eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Signature of nonlinear damping in geometric structure of a nonequilibrium process

Eun-jin Kim

Department of Mathematics and Statistics,

University of Sheffield, Sheffield, S3 7RH, UK

Rainer Hollerbach

Department of Applied Mathematics,

University of Leeds, Leeds LS2 9JT, UK

Abstract

We investigate the effect of nonlinear interaction on the geometric structure of a non-equilibrium process. Specifically, by considering a driven-dissipative system where a stochastic variable x is damped either linearly ($\propto x$) or nonlinearly ($\propto x^3$) while driven by a white noise, we compute the time-dependent probability density functions (PDFs) during the relaxation towards equilibrium from an initial non-equilibrium state. From these PDFs, we quantify the information change by the information length $\mathcal{L}$, which is the total number of statistically distinguishable states which the system passes through from the initial state to the final state. By exploiting different initial PDFs and the strength D of the white noise forcing, we show that for a linear system, $\mathcal{L}$ increases essentially linearly with an initial mean value y_0 of x as $\mathcal{L} \propto y_0$, demonstrating the preservation of a linear geometry. In comparison, in the case of a cubic damping, $\mathcal{L}$ has a power-law scaling as $\mathcal{L} \propto y_0^m$, with the exponent m depending on D and the width of the initial PDF. The rate at which information changes also exhibits a robust power-law scaling with time for the cubic damping.

I. INTRODUCTION

Many important phenomena in nature are stochastic and exhibit seemingly complex temporal behaviour, nevertheless often manifesting a remarkable universal property of the emergence of order (the so-called self-organisation) [1–4]. For a proper understanding of such phenomena, it is essential to utilize a probabilistic methodology such as a (time-dependent) Probability Density Function (PDF). Furthermore, in order to compare different systems, it is invaluable to utilize a measure which is independent of any particular realization of a system. This can very conveniently be achieved by using a geometric measure in a statistical space by assigning a metric between PDFs. There has in fact been a significant interest in defining a metric on probability (e.g. [5–10]) from theoretical and practical considerations. For instance, the Wasserstein metric which provides an exact solution to the Fokker-Planck equation [11] for a gradient flow subject to the minimisation of the energy functional (the sum of entropy and potential energy) [6] has been extensively used in the optimal transport problem [9]. Unlike the Wasserstein metric which has the unit of a physical length, a statistical distance based on the Fisher metric [12, 13] is dimensionless and represents the number of distinguishable states between two PDFs. For example, for a Gaussian distribution, statistically distinguishable states are determined by the standard deviation, which increases with the level of fluctuations; two PDFs which have the same standard deviation and differ in peak positions by less than one standard deviation are statistically indistinguishable. Previously, this fluctuation-based metric has been used mostly in equilibrium or near-equilibrium of classical systems and quite extensively in quantum systems [14–22].

Compared with a metric defined for any given two PDFs, significantly much less work has been done in the case of a time-dependent PDF in non-equilibrium systems. A continuous change in PDFs in this case necessitates defining a distance at any time by comparing two PDFs at times infinitesimally apart and the summation of these distances over time (see Section II). In our recent work [23–26], we proposed information length $\mathcal{L}(t)$ (see §2) as such a metric, which can quantify the total number of different states that the system undergoes in time. This information length was invoked as a new way of mapping out an attractor structure and a useful measure that can link stochastic processes and geometry.

For instance, by considering the relaxation of an initially non-equilibrium state localised around some state $x = y_0$ towards the equilibrium, we showed that in a stable attractor, the information length takes its minimum value for a stable equilibrium point [25] while in a chaotic attractor, it takes its minimum value for a unstable equilibrium point [23]. Interestingly, in a chaotic attractor, the property like the Lyapunov exponent was captured by a sensitive dependence of $\mathcal{L}$ on the initial condition [23]. Furthermore, [26] investigated a geodesic along which a system undergoes the minimum change in $\mathcal{L}$ and demonstrated its utility as an optimal protocol for controlling population [26]. We note that the information length is an extension of the concept of the thermodynamic length [20] to any arbitrary time-dependent PDF (which often does not take the canonical forms) in non-equilibrium systems where thermodynamic state in a strict sense does not exist. An important case of non-equilibrium processes is classical music analyzed in [24] where the information length was calculated by using time-dependent (very intermittent) PDFs that were constructed from the music MIDI file.

The purpose of this paper is to investigate how nonlinear interaction affects the time-scale of information change and geometric structure of an attractor by using information length. In order to isolate the key effect of nonlinear interaction, we consider stochastic driven-dissipative systems with linear and nonlinear damping, respectively, which correspond to the Ornstein-Uhlenbeck (O-U) process and a nonlinear diffusion model with a cubic damping [27–31] and investigate similarities and differences during their relaxation processes in statistical space. Specifically, we quantify the change in time-dependent PDFs during relaxation by using the information length [23–26] and examine the difference in geometric structure associated with the linear O-U process and nonlinear processes. In particular, we demonstrate how the information length $\mathcal{L}$ depends on the (mean) location y_0 of a narrow initial PDF and explore its scaling relation $\mathcal{L} \propto y_0^m$ with the exponent m. Interestingly, m is shown to be 1 for the O-U process regardless of the strength of a stochastic noise (diffusion coefficient) and the width of the initial PDF while m depends on the latter for the cubic process, with a power-law scaling relation.

The remainder of the paper is organized as follows. §II discusses information length and §III introduces our model. §IV and §V present analytical linear results and nonlinear

solutions. §VI and §VII provide the time evolution of the information length and attractor structure. Conclusions are found in §VIII.

II.   INFORMATION LENGTH

We consider a stochastic variable x and suppose that we can compute its time-dependent PDFs $p(x,t)$ either analytically or numerically in the case where its governing equation is known, or construct $p(x,t)$ from experimental/observational data. Defining the information length involves two steps [23–26]: First we need to compute the dynamic time unit $\tau(t)$, which is the characteristic timescale over which $p(x,t)$ temporally changes *on average* at time t. Second, we need to compute the total elapsed time in units of this $\tau(t)$. As done in [23–26], we compute τ by utilising the following second moment $\mathcal{E}$:

$$\mathcal{E} \equiv \frac{1}{[\tau(t)]^2} = \int dx \frac{1}{p(x,t)} \left[\frac{\partial p(x,t)}{\partial t} \right]^2. \tag{1}$$

We note that $\mathcal{E}$ is the root-mean-square fluctuating energy for a Gaussian PDF (see Appendix A and [26]). As defined in Eq. (1), τ has dimensions of time, and quantifies the correlation time over which $p(x,t)$ changes, thereby serving as the time unit in statistical space (see also Appendix B). Alternatively, $1/\tau$ quantifies the (average) rate of change of information in time. We recall that $\tau(t)$ in Eq. (8) is related to the second derivative of the relative entropy (or Kullback-Leibler divergence) [25]. To show this, we consider $p_1 = p(x,t_1)$ and $p_2 = p(x,t_2)$ and the relative entropy $D(p_1,p_2) = \int dx\, p_2 \ln(p_2/p_1)$. To expand $D(p_1,p_2)$ for an infinitesimally small $|t_2 - t_1|$, we compute

$$\frac{\partial}{\partial t_1} D(p_1,p_2) = -\int dx\, p_2 \frac{\partial_{t_1} p_1}{p_1}, \tag{2}$$

$$\frac{\partial^2}{\partial t_1^2} D(p_1,p_2) = \int dx\, p_2 \left[\frac{(\partial_{t_1} p_1)^2}{p_1^2} - \frac{\partial_{t_1}^2 p_1}{p_1} \right], \tag{3}$$

$$\frac{\partial}{\partial t_2} D(p_1,p_2) = \int dx\, [\partial_{t_2} p_2 + \partial_{t_2} p_2(\ln p_2 - \ln p_1)], \tag{4}$$

$$\frac{\partial^2}{\partial t_2^2} D(p_1,p_2) = \int dx\, \left[\partial_{t_2}^2 p_2 + \frac{(\partial_{t_2} p_2)^2}{p_2} + \partial_{t_2}^2 p_2(\ln p_2 - \ln p_1) \right]. \tag{5}$$

By taking the limit where $t_2 \to t_1 = t$ ($p_2 \to p_1 = p$) and by using the total probability conservation (e.g. $\int dx \partial_t p = 0$), Eqs. (2) and (4) above lead to

$$\lim_{t_2 \to t_1 = t} \frac{\partial}{\partial t_1} D(p_1,p_2) = \lim_{t_2 \to t_1 = t} \frac{\partial}{\partial t_2} D(p_1,p_2) = \int dx \partial_t p = 0, \tag{6}$$

4

while Eqs. (3) and (5) give

$$\lim_{t_2 \to t_1 = t} \frac{\partial^2}{\partial t_1^2} D(p_1, p_2) = \lim_{t_2 \to t_1 = t} \frac{\partial^2}{\partial t_2^2} D(p_1, p_2) = \int dx \frac{(\partial_t p)^2}{p} = \frac{1}{\tau(t)^2}. \tag{7}$$

See also [20] for similar derivation. Thus, the second derivative of the relative entropy gives $\mathcal{E}$, the inverse of the square of the characteristic time over which PDF changes in time.

The total accumulated change in information between the initial and final times, 0 and t respectively, is defined by measuring the total elapsed time in units of τ as:

$$\mathcal{L}(t) = \int_0^t \frac{dt_1}{\tau(t_1)} = \int_0^t dt_1 \sqrt{\int dx \frac{1}{p(x, t_1)} \left[\frac{\partial p(x, t_1)}{\partial t_1} \right]^2}. \tag{8}$$

To relate Eq. (8) to the relative entropy, we expand $D(p_1, p_2)$ for small $dt = t_2 - t_1$ by using Eqs. (6)-(7) and $D(p_1, p_1) = 0$ as

$$D(p_1, p_2) = \frac{1}{2} \left[\int dx \frac{(\partial_{t_1} p(x, t_1))^2}{p(x, t_1)} \right] (dt)^2 + O((dt)^3), \tag{9}$$

where $O((dt)^3)$ is higher order term in dt. We can then define the infinitesimal distance $dl(t_1)$ between t_1 and $t_1 + dt$ by

$$dl(t_1) = \sqrt{D(p_1, p_2)} = \frac{1}{\sqrt{2}} \sqrt{\int dx \frac{(\partial_t p(x, t_1))^2}{p(x, t_1)}} dt + O((dt)^{3/2}). \tag{10}$$

We sum $dt(t_1)$ at different times $t_1 = 0, dt, ...t - dt$ by using Eq. (10) and then take the limit of $dt \to 0$ as

$$
\begin{aligned}
l(t) &= \lim_{dt \to 0} \left[dl(0) + dl(dt) + dl(2dt) + dl(3dt) + \cdots dl(t - dt) \right] \\
&= \lim_{dt \to 0} \left[\sqrt{D(p(x, 0), p(x, dt))} + \sqrt{D(p(x, dt), p(x, 2dt))} + \cdots \sqrt{D(p(x, t - dt), p(x, t))} \right] \\
&\propto \int_0^t dt_1 \sqrt{\int dx \frac{(\partial_{t_1} p(x, t_1))^2}{p(x, t_1)}} = \mathcal{L}(t).
\end{aligned}
\tag{11}
$$

Thus, the sum of relative entropy calculated at times infinitesimally apart is the same as $\mathcal{L}$ up to a numerical factor. It is important to note that Eq. (11) or $\mathcal{L}$ depends not only on initial $p(x, 0)$ and final PDF $p(x, t)$, but also on a particular path that a system takes. Thus, in general, $l(t)^2$ in Eq. (11) is not simply proportional to the relative entropy $D(p(x, 0), p(x, t))$, which depends only on $p(x, 0)$ and $p(x, t)$. See Appendix C for an example. That is, the relative entropy does not uniquely determine $\mathcal{L}$ as it can take the

same value for two different paths.

Eq. (8) provides the total number of different states that a system passes through from the initial state with the PDF $p(x, t = 0)$ at time $t = 0$ to the final state with the PDF $p(x, t)$ at time t, establishing a distance between the initial and final PDFs in the statistical space. For example, in equilibrium where $\frac{\partial p}{\partial t} = 0$, $\mathcal{E} = 0$ and hence $\tau(t_1) \to \infty$ for all time t_1. Measuring dt_1 in units of this infinite $\tau(t_1)$ at any t_1, $dt_1/\tau(t_1) = 0$ in Eq. (8), and thus $\int_0^t dt_1/\tau(t_1) = 0$. This can be viewed as that in statistical space there is no flow of time in equilibrium. In the opposite limit, large $\mathcal{E}$ corresponds to small τ, meaning that information changes very quickly in dimensional time.

III.   MODEL

The particular model that we will explore using these information length ideas is the following Langevin equation for over-damped oscillators:

$$\frac{dx}{dt} = F(x) + \xi. \tag{12}$$

Here, x is a random variable of interest, and ξ is a white noise with a short correlation time with the following property:

$$\langle \xi(t)\xi(t') \rangle = 2D\delta(t - t'). \tag{13}$$

Here, D is the strength of the forcing. We can easily check that the dimension of D is length^2/time by using that the dimensions of ξ and $\delta(t - t')$ are length/time and 1/time, respectively. $F(x)$ is a deterministic force, which can be interpreted as the gradient of the potential $U(x)$ as $F(x) = -\frac{\partial U(x)}{\partial x}$. We compare the linear $F = -\gamma x$ ($U = \frac{\gamma}{2}x^2$) and the cubic $F = -\mu x^3$ ($U = \frac{\mu}{4}x^4$), where γ and μ have dimensions of 1/time and 1/(time $\times$ length^2), respectively. The linear system is the familiar Ornstein-Uhlenbeck (O-U) process, which has been widely used as a model for a noisy relaxation system in many areas of physical science and financial mathematics (e.g. [32]). Numerically, instead of solving Eq. (12) directly, we will consider the equivalent Fokker-Planck equation [11]

$$\frac{\partial}{\partial t}p(x, t) = \frac{\partial}{\partial x}\left[-F(x) + D\frac{\partial}{\partial x}\right]p(x, t). \tag{14}$$

6

IV. ANALYTIC LINEAR RESULTS

In this section, we provide main results for the (linear) O-U process, where we find exact analytic expressions for all quantities of interest. If the initial PDF is taken as Gaussian with the inverse temperature β_0 as

$$p(x_0, 0) = \sqrt{\frac{\beta_0}{\pi}} \exp[-\beta_0(x_0 - y_0)^2], \tag{15}$$

then the solution at any later time is [25, 26]

$$p(x, t) = \sqrt{\frac{\beta(t)}{\pi}} \exp[-\beta(t)(x - y(t))^2], \tag{16}$$

where

$$y(t) = y_0 e^{-\gamma t}, \tag{17}$$

$$\frac{1}{\beta(t)} = \frac{1}{\beta_1(t)} + \frac{e^{-2\gamma t}}{\beta_0}, \tag{18}$$

$$\frac{1}{\beta_1(t)} = \frac{2D(1 - e^{-2\gamma t})}{\gamma}. \tag{19}$$

Here $y(t) = \langle x(t) \rangle$ is the mean position of the Gaussian profile, and y_0 is its initial value. Similarly, $\beta(t)$ is the inverse temperature, and β_0 is its initial value. As t tends to infinity, $y(t) \to 0$ and $\beta(t) \to \frac{\gamma}{2D} \equiv \beta_*$. To compare initial and final equilibrium states, it is convenient also to introduce $D_0 = \frac{\gamma}{2\beta_0}$. The variance at $t = 0$ and $t \to \infty$ is then given by $\langle (x_0 - y_0)^2 \rangle = \frac{1}{2\beta_0} = \frac{D_0}{\gamma}$ and $\langle x^2 \rangle = \frac{1}{2\beta_*} = \frac{D}{\gamma}$, respectively. We note from Eqs. (18)-(19) that when $D = D_0$, $\beta(t) = \beta_0 = \frac{\gamma}{2D}$ for all time. In this case, the Gaussian simply moves from y_0 to 0 without changing its shape at all. If D is greater (lesser) than D_0, it also broadens (narrows) as it moves.

Given Eqs. (16)-(19), one can compute Eq. (1) by carrying out the analysis in Appendix D as follows:

$$\mathcal{E} = \frac{1}{\tau^2} = \frac{1}{2\beta^2}\left(\frac{d\beta}{dt}\right)^2 + 2\beta\left(\frac{dy}{dt}\right)^2 = \frac{2\gamma^2}{T^2}(r^2 + qT). \tag{20}$$

In Eq. (20), $q = \beta_0 \gamma y_0^2$, $r = 2\beta_0 D - \gamma$, and $T = 2\beta_0 D(e^{2\gamma t} - 1) + \gamma$, following the same notation as in [25]. Note that q is due to the difference in y_0 and $y(t \to \infty)$ while r is due to the difference in D_0 and D. Thus, the first term in $\mathcal{E}$ involving r represents the information

7

change due to the change in PDF width when $D_0 \neq 0$ while the second term is due to the movement of the PDF (or the mean value of x). Recalling $D_0 = \frac{\gamma}{2\beta_0}$, we can recast r, q and T in Eq. (20) as

$$q = \frac{\gamma^2 y_0^2}{2D_0}, \quad r = \gamma \left(\frac{D}{D_0} - 1 \right), \quad T = \gamma \left[\frac{D}{D_0} (e^{2\gamma t} - 1) + 1 \right]. \tag{21}$$

From Eq. (21), we can see that the dimension of q, r and T is the inverse of time. Thus, $\mathcal{E}$ and subsequently $\mathcal{L}$ are invariant under the rescaling $\gamma \to \alpha^2 \gamma$, $D_0 \to \alpha D_0$, $D \to \alpha D$ and $t \to \alpha^{-2} t$. In particular, in the long time limit $t \to \infty$, $\mathcal{L}(t \to \infty)$ becomes invariant under the rescaling $\gamma \to \alpha^2 \gamma$, $D_0 \to \alpha D_0$, $D \to \alpha D$. From Eqs. (20) and (8), we show in Appendix E that for $r \neq 0$

$$\mathcal{L} = \frac{1}{\sqrt{2}} \left[\ln \left(\frac{Y-r}{y+r} \right) \right]_{Y_i}^{Y_f} + \frac{\sqrt{2}}{r} H. \tag{22}$$

Here T_i and T_f are T evaluated at t_i and t_f respectively; Y_i and Y_f are $Y = \sqrt{r^2 + qT}$ evaluated at T_i and T_f, and H is defined as

$$H = \begin{cases} \sqrt{qr - r^2} \tan^{-1} \left(\frac{Y}{\sqrt{qr-r^2}} \right) & \text{if } qr - r^2 > 0, \\ -\frac{\sqrt{r^2-rq}}{2} \ln \left(\frac{Y-\sqrt{r^2-rq}}{Y+\sqrt{r^2-rq}} \right) & \text{if } qr - r^2 < 0. \end{cases} \tag{23}$$

In Eq. (22), the contribution from the difference in PDF width through $r \neq 0$ and that from the difference in mean value of x (e.g. PDF peaks) through $q \neq 0$ appear in both first and second terms. Thus, in order to separate their effects, it is simpler to use Eq. (20), take the limit of $q = 0$, and calculate Eq. (8)

$$\mathcal{L} = \frac{1}{\sqrt{2}} \int_{T_i}^{T_f} \left\{ \frac{1}{T} \frac{1}{T+r} |r| \right\} dT = \frac{1}{\sqrt{2}} \frac{|r|}{r} \ln \left[\frac{T}{T+r} \right]. \tag{24}$$

Note that as a metric, $\mathcal{L}$ is a non-negative quantity. Eq. (24) is the information length solely due to the change in the width of PDFs. To simplify Eq. (24), we use $T + r = 2\beta_0 D e^{2\gamma t}$ and $\beta(t) = \frac{\gamma \beta_0 e^{2\gamma t}}{T}$ (Eq. (D4)) to obtain

$$\frac{T+r}{T} = \beta(t) \frac{2D}{\gamma}. \tag{25}$$

Using Eq. (25) in Eq. (24) with $t_0 = 0$ and $t_f = t$ and $\beta(t = 0) = \beta_0$ gives us

$$\mathcal{L} = \frac{1}{\sqrt{2}} \left| \ln \frac{\beta(t)}{\beta_0} \right|. \tag{26}$$

We note that Eq. (26) can directly be computed from the first term in Eq. (20). Then, by calculating the differential entropy $S(t) = - \int dx \, p(x,t) \ln p(x,t) = \frac{1}{2} \left[1 + \ln \frac{\pi}{\beta(t)} \right]$ (with the Boltzmann constant $K_B = 1$) for $p(x,t)$ given in Eq. (16), we obtain the following entropy difference:

$$S(t) - S(0) = \frac{1}{2} \ln \frac{\beta_0}{\beta(t)}. \tag{27}$$

Thus, $\mathcal{L}$ in Eq. (26) solely due to the change in PDF width is the same as the magnitude of the change in entropy in Eq. (27) up to a constant numerical factor. In Appendix C, the relative entropy between the initial and final PDF is shown to take the form different from Eqs. (26)-(27).

In the opposite case of $r = 0$ where the initial and final PDFs have the same width, $\beta(t) = \beta_0$ for all time, and Eqs. (20) and (8) give us

$$\mathcal{L} = \frac{1}{\sqrt{2}} \int_{T_i}^{T_f} \frac{\sqrt{q}}{T^{\frac{3}{2}}} dT = -\sqrt{2q} \left[\frac{1}{\sqrt{T}} \right]_{T_i}^{T_f}. \tag{28}$$

We use that for $r = 0$, $T = \gamma e^{2\gamma t}$, $T_i = \gamma$ at $t = 0$ and simplify Eq. (28) as

$$\mathcal{L}(t) = \frac{\sqrt{\gamma} y_0}{\sqrt{D}} \left[1 - e^{-\gamma t} \right] = \frac{1}{\sqrt{D/\gamma}} \left[y_0 - y \right], \tag{29}$$

where $y = y_0 e^{-\gamma t} = \langle x \rangle$ is the mean position. Thus, $\mathcal{L}$ in Eq. (29) is the change in the mean position $y_0 - y$ between initial and time t measured in unit of the resolution $\sqrt{\frac{D}{\gamma}}$. Interestingly, this resolution $\sqrt{\frac{D}{\gamma}}$ is the standard deviation, which is the square root of the variance $\langle (x - \langle x \rangle)^2 \rangle = \frac{D}{\gamma} = \frac{1}{2\beta} = \frac{1}{2\beta_0}$. In general when $q \neq 0$ and $r \neq 0$, $\mathcal{L}$ results from the mixed contribution from the entropy change ($r \neq 0$) and the change in y ($q \neq 0$) measured in unit of the resolution. In a more technical term, β and y in Eq. (20) constitute hyperbolic geometry upon a suitable change of variables (e.g. see [26]).

V. NONLINEAR SOLUTION

For the cubic system, exact numerical solutions together with approximate analytical solutions were reported in [31]. One of the interesting results is that starting from a narrow PDF centred about y_0, a rapid initial evolution of the PDF is dominated by the O-U process with the effective friction coefficient γ_e given by

$$\gamma_e \sim \zeta \mu \langle x \rangle^2, \tag{30}$$

9

where ζ is an $O(1)$ constant, and $\langle x \rangle = y_0/\sqrt{1 + 2\mu y_0^2 t}$. This thus gives $\beta_1(t)$ in Eq. (19) as follows:

$$\frac{1}{\beta_1(t)} = \frac{2D(1 - e^{-2\gamma_e t})}{\gamma_e}, \tag{31}$$

with γ_e given by Eq. (30). This will be utilised below in understanding exact numerical results.

For a numerical solution of Eq. (14), we begin by noting that without any loss of generality any finite interval in x can always be rescaled to $x \in [-1, 1]$. If the initial condition is also restricted away from the boundaries, then solving (14) on this finite interval (with boundary conditions $p = 0$ at $x = \pm 1$) is an excellent match to an unbounded interval. By rescaling t and D, we can similarly fix $\mu = 1$, thereby reducing the number of parameters that need to be varied numerically. The numerical procedure then involves second-order finite-differencing in both space and time, using $O(10^6)$ grid-points in x, and time-steps as small as $O(10^{-7})$.

Starting from the same initial condition as before,

$$p(x_0, 0) = \frac{1}{\sqrt{2D_0\pi}} \exp[-(x_0 - y_0)^2/2D_0], \tag{32}$$

we numerically solve for $p(x, t)$ at later times, and evaluate $\mathcal{E}$ and $\mathcal{L}$. The system was solved for D and D_0 in the range 10^{-3} to 10^{-7}, and $y_0 \in [0, 0.75]$. In total 25 combinations of D and D_0 were considered, with ~ 20 y_0 values for each. In the next section, we present the resulting $\mathcal{E}$ and $\mathcal{L}$ and compare with the equivalent $\gamma = 1$ linear results (obtained either analytically, or numerically as a useful check of the code).

VI.   TIME EVOLUTION OF $\mathcal{E}$ AND $\mathcal{L}$

Fig. 1 shows the results for $\mathcal{E}$ when starting with a very narrow peak ($D_0 = 10^{-8}$) that is very far from the origin ($y_0 = 0.7$). $D = 10^{-3}$, 10^{-5} and 10^{-7}, and the two cases linear and cubic are considered. Starting with the behaviour for small times ($t \leq 10^{-4}$), called stage (i), there are two features that stand out. First, the two $D = 10^{-3}$ cases are far above

10

$D = 10^{-5}$ and 10^{-7}, and linear and cubic are the same. Second, for $D = 10^{-5}$ and 10^{-7} the two different D values follow the same curves, but the linear and cubic cases are now different, with linear being approximately four times greater than cubic. Also, at least for these early times, these four curves are all essentially independent of t.

To understand these results, we recall that $\mathcal{E}$ is a measure of $(\frac{\partial p}{\partial t})^2$, which in turn consists of two parts, the movement of the PDF (advection) by the damping force $F(x)$ and the change in width of the PDF due to diffusion $D > D_0$. Since F is different in the linear and cubic processes while D is the same, $\mathcal{E}$ would evolve similarly for both processes if dominated by diffusion (diffusion-dominated) while evolving differently if dominated by advection (advection-dominated) due to the damping force. We need to combine this knowledge with the fact that the initial evolution of $\mathcal{E}$ in stage (i) is dominated for small D by advection while for large D by diffusion. To show this in the linear case, we examine Eq. (20) at $t = 0$:

$$\mathcal{E} = 2\gamma^2 \left(\frac{D}{D_0} - 1 \right)^2 + \frac{\gamma^3 y_0^2}{D_0}, \tag{33}$$

where the first and second terms represent the effect of the diffusion and advection, respectively. Inserting $\gamma = 1$, $D_0 = 10^{-8}$ and $y_0 = 0.7$, $D = 10^{-3}$ yields $\mathcal{E} = 2 \cdot 10^{10}$, whereas $D = 10^{-5}$ and 10^{-7} both yield $\mathcal{E} = 5 \cdot 10^7$, as in Fig. 1. Thus, in stage (i), $\mathcal{E}$ exhibits
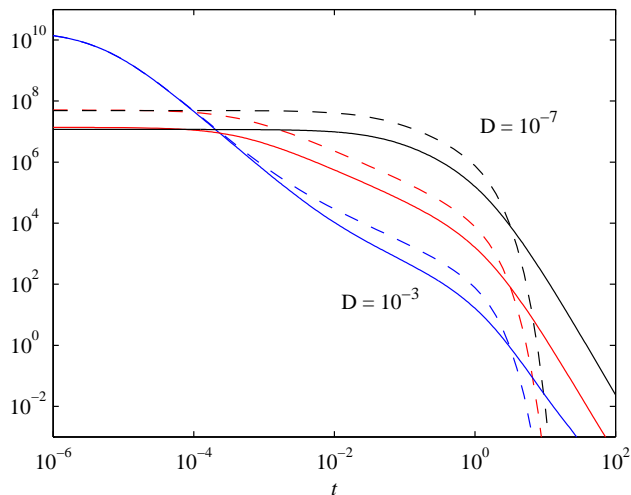

FIG. 1: $\mathcal{E}$ as functions of time t for the linear (dashed lines) and cubic (solid lines) processes. $D = 10^{-3}$, 10^{-5} and 10^{-7}, as labelled. The initial condition in each case was $D_0 = 10^{-8}$, $y_0 = 0.7$.

11

the transition from diffusion-dominated to advection-dominated as D is reduced. Similar conclusion can be obtained for the cubic case by replacing in Eq. (33) by $\gamma_e \sim \zeta\mu\langle x\rangle^2$ (Eq. (30)). The transition point, where the two terms in Eq. (33) are comparable, occurs when $D \sim y_0\sqrt{D_0/2} = 5 \cdot 10^{-5}$.

This predicted transition from advection-dominated to diffusion-dominated occurring around $D \sim y_0\sqrt{D_0/2} = 5 \cdot 10^{-5}$ is indeed observed in Fig. 1. Specifically, for $D = 10^{-3}$, $\mathcal{E}$ is dominated by diffusion and takes the same (large) value for linear and cubic. In contrast, for $D = 10^{-5}$ and 10^{-7}, $\mathcal{E}$ is dominated by the advection, and different evolutions are observed in linear and cubic processes. We can even understand why the linear curves are above the cubic curves by this factor of four: If the positions of the peaks are expected to evolve as $y_0 e^{-t}$ and $y_0/\sqrt{1 + 2y_0^2 t}$ in the two cases (setting $\gamma = \mu = 1$ in the general formulas), then the speeds at which they initially move are y_0 and y_0^3 respectively (obtained by evaluating $|\frac{d}{dt}\langle x\rangle|$ at $t = 0$ in the two cases). For $y_0 = 0.7$ the linear peak thus moves roughly twice as fast as the cubic peak, hence a factor of four in $\mathcal{E}$. Finally, the reason these curves remain independent of time up to $t \approx 10^{-4}$ is that the speeds of the peaks are essentially unchanged up to that time with constant $F(x) \sim -\gamma x_0$ and $-\mu x_0^3$ for linear and cubic, respectively; diffusion is also not yet playing an important role and $\beta(t) = \beta_0$.

For the linear case, for somewhat larger times in stage (ii), up to $t < O(1)$, $\mathcal{E}$ exhibits a power-law decrease in time. This can also be inferred from Eq. (20) by keeping the first-order correction $T \sim \gamma\left[\frac{D}{D_0}2\gamma t + 1\right] \sim \gamma\left[\frac{D}{D_0}2\gamma t\right]$ for $\frac{D_0}{2\gamma D} \ll t \ll 1$ (recall $D \gg D_0$):

$$\mathcal{E} \sim 2\gamma^2\left[\frac{1}{2t^2} + \frac{y_0^2}{Dt}\right]. \tag{34}$$

The second term in Eq. (34) is due to the peak with the variance $\langle(x - \langle x\rangle)^2\rangle \propto= 1/2\beta \propto t$ (see Eq. (17)) and $F(x) = -\gamma x \sim -\gamma y_0$ for $t < O(1)$, which gives $\mathcal{E} \propto t^{-1}$. In comparison, the final stage (iii) is due to the adjustment to the stationary PDF. This involves an exponential decrease in $\mathcal{E}$ since

$$T \sim \frac{\gamma D e^{2\gamma t}}{D_0} \ll T^2,$$

as $t \to \infty$, and thus

$$\mathcal{E} = \frac{2\gamma^2}{T^2}(r^2 + qT) \sim \frac{2\gamma^2 q}{T} \propto \frac{1}{T} \propto e^{-2\gamma t} \to 0,$$

exponentially decreasing in time as $t \to \infty$. This is physically due to the exponential decrease in peak position $y = y_0 e^{-\gamma t}$ while $\beta \sim \frac{\gamma}{2D}$. This last stage occurs around $t \approx O(1)$, independent of D.

To summarize the O-U process, for a sufficiently small $D < y_0 \sqrt{\frac{D_0}{2}}$, the relaxation of the O-U process undergoes three scaling regimes of $\mathcal{E}$ with t: i) constant, ii) power-law, and iii) exponential. The stage i) is due to the movement of the PDF; the stage ii) is due to the diffusion with $1/\beta \propto \langle (x - \langle x \rangle)^2 \rangle$ (see Eq. (17)) [e.g. due to the Brownian motion where the rms displacement increase as $t^{1/2}$]; the stage iii) is due to the exponential adjustment of the peak position as $y = y_0 e^{-\gamma t}$ in settling into the equilibrium PDF. These scalings and leading contribution from $F(x)$ and β responsible for such scalings are summarised in Table 1.

Since $\tau = \mathcal{E}^{-1/2}$ is the time unit or correlation time (over which the physical time is to be measured), our results imply three stages of i) constant, ii) power-law and iii) exponential scalings of the time unit τ. Furthermore, in the O-U process, the final stage starts at $t = O(1)$, the same for all D, suggesting the independence of the relaxation time on D. Alternately, this can be viewed as the independence of x and t in linear processes since D only affects x (dependence of PDFs).

Compared to the O-U process, the time evolution of $\mathcal{E}$ for the cubic process occurs over

TABLE I: Scalings of $\mathcal{E}$ in stages (i), (ii) and (iii) for advection-dominated case for sufficiently small $D < y_0 \sqrt{\frac{D_0}{2}}$ ($D_0 = 10^{-8}$) and physical origins of such scaling behaviour $(F(x)$ and $\beta(t))$; $3 < n < 4$.

Process	scaling/contribution	stage (i)	stage (ii)	stage (iii)
	$\mathcal{E}$	constant	t^{-1}	$e^{-\gamma t}$
Linear	$F(x)$	$-\gamma x_0$	$-\gamma x_0$	$\propto e^{-\gamma t}$
	β	β_0	$\propto t^{-1}$	$\frac{\gamma}{2D}$
	$\mathcal{E}$	constant	t^{-1}	t^{-n}
cubic	$F(x)$	$-\mu x_0^3$	$-\mu x_0^3$	$\propto t^{-1}$
	β	β_0	$\propto t^{-1}$	—

13

a much longer timescale, as seen in Fig. 1. This is due to the fact that with a cubic nonlinear damping, the equilibration of a PDF to the final equilibrium quartic exponential PDF requires the time $t \gg t_c$ where [31]

$$t_c \sim \sqrt{\frac{1}{D\mu}}. \tag{35}$$

As $t_c \propto D^{-1/2}$, the relaxation time becomes longer for smaller D, as previously noted also by [31]. To understand the evolution of $\mathcal{E}(t)$, it is useful to utilise the effective γ_e in Eq. (30). Specifically, at small and intermediate times, γ_e is almost independent of t as $\gamma_e \propto \mu y_0^2$, and thus the behaviour of $\mathcal{E}$ for the cubic process is quite similar to that of the O-U process. For stage (iii), the prediction based on Eq. (20) becomes questionable due to large fluctuations. It suffices for the purpose of this paper to conclude from Fig. 1 that $\mathcal{E}$ in stage (iii) follows power-law as $\mathcal{E} \propto t^{-n}$ ($3 < n < 4$). To summarize, for a sufficiently small D, the relaxation of the cubic process undergoes three scaling regimes of $\mathcal{E}$ with t: i) constant, ii) power-law, and iii) power-law. The stage i) is due to the movement of the PDF, similarly to linear case; the stage ii) is due to the diffusion, similar to linear case. The last stage with the power-law scaling is different from the exponential scalings in the O-U process. The scalings are summarised in Table I together with leading behaviour of $F(x)$ and β.

Our results demonstrate that nonlinear interaction promotes power-law scalings of statistical measures $\mathcal{E}$ (τ) with respect to time. Making an analogy to power-law scaling often observed in self-organising system which ensures scale-invariance, we speculate that power-law scale of statistical measures may also be induced in self-organising systems through non-linear interaction. This issue will need to be explored further in future. Furthermore, compared with the O-U process, nonlinear interaction in the cubic process results in $\mathcal{E}$ which varies much less rapidly. [That is to say, a power-law evolves much slower than exponential.] Recalling that a geodesic is a particular path with a constant $\mathcal{E}$ along the path [26] which minimizes the total $\mathcal{L}$ between given two times, we infer that the cubic process follows a path which is closer to a geodesic compared to the O-U process. Thus, we expect a smaller $\mathcal{L}$ in the cubic process than in the O-U process, and this will shortly be shown to be observed in our numerical results.

Furthermore, in comparison with the linear case where the relaxation time to the
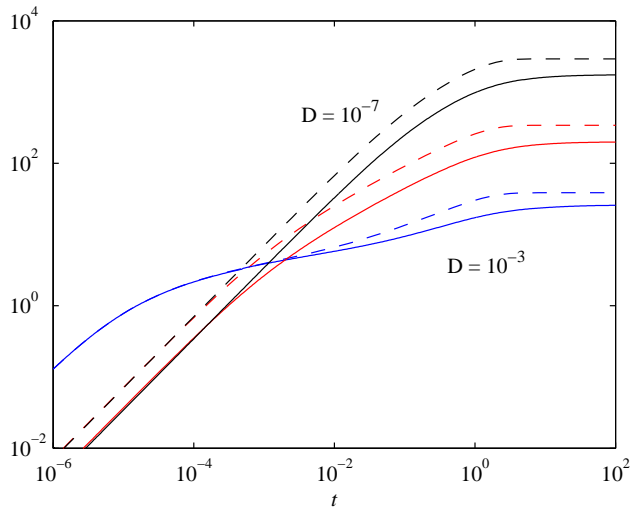
14

FIG. 2: $\mathcal{L}$ as functions of time t for the linear (dashed lines) and cubic (solid lines) processes. All parameter values as in Fig. 1.

equilibrium is independent of D, the dependence of t_c in Eq. (35) on D reflects that the diffusion affects not only x but also key transition timescale (e.g. t_c in Eq. (35)), implying a close link between x and t through non-linear interaction. We note that [31] showed that the cubic system can be linearised by introducing a nonlinear time which depends on x, which is most likely why t_c is affected by D (i.e. through x which depends on D).

Finally, Fig. 2 shows $\mathcal{L}$ for the six cases corresponding to Fig. 1. Since $\mathcal{E}$ in Fig. 1 monotonically decreases in time, the largest contribution to $\mathcal{L}$ comes from $\mathcal{E}$ at small times. The most prominent difference between the O-U and cubic processes is that the relaxation time t_c to converge to the stationary state is much longer for the cubic process and depends on D. Furthermore, $\mathcal{L}$ tends to be smaller for the cubic process, confirming our expectation above.

VII. ATTRACTOR STRUCTURE: $\mathcal{L}$ VS y_0

In the absence of a stochastic forcing, a system with either linear or cubic damping has one stable equilibrium point $x = 0$, to which all initial positions approach in the long time limit. The proximity of different x to the equilibrium point $x = 0$ can be quantified by the difference in the potential $V(x)$ ($= \frac{\gamma}{2}x^2$ and $\frac{\mu}{4}x^4$ for the linear and cubic processes,

15

respectively), or its gradient $F(x)$. In the presence of the stochastic forcing ξ, any initial value of x always tends to approach $x = 0$ for sufficiently large time, and fluctuates around it, forming an equilibrium distribution. In this case, $V(x)$ from the deterministic force does not provide an accurate measure of the difference between different initial points due to ξ.

Motivated by this, [23] considered the relaxation of an initial non-equilibrium state strongly localised around $x = y_0$ [i.e. modelled by $p(x, 0) \propto \delta(x - y_0)$] into the final equilibrium state around $x = 0$ and defined the distance between the point y_0 and $x = 0$ by the total $\mathcal{L}$ between the initial localised PDF and the final equilibrium PDF. This $\mathcal{L}$ provides a metric which quantifies the distance between $x = y_0$ and the equilibrium, serving as a useful measure to differentiate different x's in view of the proximity to the equilibrium point $x = 0$ [35]. As $\mathcal{L}$ measures different states along a path that a system passes through, it can be viewed as a 'Lagrangian/dynamic' measure of a metric. In general, when an initial PDF has a finite width [24–26], the total $\mathcal{L}$ between an initial PDF with the mean value y_0 and final equilibrium PDF was used as the distance between y_0 (mean value of x at $t = 0$) and $x = 0$ (mean value of x at $t \to \infty$). This metric consequently depends on both the strength D of the stochastic noise (which determines the width of the final equilibrium PDF) and the width of the initial PDF.

In order to elucidate the effect of nonlinear interaction on the geometric structure, we now present how this metric depends on y_0 for different D and D_0 for the O-U and cubic processes. Fig. 3 shows the results of the total $\mathcal{L}$ (in the limit $t \to \infty$) as a function of y_0, for D_0 and D equal to 10^{-3}, 10^{-5} and 10^{-7}. Focusing on the linear case first, the dependence on y_0 is clearly linear, except for small regions near $y_0 = 0$, where a sufficiently large mismatch between D and D_0 yields results dominated by diffusion rather than movement of the peak from y_0 to 0. Table 1 summarizes the slopes of these straight lines (including also additional D and D_0 values). For the simplest $D = D_0$ cases, where Fig. 3 indicates an exactly linear relationship for all y_0, the slopes clearly scale as $D^{-\frac{1}{2}}$. Above the diagonal in Fig. 3 or Table 1 ($D < D_0$) yields a greater slope than below the diagonal ($D > D_0$).

To understand these results, we examine Eqs. (22), (23) and (28). When $y_0 \neq 0$, Eqs. (22) and (23) imply that $\mathcal{L}$ in general has a complex dependence on y_0, D_0 and D. Some
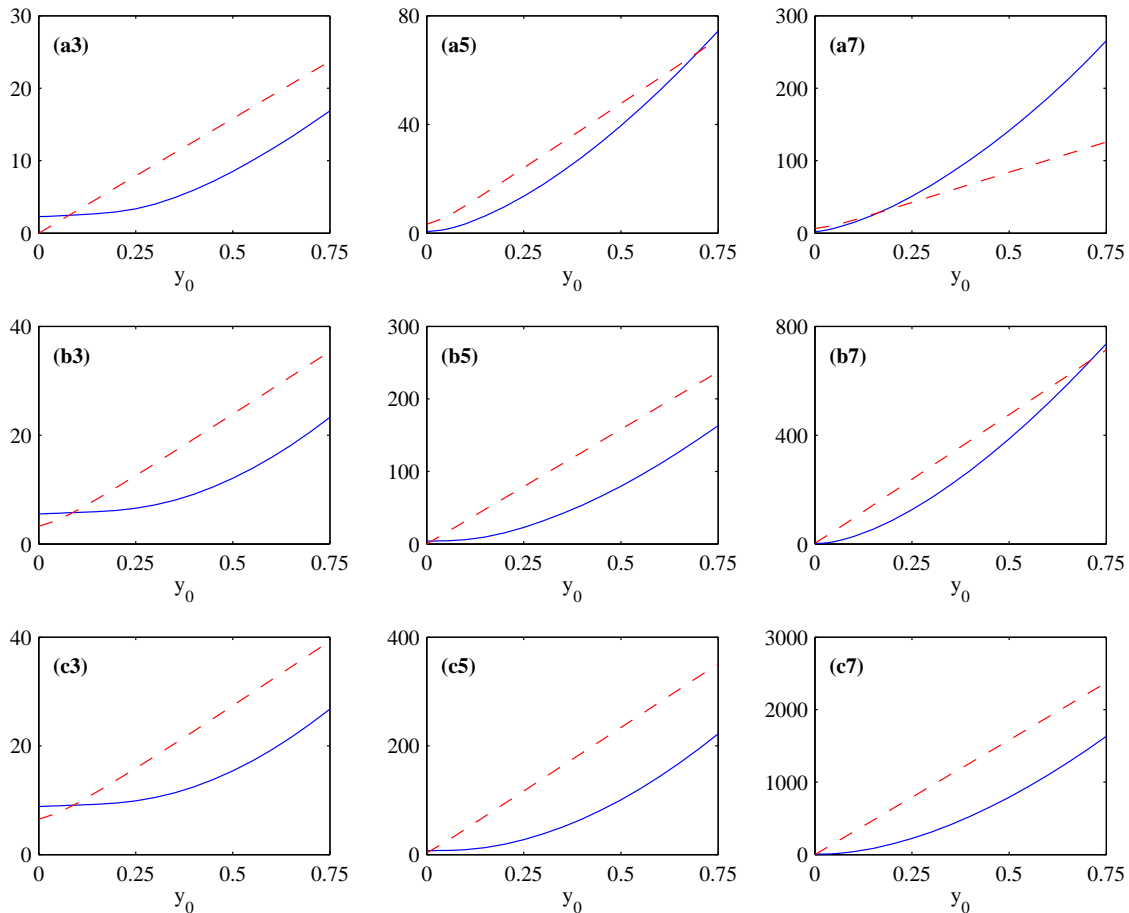
FIG. 3: $\mathcal{L}$ as functions of y_0 for the linear (dashed lines) and cubic (solid lines) processes. The nine panels (a3)-(c7) are labelled such that rows (a,b,c) correspond to $D_0 = 10^{-3}$, 10^{-5} and 10^{-7} respectively, and columns (3,5,7) correspond to $D = 10^{-3}$, 10^{-5} and 10^{-7} respectively.

simple scaling relations are however obtained when $D = D_0$, or when y_0 is sufficiently large. First, when $D = D_0$, $r = 0$; so using $T_i = \gamma$ and $T_f \to \infty$ (since $t_f \to \infty$), $q = \beta_0 \gamma y_0^2$ and $2\beta_0 = \gamma/D_0$ in Eq. (28) gives us

$$\mathcal{L} = \sqrt{\frac{2q}{\gamma}} = \frac{\sqrt{\gamma} y_0}{\sqrt{D_0}}. \tag{36}$$

Thus, when $D = D_0$, $\mathcal{L}$ has an exact linear scaling with y_0, with slope $\sqrt{\frac{\gamma}{D}}$, as seen also in Table 1 (where $\gamma = 1$).

Second, for a sufficiently large y_0 such that $q \gg 1$, a clear linear relation between $\mathcal{L}$ and y_0 is obtained, with different slopes for $D > D_0$ and $D < D_0$. When $D > D_0$ and $q \gg 1$

$(0 < r < q)$, the leading order contribution to $\mathcal{L}$ comes from H in Eq. (23) as (see Appendix F for details)

$$\left[\tan^{-1}\left(\frac{Y}{\sqrt{qr - r^2}}\right)\right]_{Y_i}^{Y_f} \sim \frac{\pi}{2} - \frac{1}{\sqrt{r_0}},$$

where $r_0 = \frac{D}{D_0} - 1$ and thus (see again Appendix F),

$$\mathcal{L} \sim \left(\frac{\pi}{2} - \sqrt{\frac{D_0}{D}}\right) \frac{\sqrt{\gamma}y_0}{\sqrt{D - D_0}} \sim \left(\frac{\pi}{2} - \sqrt{\frac{D_0}{D}}\right) \frac{\sqrt{\gamma}y_0}{\sqrt{D}}, \tag{37}$$

where $r \sim \frac{D}{D_0}$ is used for $D \gg D_0$. Thus, when $D > D_0$, $\mathcal{L}$ is determined by measuring the change in the mean position y_0 in units of $\sqrt{D}$ to leading order, and takes its maximum value $\frac{\pi\sqrt{\gamma}}{2}\frac{y_0}{\sqrt{D}}$ for a very narrow initial distribution (as $\frac{D_0}{D} \to 0$). These scalings can also be confirmed above the diagonals in Table 1.

When $D < D_0$ and $q \gg 1$ ($r < 0$), the leading order contribution to $\mathcal{L}$ comes from H in Eq. (23) as

$$\left[\ln\left(\frac{Y - \sqrt{r^2 - rq}}{Y + \sqrt{r^2 - rq}}\right)\right]_{Y_i}^{Y_f} \sim \ln\frac{2}{1 - \sqrt{|r_0|}} \sim \ln\frac{4D_0}{D},$$

where $|r_0| = 1 - \frac{D}{D_0}$ and thus,

$$\mathcal{L} \sim \frac{\sqrt{\gamma}y_0}{2\sqrt{D_0}} \ln\frac{4D_0}{D} \sim \left(\ln 2 + \ln\sqrt{\frac{D_0}{D}}\right) \frac{\sqrt{\gamma}y_0}{\sqrt{D_0}}. \tag{38}$$

Thus, when $D_0 > D$, $\mathcal{L}$ is given by y_0 measured in units of $\sqrt{D_0}/\ln\sqrt{\frac{D}{D_0}}$, increasing as $\frac{D_0}{D} \to \infty$. It is interesting to see the logarithmic correction factor $\ln\sqrt{\frac{D_0}{D}}$ to y_0 measured in units of $\sqrt{D_0}$, which is due to the narrowing of the PDF. Again, these values are quite close to the exact results in Table 1. We have checked similar results with $\mathcal{L} \propto y_0$ for different initial PDFs (quartic exponential PDFs).

In sharp contrast to the linear case, Fig. 3 shows that for the cubic case, $\mathcal{L}$ is clearly not linearly dependent on y_0. However, plotting the same data on a log-log scale, Fig. 4 shows that for sufficiently large y_0 clear power-law scalings emerge. For $y_0 < O(D^{1/4})$ $\mathcal{L}$ is dominated by diffusion, and hence largely independent of y_0. For $y_0 > O(D^{1/4})$ though, in the regime dominated by the advection by damping force F, all nine panels exhibit power-law behaviour. For sufficiently small values of D, the power-law regime

18

TABLE II: Slopes of $\mathcal{L}$ versus y_0 for the linear process, for different D and D_0 as indicated.

D_0 \ D	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
10^{-3}	31.6	60.5	95.0	131	167
10^{-4}	41.5	100	192	301	415
10^{-5}	46.1	132	316	606	951
10^{-6}	47.3	148	416	1000	1917
10^{-7}	47.5	153	467	1317	3162

TABLE III: Slopes of $\log_{10} \mathcal{L}$ versus $\log_{10} y_0$ for the cubic process, for different D and D_0 as indicated.

D_0 \ D	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
10^{-3}	1.69	1.62	1.56	1.53	1.60
10^{-4}	1.77	1.74	1.64	1.58	1.56
10^{-5}	1.74	1.85	1.76	1.70	1.59
10^{-6}	1.63	1.91	1.88	1.80	1.66
10^{-7}	1.52	1.91	1.96	1.88	1.78

$O(D^{1/4}) < y_0 < O(1)$ would also extend over arbitrarily many orders of magnitude. The slopes – that is, the power-law exponents – of these straight line portions at large y_0 are presented in Table 2. We infer asymptotic scalings $L \sim (y_0)^m$ with the exponent m around 1.5 to 1.9. This suggests that geometry is curved by the nonlinear interaction in the statistical space. What is more interesting is that this scaling of $\mathcal{L} \propto y_0^m$ has no resemblance to either the quartic potential $V(x) \propto x^4$ or its gradient $F \propto x^3$. That is, the combined action of the deterministic force and stochastic force results in a unique characteristic of the geometry of the attractor, governed by a power-law with index $m = m(D, D_0) < 2$. In comparison, $\mathcal{L} \propto y_0$ for the O-U process manifests the preservation of a linear geometry
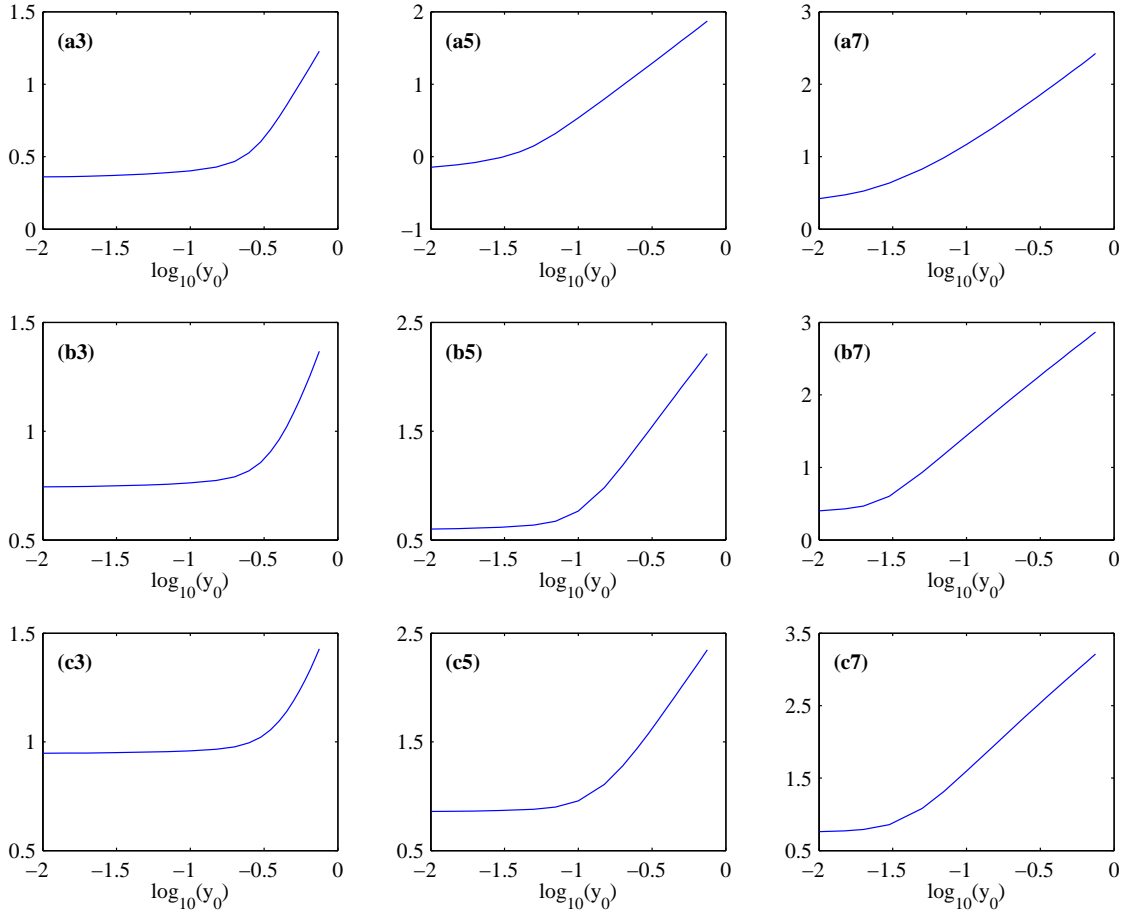
FIG. 4: As in Fig. 3, but now showing $\log_{10}(\mathcal{L})$ as functions of $\log_{10}(y_0)$, and for the cubic process only.

both by the linear damping force and the white-noise stochastic force.

To trace the origin of this power-law scaling, we again utilise the result that the dominant contribution to $\mathcal{L}$ comes from the initial/intermediate stages, where the effect of damping can be approximated by a linear friction constant γ_e in Eq. (30). Thus, we can get an estimate on the upper bound on m by replacing γ by γ_e in Eqs. (22)-(23) and taking $\langle x \rangle \sim y_0$ as follows:

$$\mathcal{L} \sim \begin{cases} \psi \frac{\sqrt{\mu}}{\sqrt{D}} y_0^2 & \text{if } D > D_0, \ 0 < r < q, \\ \phi \frac{\sqrt{\mu}}{\sqrt{D_0}} y_0^2 & \text{if } D < D_0, \ r < 0 \ , \end{cases} \tag{39}$$

for $q \gg 1$. Here ψ and ϕ are $O(1)$ constants. Eqs. (39) thus show that the power-law scaling has the upper bound as $\mathcal{L} \sim y_0^m$ where $m \leq 2$. We have checked similar power-scalings for

20

different initial PDFs (quartic exponential PDFs).

VIII. CONCLUSION

We investigated the effect of nonlinear interaction on a metric structure in a non-equilibrium process. By considering linear and nonlinear (cubic) damping, we computed the information change in the relaxation of an initial non-equilibrium state to a final equilibrium state and measured by the information length $\mathcal{L}$ the number of distinguishable states that a system undergoes during the relaxation. We explored scalings of statistical quantities of τ (the inverse of the rate of change of $\mathcal{L}$) and $\mathcal{L}$. Specifically, we illustrated that nonlinear interactions promoted temporal power-law scaling of $\tau \propto t^n$. By varying D_0 and D, we also demonstrated power-law scalings of $\mathcal{L}$ with the mean position y_0 of the initial PDF. For a linear damping, an underlying linear geometry was captured in $\mathcal{L} \propto y_0$. In comparison, the cubic damping supports a power-law relation $\mathcal{L} \propto y_0^m$, with a varying power-index $m \sim 1.5 - 1.9$, depending on D and D_0. This has to be contrasted with $m = 1$ in the linear case. This demonstrates that nonlinear interaction tends to change geometric structure of a non-equilibrium process from linear to power-law scalings.

We emphasize that $\mathcal{L}$ is path-specific and is a dynamical measure of the metric, capturing the actual statistical change that occurs during time evolution. This path-specificity would be crucial when it is desirable to control certain quantities according to the state of the system (e.g. time-dependent PDF) at any given time. An interesting example would be the treatment of large population (e.g. of bacteria, tumour cells) where the treatment should be adjusted according to the status of the population to optimize desirable outcomes while avoiding undesirable side-effects (e.g. resistance). A toy optimization problem was addressed in terms of a geodesic solution in [26]. Due to the generality of our methodology, we envision a large scope for further applications to natural phenomena to characterize non-equilibrium processes (e.g. relaxation processes). Beyond analytical/numerically solvable models, $\mathcal{L}$ can be applied to any data as long as time-dependent PDFs can be constructed from data (e.g. see [24]). Such application of $\mathcal{L}$ to data (EGG, EKG) is currently underway. Exploration of different metrics would also be of great interest.

21

Appendix A: Fluctuating Hamiltonian $\mathcal{E}$

To appreciate the relation between $\mathcal{E}$ and fluctuating energy, we express the PDF $p(x, t)$ as

$$p(x, t) = \sqrt{\frac{\beta}{\pi}} e^{-S_A} \equiv e^{-S_A + \mathcal{F}}. \tag{A1}$$

Here, $\mathcal{F} = \frac{1}{2} \ln \frac{\beta}{\pi}$ is the free energy; S_A is the effective action which can be related to the Hamiltonian H of the stochastic system (see [33]) as

$$H = -\frac{\partial S_A}{\partial t}, \tag{A2}$$

which is a stochastic analogy to the Hamilton-Jacobi relation [33, 34]. Specifically, it was shown in [33] by a path integral formulation that H is given in terms of

$$H(t) = -\frac{\partial S_A}{\partial t} = \frac{D}{2} \Pi^2 - \mu \Pi x$$

where Π is the conjugate momentum. Note that Π stems from the stochastic noise. Taking the time derivative of Eq. (A1) gives us

$$\frac{\partial p(x, t)}{\partial t} = (\dot{\mathcal{F}} + H) p(x, t), \tag{A3}$$

where $\dot{\mathcal{F}} = \frac{d\mathcal{F}}{dt}$. First, we integrate both sides of Eq. (A3) over x and use the conservation of the total probability as follows:

$$0 = \int dx \frac{\partial p}{\partial t} = \int dx (\dot{\mathcal{F}} + H) p(x, t) = \dot{\mathcal{F}} + \langle H \rangle, \tag{A4}$$

where $\langle H \rangle$ is the mean (average) value of the Hamiltonian. Therefore,

$$\dot{\mathcal{F}} = -\langle H \rangle. \tag{A5}$$

That is, the mean value of the Hamiltonian compensates for the change in free energy to conserve the total probability. We now compute the second moment which is related to $\mathcal{E}$ in Eq. (20) as

$$\begin{aligned} \mathcal{E} = \int dx \frac{1}{p} \left(\frac{\partial p}{\partial t} \right)^2 &= \int dx (H + \dot{\mathcal{F}})^2 p(x, t) \\ &= \langle (H + \dot{\mathcal{F}})^2 \rangle = \langle (\delta H)^2 \rangle, \end{aligned} \tag{A6}$$

where $\delta H = H - \langle H \rangle = H + \dot{\mathcal{F}}$ is the fluctuating Hamiltonian. By using Eq. (A5), it is interesting to observe that

$$\langle (\delta H)^2 \rangle = \langle H^2 \rangle + 2 \langle H \rangle \dot{\mathcal{F}} + \dot{\mathcal{F}}^2 = \langle H^2 \rangle - \langle H \rangle^2.$$

Appendix B: Physical meaning of $\mathcal{L}$

In this Appendix, we make an analogy to a deterministic system to elucidate the key concepts of τ and $\mathcal{L}$ in Eqs. (1) and (8). Specifically, we consider the case where an object is not moving but its length changes according to the time-dependent function $l(t)$. For this deterministic function $l(t)$, the easiest way of extracting the characteristic timescale $\tau(t)$ of $l(t)$ is by computing

$$\frac{1}{\tau(t)} = \frac{1}{l}\left(\frac{dl}{dt}\right). \tag{B1}$$

By using Eq. (B1) in Eq. (8), we can then measure the total time between initial $t = 0$ and final time t in unit of $\tau(t)$ as

$$\mathcal{L}(t) = \int_0^t \frac{dt_1}{\tau(t_1)}. \tag{B2}$$

For example, if we take $l(t) = Ae^{\lambda t}$ where $A > 0$ and $\lambda > 0$ are constant, then $\tau(t) = \lambda^{-1}$, thus Eq. (B2) gives

$$\mathcal{L}(t) = \int_0^t \frac{dt_1}{\tau(t_1)} = \int_0^t dt_1 \lambda = \lambda t = \ln\left(\frac{Ae^{\lambda t}}{A}\right) = \ln\left(\frac{l(t)}{l(0)}\right) \tag{B3}$$

We realise that $\frac{l(t)}{l(0)}$ in Eq. (B3) is just the total number of a segment of (initial) length $l(0)$ within the final length $l(t)$ and that Eq. (B3) is nothing more than the entropy (by using $k_B = 1$). Thus, $\mathcal{L}(t)$ characterises the change in entropy (amount of disorder) over time t when the object has no mean motion.

Switching back to the stochastic case with the time-dependent PDF $p(x, t)$, we now consider the rate at which $p(x, t)$ changes in time to extract the timescale of $p(x, t)$ as

$$\frac{1}{\overline{\tau}(x, t)} = \frac{1}{p(x, t)}\frac{\partial p(x, t)}{\partial t}. \tag{B4}$$

As can clearly be seen from Eq. (9), the characteristic timescale $\overline{\tau}(x, t)$ depends not only on t but also x. To obtain the dynamic time-unit $\tau(t)$ independent of x, we can take an average of Eq. (B4) over x as

$$\frac{1}{\tau(t)} \equiv \int dx\, p(x, t)\frac{1}{\overline{\tau}(x, t)} = \int dx\, p(x, t)\frac{1}{p(x, t)}\frac{\partial p(x, t)}{\partial t} = \int dx\, \frac{\partial p(x, t)}{\partial t} = 0, \tag{B5}$$

where the last equality follows from the total probability conservation. Therefore, in order to obtain a non-zero $\tau(t)$, we can consider squaring Eq. (B4) before taking the average over x:

$$
\begin{aligned}
\frac{1}{(\tau(t))^2} &\equiv \int dx\, p(x,t) \frac{1}{(\overline{\tau}(x,t))^2} = \int dx\, p(x,t) \frac{1}{p(x,t)^2} \left(\frac{\partial p(x,t)}{\partial t} \right)^2 \\
&= \int dx\, \frac{1}{p(x,t)} \left(\frac{\partial p(x,t)}{\partial t} \right)^2 ,
\end{aligned} \tag{B6}
$$

obtaining Eq. (1) in the main text. We note that Eq. (B6) corresponds to the second time derivative of relative entropy (or Kullback-Leibler divergence), as shown in Eq. (7).

Appendix C: Comparison between $\mathcal{L}$ in Eq. (26) [(27)] and entropy.

To demonstrate that $\mathcal{L}$ take the form different form from the relative entropy, it is valuable to consider $p_1 = p(x, t_1)$ and $p_2 = p(x, t_2)$ that have the same zero mean value but different width with inverse temperature β_1 and β_2

$$
p_1 = \sqrt{\frac{\beta_1}{\pi}} e^{-\beta_1 x^2}, \quad p_2 = \sqrt{\frac{\beta_2}{\pi}} e^{-\beta_2 x^2} . \tag{C1}
$$

We can then easily compute the relative entropy between p_1 and p_2 as

$$
\begin{aligned}
D(p_1, p_2) &= \int dx\, p_2 \ln\left(p_2/p_1 \right) \\
&= \int dx\, p_2 \ln\left(p_2 \right) - \int dx\, p_2 \ln\left(p_1 \right) \\
&= \int dx\, p_2 \left[\ln \sqrt{\frac{\beta_2}{\pi}} - \beta_2 x^2 \right] - \int dx\, p_2 \left[\ln \sqrt{\frac{\beta_1}{\pi}} - \beta_1 x^2 \right] \\
&= \ln \sqrt{\frac{\beta_2}{\pi}} - \beta_2 \langle x^2 \rangle_2 - \left[\ln \sqrt{\frac{\beta_1}{\pi}} + \beta_1 \langle x^2 \rangle_2 \right] \\
&= \frac{1}{2} \ln \frac{\beta_2}{\beta_1} - \frac{1}{2} \left[1 - \frac{\beta_1}{\beta_2} \right] .
\end{aligned} \tag{C2}
$$

Here, $\langle x^2 \rangle_2 = \int dx\, p_2 x^2 = \frac{1}{2\beta_2}$ was used. While the first term in Eq. (C2) appears to be similar to Eqs. (26) or (27), the second term inside the square brackets takes different form. We can now show that the integral of the square root of Eq. (C2) for small $|t_2 - t_1|$ becomes

24

similar to Eqs. (26) or (27). To this end, we expand terms in Eq. (C2) by letting $\beta_2 = \beta_1 + \delta$:

$$
\begin{aligned}
D(p_1, p_2) &= \frac{1}{2}\ln\left[1 + \frac{\delta}{\beta_1}\right] - \frac{1}{2}\left[1 - \frac{\beta_1}{\beta_1 + \delta}\right] \\
&= \frac{1}{2}\left[\frac{\delta}{\beta_1} - \frac{1}{2}\frac{\delta^2}{\beta_1^2}\right] - \frac{1}{2}\left[\frac{\delta}{\beta_1} - \frac{\delta^2}{\beta_1^2}\right] + O\left(\frac{\delta^3}{\beta_1^3}\right) \\
&= \frac{1}{4}\frac{\delta^2}{\beta_1^2} + O\left(\frac{\delta^3}{\beta_1^3}\right),
\end{aligned}
\tag{C3}
$$

where $\ln(1+x) = x - \frac{1}{2}x^2 + O(x^3)$ was used. By taking a square root of Eq. (C3), writing $\delta = \dot{\beta}_1 dt$, and then summing over time in the limit $\delta t \to 0$, we obtain

$$
\int_0^t dt\, \frac{1}{2}\frac{\dot{\beta}_1}{\beta_1} = \frac{1}{2}\ln\frac{\beta(t)}{\beta(0)},
\tag{C4}
$$

which is the same as the entropy change in Eq. (27).

Appendix D: Derivation of Eq. (20)

From $p(x,t)$ in Eq. (13), we obtain

$$
\frac{\partial p}{\partial t} = \left[\dot{\beta}\left(\frac{1}{2\beta} - (x - y)^2\right) + 2\beta(x - y)\dot{y}\right]p,
\tag{D1}
$$

where we recall $y = \langle x \rangle = y_0 e^{-\gamma t}$. Using Eq. (D1) in Eq. (1) gives us

$$
\begin{aligned}
\frac{1}{[\tau(t)]^2} &= \int dx\left[\left(\frac{1}{2\beta} - (x - y)^2\right)\dot{\beta} + 2\beta(x - y)\dot{y}\right]^2 p \\
&= \dot{\beta}^2\left[\left(\frac{1}{2\beta}\right)^2 - \frac{1}{\beta}\langle(x - y)^2\rangle + \langle(x - y)^4\rangle\right] + 4\beta^2\langle(x - y)^2\rangle\dot{y}^2 \\
&= \frac{1}{2\beta(t)^2}\left(\frac{d\beta}{dt}\right)^2 + 2\beta\left(\frac{dy}{dt}\right)^2.
\end{aligned}
\tag{D2}
$$

Here, $\dot{\beta} = \frac{d\beta}{dt}$ and $\dot{y} = \frac{dy}{dt}$; we used $\langle(x - y)^2\rangle = \frac{1}{2\beta}$ and $\langle(x - y)^4\rangle = 3\left(\frac{1}{2\beta}\right)^2$. To obtain the last equation in Eq. (20), it is useful to express β in Eq. (18) in the following form:

$$
\beta = \frac{1}{\frac{1}{\beta_1} + \frac{1}{\beta_0}e^{-2\gamma t}} = \frac{1}{\frac{2D(1 - e^{-2\gamma t})}{\gamma} + \frac{1}{\beta_0}e^{-2\gamma t}}
\tag{D3}
$$

$$
= \frac{\gamma\beta_0 e^{2\gamma t}}{T}.
\tag{D4}
$$

where we used $T = 2\beta_0 D(e^{2\gamma t} - 1) + \gamma$. By differentiating Eq. (D3), we then obtain

$$
\begin{aligned}
\dot{\beta} &= -2\gamma\beta^2 e^{-2\gamma t}\left[\frac{2D}{\gamma} - \frac{1}{\beta_0}\right] \\
&= -2\beta^2 e^{-2\gamma t}[2\beta_0 D - \gamma]\frac{1}{\beta_0},
\end{aligned}
\tag{D5}
$$

25

Eqs. (D4)-(D5) and $r = 2\beta_0 D - \gamma$ then give us

$$\frac{\dot{\beta}^2}{2\beta^2} = 2\gamma^2 r^2 \frac{1}{T^2}. \tag{D6}$$

Similarly, using $\dot{y} = \frac{d}{dt}(y_0 e^{-\gamma t}) = -\gamma y_0 e^{-\gamma t}$, $T = 2\beta_0 D(e^{2\gamma t} - 1) + \gamma$ and $q = \beta_0 \gamma y_0^2$, we obtain

$$2\beta \dot{y}^2 = 2q\gamma^2 \frac{1}{T}. \tag{D7}$$

Finally, using Eqs. (D6)-(D7) in Eq. (D2) gives us Eq. (20).

Appendix E: Derivation of Eqs. (22) and (23)

By using Eqs. (20) and (21) in Eq. (8), we obtain

$$\mathcal{L} = \frac{1}{\sqrt{2}} \int_{T_i}^{T_f} \left\{ \frac{1}{T} \frac{1}{T+r} \sqrt{r^2 + qT} \right\} dT, \tag{E1}$$

To compute Eq. (E1), we let $Y = \sqrt{r^2 + qT}$ and recast it as

$$\begin{aligned}
\mathcal{L} &= \frac{\sqrt{2}}{r} \int_{Y_i}^{Y_f} \left\{ \frac{r^2}{Y^2 - r^2} + \frac{qr - r^2}{Y^2 + qr - r^2} \right\} dY \\
&= \frac{1}{\sqrt{2}} \left[\ln\left(\frac{Y-r}{Y+r} \right) \right]_{Y_i}^{Y_f} + \frac{\sqrt{2}}{r} H,
\end{aligned} \tag{E2}$$

where Y_i and Y_f are Y evaluated at T_i and T_f, and H is defined as

$$H = \int_{Y_i}^{Y_f} \frac{qr - r^2}{Y^2 + qr - r^2} dY. \tag{E3}$$

Eq. (E3) is to be evaluated separately for two cases: $q \geq r$ and $q < r$. First, for $q \geq r$, we use $Y = \sqrt{qr - r^2} \tan\theta$ in Eq. (E3) to obtain

$$H = \sqrt{qr - r^2} \int \frac{\sec^2\theta}{\tan^2\theta + 1} d\theta \tag{E4}$$

$$= \sqrt{qr - r^2} \left[\tan^{-1}\left(\frac{Y}{\sqrt{qr - r^2}} \right) \right]_{Y_i}^{Y_f}. \tag{E5}$$

Secondly, in the $q < r$ case, we let $Y = \sqrt{|qr - r^2|} \sec\theta = \sqrt{r^2 - qr} \sec\theta$ ($\cos\theta = \frac{\sqrt{r^2 - qr}}{Y}$) to obtain

$$H = -\sqrt{r^2 - qr} \int \frac{1}{\sin\theta} d\theta$$

$$= -\frac{\sqrt{r^2 - qr}}{2} \left[\ln\left(\frac{Y - \sqrt{r^2 - qr}}{Y + \sqrt{r^2 - qr}} \right) \right]_{Y_i}^{Y_f}. \tag{E6}$$

26

We note that Eq. (E2) is continuous across $q = r$. In summary, Eqs. (E5) and (E6) lead to Eq. (23) in the text; Eq. (E2) gives Eq. (22) in the text.

Appendix F: Derivation of Eq. (37)

In this Appendix, we show the main steps leading to Eq. (37) when $D > D_0$, $q \gg 1$, and $0 < r < q$. In this case, we note that H in Eq. (22) is given by the first line in Eq. (23) and thus

$$\frac{H}{r} = \sqrt{\frac{q}{r} - 1} \, \tan^{-1} \left(\frac{Y}{\sqrt{qr - r^2}} \right), \tag{F1}$$

where

$$Y = \sqrt{qT + r^2}, \; q = \frac{\gamma^2 y_0^2}{2D_0}, \; r = \gamma \left(\frac{D}{D_0} - 1 \right), \; T = \gamma \left[\frac{D}{D_0} (e^{2\gamma t} - 1) + 1 \right].$$

We evaluate Eq. (F1) at $t = 0$ and $t \to \infty$ to compute the second term on the right-hand side of Eq. (22). To this end, first, we approximate the argument of the arctan function in Eq. (F1) for large $q \gg 1$ as

$$\frac{Y}{\sqrt{qr - r^2}} = \sqrt{\frac{\frac{qT}{r^2} + 1}{\frac{q}{r} - 1}} \sim \sqrt{\frac{T}{r}} \sim \sqrt{\frac{D_0}{D - D_0} \left[\frac{D}{D_0} (e^{2\gamma t} - 1) + 1 \right]}. \tag{F2}$$

In the long time limit as $t \to \infty$, we let $\phi \equiv \frac{D_0}{D - D_0} \left[\frac{D}{D_0} (e^{2\gamma t} - 1) \right] \; (\to \infty)$, and evaluate Eq. (F2) as

$$\frac{Y}{\sqrt{qr - r^2}} \bigg|_{t \to \infty} \sim \sqrt{\phi + \frac{D_0}{D - D_0}}. \tag{F3}$$

On the other hand, at $t = 0$, Eq. (F2) is simplified as

$$\frac{Y}{\sqrt{qr - r^2}} \bigg|_{t=0} \sim \sqrt{\frac{D_0}{D - D_0}}. \tag{F4}$$

We now evaluate $\sqrt{\frac{q}{r} - 1}$ in front of the arctan function in Eq. (F1) at $t = 0$ and $t \to \infty$, which in both limits becomes:

$$\sqrt{\frac{q}{r} - 1} \sim \sqrt{\frac{\gamma y_0^2}{2D} - 1} \sim \sqrt{\frac{\gamma}{2D}} y_0. \tag{F5}$$

Thus, by putting Eqs. (F3)-(F5) in Eq. (F1), we obtain

$$\left[\frac{H}{r} \right]_{t=0}^{\infty} \sim y_0 \sqrt{\frac{\gamma}{2D}} \left[\tan^{-1} \sqrt{\phi + \frac{D_0}{D - D_0}} - \tan^{-1} \sqrt{\frac{D_0}{D - D_0}} \right] \sim y_0 \sqrt{\frac{\gamma}{2D}} \left(\frac{\pi}{2} - \sqrt{\frac{D_0}{D}} \right), \tag{F6}$$

by using $\phi \to \infty$, $D_0 \ll D$, and $\tan^{-1} x \sim x$ for $x \ll 1$. In comparison with the contribution from the second term involving H on the right-hand side of Eq. (22), the contribution from the first term involving the logarithmic function can be shown to be negligible by following similar analysis as above. Therefore, multiplying Eq. (F6) by $\sqrt{2}$ gives Eq. (37) in the main text.

---

[1] H. Haken, *Information and self-organization: A macroscopic approach to complex systems* (Berlin: Springer) (2006).

[2] W.B. Cannon, Organization for physiological homeostasis, *Physiological Rev.* **9** 399-431 (1929).

[3] E. Kim and P.H. Diamond, Zonal flows and transient dynamics of the L-H transition, *Phys. Rev. Lett.* **90** 185006 (2003).

[4] K. Srinivasan and W.R. Young, Zonostrophic instability, *J. Atmos. Sci.* **69** 1633-1656 (2006).

[5] A.L. Gibbs and F.E. Su, On choosing and bounding probability metrics, *Int. Stat. Rev.* **70** 419-435 (2002).

[6] R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker-Planck equation, *SIAM J. Math. Anal.* **29** 1-17 (1998).

[7] A. Takatsu, Wasserstein geometry of Gaussian measures, *Osaka J. Math.* **48** 1005-1026 (2011).

[8] J. Lott, Some geometric calculations on Wasserstein space, *Commun. Math. Phys.* **277** 423-437 (2008).

[9] W. Gangbo and R.J. McCann, Optimal maps in Monge's mass transport problem, *C. R. Acad. Sci. Paris* **321** 1653-1658 (1995).

[10] S. Ferradans, G.-S. Xia, G. Peyré and J.-F. Aujol, Static and dynamic texture mixing using optimal transport, *Lecture Notes Comp. Sci.* **7893** 137-148 (2013).

[11] H. Risken, *The Fokker-Planck equation: methods of solutions and applications* (Berlin: Springer) (2013).

[12] W.K. Wootters, Statistical distance and Hilbert space, *Phys. Rev. D* **23** 357-362 (1981).

[13] M.T. Martin, A. Plastron and O.A. Ross, Statistical complexity and disequilibrium, *Phys. Lett. A* **311** 126-132 (2003).

[14] G. Ruppeiner, Thermodynamics: a Riemannian geometric model, *Phys. Rev. A* **20** 1608-1613 (1979).

[15] F. Schlögl, Thermodynamic metric and stochastic measures, *Z. Phys. B Cond. Matt.* **59** 449-454 (1985).

[16] E.H. Feng and G.E. Crooks, Far-from-equilibrium measurements of thermodynamic length, *Phys. Rev. E* **79** 012104 (2009).

[17] S.L. Braunstein and C.M. Caves, Statistical distance and the geometry of quantum states, *Phys. Rev. Lett.* **72** 3439-3443 (1994).

[18] H. Strobel, W. Muessel, D. Linnemann, T. Zibold, D.B. Hume, L. Pezzé, A. Smerzi and M.K. Oberthaler, Fisher information and entanglement of non-Gaussian spin states, *Science* **345** 424-427 (2014).

[19] J. Nulton, P. Salamon, B. Andresen and Q. Anmin, Quasistatic processes as step equilibrations, *J. Chem. Phys.* **83** 334-338 (1985).

[20] G.E. Crooks, Measuring thermodynamic length, *Phys. Rev. Lett.* **99** 100602 (2007).

[21] D.A. Sivak and G.E. Crooks, Thermodynamic metrics and optimal paths, *Phys. Rev. Lett.* **108** 190602 (2012).

[22] P. Salamon, J.D. Nulton, G. Siragusa, A. Limon, D. Bedeaux, and S. Kjelstrup, A simple example of control to minimize entropy production, *J. Non-Equilib. Thermodyn.* **27** 45-55 (2002).

[23] S.B. Nicholson and E. Kim, Investigation of the statistical distance to reach stationary distributions, *Phys. Lett. A* **379** 83-88 (2015).

[24] S.B. Nicholson and E. Kim, Structures in sound: analysis of classical music using the information length, *Entropy* **18** 258 (2016).

[25] J. Heseltine and E. Kim, Novel mapping in non-equilibrium stochastic processes, *J. Phys. A* **49** 175002 (2016).

[26] E. Kim, U. Lee, J. Heseltine and R. Hollerbach, Geometric structure and geodesic in a solvable model of nonequilibrium process, *Phys. Rev. E* **93** 062127 (2016).

[27] E. Kim, H.-L. Liu and J. Anderson, Probability distribution function for self-organization of shear flows, *Phys. Plasmas* **16** 0552304 (2009).

[28] A.P.L. Newton E. Kim and H.-L. Liu, On the self-organizing process of large scale shear flows, *Phys. Plasmas* **20** 092306 (2013).

[29] E.T. Lu, Avalanches in continuum driven dissipative systems, *Phys. Rev. Lett.* **74** 2511-2514 (1995).

[30] M.H. Pinsonneault, S.D. Kawaler, S. Sofia and P. Demarque, Evolutionary models of the rotating Sun, *Astrophys. J.* **338** 424-452 (1998).

[31] E. Kim and R. Hollerbach, Time-dependent probability density function in cubic stochastic processes, *Phys. Rev. E* **94** 052118 (2016).

[32] F. Klebaner, *Introduction to stochastic calculus with applications* (London: Imperial College Press) (2016).

[33] E. Kim and S. Nicholson, Complementary relations in non-equilibrium stochastic processes, Phys. Lett. A, **379**, 1613 (2015).

[34] R.P Feynman and A.R. Hibbs, Quantum Mechanics and Path Integral, Chap. 2, McGraw-Hill Book Company (1965); R.P. Feynman, Statistical Mechanics, Chap. 3, W.A. Benjamin, Inc (1972).

[35] The difference between different initial points (y_0's) is then quantified by the difference in corresponding $\mathcal{L}$'s.