PDF equations for random dynamical systems

Consider the following n-dimensional dynamical system

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}) \\ \boldsymbol{x}(0;\omega) = \boldsymbol{x}_0(\omega) \end{cases} \tag{1}$$

where $\boldsymbol{x}_0(\omega)$ is a random initial state with joint PDF $p_0(\boldsymbol{x})$. We are interested in studying the statistical properties of the solution to (1) using probability density function (PDF) methods. As we shall see hereafter, systems of the form (1) include systems in which we have random variables at appearing the right hand side of the ODE, i.e., systems with random parameters.

Systems with random parameters. It is straightforward to show that a non-autonomous system of the form

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{G}(\boldsymbol{x}, \boldsymbol{\xi}(\omega), t) \\ \boldsymbol{x}(0;\omega) = \boldsymbol{x}_0(\omega) \end{cases} \tag{2}$$

can be transformed into an autonomous system evolving form a random initial state. To this end, we define the phase variables of $z(t) = t$ and $\boldsymbol{y}(t) = \boldsymbol{\xi}(\omega)$ rewrite (2) as

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{G}(\boldsymbol{x}, \boldsymbol{y}, z) \\ \dfrac{d\boldsymbol{y}}{dt} = \boldsymbol{0} \\ \dfrac{dz}{dt} = 1 \\ \boldsymbol{x}(0;\omega) = \boldsymbol{x}_0(\omega), \qquad \boldsymbol{y}(0;\omega) = \boldsymbol{\xi}(\omega), \qquad z(0,\omega) = 0. \end{cases} \tag{3}$$

Remarkably, system of the form (2) include also dynamical systems driven by finite-dimensional random processes, i.e., random processes that can be represented in terms of series expansions involving a finite number of random variables.

Example: An simple example of a system of the form (2) is

$$\frac{dx}{dt} = f(x) + \eta(t;\omega) \tag{4}$$

i.e., a scalar ODE driven by *colored random noise* $\eta(t;\omega)$ [12, 21, 17]. Let us represent $\eta(t;\omega)$ as a truncated Karhunen-Loève series expansion (see [21] for an application to cancer modeling)

$$\eta(t;\omega) \simeq \sum_{k=1}^{M} \sqrt{\lambda}_k \xi_k(\omega) \psi_k(t) \tag{5}$$

involving a finite number of uncorrelated random variables $\{\xi_1, \ldots, \xi_M\}$. We shall call M the *dimensionality*

of the noise process[1] The adjective "colored" refers to the fact that the Fourier power spectral density of the random noise $\eta(t;\omega)$ is in general not flat as in the case of white noise[2]. The power spectral density is the inverse Fourier transform of the temporal auto-correlation function of the noise, i.e.,

$$\mathbb{E}\left\{f(t;\omega)f(t';\omega)\right\} = \sum_{k=1}^{M} \lambda_k \psi_k(t)\psi_k(t'). \tag{7}$$

Remark: A random dynamical systems is a systems driven by a finite number of random variables. An example is the system (4)-(5), in which the the random input process is finite-dimensional (M finite). On the other hand, a "stochastic dynamical system" is usually driven by infinite-dimensional random processes, i.e., processes that can be represented in terms of an infinite (countable or uncountable) number of random variables. An example is the ODE (4) if we choose $\eta(t;\omega)$ to be, e.g., Gaussian white noise process (derivative of a Wiener process). In this case, it is more appropriate to write the ODE as

$$dx = f(x)dt + d\zeta(t), \tag{8}$$

where $d\zeta(t)$ denotes the increment of a Wiener process.

Liouville equation. Let $\boldsymbol{x}(t;\boldsymbol{x}_0)$ be the flow generated by (1). The PDF of $\boldsymbol{x}(t;\boldsymbol{x}_0)$, i.e., the solution of (1) at time t, satisfies the Liouville equation

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} + \nabla\cdot(\boldsymbol{f}(\boldsymbol{x})p(\boldsymbol{x},t)) = 0, \qquad\qquad p(\boldsymbol{x},0) = p_0(\boldsymbol{x}), \tag{9}$$

where $p_0(x)$ is the PDF of the random initial state $\boldsymbol{x}_0(\omega)$. To derive equation (9), consider the characteristic function representation of $p(\boldsymbol{x},t)$

$$\phi(\boldsymbol{a},t) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i\boldsymbol{a}\cdot\boldsymbol{x}} p(\boldsymbol{x},t)d\boldsymbol{x} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i\boldsymbol{a}\cdot\boldsymbol{x}(t;\boldsymbol{x}_0)} p(\boldsymbol{x}_0)d\boldsymbol{x}_0. \tag{10}$$

Differentiating (10) with respect to t yields

$$\begin{aligned}
\frac{\partial\phi(\boldsymbol{a},t)}{\partial t} &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} i\boldsymbol{a}\cdot\frac{\partial\boldsymbol{x}(t,\boldsymbol{x}_0)}{\partial t} e^{i\boldsymbol{a}\cdot\boldsymbol{x}(t;\boldsymbol{x}_0)} p(\boldsymbol{x}_0)d\boldsymbol{x}_0 \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} i\boldsymbol{a}\cdot\boldsymbol{f}\left(\boldsymbol{x}(t,\boldsymbol{x}_0)\right) e^{i\boldsymbol{a}\cdot\boldsymbol{x}(t;\boldsymbol{x}_0)} p(\boldsymbol{x}_0)d\boldsymbol{x}_0 \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} i\boldsymbol{a}\cdot\boldsymbol{f}\left(\boldsymbol{x}\right) e^{i\boldsymbol{a}\cdot\boldsymbol{x}} p(\boldsymbol{x},t)d\boldsymbol{x} \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{\partial}{\partial\boldsymbol{x}}\left(e^{i\boldsymbol{a}\cdot\boldsymbol{x}}\right)\cdot\boldsymbol{f}\left(\boldsymbol{x}\right) p(\boldsymbol{x},t)d\boldsymbol{x} \\
&= -\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i\boldsymbol{a}\cdot\boldsymbol{x}}\nabla\cdot\left(\boldsymbol{f}\left(\boldsymbol{x}\right) p(\boldsymbol{x},t)\right)d\boldsymbol{x}.
\end{aligned} \tag{11}$$

[1]To sample realizations of the random process (5) we need to sample the random variables $\{\xi_1,\ldots,xi_M\}$. Such random variables are (by construction) orthonormal, i.e., they are uncorrelated and have variance one:

$$\mathbb{E}\{\xi_i(\omega)\xi_j(\omega)\} = \delta_{ij}. \tag{6}$$

If $\{\xi_1,\ldots,xi_M\}$ are jointly Gaussian then we know that (6) is necessary and sufficient for independence. Hence, in this case sampling the joint PDF $\{\xi_1,\ldots,xi_M\}$ reduces to sampling the PDF of an independent set of one-dimensional Gaussian PDFs with zero mean and variance one. More generally, the joint PDF of $\{\xi_1,\ldots,xi_M\}$ can be sampled using Markov Chain Monte Carlo methods, e.g., the Metropolis-Hastings algorithm or the Gibbs sampling algorithm.

[2]A flat power spectral density implies that all frequencies contribute equally to the signal. The adjective "white" follows from an analogy the power spectrum of visible colors, in which the color white has all visible frequencies contributing equally. Stochastic ODEs driven by Gaussian white noise, and corresponding models are discussed extensively in the course AM216. If the power spectral density of a random signal decays with the frequency ν as $1/\nu^{\alpha}$ ($\alpha\in[1,2]$) then the noise is called "pink".

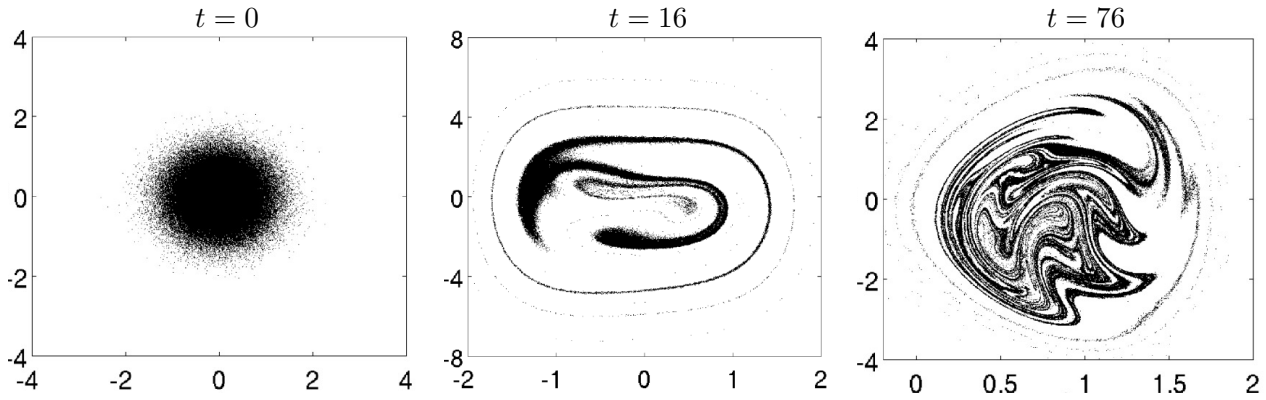

Figure 1: Point clouds corresponding to a jointly Gaussian initial PDF advected by the flow map generated by the Duffing equation. The x-axis corresponds to x while the y-axis represents $\dot{x}$. We plot the joint PDF of $x(t)$ and $\dot{x}(t)$ at different times.

In the last step we used integration by parts and the fact that the PDF $p(\boldsymbol{x},t)$ decays to zero at infinity sufficiently fast. By combining (10) and (11) we obtain

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i\boldsymbol{a}\cdot\boldsymbol{x}}\left[\frac{\partial p(\boldsymbol{x},t)}{\partial t}+\nabla\cdot\left(\boldsymbol{f}\left(\boldsymbol{x}\right)p(\boldsymbol{x},t)\right)\right]d\boldsymbol{x}=0,\quad \text{for all } \boldsymbol{a}\in\mathbb{R}^n, \tag{12}$$

which implies that the function between square bracket must be equal to zero for all $\boldsymbol{x}$ and all t. This proves the Liouville equation (9).

Note that from a mathematical viewpoint, the Liouville equation (9) is a linear hyperbolic conservation law in as many variables as the dimension of the system (1). Therefore, computing its solution can be challenging due to high-dimensionality (PDE in n independent variables), normalization and positivity constraints of the solution (the solution is a PDF), as well as potential multiple scales (The PDF is a hyperbolic conservation law). Related to the last point, in Figure 1 we show what happens to a jointly Gaussian initial state when samples from such PDF are evolved forward in time by the flow map generated by the 2D Duffing oscillator [1]

$$\frac{d^2x}{dt^2}=-x-\frac{1}{50}\frac{dx}{dt}-5x^3+8\cos\left(\frac{t}{2}\right). \tag{13}$$

By using the method of characteristics, it is straightforward to obtain the following formal solution to the Liouville equation (9)

$$p(\boldsymbol{x},t)=p_0\left(\boldsymbol{x}_0(\boldsymbol{x},t)\right)\exp\left(-\int_0^t\nabla\cdot\boldsymbol{f}\left(\boldsymbol{x}(\tau,\boldsymbol{x}_0)\right)d\tau\right), \tag{14}$$

where $\boldsymbol{x}_0(\boldsymbol{x},t)$ denotes the inverse flow map generated by (1). Equation (14) follows from the well-known

characteristic system[3]

$$\begin{cases} \dfrac{d\boldsymbol{x}(t,\boldsymbol{x}_0)}{dt} = \boldsymbol{f}\left(\boldsymbol{x}(t,\boldsymbol{x}_0)\right) \\ \boldsymbol{x}(0,\boldsymbol{x}_0) = \boldsymbol{x}_0 \\ \dfrac{dp(\boldsymbol{x}(t,\boldsymbol{x}_0),t)}{dt} = -p(\boldsymbol{x}(t,\boldsymbol{x}_0),t)\nabla\cdot\boldsymbol{f}\left(\boldsymbol{x}(t,\boldsymbol{x}_0)\right) \\ p(\boldsymbol{x}(0,\boldsymbol{x}_0),0) = p_0(\boldsymbol{x}_0) \end{cases} \tag{16}$$

Example: The Liouville equation corresponding to the system (4)-(5) is

$$\frac{\partial p(x,\boldsymbol{y},t)}{\partial t} + \frac{\partial}{\partial x}\left(f(x)p(x,\boldsymbol{y},t)\right) + \frac{\partial p(x,\boldsymbol{y},t)}{\partial x}\sum_{k=1}^{M}\sqrt{\lambda}_k y_k \psi_k(t) = 0. \tag{17}$$

If $x_0(\omega)$ and $\boldsymbol{\xi}$ are statistically independent then the initial PDF can be factored as $p(x_0,\boldsymbol{y},0) = p_{x_0}(x_0)p_{\boldsymbol{\xi}}(\boldsymbol{y})$. It is important to emphasize that the joint PDF equation involves both the state variable $x(t,\omega)$ and the variables y_k representing the variables ξ_k in the noise process (5).

Example: The Liouville equation corresponding to the three-dimensional dynamical system

$$\dot{x}_1 = x_1 x_3. \qquad \dot{x}_2 = -x_2 x_3, \qquad \dot{x}_3 = -x_1^2 + x_2^2. \tag{18}$$

is

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} = -\frac{\partial}{\partial x_1}\left(x_1 x_3 p(\boldsymbol{x},t)\right) + \frac{\partial}{\partial x_2}\left(x_2 x_3 p(\boldsymbol{x},t)\right) + \frac{\partial}{\partial x_3}\left((x_1^2 - x_2^2)p(\boldsymbol{x},t)\right). \tag{19}$$

Reduced-order PDF equations for dynamical systems. The Liouville equation (9) describes the exact dynamics of the joint PDF of state variables $\boldsymbol{x}(t)$. In most cases, however, we are only interested in a smaller subset of such variables, e.g., in the scalar quantity of interest

$$z(t,\omega) = u(\boldsymbol{x}(t,\boldsymbol{x}_0(\omega))) \qquad \text{(phase space function)}. \tag{20}$$

We have seen that the probability density function of such phase space function can be written as

$$p(z,t) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\delta\left(z-u(\boldsymbol{x})\right)p(\boldsymbol{x},t)d\boldsymbol{x} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\delta\left(z-u(\boldsymbol{x}(t,\boldsymbol{x}_0))\right)p(\boldsymbol{x}_0)d\boldsymbol{x}_0, \tag{21}$$

where $\delta(\cdot)$ is the Dirac's delta function (see [8, 20, 14]) and z is the phase space variable representing $u(\boldsymbol{x}(t))$. Multiplying the Liouville equation (9) by $\delta\left(z-u(\boldsymbol{x})\right)$ and integrating over all phase variables yields

$$\frac{\partial p(z,t)}{\partial t} + \frac{1}{2\pi}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{ia(z-u(\boldsymbol{x}))}\nabla\cdot\left(\boldsymbol{f}(\boldsymbol{x})p(\boldsymbol{x},t)\right)d\boldsymbol{x}da = 0, \tag{22}$$

[3]Note that (9) can be written as

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} + \boldsymbol{f}(\boldsymbol{x})\cdot\nabla p(\boldsymbol{x},t) = -p(\boldsymbol{x},t)\nabla\cdot\boldsymbol{f}(\boldsymbol{x}) \qquad p(\boldsymbol{x},0) = p_0(\boldsymbol{x}). \tag{15}$$

Applying the method of characteristics to (15) yields the ODE system (16). In practice, the PDF $p(\boldsymbol{x},t)$ is computed using (16) along each characteristic curve. Clearly, this is computationally challenging in high-dimensions.

$$p(x_1,t)=\int_{-\infty}^{\infty}p(x_1,x_2,t)dx_2$$

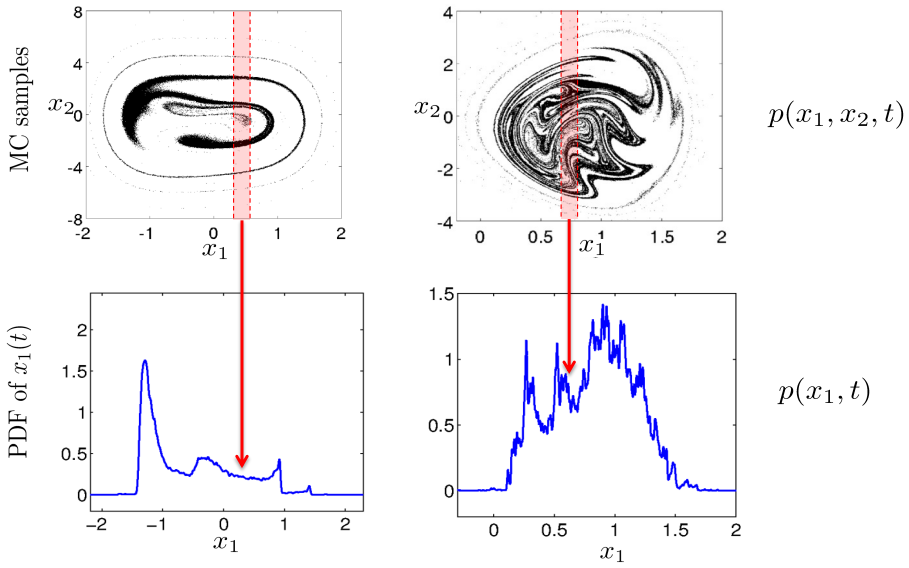

Figure 2: Regularization of PDFs by integration/marginalization. The PDF of $x_1(t)$ at one specific location is obtained by summing up the probability mass within the strips highlighted in red. The figures at the top represent the joint PDF of x_1 and x_2, i.e., x and $\dot{x}$ in the Duffing equation (13) at different times (see also Figure 1).

where we used the Fourier representation of the Dirac delta function $\delta\left(z-u(\boldsymbol{x})\right)$. In general, equation (22) is *unclosed* in the sense that there are terms at the right hand side that cannot be represented or computed based on $p(z,t)$ alone. If we set $u(\boldsymbol{x}(t))=x_k(t)$, i.e., the quantity of interest is the k-th component of the dynamical system (1), then (22) reduces to

$$\frac{\partial p(x_k,t)}{\partial t}+\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial}{\partial x_k}\left(f_k(\boldsymbol{x})p(\boldsymbol{x},t)\right)dx_1\ldots dx_{k-1}dx_{k+1}\ldots dx_N=0. \tag{23}$$

The specific form of this equation depends on the vector field $\boldsymbol{f}(\boldsymbol{x})$.

Remark: Low-dimensional marginals of high-dimensional PDF are usually smoother functions than the original PDF. This is illustrated in Figure 2 with reference to the Duffing equation (13). Hence, deriving and solving low-dimensional PDF equations for quantities of interest, has advantages relative to full Liouville equation. In particular: 1) the PDF equation for the quantity of interest is low dimensional, 2) we expect the solution to a reduced-order PDF equations to relatively smooth because of the "regularization by integration" effect.

BBGKY hierarchy. By integrating the Liouville equation (9) with respect to different phase variables it is possible to derive a hierarchy of PDEs known as *Bogoliubov-Born-Green-Kirkwood-Yvon (BBGKY) hierarchy* involving PDFs with an increasing number of phase variables. The first set of PDEs is (23), and it clearly depends on PDFs with a larger number of variables, unless $f_k(\boldsymbol{x})$ depends only on x_k (in which case the system (1) is uncoupled). Hereafter we provide specific examples of BBGKY hierarchies.

Example: Consider the Kraichnan-Orszag three-mode problem [13, 24]

$$\dot{x}_1 = x_1 x_3. \qquad \dot{x}_2 = -x_2 x_3, \qquad \dot{x}_3 = -x_1^2 + x_2^2. \tag{24}$$

The associated Liouville equation is

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} = -\frac{\partial}{\partial x_1}\left(x_1 x_3 p(\boldsymbol{x},t)\right) + \frac{\partial}{\partial x_2}\left(x_2 x_3 p(\boldsymbol{x},t)\right) + \frac{\partial}{\partial x_3}\left((x_1^2 - x_2^2) p(\boldsymbol{x},t)\right). \tag{25}$$

Suppose we are interested in the PDF of the first component of the system, i.e., set $u(\boldsymbol{x}(t)) = x_1(t)$ in equation (20). By integrating (25) with respect to x_2 and x_3, and assuming that $p(\boldsymbol{x},t)$ decays fast enough at infinity, we obtain

$$\frac{\partial p(x_1,t)}{\partial t} = -\frac{\partial}{\partial x_1}\int_{-\infty}^{\infty} x_1 x_3 p(x_1,x_3,t) dx_3. \tag{26}$$

From this equation we see that the evolution of $p(x_1,t)$ depends on an integral involving $p(x_1,x_3,t)$. Hence, to compute $p(x_1,t)$ we need to know what $p(x_1,x_3,t)$ is. The evolution equation for $p(x_1,x_3,t)$ can be obtained by integrating (25) with respect to x_2, i.e.,

$$\frac{\partial p(x_1,x_3,t)}{\partial t} = -\frac{\partial}{\partial x_1}\left(x_1 x_3 p(x_1,x_3,t)\right) + x_1^2 \frac{\partial}{\partial x_3}\left(x_3 p(x_1,x_3,t)\right) - \frac{\partial}{\partial x_3}\int_{-\infty}^{\infty} x_2^2 p(x_1,x_2,x_3,t) dx_2. \tag{27}$$

The PDE system (26)-(27) represents the first two levels of a BBGKY hierarchy. Note that the hierarchy be closed only at the level of the Liouville equation (25). Indeed, the integral at the right hand side of (27) involves $p(x_1,x_2,x_3,t)$, which is unknown unless we solve (25).

At this point we notice that we can represent the term involving $p(x_1,x_3,t)$ in (26) in a different way. Specifically, we can write the joint PDF of $x_1(t)$ and $x_3(t)$ at time t as

$$p(x_1,x_3,t) = p(x_1,t)p(x_3|x_1,t), \tag{28}$$

where $p(x_3|x_1,t)$ is the conditional probability density of $x_3(t)$ given $x_1(t)$. A substitution of (28) into (26) yields

$$\frac{\partial p(x_1,t)}{\partial t} = -\frac{\partial}{\partial x_1}\left(x_1 p(x_1,t)\mathbb{E}[x_3(t)|x_1(t)]\right), \tag{29}$$

where

$$\mathbb{E}[x_3(t)|x_1(t)] = \int_{-\infty}^{\infty} x_3 p(x_3|x_1,t) dx_3 \qquad \text{(conditional expectation of } x_3(t) \text{ given } x_1(t)\text{)}. \tag{30}$$

As we shall see hereafter, $\mathbb{E}[x_3(t)|x_1(t)]$ can be estimated from sample trajectories of (24).

Note that the reduced-order PDF equation (29) is a scalar conservation law where the (compressible) advection velocity field is equal to $x_1\mathbb{E}[x_3(t)|x_1(t)]$. It is important to emphasize that the "innocent-looking" PDE (26) is actually a PDE involving derivatives of $p(x_1,t)$ up to order *infinity* in the phase variable x_1. In fact, by using Kubo's cumulant expansion [9] of the joint characteristic function of $x_3(t)$ and $x_1(t)$ (i.e. Eq. (156) in lecture notes 1)

$$\phi(a_1,a_3,t) = \phi(a_1,t)\phi(a_3,t)\exp\left[\sum_{j,k=1}^{\infty}\left\langle x_1^j(t)x_3^k(t)\right\rangle_c \frac{(ia_1)^j(ia_3)^k}{j!k!}\right], \tag{31}$$

and the correspondence

$$p(x_1,x_3,t) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(a_1x_1+a_3x_3)}\phi(a_1,a_3,t)da_1da_3 \tag{32}$$

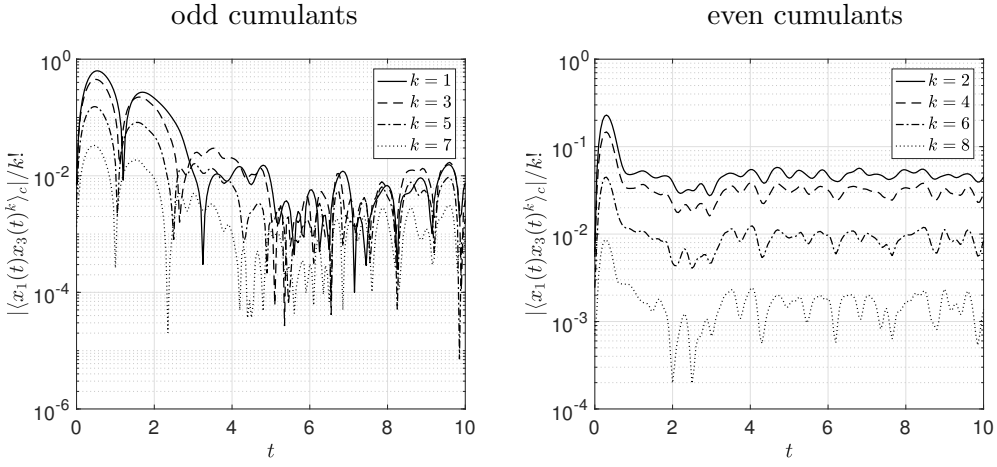

Figure 3: Kraichnan-Orszag three mode problem. Absolute values of the first 8 rescaled cumulants $\langle x_1(t)x_3(t)^k\rangle_c/k!$. The initial condition $x_i(0)$ $(i=1,2,3)$ in (24) is set to be i.i.d. Gaussian with mean and variance 1. We estimated the cumulants numerically by using Monte Carlo (50000 sample paths) and ensemble averages. It is seen that the odd cumulants decay slowly with k, suggesting that the cumulant expansion (33) cannot be truncated at low order. This implies that any reasonably accurate approximation of the reduced-order equation (35) involves high-order derivatives of $p(x_1,t)$ with respect to x_1.

we can prove that

$$\int_{-\infty}^{\infty} x_3 p(x_3,x_1,t)dx_3 = \mathbb{E}[x_3(t)]p(x_1,t) + \sum_{k=1}^{\infty}(-1)^{k+1}\frac{\langle x_1(t)x_3(t)^k\rangle_c}{k!}\frac{\partial^k p(x_1,t)}{\partial x_1^k}, \tag{33}$$

where $\langle x_1(t)x_3(t)^k\rangle_c$ are cumulant averages[4]. A substitution of (33) into (26) yields the infinite-order PDE

$$\frac{\partial p(x_1,t)}{\partial t} = -\mathbb{E}[x_3(t)]\frac{\partial\left(x_1 p(x_1,t)\right)}{\partial x_1} + \sum_{k=1}^{\infty}(-1)^{k+1}\frac{\langle x_1(t)x_3(t)^k\rangle_c}{k!}\frac{\partial^{k+1}\left(x_1 p(x_1,t)\right)}{\partial x_1^{k+1}}. \tag{35}$$

As shown in Figure 3, the rescaled cumulants $\langle x_1(t)x_3(t)^k\rangle_c/k!$ decay slowly with k, suggesting that the cumulant expansion (33) cannot be truncated at low-order. This implies that any reasonably accurate approximation of the reduced-order PDF equation (35) involves high-order derivatives of $p(x_1,t)$ with respect to x_1. The data-driven cumulant expansion approach we just described relies on computing sample paths of (24), estimating the cumulant averages $\langle x_1(t)x_3(t)^k\rangle_c$ using ensemble averaging, and then solving the PDE (35). Clearly this is not practical since such PDE potentially involves high-order derivatives of $p(x_1,t)$ with respect to x_1. Another approach relies on estimating the conditional expectation (30) directly from data and then solving the hyperbolic conservation law (29), which is a first-order linear PDE.

Example: Consider the following N-dimensional nonlinear dynamical system

$$\frac{dx_i}{dt} = -\sin(x_{i+1})x_i - Ax_i + F, \qquad i=1,...,N, \tag{36}$$

where $x_{N+1}(t) = x_1(t)$ (periodic boundary conditions). Depending on the value of F, A and on the number of phase variables N, this system can exhibit different behaviors. In Figure 4 we plot a 2D section of the

[4]The cumulant averages appearing in equation (33) are defined as

$$\langle x_1(t)x_3(t)^k\rangle_c = \mathbb{E}[x_1(t)x_3(t)^k] - \mathbb{E}\left[x_1(t)\right]\mathbb{E}[x_3(t)^k]. \tag{34}$$

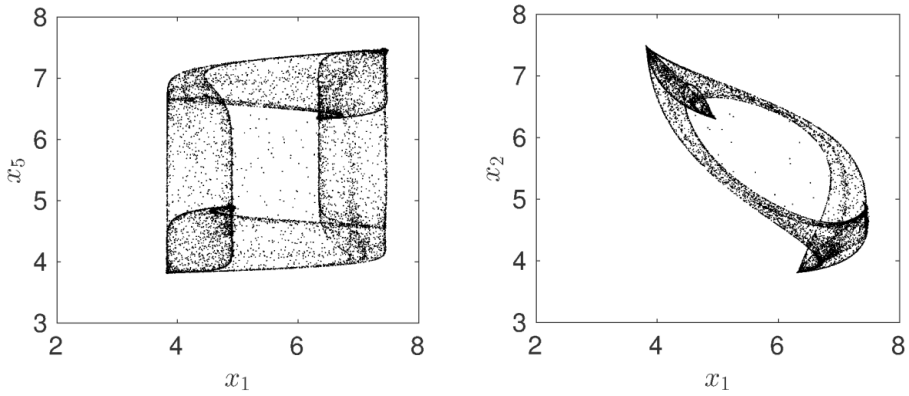

Figure 4: 2D sections of the point clouds generated by the dynamical system (36) at $t = 14$. The random initial condition samples are taken from a Gaussian distribution.

point cloud we obtain at $t = 14$ by sampling the initial condition from a Gaussian random vector. Here we set $F = 10$, $A = 0.2$ and $N = 1000$. The Liouville transport equation associated with (36) is

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} \left[(F - \sin(x_{i+1})x_i - Ax_i)\, p(\boldsymbol{x},t) \right]. \tag{37}$$

This PDF is very hard to solve numerically because of the very high number of phase variables. The evolution equation for the PDF of each phase variable $x_i(t)$ can be obtained by integrating (37) with respect to all other variables. This yields the unclosed equation

$$\frac{\partial p(x_i,t)}{\partial t} = -\frac{\partial}{\partial x_i} \int_{-\infty}^{\infty} \left[(F - \sin(x_{i+1})x_i - Ax_i)\, p(x_i, x_{i+1}, t) \right] dx_{i+1}. \tag{38}$$

We can write (38) equivalently as

$$\frac{\partial p(x_i,t)}{\partial t} = -F\frac{\partial p(x_i,t)}{\partial x_i} + A\frac{\partial (x_i p(x_i,t))}{\partial x_i} - \frac{\partial}{\partial x_i} x_i \int_{-\infty}^{\infty} \sin(x_{i+1}) p(x_i, x_{i+1}, t) dx_{i+1}. \tag{39}$$

Note that all equations for $p(x_i,t)$ have the same structure, independently of the index i. This means that if the random initial state $\boldsymbol{x}_0$ has i.i.d. components, then the evolution of each $p(x_i,t)$ does not depend on i, i.e., it is the same for all $i = 1, ..., N$. A similar conclusion holds for the joint distributions $p(x_i, x_{i+1}, t)$, which satisfy the equations

$$\begin{aligned} \frac{\partial p(x_i, x_{i+1}, t)}{\partial t} = &- \frac{\partial}{\partial x_i} \left[(F - \sin(x_{i+1})x_i - Ax_i)\, p(x_i, x_{i+1}, t) \right] - \\ &\frac{\partial}{\partial x_{i+1}} \int_{-\infty}^{\infty} \left[(F - \sin(x_{i+2})x_{i+1} - Ax_{i+1})\, p(x_i, x_{i+1}, x_{i+2}, t) \right] dx_{i+2}. \end{aligned} \tag{40}$$

The PDE system (39)-(40) represents the first two levels of the BBGKY hierarchy corresponding to (36).

Let us set $i = 1$ in equation (39) and express the integral in terms of the conditional expectation of $\sin(x_2(t))$ given $x_1(t)$. This yields

$$\frac{\partial p(x_1,t)}{\partial t} = \frac{\partial}{\partial x_1} \left(x_1 p(x_1,t) \mathbb{E}\left[\sin(x_2(t)) | x_1(t) \right] \right) + \frac{\partial}{\partial x_1} \left[(Ax_1 - F) p(x_1,t) \right], \tag{41}$$

where

$$\mathbb{E}\left[\sin(x_2(t)) | x_1(t) \right] = \int_{-\infty}^{\infty} \sin(x_2) p(x_2 | x_1, t) dx_2. \tag{42}$$

Example: Consider the Liouville equation (17) corresponding to (4)-(5). Is it possible to derive an evolution equation for $p(x,t)$, i.e., integrate the variables $\boldsymbol{y}$ representing $(\xi_1,\ldots,\xi_M)$ in the KL expansion of the noise (5)? By applying the marginalization rule

$$p(x,t)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x,\boldsymbol{y},t)d\boldsymbol{y} \tag{43}$$

to the Liouville equation (17) we obtain

$$\frac{\partial p(x,t)}{\partial t}+\frac{\partial}{\partial x}\left(f(x)p(x,t)\right)+\sum_{k=1}^{M}\sqrt{\lambda_k}\psi_k(t)\frac{\partial}{\partial x}\int_{-\infty}^{\infty}y_k p(x,y_k,t)dy_k=0. \tag{44}$$

Note that the PDF $p(x,t)$ depends on M joint PDFs $p(x,y_k,t)$. Therefore (44) is an unclosed PDF equation. We can of course derive an evolution equation for each $p(x,y_k,t)$ as

$$\frac{\partial p(x,y_k,t)}{\partial t}+\frac{\partial}{\partial x}\left(f(x)p(x,y_k,t)\right)+\frac{\partial p(x,y_k,t)}{\partial x}\sqrt{\lambda_k}y_k\psi_k(t)+\sum_{\substack{j=1\\ j\neq k}}^{M}\sqrt{\lambda_j}\psi_j(t)\frac{\partial}{\partial x}\int_{-\infty}^{\infty}y_j p(x,y_j,y_k,t)dy_j=0. \tag{45}$$

These are additional M unclosed PDEs involving $p(x,y_j,y_k,t)$. At this point we could derive the evolution equation for the joint PDF $p(x,y_j,y_k,t)$, and go on and on. The BBGKY hierarchy is formally closed only at the level of the Liouville equation, unless the system has a special structure, or a *closure approximation* is introduced. For instance, if $p(x,y_j,y_k,t)$ can be factored in terms of lower-order PDFs as

$$p(x,y_j,y_k,t)\simeq p(x,y_k,t)p(y_j) \tag{46}$$

then (44)-(45) is a closed system of PDEs.

A substitution of

$$p(x,y_k,t)=p(y_k|x,t)p(x,t), \tag{47}$$

where $p(y_k|x,t)$ is the conditional PDF of y_k given $x(t;\omega)$, into (44) yields the low dimensional PDE

$$\frac{\partial p(x,t)}{\partial t}+\frac{\partial}{\partial x}\left(f(x)p(x,t)\right)+\sum_{k=1}^{M}\psi_k(t)\frac{\partial}{\partial x}\left(p(x,t)\mathbb{E}\{\xi_k(\omega)|x(t;\omega)=x\}\right). \tag{48}$$

Here,

$$\mathbb{E}\{\xi_k(\omega)|x(t;\omega)=x\}=\int_{-\infty}^{\infty}y_k p(y_k|x,t)dy_k \tag{49}$$

is the conditional expectation of $\xi_k(\omega)$ given $x(t;\omega)=x$. We shall see hereafter that such conditional expectation can be estimated from sample trajectories of (4)-(5).

Data-driven closure approximation of BBGKY hierarchies. Computing conditional expectations from data or sample trajectories is a key step in determining accurate closure approximations of reduced-order PDF equations. A major challenge to fitting a conditional expectation is ensuring accuracy and stability. More importantly, the estimator must be flexible and effective for a wide range of numerical applications. Let us briefly recall what conditional expectations are and, more importantly, how to compute them based on sample paths of (1). To this end, consider the random processes $x_1(t)$ and $x_3(t)$ defined by the dynamical system (24) evolving from a random initial state. The conditional expectation of $x_3(t)$ given $x_1(t)$ is defined mathematically in equation (30). The geometric meaning of such conditional expectation is illustrated in Figure 5. We first compute sample trajectories of (24) – see Figure 5(a) – by sampling the initial condition and evolving it forward in time. We then project the solution samples we obtain at time t

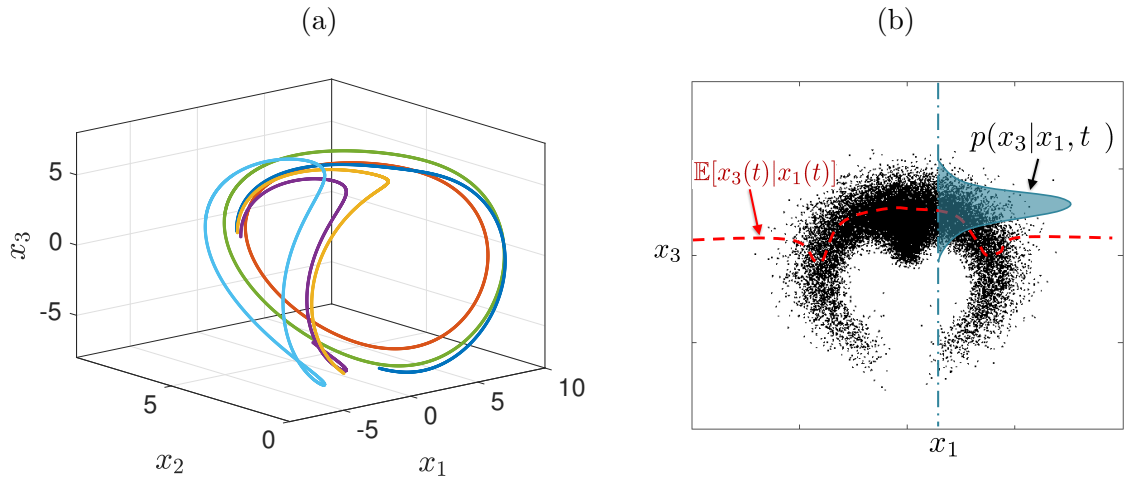

Figure 5: Kraichnan-Orszag three mode problem. (a) Sample trajectories of (24) corresponding to random samples projected on the plane (x_1, x_3). For each value of x_1, the conditional PDF $p(x_3|x_1, t)$ can be estimated based on samples sitting on or lying nearby the vertical dashed line. The conditional expectation $\mathbb{E}[x_3(t)|x_1(t)]$ is the mean of such conditional PDF.

on the plane (x_1, x_3), to obtain the scatter plot in Figure 5(b). For each value of x_1, the conditional PDF $p(x_3|x_1, t)$ can be estimated based on all samples sitting on or lying nearby the vertical dashed line. The conditional expectation $\mathbb{E}[x_3(t)|x_1(t)]$ is the mean of such conditional PDF.

In this section, we present two different approaches to estimate conditional expectations from data based on *moving averages* and *smoothing splines*. The moving average estimate is obtained by first sorting the data into bins and then computing the average within each bin. With such averages available, we can construct a smooth interpolant using the average value within each bin. Important factors affecting the bin average are the bin size (the number of samples in each bin) and the interpolation method used in the final step. Another approach to estimate conditional expectations uses smoothing splines. This approach seeks to minimize a penalized sum of squares. A smoothing parameter determines the balance between smoothness and goodness-of-fit in the least-squares sense [3]. The choice of smoothing parameter is critical to the accuracy of the results. Specifying the smoothing parameter a priori generally yields poor estimates [15]. Instead, cross-validation and maximum likelihood estimators can guide the choice the optimal smoothing value for the data set [23]. Such methods can be computationally intensive, and is not recommended when the spline estimate is performed at each time step. Other techniques to estimate conditional expectations can be built upon deep-neural nets.

In Figure 6 we compare the performance of the moving average and smoothing splines approaches in approximating the conditional expectation of two jointly Gaussian random variables. Specifically, we consider the joint distribution

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)}{\sigma_1^2}\frac{(x_2-\mu_2)}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2)}\right]\right) \tag{50}$$

with parameters $\rho = 3/4$, $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1 = 1$ $\sigma_2 = 2$. As is well known [14], the conditional expectation

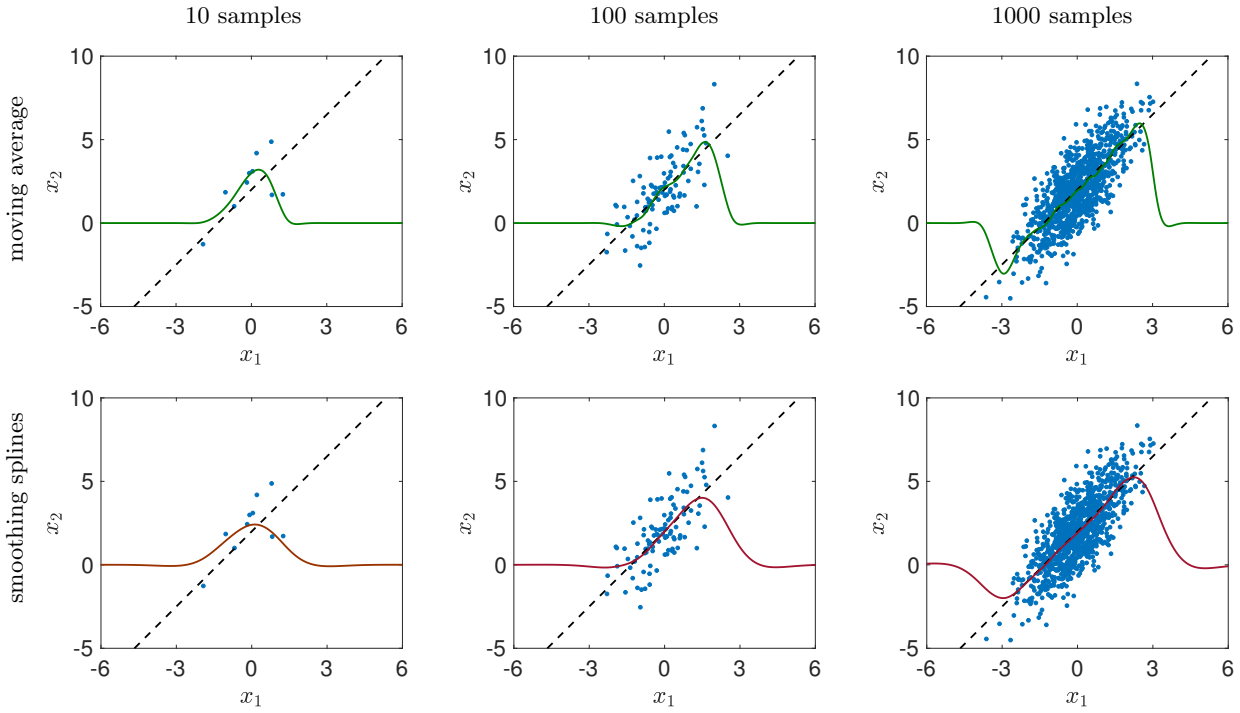

Figure 6: Numerical estimation of the conditional expectation (52) for different number of samples of (50). Shown are results obtained with moving averages (first row) and cubic smoothing splines (second row). It is seen that both methods converge to the correct conditional expectation in the active region as we increase the number of samples.

of x_2 given x_1 can be expressed as[5]

$$\mathbb{E}[x_2|x_1] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1) = 2 + \frac{3}{2}x_1. \tag{52}$$

Such expectation is plotted in Figure 6 (dashed line), together with the plots of the estimates we obtain with the moving average and the smoothing spline approaches for different numbers of samples. It is seen that both methods converge to the correct conditional expectation as we increase the number of samples. Both estimators are parametric, i.e., they require setting suitable parameters to compute the expectation, e.g., the width of the moving average window in the moving average approach, or the smoothing parameter in the cubic spline.

If the joint PDF of x_1 and x_2 is not compactly supported, then the conditional expectation is defined on the whole real line. It is computationally challenging to estimate (52) in regions where the PDF is very small. At the same time, if we are not interested in rare events (i.e., tails of probability densities), then resolving the dynamics in such regions of small probability is not really needed. This means that if we have available a sufficient number of sample trajectories, then we can identify the regions of the phase space where dynamics is happening with high probability, and approximate the conditional expectation only within such regions. Outside the active regions, we can set the expectation equal to zero. However, keep in mind that if the joint PDF of x_1 and x_2 is compactly supported, e.g. uniform on the square $[0,1]^2$, then conditional expectation is undefined outside the support of the joint PDF.

[5]Given two random variables with joint PDF $p(x_1, x_2)$, the conditional expectation of x_2 given x_1 is defined as

$$\mathbb{E}[x_2|x_1] = \int_{-\infty}^{\infty} x_2 p(x_2|x_1)dx_2 = \frac{1}{p(x_1)}\int_{-\infty}^{\infty} x_2 p(x_1, x_2)dx_2, \tag{51}$$

where $p(x_1)$ is the marginal of $p(x_1, x_2)$ with respect to x_2.

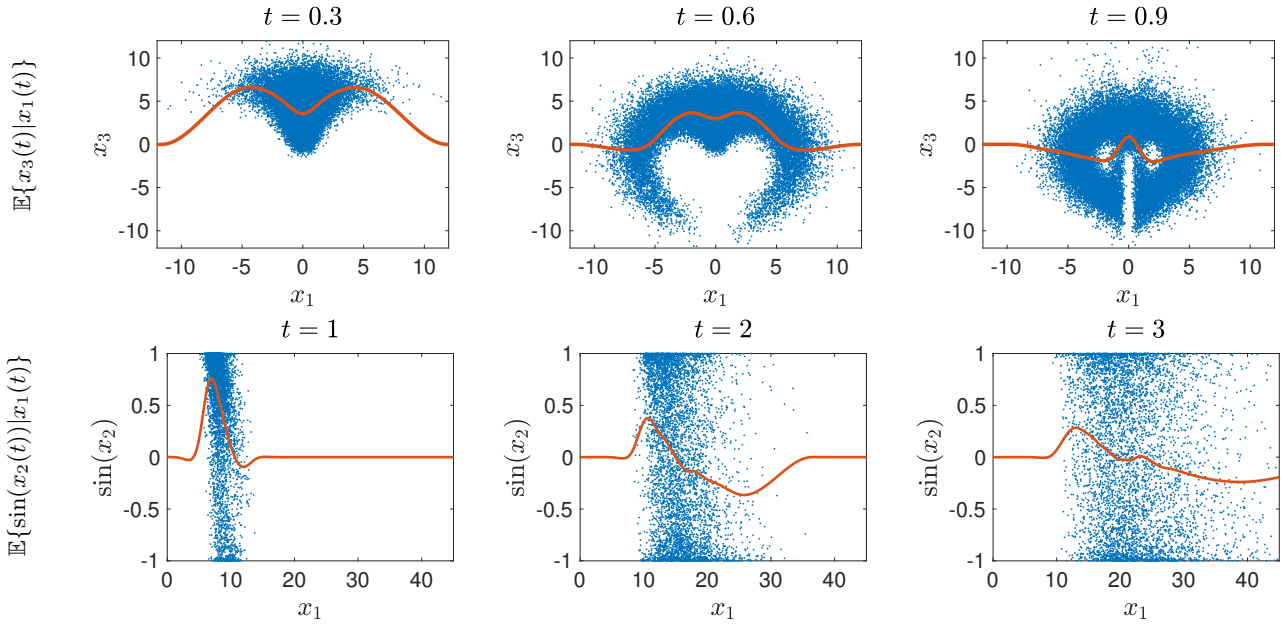

Figure 7: Data-driven estimates of the conditional expectations (30) and (42) defining the reduced-order PDF models (29) and (41).

In Figure 7, we summarize the results we obtain by applying the smoothing spline conditional expectation estimator to the dynamical systems (24) and (36). In figure 8 and figure 9 we provide numerical simulation result for (29) and (41), respectively.

PDF equations for nonlinear PDEs evolving from random initial conditions

The procedure we used to derive reduced-order PDF equations for dynamical systems can be extended to PDEs evolving from random initial states, or PDEs with random forcing (see, e.g., [10, 7]). To describe the method, consider the prototype problem of a one dimensional heat equation evolving from a random initial state

$$\frac{\partial u(x,t;\omega)}{\partial t} = \kappa^2 \frac{\partial^2 u(x,t;\omega)}{\partial x^2}, \qquad u(x,0;\omega) = u_0(x;\omega). \tag{53}$$

We have seen in Chapter 1 that the Hopf functional[6]

$$\Phi([\theta],t) = \mathbb{E}\left\{\exp\left[i\int_{-\infty}^{\infty} u(x,t;\omega)\theta(x)dx\right]\right\} \tag{54}$$

provides full statistical information on $u(x,t;\omega)$ at each time t. This includes, e.g., multi-point statistical moments such as

$$\mathbb{E}\left\{u(x_i,t;\omega)u(x_j,t;\omega)\right\} \quad \text{and} \quad \mathbb{E}\left\{u(x_i,t;\omega)u(x_j,t;\omega)u(x_k,t;\omega)\right\}, \tag{55}$$

or multi-point probability density functions. It is possible to derive an evolution equation for the Hopf functional corresponding to the solution of (53). To this end, let us differentiate (54) with respect to time

[6] In (54) we assumed that the spatial domain for the heat equation (53) is $\mathbb{R}$. If the spatial domain is a compact subset $\mathbb{R}$, say $[0, 2\pi]$, then the domain on which the integral in (54) is evaluated changes accordingly.

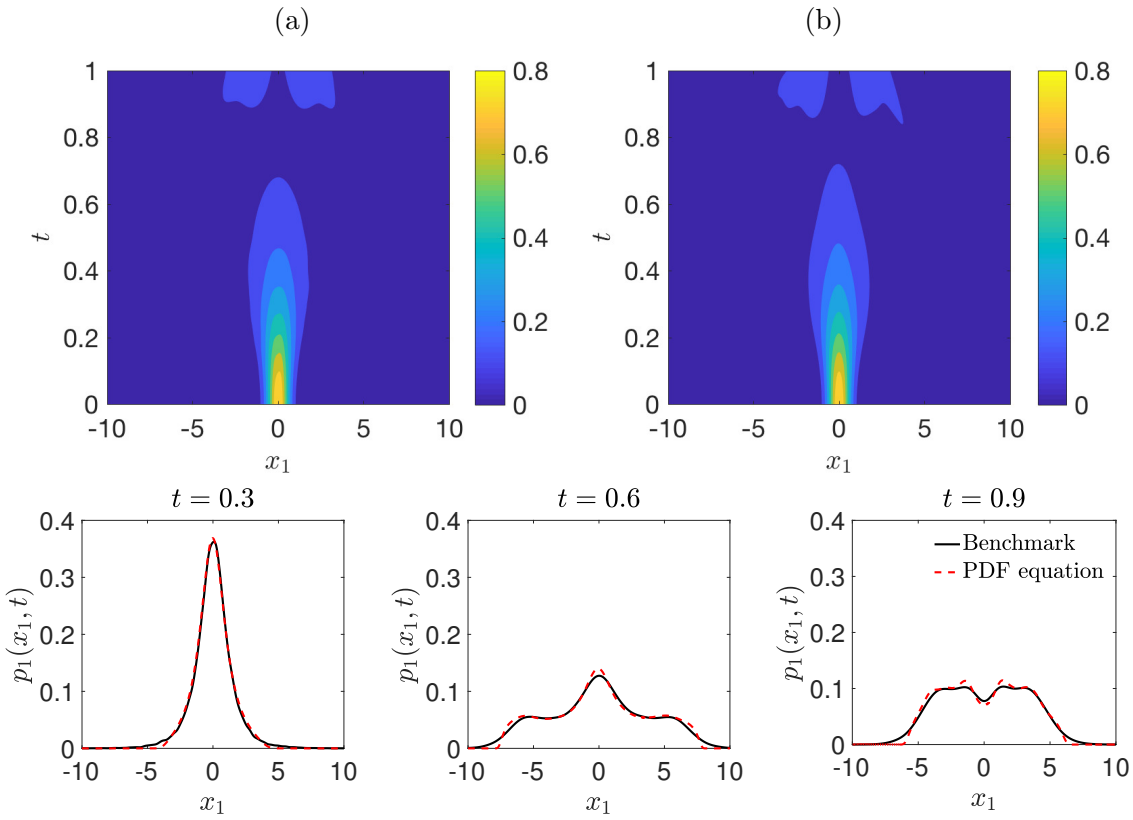

Figure 8: Kraichnan-Orszag three-mode problem. (a) Accurate kernel density estimate of $p_1(x_1,t)$ based on 30000 sample trajectories. (b) Numerical solution of (29) obtained by estimating $\mathbb{E}[x_3(t)|x_1(t)]$ with 5000 sample trajectories.

to obtain

$$\begin{aligned}\frac{\partial\Phi([\theta],t)}{\partial t} =&\mathbb{E}\left\{\exp\left[i\int_{-\infty}^{\infty}u(x,t;\omega)\theta(x)dx\right]i\int_{-\infty}^{\infty}\frac{\partial u(x,t;\omega)}{\partial t}\theta(x)dx\right\}\\ =&i\kappa^2\int_{-\infty}^{\infty}\mathbb{E}\left\{\exp\left[i\int_{-\infty}^{\infty}u(x,t;\omega)\theta(x)dx\right]\frac{\partial^2u(x,t;\omega)}{\partial x^2}\right\}\theta(x)dx\\ =&i\kappa^2\int_{-\infty}^{\infty}\frac{\partial^2}{\partial x^2}\left(\mathbb{E}\left\{\exp\left[i\int_{0}^{2\pi}u(x,t;\omega)\theta(x)dx\right]u(x,t;\omega)\right\}\right)\theta(x)dx\\ =&i\kappa^2\int_{-\infty}^{\infty}\frac{\partial^2}{\partial x^2}\left(\frac{\delta\Phi([\theta],t)}{\delta\theta(x)}\right)\theta(x)dx,\end{aligned}\tag{56}$$

where $\delta\Phi([\theta],t)/\delta\theta(x)$ denotes the first-order functional derivative of the nonlinear functional (54) (see [19] or [5, p. 309]). Technically speaking, equation (56) is a *functional-differential* equation (FDE) as it involves derivatives with respect to functions and derivatives with respect to independent variables x and t. The solution to (56) is a time-dependent nonlinear functional, i.e., a nonlinear operator from a space of functions into $\mathbb{C}$. The functional differential equation (56) is essentially an infinite-dimensional PDE, i.e., a PDE in an infinite number of independent variables which may be approximated by a PDE in a finite (though very large) number of variables using the functional methods described in [19].

The Hopf equation (56) plays the same role for the heat equation (53) as the Fourier transform of the Liouville equation (9) does for finite-dimensional dynamical systems (1).

Example: Another well-known example of a FDE involves the characteristic functional of the solution to

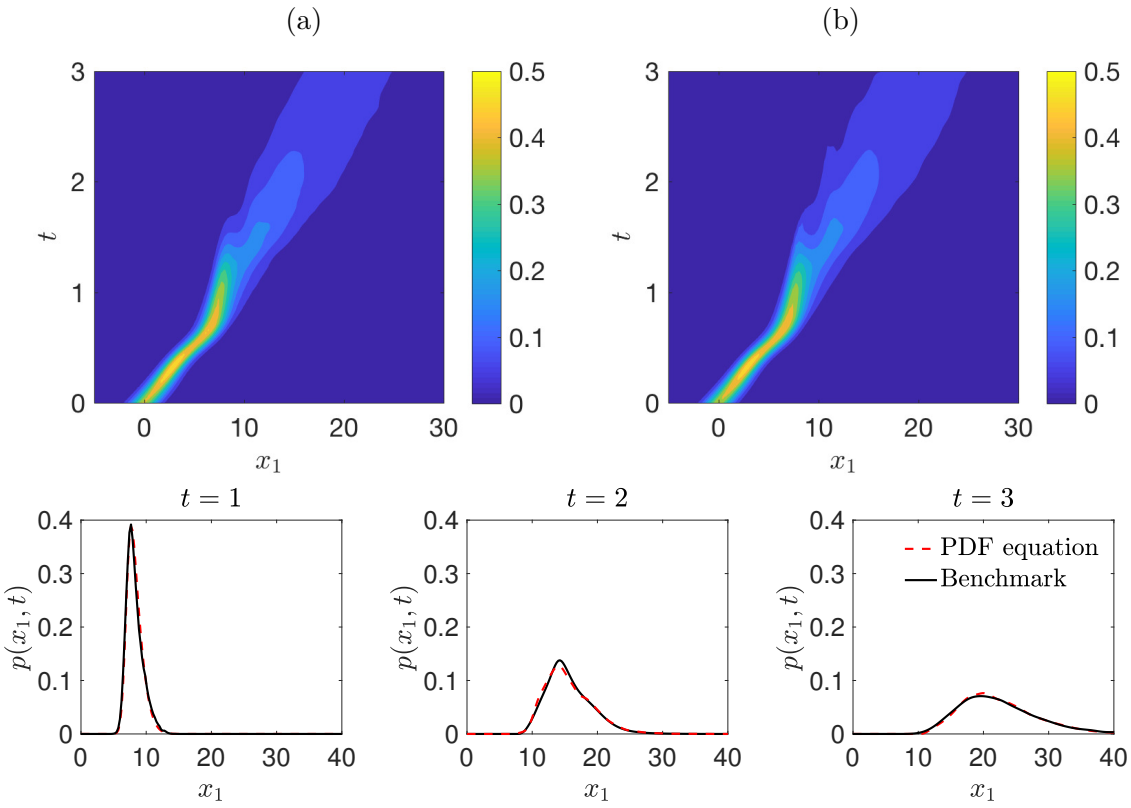

Figure 9: Nonlinear dynamical system (36). (a) Accurate kernel density estimate [2] of $p(x_1, t)$ based on 20000 sample trajectories. (b) Data-driven solution of the transport equation (41). We estimated the conditional expectation $\mathbb{E}\left[\sin(x_2(t))|x_1(t)\right]$ based on 5000 sample trajectories of (36) (see Figure 7).

the Navier-Stokes equations

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla p + \nu \nabla^2 \boldsymbol{u} \qquad \nabla \cdot \boldsymbol{u} = 0. \tag{57}$$

Such a FDE can be written as [6, 11, 19]

$$\frac{\partial \Phi([\boldsymbol{\theta}], t)}{\partial t} = \sum_{k=1}^{3} \int_V \theta_k(\boldsymbol{x}) \left(i \sum_{j=1}^{3} \frac{\partial}{\partial x_j} \frac{\delta^2 \Phi([\boldsymbol{\theta}], t)}{\delta \theta_k(\boldsymbol{x}) \delta \theta_j(\boldsymbol{x})} + \nu \nabla^2 \frac{\delta \Phi([\boldsymbol{\theta}], t)}{\delta \theta_k(\boldsymbol{x})} \right) d\boldsymbol{x}, \tag{58}$$

where

$$\Phi([\theta], t) = \mathbb{E}\left\{ \exp\left[i \int_V \boldsymbol{u}(\boldsymbol{x}, t; \omega) \cdot \boldsymbol{\theta}(\boldsymbol{x}) d\boldsymbol{x} \right] \right\}. \tag{59}$$

Here, $\boldsymbol{u}(x, t; \omega)$ represents a stochastic solution to the Navier-Stokes equation (57) corresponding to a random initial state, and $\mathbb{E}\{\cdot\}$ is the expectation over the probability measure of such random initial state. Equation (58) was deemed by Monin and Yaglom ([11, Ch. 10]) to be *"the most compact formulation of the general turbulence problem"*, which is the problem of determining the statistical properties of the velocity field generated by the Navier-Stokes equations given statistical information on the initial state[7].

Remark: Clearly, if we discretize the PDE (53) or (57) in the spatial domain, e.g., with finite-differences, then we obtain a system of ODEs which can be handled with the mathematical tools we discussed in the

[7] In equations (58)-(59), $V \subseteq \mathbb{R}^3$ is a periodic box, $\boldsymbol{\theta}(\boldsymbol{x}) = (\theta_1(\boldsymbol{x}), \theta_2(\boldsymbol{x}), \theta_3(\boldsymbol{x}))$ is a vector-valued (divergence-free) function, and $\delta / \delta \theta_j(\boldsymbol{x})$ denotes the first-order functional derivative.

previous section. In particular, it is possible to derive a Liouville equation for such finite-dimensional ODE system and correspondingly a BBGKY hierarchy for the solution evaluated e.g., at the spatial grid points. An interesting question is how to compute statistical properties at spatial locations that do not coincide with the grid points. For example, is it possible to "interpolate" the joint characteristic function of the solution $u(x,t;\omega)$ at n spatial nodes $\{x_k\}$ and obtain an approximation of the joint characteristic at a different set of m nodes? To answer this question, consider the 2-point characteristic function

$$\phi_2(a_1,a_2,t) = \mathbb{E}\left\{e^{ia_1u(x_1,t;\omega)+ia_2u(x_2,t;\omega)}\right\}. \tag{60}$$

Let x^* be a point in between x_1 and x_2. Assuming that $u(x,t;\omega)$ is a smooth solution to a PDE, we can construct an interpolant for $u(x^*,t,\omega)$, e.g., a linear interpolant as

$$u(x^*,t;\omega) = u(x_1,t;\omega)\ell_1(x^*) + u(x_2,t;\omega)\ell_2(x^*) \tag{61}$$

where $\ell_1(x) = (x-x_2)/(x_1-x_2)$ and $\ell_2(x) = (x-x_1)/(x_2-x_1)$ are Lagrange characteristic polynomials. This representation allows us to represent the three-point joint characteristic function of $u(x,t;\omega)$ at x_1, x_2 and x^* as

$$\begin{aligned}\phi_3(a_1,a_2,a_3,t) =& \mathbb{E}\left\{e^{ia_1u(x_1,t;\omega)+ia_2u(x_2,t;\omega)+ia_3u(x^*,t;\omega)}\right\}\\ =& \mathbb{E}\left\{e^{ia_1u(x_1,t;\omega)+ia_2u(x_2,t;\omega)+ia_3(u(x_1,t;\omega)\ell_1(x^*)+u(x_2,t;\omega)\ell_2(x^*))}\right\}\\ =& \mathbb{E}\left\{e^{i(a_1+a_3\ell_1(x^*))u(x_1,t;\omega)+i(a_2+a_3\ell_2(x^*))u(x_2,t;\omega)}\right\}\\ =& \phi_2(a_1+a_3\ell_1(x^*),a_2+a_3\ell_2(x^*),t).\end{aligned} \tag{62}$$

This expression provides an approximation of the three-point characteristic function in terms of the two point characteristic function. Of course the method can be generalized to n point characteristic functions. If the spatial discretization is sufficiently fine, and the interpolants are accurate, we can represent the $2n$, $3n$, etc., characteristic functions in terms of one *core* characteristic function e.g., involving the solution at n spatial points.

Lundgren-Monin-Novikov (LMN) hierarchy. We are interested in deriving the PDF equation governing the PDF of $u(x,t)$. To this end, consider the characteristic function

$$\phi(a,x,t) = \mathbb{E}\left\{e^{iau(x,t;\omega)}\right\}, \tag{63}$$

and differentiate it with respect to time to obtain

$$\begin{aligned}\frac{\partial\phi(a,x,t)}{\partial t} =& ia\mathbb{E}\left\{\frac{\partial u(x,t;\omega)}{\partial t}e^{iau(x,t;\omega)}\right\}\\ =& ia\kappa^2\mathbb{E}\left\{\frac{\partial^2 u(x,t;\omega)}{\partial x^2}e^{iau(x,t;\omega)}\right\}\\ =& ia\kappa^2\lim_{y\to x}\mathbb{E}\left\{\frac{\partial^2 u(y,t;\omega)}{\partial y^2}e^{iau(x,t;\omega)}\right\}\\ =& ia\kappa^2\lim_{y\to x}\frac{\partial^2}{\partial y^2}\mathbb{E}\left\{u(y,t;\omega)e^{iau(x,t;\omega)}\right\}.\end{aligned} \tag{64}$$

Recalling that the two-point characteristic function is defined as

$$\phi(a,b,x,y,t) = \mathbb{E}\left\{e^{iau(x,t;\omega)+ibu(y,t;\omega)}\right\} \tag{65}$$

we see that we can write the term at the right hand side of (64) as

$$i\mathbb{E}\left\{u(y,t;\omega)e^{iau(x,t;\omega)}\right\} = \lim_{b\to 0}\frac{\partial}{\partial b}\phi(a,b,x,y,t). \tag{66}$$

Substituting into (66) into (64) yields

$$\frac{\partial\phi(a,x,t)}{\partial t} = \kappa^2 \lim_{b\to 0}\lim_{y\to x}\frac{\partial^2}{\partial y^2}a\frac{\partial\phi(a,b,x,y,t)}{\partial b}. \tag{67}$$

Next, we transform this equation for the characteristic function to an equation for the PDF. To this end, we first recall that

$$\phi(a,x,t) = \int_{-\infty}^{\infty} e^{iau}p(u,x,t)du, \qquad \phi(a,b,x,y,t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iau+ibv}p(u,v,x,y,t)dudv. \tag{68}$$

This allows us to write the right hand side of (67) as

$$a\frac{\partial\phi(a,b,x,y,t)}{\partial b} = ia\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iau+ibv}vp(u,v,x,y,t)dudv. \tag{69}$$

Taking the limit

$$\begin{aligned}\lim_{b\to 0} a\frac{\partial\phi(a,b,x,y,t)}{\partial b} =& ia\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ve^{iau}p(u,v,x,y,t)dudv \\ =& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v\frac{\partial\left(e^{iau}\right)}{\partial u}p(u,v,x,y,t)dudv \\ =& -\int_{-\infty}^{\infty} e^{iau}\int_{-\infty}^{\infty} v\frac{\partial p(u,v,x,y,t)}{\partial u}dudv.\end{aligned} \tag{70}$$

Hence,

$$\frac{\partial p(u,x,t)}{\partial t} = -\kappa^2\lim_{y\to x}\frac{\partial^2}{\partial y^2}\int_{-\infty}^{\infty} v\frac{\partial p(u,v,x,y,t)}{\partial u}dv. \tag{71}$$

In other words, the dynamics of the one-point PDF $p(u,x,t)$ (PDF of the solution at location x and time t) depends on the joint PDF of $u(x,t;\omega)$ and $u(y,t,\omega)$ through some quite unusual limit. Equation (71) is the first equation PDF of an infinite hierarchy known as *Lundgren-Monin-Novikov (LMN) hierarchy* [10, 4, 22], first developed by Thomas Lundgren to study the statistical properties of turbulence. The second equation of the LMN hierarchy is an equation for the time derivative of the two point PDF $p(u,v,x,y,t)$. It was shown in [7] that the Hopf functional equation is completely equivalent to the LMN hierarchy.

Remark: Why do we get a closure problem for the one-point one-time PDF equation of the solution to the heat equation? The reason is rather simple, and can be understood by recalling that the analytical solution of (53) in an infinite domain is

$$u(x,t;\omega) = \int_{-\infty}^{\infty}\mathscr{G}\left(x,t|x',t'\right)u(x',t')dx' \qquad t\geq t', \tag{72}$$

where

$$\mathscr{G}\left(x,t|x',t'\right) = \frac{1}{[4\pi\kappa^2(t-t')]^{1/2}}\exp\left(-\frac{(x-x')^2}{4\kappa^2(t-t')}\right) \tag{73}$$

is the heat kernel, i.e., the Green function of the diffusion equation on the real line. For an infinitesimal time increment Δt we have that the random variable $u(x,t+\Delta t;\omega)$ (for fixed x) depends on all random

variables at the previous time step $u(x',t;\omega)$ (arbitrary $x' \in \mathbb{R}$). To see this more clearly, consider the following quadrature approximation of the integral in (72) (e.g., Hermite quadrature)

$$u(x_p,t+\Delta t;\omega) = \sum_{j=1}^{M} \frac{w_j}{[4\pi\kappa^2\Delta t]^{1/2}} \exp\left(-\frac{(x_p-x_j)^2}{4\kappa^2\Delta t}\right) u(x_j,t;\omega), \tag{74}$$

where x_j are Gauss-Hermite nodes, and w_j are quadrature weights. Clearly, (74) represents a mapping from M random variables $\{u(x_1,t;\omega),\ldots,u(x_M,t;\omega)\}$ into one random variable $u(x_p,t+\Delta t;\omega)$. We know that the PDF of $u(x_p,t+\Delta t;\omega)$ can be computed if and only if the joint PDF of the random vector $\{u(x_1,t;\omega),\ldots,u(x_M,t;\omega)\}$ is available. In other words, the fact that the solution (72) is *non-local* in space implies that the statistical properties at some fixed spatial point x and time $t+\Delta t$ are determined by the joint statistics at all points x' at a previous time instant. Hence, a closed equation for the one-point PDF cannot exist.

By using similar methods, it is possible to derive LMN PDF hierarchies corresponding to rather general nonlinear PDEs, e.g., the Navier-Stokes equation (57), evolving from random initial states (see [10]).

Data-driven closure approximation of LMN hierarchies. The integral at the right hand side of (71) can be written in terms of a conditional expectation of $u(y,t;\omega)$ given $u(x,t;\omega)$. A substitution of the identity

$$p(u,v,x,y,t) = p(v,y,t|u,x,t)p(u,x,t) \tag{75}$$

into (71) yields

$$\begin{aligned}\frac{\partial p(u,x,t)}{\partial t} &= -\kappa^2 \lim_{y\to x}\frac{\partial^2}{\partial y^2}\int_{-\infty}^{\infty} v p(v,y,t|u,x,t)\frac{\partial p(u,x,t)}{\partial u}dv \\ &= -\kappa^2 \frac{\partial p(u,x,t)}{\partial u}\lim_{y\to x}\frac{\partial^2}{\partial y^2}\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\}.\end{aligned} \tag{76}$$

As before, if we estimate the conditional expectation $\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\}$ from sample paths of (53) then we can solve (76) as we did in the case of data-driven closures for BBGKY hierarchies. The development of efficient methods for data-driven estimation of conditional expectations such as $\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\}$ nearby $x=y$ is (to my knowledge) an open problem.

Example: Consider the Kuramoto-Sivashinsky equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial^2 u}{\partial x^2} + \nu\frac{\partial^4 u}{\partial x^4} = 0 \tag{77}$$

By using the methods we just outlined it can be shown that the first equation of the LMN hierarchy is

$$\begin{aligned}\frac{\partial p(u,x,t)}{\partial t} + \int_{-\infty}^{u}\frac{\partial p(u',x,t)}{\partial x}du' + u\frac{\partial p(u,x,t)}{\partial x} = \\ -\lim_{y\to x}\left[\frac{\partial^2}{\partial y^2}\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\} + \nu\frac{\partial^4}{\partial y^4}\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\}\right].\end{aligned} \tag{78}$$

Hence, once again, the PDF equation can be closed by estimating $\mathbb{E}\left\{u(y,t;\omega)|u(x,t;\omega)\right\}$ from data.

Appendix A: Fokker-Planck equation and generalized Fokker-Planck equations

Consider the following stochastic ODE

$$d\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{X},t)dt + \boldsymbol{M}(\boldsymbol{X},t)d\boldsymbol{\zeta}(t;\omega), \qquad \boldsymbol{X}(0;\omega) = \boldsymbol{X}_0(\omega). \tag{79}$$

where $\boldsymbol{\zeta}(t)$ is a vector-valued m-dimensional random process with known statistical properties, and $\boldsymbol{M}(\boldsymbol{X},t)$ is a $n\times m$ matrix of functions. We have seen at the beginning of this Chapter that if $\boldsymbol{\zeta}(t)$ is finite-dimensional (i.e., it can be represented in terms of a finite-number of random variables) then it is possible to derive an exact transport equation (i.e., (9)) for the joint PDF of $\boldsymbol{X}(t;\omega)$ and all random variables representing $\boldsymbol{\zeta}(t;\omega)$. By integrating out the phase variables corresponding to the noise, i.e., by marginalizing the Liouville equation with respect to the phase variables representing the noise, it is straightforward to obtain an evolution equation for the PDF of $\boldsymbol{X}(t;\omega)$ alone. Such an equation represents the first equation of a BBGKY hierarchy and it is usually not closed, meaning that it involves quantities that cannot be computed just based on the PDF of $\boldsymbol{X}(t;\omega)$. However, there are cases in which the integration of the noise can be carried out exactly, and a closed equation for the PDF of $\boldsymbol{X}(t;\omega)$ can be derived. Perhaps the most famous example is the case where $\boldsymbol{x}i(t)$ is a Wiener process. In this case it was shown in [16] that $p(\boldsymbol{x},t)$ satisfies the Fokker-Plank equation

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} = -\sum_{k=1}^{n}\frac{\partial}{\partial x_k}\left(G_k(\boldsymbol{x},t)p(\boldsymbol{x})\right) + \frac{1}{2}\sum_{i,k=1}^{n}\frac{\partial^2}{\partial x_i\partial x_k}\left(\sum_{j=1}^{m}M_{ij}(\boldsymbol{x},t)M_{kj}(\boldsymbol{x},t)p(\boldsymbol{x},t)\right). \tag{80}$$

Let us denote by $\mathscr{K}(\boldsymbol{x},t)$ the Kolmogorov operator defining the right hand side of (80), i.e.,

$$\mathscr{K}(\boldsymbol{x},t)p(\boldsymbol{x},t) = -\sum_{k=1}^{n}\frac{\partial}{\partial x_k}\left(G_k(\boldsymbol{x},t)p(\boldsymbol{x})\right) + \frac{1}{2}\sum_{i,k=1}^{n}\frac{\partial^2}{\partial x_i\partial x_k}\left(\sum_{j=1}^{m}M_{ij}(\boldsymbol{x},t)M_{kj}(\boldsymbol{x},t)p(\boldsymbol{x},t)\right). \tag{81}$$

It is shown in [16, p. 86] that $\mathscr{K}$ represents the first-order term in the short-time expansion of the transition density

$$p_{t+dt|t}(\boldsymbol{x},t+dt|\boldsymbol{y},t) = \left[I + \mathscr{K}(\boldsymbol{x},t)dt + \mathcal{O}(dt^2)\right]\delta(\boldsymbol{x}-\boldsymbol{y}), \tag{82}$$

where $\delta(\cdot)$ is the multivariate Dirac delta function. The transition density (82) allows us to compute the PDF $p(\boldsymbol{x},t)$ of the random vector $\boldsymbol{X}(t;\omega)$ appearing in (79) given the PDF $p(\boldsymbol{x},s)$ of $\boldsymbol{X}(s;\omega)$ at any time $s\leq t$

$$p(\boldsymbol{x},t) = \int p_{t|s}(\boldsymbol{x},t|\boldsymbol{y},s)p(\boldsymbol{y},s)d\boldsymbol{y}. \tag{83}$$

We emphasize that the PDE governing the PDF of the solution to (4) depends substantially on the statistical properties of the noise $\boldsymbol{\zeta}(t)$. For instance, if we replace the Wiener process $\boldsymbol{\zeta}(t;\omega)$ in (4) with a Lévy random walk then the PDF equation for $\boldsymbol{X}(t;\omega)$ comes with a *fractional Laplace operator* [18], i.e., it is a fractional PDE.

Similarly, for weakly colored random noise, i.e., noise with short temporal correlation, it is possible to leverage the quasi-Markovian nature of the system, and integrate out the noise, e.g., using functional integration [12, 21].

References

[1] C. Bonatto, J. A. C. Gallas, and Y. Ueda. Chaotic phase similarities and recurrences in a damped-driven Duffing oscillator. *Phys. Rev. E*, 77:026217(1–5), 2008.

[2] Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38(5):2916–2957, 2010.

[3] P. Craven and G. Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4):377–403, 1979.

[4] R. Friedrich, A. Daitche, O. Kamps, J. Lülff, M. Voβkuhle, and M. Wilczek. The Lundgren-Monin-Novikov hierarchy: kinetic equations for turbulence. *Comptes Rendus Physique*, 13(9-10):929–953, 2012.

[5] P. Hänggi. Colored noise in continuous dynamical system. In F. Moss and P. V. E. McClintock, editors, *Noise in nonlinear dynamical systems (Vol. 1)*, pages 307–347. Cambridge Univ. Press, 1989.

[6] E. Hopf. Statistical hydromechanics and functional calculus. *J. Rat. Mech. Anal.*, 1(1):87–123, 1952.

[7] I. Hosokawa. Monin-Lundgren hierarchy versus the Hopf equation in the statistical theory of turbulence. *Phys. Rev. E*, 73:067301(1–4), 2006.

[8] A. I. Khuri. Applications of Dirac's delta function in statistics. *Int. J. Math. Educ. Sci. Technol.*, 35(2):185–195, 2004.

[9] R. Kubo. Generalized cumulant expansion method. *Journal of the Physical Society of Japan*, 17(7):1100–1120, 1962.

[10] T. S. Lundgren. Distribution functions in the statistical theory of turbulence. *Phys. Fluids*, 10(5):969–975, 1967.

[11] A. S. Monin and A. M. Yaglom. *Statistical Fluid Mechanics, Volume II: Mechanics of Turbulence.* Dover, 2007.

[12] F. Moss and P. V. E. McClintock, editors. *Noise in nonlinear dynamical systems. Volume 1: theory of continuous Fokker-Planck systems.* Cambridge Univ. Press, 1995.

[13] S. A. Orszag and L. R. Bissonnette. Dynamical properties of truncated Wiener-Hermite expansions. *Physics of Fluids*, 10(12):2603–2613, 1967.

[14] A. Papoulis. *Probability, random variables and stochastic processes.* McGraw-Hill, third edition, 1991.

[15] S. B. Pope and R. Gadh. Fitting noisy data using cross-validated cubic smoothing splines. *Communications in Statistics - Simulation and Computation*, pages 349–376, 1988.

[16] H. Risken. *The Fokker-Planck equation: methods of solution and applications.* Springer-Verlag, second edition, 1989. Mathematics in science and engineering, vol. 60.

[17] K. Sobczyk. *Stochastic differential equations: with applications to physics and engineering.* Springer, 2001.

[18] J. P. Taylor-King, R. Klages, S. Fedotov, and R. A. Van Gorder. Fractional diffusion equation for an n-dimensional correlated Lévy walk. *Phys. Rev. E*, 94:012104, 2016.

[19] D. Venturi and A. Dektor. Spectral methods for nonlinear functionals and functional differential equations. *Res. Math. Sci.*, 8(27):1–39, 2021.

[20] D. Venturi and G. E. Karniadakis. Convolutionless Nakajima-Zwanzig equations for stochastic analysis in nonlinear dynamical systems. *Proc. R. Soc. A*, 470(2166):1–20, 2014.

[21] D. Venturi, T. P. Sapsis, H. Cho, and G. E. Karniadakis. A computable evolution equation for the joint response-excitation probability density function of stochastic dynamical systems. *Proc. R. Soc. A*, 468(2139):759–783, 2012.

[22] M. Waclawczyk, N. Staffolani, M. Oberlack, A. Rosteck, M. Wilczek, and R. Friedrich. Statistical symmetries of the Lundgren-Monin-Novikov hierarchy. *Phys. Rev. E*, 90:013022(1–11), 2014.

[23] G. Wahba. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Annals of Statistics*, 13(4):1378–1402, 1985.

[24] X. Wan and G. E. Karniadakis. Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM J. Sci. Comput.*, 28(3):901–928, 2006.