1 = W_t$ (dashed/solid white contours for negative/positive vertical velocity, respectively) and buoyancy field $Y_t^2 = B_t$ (red is positive; blue is negative) following a negative fluctuation in the buoyancy at the bottom boundary. (b) Partition of the vertical domain into three zones $Z \in \{z_1, z_2, z_3\}$ defined by $f_{X|Z}$, where $f_{X|Z}(x|z_2)$ (the central zone) is proportional to the horizontally and time-averaged vertical buoyancy flux $W_t B_t$. In the boundary-layer zones above and below the bottom and top boundary, $f_{X|Z}$ is proportional to the horizontally and time-averaged diffusive buoyancy flux $-\alpha_2 \partial_{X^3} B_t$. The constant of proportionality $\mathbb{P}\{Z = z\}$ for each zone is chosen to ensure that the sum of $f_{X|Z} \mathbb{P}$ over the zones is the uniform distribution (black line) (cf. (6.1) in § 6). <https://www.cambridge.org/S0022112025106319/JFM-Notebooks/files/fig11-12/fig11-12.ipynb>.

The data for this example were obtained from direct numerical simulation of the Boussinesq equations (7.8) on a two-dimensional and horizontally periodic domain of unit height, as shown in figure 11(a). A spatially uniform buoyancy equal to $-1/2$ is imposed on the top boundary. On the bottom boundary, the imposed buoyancy is spatially uniform but follows an Ornstein–Uhlenbeck process in time, as described in § 7.2. The Ornstein–Uhlenbeck process has a mean of $1/2$, unit variance and a time scale (corresponding to $1/a$ in § 7.2) of 2 units. The governing equations can therefore be regarded as non-dimensionalised on the domain height and mean buoyancy difference between the horizontal boundaries. In terms of these scales, the Rayleigh number is 10^7 , which corresponds to a (non-dimensionalised) viscosity and diffusivity of $\alpha_1 = \sqrt{10^7}$ and $\alpha_2 = \sqrt{10^7}$, respectively, for a Prandtl number of unity.

The governing equations are approximated numerically using the pseudo-spectral method provided by Dedalus (Burns *et al.* 2020). Fourier and Chebyshev bases are used for the horizontal (X^1) and vertical (X^3) directions, with 512 and 128 grid points, respectively. The grid points are spaced uniformly in the horizontal direction and non-uniformly in the vertical direction, where they correspond to Gauss–Chebyshev quadrature nodes. A single simulation is run for 4000 time units, where one time unit corresponds to a typical turnover time. The probability densities discussed later and presented in figure 12 were constructed from field data sampled at regular time intervals of 2 units following a discarded transitory period of 100 time units.

Following §§ 3.4 and 6, the joint probability density for each zone evolves according to $\partial_t f_{Y|Z} = (\mathcal{L}^\dagger - V) f_{Y|Z}$, where $\mathcal{L}$ and V are specific to each zone. For example, the source/sink term is

$$V(Y_t|Z, t) := -\mathbb{E}[U_t \cdot \nabla \log f_{X|Z}|Y_t, Z] \quad (7.11)$$

and the drift coefficient is

$$D_1(Y_t|Z, t) := \begin{pmatrix} b \\ 0 \end{pmatrix} - \mathbb{E} \left[\begin{pmatrix} \partial_{X^3} P_t \\ 0 \end{pmatrix} \middle| Y_t, Z \right] - \underbrace{\mathbb{E} \left[\begin{pmatrix} \alpha_1 \nabla W_t \cdot \nabla \log f_{X|Z} \\ \alpha_2 \nabla B_t \cdot \nabla \log f_{X|Z} \end{pmatrix} \middle| Y_t, Z \right]}_{D_1^{\text{wall}} + D_1^{\text{zone}}}, \quad (7.12)$$

where D_1^{wall} and D_1^{zone} distinguish between transport across the bottom and top boundaries ($X^3 \in \{0, 1\}$), which was discussed in the example in § 7.2, from transport between zones, labelled D_1^{zone} . In each of these expressions, it is important to appreciate that the expectation is conditioned, and therefore particular, to a zone Z . Integration with respect to w of the governing equation for $f_{Y|Z}$ recovers an equation for $f_{B|Z}$, the marginal probability density for the buoyancy in each zone:

$$\begin{aligned}\partial_t f_{B|Z} &= \int_{\mathbb{R}} (\mathcal{L}^\dagger - V) f_{Y|Z} dw = (\overline{\mathcal{L}^\dagger} - \overline{V}) f_{B|Z} \\ &= -\overline{V} f_{B|Z} - \partial_b ((\overline{D}_1^{wall} + \overline{D}_1^{zone}) f_{B|Z}) + \partial_b^2 (\overline{D}_2 f_{B|Z})\end{aligned}\quad (7.13)$$

for which $\overline{\square}$ denotes the projection of the operator $\square$ defined by the integral in (7.13).

Figure 11(a) displays a snapshot of the vertical velocity $Y_t^1 := W_t$ and the buoyancy $Y_t^2 := B_t$, which correspond to the fields displayed in the example derived from the Lorenz equations presented in § 2, a short time after a negative fluctuation in the buoyancy of the bottom boundary. Figure 11(b) displays the three sample distributions used to define the control volume zones z_1 (bottom boundary layer), z_2 (bulk central region) and z_3 (top boundary layer). For z_1 , $f_{X|Z}$ is proportional to the horizontally and time-averaged diffusive buoyancy flux $-\alpha_2 \partial_{X^3} \mathbb{E}[B_t | X^3]$ in the lower half of the domain and zero above. Conversely, for z_3 , $f_{X|Z}$ is proportional to $-\alpha_2 \partial_{X^3} \mathbb{E}[B_t | X^3]$ in the upper half of the domain and zero below. In the central region ($Z = z_2$), $f_{X|Z}$ is proportional to the mean convective buoyancy flux $\mathbb{E}[W B | X^3]$. For a given zone z , the constant of proportionality for each $f_{X|Z}$ can be expressed as the total (diffusive plus convective) volume-integrated vertical buoyancy flux divided by $\mathbb{P}\{Z = z\}$, which implies that $\mathbb{P}\{Z = z\}$ corresponds to the fraction of the volume-integrated vertical buoyancy flux over a given zone, leading to the partition of unity by $f_{X|Z} \mathbb{P}\{Z = z\}$ illustrated in figure 11(b).

Figure 12 displays stationary joint and marginal distributions of Y_t , along with contributions to $-V f_{Y|Z}$ and $-\partial_y D_1^{zone} f_{Y|Z}$ arising from transport between the control volumes. The range of buoyancy in zone z_1 (figure 12a) is larger than it is in zone z_3 due to the stochastic boundary condition. Indeed, the marginal distribution $f_{B|Z}$ in zone z_3 is noticeably skewed; buoyancy in the top zone does not appear to fall below the fixed value of $-1/2$ imposed on the top boundary, but reaches values up to $1/2$ due to exchange with the bulk zone below. The projected convective transport $-\overline{V} f_{B|Z}$ and the diffusive transport $-\partial_b (\overline{D}_1^{zone} f_{B|Z})$ for both boundary layer zones are the same order of magnitude and exhibit similar behaviour. In zone z_1 , fluid with relatively large ($b \approx 0.5$) buoyancy is transported out of the zone, whereas fluid with buoyancy that is close to the volume average ($b \approx 0$) is transported into the zone. In this regard, it is interesting that fluid with a buoyancy ($b < -0.5$), which must have been created in the vicinity of a large negative fluctuation in the buoyancy imposed on the bottom boundary (see figure 11a), is transported out of the zone. Similar remarks apply to zone z_3 , except that $-V f_{B|Z}$, which accounts for convective transport, tends to zero as b tends to the value -0.5 of buoyancy imposed on the top boundary. In contrast, transport from zone z_3 into zone z_2 (the central region) is negative and large in magnitude in the vicinity of the buoyancy $b = -0.5$ of the top boundary.

Zones z_1 and z_3 interact with the central bulk zone z_2 , whose joint probability density for vertical velocity W_t and buoyancy B_t is displayed in figure 12(b). The vector field depicted with blue arrows represents the flux of probability density $D_1^{zone} f_{Y|Z}$ due to diffusion to the boundary layers, whose divergence corresponds to a sink in the budget for $f_{Y|Z}$. For this particular problem, the vector field resembles a saddle node, converging with respect

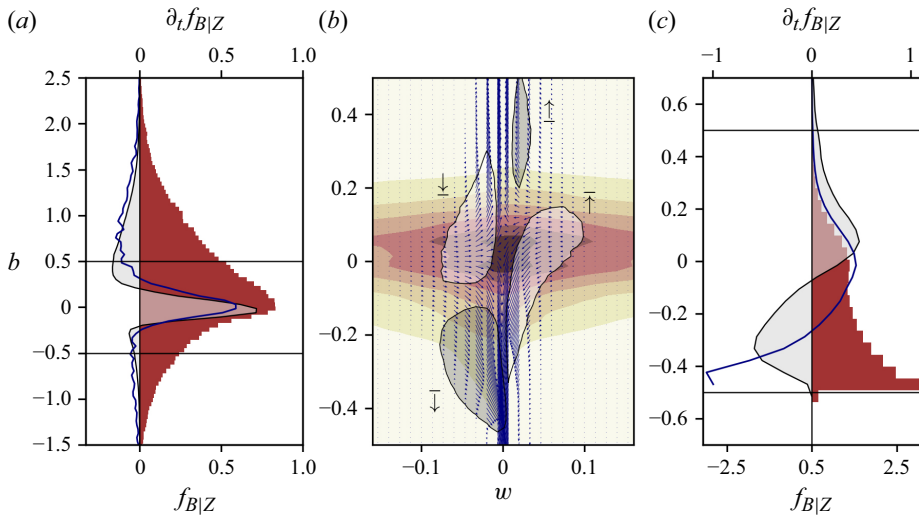


Figure 12. (a), (b) and (c) Information relating to the joint density $f_{Y|Z}$ conditioned on zone z_1 (bottom boundary layer), z_2 (bulk central zone) and z_3 (top boundary layer), respectively. Panel (b) displays the joint density $f_{Y|Z}$ (red contours) along with the flux $\mathbf{D}_1^{zone} f_{Y|Z}$ (blue arrows) and isoregions of $-V f_{Y|Z}$ corresponding to $(-\infty, -0.1]$ (light grey and labelled $\downarrow, \uparrow$) and $[0.1, \infty)$ (dark grey and labelled $\uparrow, \downarrow$). Panels (a) and (c) display the marginal distribution for buoyancy $f_{B|Z}$ (red bars) along with the source/sink term $-V f_{B|Z}$ (light grey) and zonal flux term $-\partial_b(\overline{D}_1^{zone} f_{B|Z})$ (blue line) in (7.13). The horizontal lines in panels (a) and (c) demarcate the range of the vertical axis in panel (b). <https://www.cambridge.org/S0022112025106319/JFM-Notebooks/files/fig11-12/fig11-12.ipynb>.

to vertical velocity (w) and diverging with respect to buoyancy (b). This structure is due to the fact that the boundary layers are sinks of velocity/momentum and sources of buoyancy.

The light and dark shaded regions in figure 12(b) correspond to negative (velocity directed out of the control volume) and positive (velocity directed into the control volume) contributions to $\partial_t f_{Y|Z}$ from $-V f_{Y|Z}$, respectively. Positively buoyant fluid tends to be transported by positive vertical velocity from the bottom boundary layer ($\uparrow$), while negatively buoyant fluid tends to be transported by negative vertical velocity from the top boundary layer ($\downarrow$). Both of these processes correspond to positive forcing of $f_{Y|Z}$ (i.e. $-V f_{Y|Z} > 0$). The remaining negative regions identified in figure 12(b) typically correspond to mixed fluid, with buoyancy close to zero, transported from the bulk into the bottom or top boundary layer (denoted $\downarrow$ and $\uparrow$, respectively).

7.5. Available potential energy

Available potential energy is the part of potential energy of a body of fluid that is theoretically ‘available’ for conversion into kinetic energy. The other part, known as background potential energy, cannot be converted into kinetic energy and is associated with a stable equilibrium (Lorenz 1955). For example, kinetic energy cannot typically be extracted from a stably stratified environment (in which a fluid’s density decreases with height) and therefore possesses zero available potential energy.

When divided by the volume of the domain, the potential energy of the fluid modelled in § 7.3 corresponds to the expectation of the product $-B_t Z$ (see, for example Winters *et al.* 1995):

$$-\mathbb{E}[B_t Z] = - \int_{\mathbb{R}^n} b z f_Y(\mathbf{y}, t) d\mathbf{y}, \quad (7.14)$$

where $f_Y(\mathbf{y}, t)$ is the joint density for the variables $\mathbf{Y}_t := (W_t, B_t, Z)^\top$ from the example in § 7.3 evaluated at $\mathbf{y} := (w, b, z)^\top$ and the minus sign accounts for relatively warm parcels of fluid having greater potential energy when they are moved downwards. The so-called ‘reference’ buoyancy profile $b^*: \mathbb{R} \rightarrow \mathbb{R}$ in available potential energy theory is a volume-preserving function of height that minimises potential energy (see e.g. Winters *et al.* 1995). ‘Volume-preserving’ in this context means that the mapping b^* does not change the marginal distribution of buoyancy, which, in the present context, means that $\mathbb{E}[g \circ b^*(Z)] = \mathbb{E}[g(B_t)]$ for any observable g .

The potential energy (7.14) is minimised by placing positively buoyant parcels at the top of the domain and negatively buoyant parcels at the bottom of the domain. The required mapping b^* is monotonic and corresponds to

$$b^* := F_B^{-1} \circ F_Z, \quad (7.15)$$

where F_B and F_Z are the marginal cumulative distribution functions for B_t and Z , respectively. In the field of optimal transport, b^* minimises the expected squared distance between two variables and has been known for a long time (Knott & Smith 1984; McCann 1995). The reference state allows the available potential energy to be computed as the difference between the actual potential energy and the (minimum) potential energy associated with the reference state:

$$\mathbb{E}[b^*(Z)Z] - \mathbb{E}[B_t Z] = \int_{\mathbb{R}^n} (b^*(z) - b) z f_Y(\mathbf{y}, t) d\mathbf{y} \geq 0. \quad (7.16)$$

As described by Craske (2021), available potential energy can be decomposed into contributions from subvolumes of a domain. Each contribution consists of ‘inner’ and ‘outer’ parts, corresponding to the potential energy that can be released within a subvolume and by combining the subvolume with its parent domain, respectively. The approach generalises to subvolumes that are specified by the conditional sample distributions $f_{X|Z}$ described in § 6. The resulting constructions formally resemble Bayes’ theorem due to their connection with probability theory.

A key related quantity in stratified turbulence is the horizontally averaged vertical buoyancy flux $\mathbb{E}[W_t B_t | Z = z]$, which is responsible for the reversible conversion of available potential energy to and from kinetic energy (see, for example, Caulfield 2020). In the averaged Boussinesq equations, the horizontally averaged vertical buoyancy flux requires closure, but from the perspective of the joint density for W_t and B_t , conditional on Z , it is known exactly in terms of independent and dependent variables:

$$\mathbb{E}[W_t B_t | Z = z] = \iint_{\mathbb{R}^2} w b f_{BW|Z}(b, w | z, t) dw db, \quad \forall z : f_Z(z) \neq 0, \quad (7.17)$$

which suggests that the evolution equations for f_Y might be a useful perspective to adopt for modelling stratified turbulence and making use of available energetics in a prognostic capacity.

8. Closure

As described in § 1, the primary aim of this work was to develop a framework for analysing heterogeneous and stochastic flow observables over arbitrary control volumes using joint probability distributions. Although the forward equation (4.2) can be used diagnostically to analyse data, V , $\mathbf{D}_1$ and $\mathbf{D}_2$ involve unclosed terms that would need to be modelled for (3.19) to be used prognostically. Modelling in this regard must account for the effects of averaging over states that are not measurable in the variable $\mathbf{Y}_t$. As with any coarse-grained representation, one's aim is to forecast a marginal distribution f_Y that is a good approximation to the corresponding projection of the full (infinite dimensional) state of the system.

In the context of local p.d.f. methods (describing a single Eulerian or Lagrangian point), a variety of models to address the closure problem are reviewed by Pope (2000) and Fox (2003). Unlike their moment-based counterparts, for which closures are typically sought for fluxes or correlations such as $\mathbb{E}[WB|\mathbf{X}=\mathbf{x}]$ from § 7.3, the modelling of irreversible mixing by molecular diffusion (i.e. $\mathbf{D}_2$) becomes the central issue in local probability density methods (Pope 2000).

If $-\mathbf{D}_2 \geq 0$ (as discussed in § 5), 'diffusion' in (3.20) is negative and results in greater certainty as time increases regarding the observable $\mathbf{Y}_t$. A simple approach to modelling this term is to replace it with a mean-restoring drift, for which the contribution to $\mathbf{D}_1$ would be proportional to $\mathbf{y} - \mathbb{E}[\mathbf{Y}]$, known as 'interaction by exchange with the mean' (Villiermaux & Devillon 1972; Dopazo & O'Brien 1974). The closure is exact for a statistically homogeneous Gaussian field (Pope 2000). Unfortunately, it has the undesirable effect in one dimension of preserving the shape of other distributions too, which contradicts the expected relaxation towards a Gaussian distribution. This particular issue is overcome by 'mapping closures' (Chen, Chen & Kraichnan 1989; Pope 1991; Gao & O'Brien 1991), which evolve a map from statistically homogeneous Gaussian fields to surrogate fields that have the same p.d.f. as $\mathbf{Y}_t$. Their assumption is that conditional expectations of the derivatives of the surrogate fields are the same as those of $\mathbf{Y}_t$. Chen *et al.* (1989) suggest that the mapping closure would benefit from the inclusion of spatial derivatives in $\mathbf{Y}_t$, which would come at the expense of complicating the structure of the various mapping dependencies between the fields (Pope 1991). Both closures discussed in this paragraph result in a nonlinear forward equation for probability density (see, e.g. Frank 2005), in view of their assumed dependence of $\mathbf{D}_1$ and $\mathbf{D}_2$ on f_Y .

Regarding the random variables $\mathbf{Y}_t$ as particles in phase space, evolving according to a stochastic equation to produce the distribution predicted by (4.2), provides an efficient means of dealing with a large number of observables and affords a different perspective for closure. In this vein, the closure described by Curl (1963) randomly selects and merges pairs of particles (at a rate specified as the mixing time scale). In the context of models for stratified turbulence, Kerstein (1999) and, recently, Petropoulos, de Bruyn Kops & Caulfield (2025) also adopt a particle perspective, the latter accounting for competing effects of turbulent fluctuations and buoyancy with a resetting process.

Heterogeneous flows are more challenging for closures and involve joint p.d.f.s over a larger number of dimensions. In this regard, the flexibility of the framework described in the present work should prove useful. It can accommodate arbitrary coordinates as observables and allows one to decompose a domain into overlapping control volumes that are themselves defined as probability densities. As illustrated in § 6, a coarse grained approach to a heterogeneous flow might decompose a domain into constituent parts that are approximately homogeneous and, therefore, more amenable to the closures discussed. An outstanding challenge would then be modelling of the boundary fluxes between the control volumes in $\mathbf{D}_1$ and/or V (see § 4 and the example in § 7.4), for which the concepts relating to available potential energy discussed in § 7.5 might be instructive.

Like the formulation presented here, the p.d.f.s of Villiermaux (2019) describe the heterogeneous distribution of concentration over a control volume. Therefore, the simplified p.d.f. models characterising the evolution of lamellae discussed by Villiermaux (2019) effectively provide a closure to the forward equation. Conversely, such models could be examined in the light of the forward equation (4.2) to establish a better understanding of the diffusivity D_2 .

9. Conclusions

We have derived and explained an equation that governs the evolution of the joint probability distribution f_Y of a set of stochastically forced multicomponent observables Y drawn randomly from points in space that are distributed according to f_X . Our use of the distribution f_X to describe space generalises the classical definition of a control volume by relaxing the notion of set membership in accounting for the possibility of gradual transitions between regions (cf. § 6 and figure 7). The generality of the approach clarifies the role of boundary fluxes and accommodates a wide range of applications, including those for which properties of a flow are used to define control volumes.

Fluxes of multicomponent scalars over the bounding region of the sample distribution f_X contribute to drift in probability space. The factor $\nabla \log f_X$ that multiplies such fluxes defines the normal direction of the bounding region, as well as its relative size, in transforming the measure with respect to which conditional expectations are taken. In contrast, irreversible mixing leads to diffusion in probability space. Although the sign of the resulting diffusion coefficient is typically negative semidefinite, representing the homogenising effect of irreversible mixing, circumstances for which the coefficient could be sign indefinite were explored in § 5.

The approach can be applied to control volumes defined by any sample distribution, shape, size and dimension, and readily accommodates an arbitrary number of flow observables. As demonstrated in § 7.3, coordinates can be regarded as observables, which means that the equations derived in § 3 can be applied without modification to slices of a domain parametrised by a coordinate, or, for example, geopotential height. In this regard, the work illustrates that the equations governing global constructions of available and background potential energy (Winters *et al.* 1995) are specific examples of the more general principles that determine the evolution of probability distributions.

In addition to unifying a wide range of possible applications, an advantage in formulating the problem in terms of the density f_X , rather than traditional sets, is that $\nabla \log f_X$ implicitly accounts for the size and orientation of the bounding region without requiring further factors or notation. The use of f_X also proves useful in producing relatively smooth boundary information from limited data from simulations, from which evaluation of fluxes across an arbitrary surface is challenging and subject to large uncertainty.

As explained in § 6, the sample distribution f_X can be treated as a mixture distribution that decomposes into a sum of conditional distributions $f_{X|Z}$. Using the conditional distributions to describe (sub)regions of a flow that are approximately homogeneous might provide a viable means of generating bulk stochastic models for heterogeneous flows using the closures discussed in § 8. A particularly attractive possibility in this regard would be the use of time-dependent distributions defined by a flow observable (an example would be the use of enstrophy to discriminate between turbulent/non-turbulent regions of van Reeuwijk *et al.* (2021)). For such cases, the effect that the time-dependence of f_X would have on the evolution of f_Y can be readily interpreted using the Feynman–Kac formula, as shown in Appendix C.1.

The present work is aimed at engineers and physicists who wish to diagnose or model bulk transport and mixing processes from a probabilistic perspective. The resulting equations are nevertheless of mathematical interest and perhaps worthy of study in their own right. If the terms V , D_1 and D_2 requiring closure are, in general, regarded as a function of the joint density f_Y , the resulting partial differential equation is a nonlinear Fokker–Planck equation, reminiscent of mean field theory (Frank 2005; Barbu & Röckner 2024) and the associated McKean–Vlasov stochastic differential equation (McKean 1966).

Supplementary material. Computational Notebook files are available as supplementary material at <https://doi.org/10.1017/jfm.2025.10631> and online at <https://www.cambridge.org/S0022112025106319/JFM-Notebooks>

Funding. This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/V033883/1] as part of the [D*]stratify project.

Declaration of interests. The authors report no conflict of interest.

Data availability statement. The data that support the findings of this study are openly available in the JFM notebooks at <https://www.cambridge.org/S0022112025106319/JFM-Notebooks>.

Appendix A. The transformation of probability density

A function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^n$ that assigns to coordinates $X \in \mathbb{R}^d$ the value of the dependent variables $Y \in \mathbb{R}^n$, transforms a distribution over $\mathbb{R}^d$ into a distribution over $\mathbb{R}^n$. We will start by assuming that $n = d$ and discuss what happens when $n \neq d$ at the end of this section.

The joint probability density $f_Y : \mathbb{R}^n \rightarrow \mathbb{R}$ (if it exists) must provide a consistent means of computing the expectation of an observable $g : \mathbb{R}^n \rightarrow \mathbb{R}$:

$$\mathbb{E}[g] = \int_{\mathbb{R}^n} g(y) f_Y(y, t) dy = \int_{\mathbb{R}^d} g \circ \varphi(x) f_X(x) dx. \quad (\text{A1})$$

Notice that if $\varphi(x)$ is constant over subsets of $\mathbb{R}^d$ for which $f_X(x) \neq 0$, then such a density f_Y does not exist. In that case, $f_Y dy$ would need to be replaced with the pushforward measure $\nu(dy)$ (Chung 2001; Bogachev 2007) to account for individual values y that occur over subsets of $\mathbb{R}^d$ of finite size (referred to as ‘atoms’ of ν (Chung 2001)). If, in the cases where the density f_Y does exist, we pick g as the indicator function for the sub-codomain $\mathcal{Y}(y) : \{y' \in \mathbb{R}^n : y^i \leq y^i\}$, then

$$f_Y(y) = \frac{\partial^n}{\partial y^1 \dots \partial y^n} \int_{\varphi^{-1}(\mathcal{Y}(y))} f_X(x) dx, \quad (\text{A2})$$

where $\varphi^{-1}(\mathcal{Y}(y)) := \{X \in \mathbb{R}^d : \varphi(X) \in \mathcal{Y}(y)\}$ is the preimage of $\mathcal{Y}$, which accounts for the possibility that several distinct values X might map onto the same value Y . If, however, $\varphi|_{\mathcal{X}}$ is the restriction of φ to a subdomain $\mathcal{X} \subset \mathbb{R}^d$ over which φ is invertible, then

$$f_{Y|\mathcal{X}}(y) = f_X(\varphi|_{\mathcal{X}}^{-1}(y)) \left| \frac{\partial \varphi|_{\mathcal{X}}^{-1}}{\partial y} \right|, \quad (\text{A3})$$

where $f_{Y|\mathcal{X}}$ is the probability density conditional on $X \in \mathcal{X}$. The density f_Y can be calculated from (A3) using Bayes’ theorem to account for a set $\mathcal{S} \ni \mathcal{X}$ of disjoint subdomains:

$$f_Y(y) = \sum_{\mathcal{X} \in \mathcal{S}} f_{Y|\mathcal{X}}(y, t) \mathbb{P}(\mathcal{X}). \quad (\text{A4})$$

When $n < d$, the sample space is large in comparison with n and $(d - n)$ auxiliary variables, such as $X_{n+1}, X_{n+2}, \dots, X_d$ need to be appended to $\mathbf{Y}$ to construct the Jacobian in (A3). The joint distribution (A4) can then be integrated with respect to the auxiliary variables to produce the (marginal) distribution that was originally sought. If $d < n$, however, the sample space is relatively small. Auxiliary dimensions $X_{d+1}, X_{d+2}, \dots, X_n$, with regards to which $\mathbf{Y}$ is assumed constant, can be appended to $\mathbf{X}$, which will produce Dirac measures in the resulting distribution, as discussed in § 7. In general, $f_{\mathbf{Y}}$ in this case will be ‘sparse’ in the sense that the (possibly fractal) dimension associated with the singular/non-zero parts of the distribution will be less than n . This idea can be made precise by considering the information dimension of $\mathbf{Y}$ (Rényi 1959).

Appendix B. Lorenz (1963) model

Lorenz’s model for convection (Lorenz 1963) is a truncated solution of the Boussinesq equations for which the vertical velocity Y_t^1 and buoyancy field Y_t^2 , relative to linear conduction, on a horizontally periodic domain $\mathcal{X} := [0, 2\pi/k) \times [0, 1] \ni \mathbf{X} := (X^1, X^2)$, are

$$Y_t^1 = \varphi^1(X^1, X^2) := \frac{\sqrt{2}}{\pi} (k^2 + \pi^2) a_1(t) \cos(kX^1) \sin(\pi X^2), \quad (\text{B1})$$

$$Y_t^2 = \varphi^2(X^1, X^2) := \frac{\sqrt{2}}{\pi r} a_2(t) \cos(kX^1) \sin(\pi X^2) - \frac{1}{\pi r} a_3(t) \sin(2\pi X^2), \quad (\text{B2})$$

where r is a renormalised Rayleigh number. The amplitudes $\mathbf{a} := (a_1, a_2, a_3)^\top$ evolve in time according to

$$\frac{da_1}{dt} = s(a_2 - a_1), \quad \frac{da_2}{dt} = ra_1 - a_2 - a_1a_3, \quad \frac{da_3}{dt} = a_1a_2 - ba_3, \quad (\text{B3})$$

where $b := 4\pi^2(k^2 + \pi^2)^{-1}$ characterises the aspect ratio of the domain and s is the Prandtl number. Following Appendix A, to construct the density $f_{\mathbf{Y}}(-, t) : \mathbb{R}^n \rightarrow \mathbb{R}$, it is sufficient to consider half the horizontal domain (i.e. $X^1 \in [0, \pi/k)$) and the subdomains $\mathcal{X}_1 := [0, \pi/k) \times ((0, 1/4) \cup (3/4, 1))$ and $\mathcal{X}_2 := [0, \pi/k) \times (1/4, 3/4)$. Over $\mathcal{X}_1$ and $\mathcal{X}_2$, the Jacobian $\partial\varphi/\partial\mathbf{X}$ is non-singular, which means that φ has a single-valued inverse when its domain is restricted to $\mathcal{X}_1$ and $\mathcal{X}_2$, and can be found by manipulation of (B1)–(B2).

Appendix C. Feynman–Kac formula

As discussed in § 3, solutions of the backward Kolmogorov equation give the expected value of $g(\mathbf{Y}_t)$ given that $\mathbf{Y}_s = \mathbf{y}$, where $s \leq t$. A more general relationship comes from considering the expectation of functionals that include integrals with respect to time, which is the subject of the Feynman–Kac formula (Kac 1949).

In particular, the conditional expectation

$$u(\mathbf{y}, s) = \mathbb{E} \left[\exp \left(- \int_s^t V(\mathbf{Y}_\tau, \tau) d\tau \right) g(\mathbf{Y}_t) | \mathbf{Y}_s = \mathbf{y} \right] \quad (\text{C1})$$

satisfies (see, for example, Shreve 2004, Theorem 6.4.3)

$$- \partial_s u(\mathbf{y}, s) = (\mathcal{L} - V)u(\mathbf{y}, s), \quad (\text{C2})$$

which, in comparison with (3.16), contains the additional decay rate $V(\mathbf{y}, s)$. Following the same steps as those leading to (3.19), the forward equation corresponding to (C2) is $\partial_t f_{\mathbf{Y}} = (\mathcal{L}^\dagger - V)f_{\mathbf{Y}}$, where it is seen that $V(\mathbf{y}, t)$ affects neither the drift nor diffusion

associated with Y_t , but rather the relative weight or importance ascribed to different trajectories in the calculation of $\mathbb{E}[g(Y_t)|Y_s = y]$. As noted in § 4, if $\mathbb{E}[V(Y_t, t)] = 0$, then f_Y remains a legitimate probability density in spite of the forcing term $-Vf_Y$.

C.1. Time dependent sample distributions

In § 3, the spatial sampling distribution f_X was treated as being independent of time. However, in some circumstances, f_X might depend on time, which would be the case if it were defined in terms of the state observables Y , for example, such that $f_X : \mathbb{R}^d \times \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$. To account for a possible time-dependence of f_X , let

$$V(Y_\tau, \tau) := -\mathbb{E} \left[\frac{d}{d\tau} \log f_X(X, Y_\tau, \tau) | Y_\tau \right] + \dots \quad (\text{C3})$$

where ‘ $\dots$ ’ is the contribution to V in (4.2) from advective transport across the boundary. To understand the meaning of the first term on the right-hand side, we evaluate (C1):

$$u(y, s) = \mathbb{E} \left[\frac{f_X(X, Y_t, t)}{f_X(X, Y_s, s)} g(Y_t) | Y_s = y \right] = \tilde{\mathbb{E}} [g(Y_t) | Y_s = y], \quad (\text{C4})$$

where $\tilde{\mathbb{E}}$ denotes the expectation with respect to the sample distribution at time t . Equation (C4) therefore states the relationship between the expectation $\mathbb{E}$ under $f_Y(y, s)$ and the expectation $\tilde{\mathbb{E}}$ under $f_Y(y, t)$, which corresponds to the rule for transforming conditional expectations under a change of measure (see, for example, Shreve 2004, Lemma 5.2.2). Note, in particular, that if f_X is a probability density, then (C3) implies that $\mathbb{E}[V(Y_t, t)] = 0$ and, therefore, that the $f_Y(y, t)$ obtained from $\partial_t f_Y = (\mathcal{L}^\dagger - V)f_Y$ and used to define the expectation $\tilde{\mathbb{E}}$ in (C4) would also be a probability density.

If the sample distribution is advected with the flow, then its material derivative is zero, such that $\partial_t f_X = -U \cdot \nabla f_X$ and

$$V(Y_\tau, \tau) := \mathbb{E} [U \cdot \nabla \log f_X | Y_\tau] + \dots, \quad (\text{C5})$$

which exactly balances the contribution to V labelled ‘ $\dots$ ’ from (4.2) due to advection over the sample distribution’s boundary, resulting in $V \equiv 0$. More generally, if a sample distribution is defined to evolve according to $\partial_t f_X = -U_X \cdot \nabla f_X$, where U_X is not necessarily the same as the flow velocity, the resulting expression for V would be

$$V(Y_\tau, \tau) := \mathbb{E} [(U_X - U) \cdot \nabla \log f_X | Y_\tau]. \quad (\text{C6})$$

Equation (C6) highlights the generality of the framework in accommodating Eulerian control volumes ($U_X = 0$) and Lagrangian control volumes ($U_X = U$) as particular cases.

REFERENCES

- ANDRIAN, V. & CRASKE, J. 2023 Stochastic models of ventilation driven by opposing wind and buoyancy. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* **479**(2276), 20230170.
- BARBU, V. & RÖCKNER, M. 2024 *Nonlinear Fokker-Planck Flows and Their Probabilistic Counterparts*. 1st edn. Springer Nature Switzerland.
- BHAGAT, R.K., DALZIEL, S.B., DAVIES WYKES, M.S. & LINDEN, P.F. 2024 Building ventilation: the consequences for personal exposure. *Annu. Rev. Fluid Mech.* **56** (1), 405–434.
- BLEKHERMAN, G., PARRILO, P. & THOMAS, R. 2012 *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics.
- BOGACHEV, V.I. 2007 Measure theory. In *Measure Theory*, vol. 1. Springer.
- BRUNTON, S.L., BUDIŠIĆ, M., KAISER, E. & KUTZ, J.N. 2022 Modern Koopman theory for dynamical systems. *SIAM Rev.* **64** (2), 229–340.
- BURNS, K.J., VASIL, G.M., OISHI, J.S., LECOANET, D. & BROWN, B.P. 2020 Dedalus: a flexible framework for numerical simulations with spectral methods. *Phys. Rev. Res.* **2**, 023068.

- CAULFIELD, C.P. 2020 Open questions in turbulent stratified mixing: do we even know what we do not know? *Phys. Rev. Fluids* **5**, 110518.
- CHEN, H., CHEN, S. & KRAICHNAN, R.H. 1989 Probability distribution of a stochastically advected scalar field. *Phys. Rev. Lett.* **63**, 2657–2660.
- CHUNG, K.L. 2001 *A Course in Probability Theory*. 3rd edn. Academic Press.
- CRASKE, J. 2021 Decomposition of available potential energy for networks of connected volumes. *J. Fluid Mech.* **920**, A17.
- ČRNJARIĆ-ŽIĆ, N., MAČEŠIĆ, S. & MEZIĆ, I. 2020 Koopman operator spectrum for random dynamical systems. *J. Nonlinear Sci.* **30**, 2007–2056.
- CURL, R.L. 1963 Dispersed phase mixing: i. theory and effects in simple reactors. *AIChE J.* **9** (2), 175–181.
- DOMBRE, T., FRISCH, U., GREENE, J.M., HÉNON, M., MEHR, A. & SOWARD, A.M. 1986 Chaotic streamlines in the ABC flows. *J. Fluid Mech.* **167**, 353–391.
- DOPAZO, C. & O'BRIEN, E.E. 1974 An approach to the autoignition of a turbulent mixture. *Acta Astronaut.* **1** (9), 1239–1266.
- ELLISON, T.H. & TURNER, J.S. 1959 Turbulent entrainment in stratified flows. *J. Fluid Mech.* **6** (03), 423–448.
- FISCHER, H.B., LIST, E.J., KOH, R.C.Y., IMBERGER, J. & BROOKS, N. 1979 *Mixing in Inland and Coastal Waters*. Academic Press.
- FOX, R.O. 2003 *Computational Models for Turbulent Reacting Flows*. Cambridge University Press.
- FRANK, T.D. 2005 *Nonlinear Fokker-Planck Equations: Fundamentals and Applications*. Springer.
- GAO, F. & O'BRIEN, E.E. 1991 A mapping closure for multispecies Fickian diffusion. *Phys. Fluids A: Fluid Dyn.* **3** (5), 956–959.
- GARDINER, C. 1990 *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences*. Springer.
- GASPARD, P. 1998 *Chaos, Scattering and Statistical Mechanics*. Cambridge University Press.
- HOPF, E. 1952 Statistical hydrodynamics and functional calculus. *J. Ration. Mech. Anal.* **1**, 87–123.
- HUNT, J.C.R. 1985 Turbulent diffusion from sources in complex flows. *Annu. Rev. Fluid Mech.* **17** (1), 447–485.
- KAC, M. 1949 On distributions of certain Wiener functionals. *Trans. Am. Math. Soc.* **65** (1), 1–13.
- KERSTEIN, A.R. 1999 One-dimensional turbulence: model formulation and application to homogeneous turbulence, shear flows, and buoyant stratified flows. *J. Fluid Mech.* **392**, 277–334.
- KLUS, S., KOLTAI, P. & SCHÜTTE, C. 2016 On the numerical approximation of the Perron-Frobenius and Koopman operator. *J. Comput. Dyn.* **3** (1), 51–79.
- KNOTT, M. & SMITH, C.S. 1984 On the optimal mapping of distributions. *J. Optimiz. Theory Appl.* **43**, 39–49.
- KOLLMANN, W. 1990 The pdf approach to turbulent flow. *Theor. Comput. Fluid Dyn.* **1** (5), 249–285.
- KOOPMAN, B.O. 1931 Hamiltonian systems and transformation in Hilbert space. *Proc. Natl Acad. Sci.* **17** (5), 315–318.
- LASOTA, A. & MACKEY, M.C. 1998 *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. Springer New York.
- LEWIS, R.M. & KRAICHNAN, R.H. 1962 A space-time functional formalism for turbulence. *Commun. Pure Appl. Maths* **15** (4), 397–411.
- LI, J., QIAN, Z. & ZHOU, M. 2022 On the transport equation for probability density functions of turbulent vorticity fields. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* **478**(2257), 20210534.
- LI, T.-Y. 1976 Finite approximation for the Frobenius–Perron operator. a solution to Ulam’s conjecture. *J. Approx. Theory* **17** (2), 177–186.
- LINDSAY, B.G. 1995 Mixture models: Theory, geometry and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics* **5**, pp. i–163.
- LORENZ, E.N. 1955 Available potential energy and the maintenance of the general circulation. *Tellus* **7** (2), 157–167.
- LORENZ, E.N. 1963 Deterministic nonperiodic flow. *J. Atmos. Sci.* **20** (2), 130–141.
- LUNDGREN, T.S. 1967 Distribution functions in the statistical theory of turbulence. *Phys. Fluids* **10** (5), 969–975.
- LÜLFF, J., WILCZEK, M. & FRIEDRICH, R. 2011 Temperature statistics in turbulent Rayleigh–Bénard convection. *New J. Phys.* **13** (1), 015002.
- LÜLFF, J., WILCZEK, M., STEVENS, R.J.A.M., FRIEDRICH, R. & LOHSE, D. 2015 Turbulent Rayleigh–Bénard convection described by projected dynamics in phase space. *J. Fluid Mech.* **781**, 276–297.
- MARGULES, M. 1903 Über die Energie der Stürme. *Jahrb. Zent.-Anst. für Meteorol. Und Erdmagnet* **48**, 1–26.
- MCCANN, R.J. 1995 Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** (2), 309–323.

- McCOMB, W.D. 1990 *The Physics of Fluid Turbulence*. Clarendon Press.
- McKEAN, H.P. 1966 A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl Acad. Sci. USA* **56** (6), 1907–1911.
- MEZIĆ, I. 2005 Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.* **41**, 309–325.
- MEZIĆ, I. 2013 Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech.* **45** (1), 357–378.
- MINIER, J.-P. & POZORSKI, J. 1997 Derivation of a PDF model for turbulent flows based on principles from statistical physics. *Phys. Fluids* **9** (6), 1748–1753.
- MONIN, A.S. & YAGLOM, A.M. 2013 *Statistical Fluid Mechanics, Volume II: Mechanics of Turbulence*. Dover Publications
- PAVLIOTIS, G.A. 2014 *Stochastic Processes and Applications Diffusion Processes, the Fokker-Planck and Langevin Equations*. 1st edn. Springer New York.
- PETROPOULOS, N., DE BRUYN KOPS, S.M. & CAULFIELD, C.P. 2025 Modelling dispersion in stratified turbulent flows as a resetting process. *J. Fluid Mech.* **1002**, A48.
- POPE, S.B. 1985 PDF methods for turbulent reactive flows. *Prog. Energ. Combust.* **11** (2), 119–192.
- POPE, S.B. 1981 Transport equation for the joint probability density function of velocity and scalars in turbulent flow. *Phys. Fluids* **24** (4), 588–596.
- POPE, S.B. 1991 Mapping closures for turbulent mixing and reaction. *Theor. Comput. Fluid Dyn.* **2** (5–6), 255–270.
- POPE, S.B. 1994 Lagrangian PDF methods for turbulent flows. *Annu. Rev. Fluid Mech.* **26** (1), 23–63.
- POPE, S.B. 2000 *Turbulent Flows*. Cambridge University Press.
- VAN REEUWIJK, M., VASSILICOS, J. C. & CRASKE, J. 2021 Unified description of turbulent entrainment. *J. Fluid Mech.* **908**, A12.
- RÉNYI, A. 1959 On the dimension and entropy of probability distributions. *Acta Mathematica Academiae Scientiarum Hungarica* **10**, 193–215.
- RUELLE, D. 2004 *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*. Cambridge University Press.
- SHREVE, S.E. 2004 *Stochastic Calculus for Finance. II, Continuous-Time Models*. Springer.
- TAILLEUX, R. 2013 Available potential energy and exergy in stratified fluids. *Annu. Rev. Fluid Mech.* **45** (1), 35–58.
- TSENG, Y.-H. & FERZIGER, J.H. 2001 Mixing and available potential energy in stratified flows. *Phys. Fluids* **13** (5), 1281–1293.
- VILLERMAUX, E. 2019 Mixing versus stirring. *Annu. Rev. Fluid Mech.* **51** (2019), 245–273.
- VILLERMAUX, J. & DEVILLON, J.C. 1972 Représentation de la coalescence et de la redispersion des domaines de ségrégation dans un fluide par un modèle d’interaction phénoménologique. In *Proceedings of the 2nd International symposium on chemical reaction engineering*, vol. 26, pp. 1–13. Elsevier New York.
- VOURIOT, C.V.M., HIGTON, T.D., LINDEN, P.F., HUGHES, G.O., VAN REEUWIJK, M. & BURRIDGE, H.C. 2023 Uniformly distributed floor sources of buoyancy can give rise to significant spatial inhomogeneities within rooms. *Flow* **3**, E18.
- WINTERS, K.B., LOMBARD, P.N., RILEY, J.J. & D’ASARO, E.A. 1995 Available potential energy and mixing in density-stratified fluids. *J. Fluid Mech.* **289**, 115–128.
- WUNSCH, S. & KERSTEIN, A.R. 2005 A stochastic model for high-Rayleigh-number convection. *J. Fluid Mech.* **528**, 173–205.
- ZELDOVICH, Y.B. 1937 The asymptotic law of heat transfer at small velocities in the finite domain problem. *Zh. Eksp. Teoret. Fiz* **7** (12), 1466–1468.