Information Geometry of Nonlinear Stochastic Systems

Rainer Hollerbach [1,*] **, Donovan Dimanche** [2,3] **and Eun-jin Kim** [2]

[1] Department of Applied Mathematics, University of Leeds, Leeds LS2 9JT, UK
[2] School of Mathematics and Statistics, University of Sheffield, Sheffield S3 7RH, UK;
    donovan.dimanche@gmail.com (D.D.); e.kim@sheffield.ac.uk (E.K.)
[3] Institut National des Sciences Appliquées de Rouen, 76801 Saint-Étienne-du-Rouvray CEDEX, France
[*] Correspondence: rh@maths.leeds.ac.uk; Tel.: +44-113-343-5134

check for
updates

Abstract: We elucidate the effect of different deterministic nonlinear forces on geometric structure of stochastic processes by investigating the transient relaxation of initial PDFs of a stochastic variable x under forces proportional to $-x^n$ ($n = 3, 5, 7$) and different strength D of δ-correlated stochastic noise. We identify the three main stages consisting of nondiffusive evolution, quasi-linear Gaussian evolution and settling into stationary PDFs. The strength of stochastic noise is shown to play a crucial role in determining these timescales as well as the peak amplitude and width of PDFs. From time-evolution of PDFs, we compute the rate of information change for a given initial PDF and uniquely determine the information length $\mathcal{L}(t)$ as a function of time that represents the number of different statistical states that a system evolves through in time. We identify a robust geodesic (where the information changes at a constant rate) in the initial stage, and map out geometric structure of an attractor as $\mathcal{L}(t \to \infty) \propto \mu^m$, where μ is the position of an initial Gaussian PDF. The scaling exponent m increases with n, and also varies with D (although to a lesser extent). Our results highlight ubiquitous power-laws and multi-scalings of information geometry due to nonlinear interaction.

1. Introduction

There is increasing interest in a metric on probability from theoretical and practical considerations, with different metrics proposed depending on the question of interest (e.g., [1–10] and further references therein). Theoretically, the assignment of an appropriate metric to probability enables us to mathematically quantify the difference among different Probability Density Functions (PDFs), providing a beautiful conceptual link between a stochastic process and geometry. At a practical level, it can be utilized for optimising various desired outcomes. For instance, the Wasserstein metric has been extensively studied for the optimal transport problem to minimize transport cost, typically taken to increase quadratically with distance between two locations [1]. The Fisher (also called Fisher-Rao) metric has recently been used for optimization, including the minimization of entropy production [11], parameter estimation [12], controlling population [13], understanding the arrow of time [14], and analysing the convexity of the relative entropy [15].

However, compared with the Wasserstein metric, whose application has established itself as a branch of applied mathematics, the geometric structure associated with the information change in the Fisher metric and its utility have been explored much less. Unlike the Wasserstein distance, the Fisher metric yields a hyperbolic geometry in the upper half-plane (e.g., [9,10]) where the distance is measured in units of the width of the PDF. That is, the distance in the Fisher metric is dimensionless and represents the number of different statistical states. Consequently, for a Gaussian PDF, statistically distinguishable

states are determined by the standard deviation; two PDFs that have the same standard deviation and differ in peak positions by less than one standard deviation are statistically indistinguishable (e.g., [11,16–22]).

By extending the Fisher metric to time-dependent problems, we recently introduced a system-independent way of quantifying information change associated with time-evolution of PDFs [13,23–30]. (Note that we use information for statistically different states, refraining from the debate on the exact definition of information, e.g., [5,31].) The key idea is to define an infinitesimal distance at any time by comparing two PDFs at adjacent times and sum these distances. The total distance gives us the number of statistically different states that a system passes through in time, and is called information length $\mathcal{L}$. While the detailed derivation of $\mathcal{L}$ is given in [5,13,16,23–30], it is useful to highlight that $\mathcal{L}$ is a measure of the total elapsed time in units of a dynamical timescale for information change. To show this, we define the dynamical time $\tau(t)$ as follows:

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

Here, $\tau(t)$ is the characteristic timescale over which the information changes. Having units of time, $\tau(t)$ quantifies the correlation time of a PDF. Alternatively, $1/\tau$ quantifies the (average) rate of change of information in time. A particular path which gives a constant valued $\mathcal{E}$ is a geodesic along which the information propagates at the same speed [13]. $\mathcal{L}(t)$ is then defined by measuring the total elapsed time t in units of τ as:

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

$\mathcal{L}(t)$ is a Lagrangian quantity (unlike entropy or relative entropy), uniquely defined as a function of time t for a given initial PDF, and represents the total number of statistically distinguishable states that a system evolves through. It thus provides a very convenient methodology for measuring the distance between $p(x,t)$ and $p(x,0)$ continuously in time for a given $p(x,0)$. One of the utilities of $\mathcal{L}(t)$ is to quantify the proximity of any initial PDF to a final attractor of a dynamical system. We note that for a linear process, we can express the information length in terms of a metric tensor g_{ij} using the two parameters $\beta = 1/2\langle (x - \langle x \rangle)^2 \rangle$ and $\langle x \rangle$ (e.g., [13]). However, even in the case where the control parameters are not known, we can still define $\mathcal{L}$ as long as we can compute time-dependent PDFs (e.g., from data [24]).

Traditionally, concepts such as bifurcations, Lyapunov exponent or a distance to an equilibrium point are commonly used to understand dynamical systems (e.g., see [32]). An intriguing question arises as to how to define a distance between a point, say x, and an equilibrium which is not a point but a limit cycle, or chaotic attractor. One interesting possibility is to consider a narrow initial PDF around x and to measure the total information length $\mathcal{L}(t \to \infty) = \mathcal{L}_\infty$ as the initial PDF evolves toward the equilibrium PDF. $\mathcal{L}_\infty$ offers a Lagrangian distance that depends on the trajectory/history of the system (e.g., time-dependent PDF), being uniquely defined as a function of time for a given initial PDF. This enables us to map out the attractor structure by measuring $\mathcal{L}_\infty$ for different locations of a narrow initial PDF. In a chaotic system, $\mathcal{L}_\infty$ changes abruptly when a different initial PDF is used, as shown in Figure 4 in [24], where the very spiky curve represents a sensitive dependence of $\mathcal{L}_\infty$ on $x(t = 0)$, the location of a very narrow initial PDF. This sensitive dependence of $\mathcal{L}_\infty$ on $x(t = 0)$ means that a small change in the initial condition causes a large difference in a path that a system evolves through and thus $\mathcal{L}_\infty$. This is quite similar to the sensitive dependence of the Lyapunov exponent on the initial condition. That is, our $\mathcal{L}$ provides a new methodology to test chaos. Furthermore, Figure 4 in [24] shows small $\mathcal{L}_\infty$ for unstable points, demonstrating that unstable points are more similar to chaotic attractors.

The purpose of this paper is to consider the effect of different orders of nonlinear interaction on the geometric structure by considering the case when the equilibrium is a stable point [26,33]. The remainder of this paper is organised as follows: Section 2 introduces the basic model. Section 3 derives exact analytic solutions in the absence of stochastic noise, as well as asymptotic scalings for the timescales, peak amplitudes and widths once noise plays a role. Section 4 presents numerical results, and shows how they compare with the analytic scalings. Section 5 summarizes the results.

2. Model

We consider the following nonlinear Langevin equation:

$$\frac{dx}{dt} = -\gamma x^n + \xi. \tag{3}$$

Here, x is a random variable and ξ is a stochastic forcing, which for simplicity can be taken as a short-correlated (or δ-correlated, white-noise) random forcing as follows:

$$\langle \xi(t)\xi(t') \rangle = 2D\delta(t - t'), \tag{4}$$

where the angular brackets represent the average over ξ, $\langle \xi \rangle = 0$, and D is the strength of the forcing. The parameter γ is a positive constant. n is the order of nonlinearity, which we take to be an odd integer to make $x = 0$ an attractor, and also preserve a reflectional symmetry (under $x \rightarrow -x$) of the system.

We note that there are two possible interpretations of Equation (3). Specifically, consider the linear case where $\frac{\partial V}{\partial x} \propto x$. The first interpretation is to view x as a velocity, in which case the force term would give a frictional force (e.g., Equation (1.2) in [34]) as $\frac{dv}{dt} = -\gamma v + \xi$, with γ as a frictional (damping) constant. The second interpretation is to take the overdamped limit of the coupled equations for x and v by dropping d^2x/dt^2 (e.g., Equation (3.130) in [34]) to obtain one equation for x from the first line in Equation (3.130). In this case, x represents the position and γ^{-1} would be the frictional constant, and a harmonic potential would correspond to $n = 1$.

For a deterministic system with $\xi = 0$, the solution to Equation (3) is readily obtained as

$$x(t) = \frac{x_0}{[1 + \gamma(n-1)tx_0^{n-1}]^{\frac{1}{n-1}}}, \tag{5}$$

where x_0 is the initial condition $x(t = 0)$.

The corresponding Fokker–Planck (FP) equation is [34]

$$\partial_t p(x,t) = \partial_x(\gamma x^n p) + D\partial_{xx}p. \tag{6}$$

We first discuss some analytic limiting cases in Section 3, and then present numerical solutions in Section 4.

3. Analytic Solutions

3.1. Exact Solutions for $D = 0$

In the absence of the stochastic noise $D = 0$, the diffusion term in the Fokker–Planck Equation (6) is zero. In this case, the PDF does not have a stationary solution, but continues to change in time due to the linear/nonlinear force. For instance, for $n = 1$, the PDF's width becomes exponentially narrow in time as the fluctuation (as well as the mean value) decreases exponentially due to $\gamma > 0$ (see below). To obtain an analytic solution to Equation (6), we use the method of characteristics. Specifically, we rewrite Equation (6) with $D = 0$ in terms of the total derivative along the characteristic as

$$\frac{dp}{dt} \equiv \frac{\partial p}{\partial t} + \frac{dx}{dt}\frac{\partial p}{\partial x} = n\gamma x^{n-1}p. \tag{7}$$

Here, the characteristic is given by

$$\frac{dx}{dt} = -\gamma x^n, \tag{8}$$

which is Equation (3) without the stochastic noise ξ. Thus, the characteristic with the initial condition $x(t = 0) = x_0$ satisfies Equation (5), which can also be written as

$$x_0 = \frac{x(t)}{[1 - \gamma(n-1)t[x(t)]^{n-1}]^{\frac{1}{n-1}}}. \tag{9}$$

Along these characteristics, we rewrite Equation (7) as

$$\frac{dp}{p} = n\gamma x^{n-1}dt = n\gamma \frac{x_0^{n-1}dt}{[1 + \gamma(n-1)tx_0^{n-1}]} = \frac{n}{n-1}d\left[\ln\left[1 + \gamma(n-1)tx_0^{n-1}\right]\right]. \tag{10}$$

The integration of Equation (10) then gives us

$$p(x,t) = p(x_0,0)[1 + \gamma(n-1)tx_0^{n-1}]^{\frac{n}{n-1}} = p(x_0,0)[1 - \gamma(n-1)tx^{n-1}]^{-\frac{n}{n-1}}. \tag{11}$$

For simplicity, we consider an initial Gaussian PDF localised around μ as

$$p(x_0,0) = \sqrt{\frac{\beta_0}{\pi}}e^{-\beta_0(x_0-\mu)^2}. \tag{12}$$

Then, we can find $p(x,t)$ at a later time from Equations (9), (11) and (12) as

$$p(x,t) = \sqrt{\frac{\beta_0}{\pi}}\,\phi^n\,e^{-\beta_0(x\phi-\mu)^2}, \tag{13}$$

where

$$\phi = [1 - \gamma(n-1)tx^{n-1}]^{-\frac{1}{n-1}}. \tag{14}$$

The maximum amplitude of p occurs at that x where $x\phi - \mu = 0$ in Equation (13), with the value

$$p_{\max}(t) = \sqrt{\frac{\beta_0}{\pi}}\left[1 + \gamma(n-1)t\mu^{n-1}\right]^{\frac{n}{n-1}}. \tag{15}$$

The linear case $n = 1$ can be obtained by taking the limit $n \to 1$ in Equations (13) and (14). One finds that $\phi = e^{\gamma t}$ so that

$$\lim_{n\to 1}p(x,t) = \sqrt{\frac{\beta}{\pi}}e^{-\beta(x-\mu e^{-\gamma t})^2}, \tag{16}$$

where $\beta = \beta_0 e^{2\gamma t}$. That is, the width of the PDF ($\propto \beta^{-1/2}$) as well as the peak position decrease exponentially. However, the linear case can be solved analytically even when $D \neq 0$ [33], so we are here more interested in the nonlinear cases $n = 3, 5, 7$.

Given $p(x,t)$ in Equation (13), we can further calculate $\mathcal{E}(t)$ as follows

$$\begin{aligned}
\mathcal{E}(t) &= \gamma^2 \int dx\,(x\phi)^{2(n-1)}\left[n - 2\beta_0 x\phi(x\phi - \mu)\right]^2 p(x,t) \\
&= \gamma^2 \int dx_0\,x_0^{2(n-1)}\left[n - 2\beta_0 x_0(x_0 - \mu)\right]^2 p(x_0,0),
\end{aligned} \tag{17}$$

where we use $x_0 = x\phi$, $\frac{dx_0}{dx} = \phi^n$, $p(x_0,0) = p(x,t)\frac{dx}{dx_0}$, and $p(x_0,0)$ is the initial Gaussian PDF in Equation (12). Interestingly, Equation (17) is independent of time, with constant τ. That is, without a stochastic noise, the evolution of PDFs follows a geodesic due to the scaling relation satisfied along

the characteristic. For example, we can show that, for $n = 1$, $\mathcal{E} = 2\gamma^2(1 + \beta_0\mu^2)$ and for $n = 3$, $\mathcal{E} = \gamma^2\mu^4 \left[2\beta_0\mu^2 + 24 + \frac{99}{2\beta_0\mu^2} + \frac{21}{2(\beta_0\mu^2)^2}\right]$. From these analyses, we can infer that for sufficiently large β_0 such that $\beta_0\mu^{n-1} \gg 1$, $\mathcal{E} \propto \beta_0\mu^{2n}$ to leading order.

3.2. Approximate Solutions for $D \neq 0$

According to the previous results, the peak amplitudes increase indefinitely in time. The corresponding widths also decrease, and eventually become so narrow that any non-zero D will start to play a significant role, and will start to broaden the PDFs again. However, during this phase of the evolution, the PDF width is still smaller than its mean position, so that we can use a quasi-linear analysis to approximate Equation (3) to leading order in $O(x'/\langle x \rangle)$ as

$$\frac{dx'}{dt} \approx -n\gamma\langle x(t)\rangle^{n-1}x' + \xi, \tag{18}$$

where $x'(t) = x(t) - \langle x(t)\rangle$ is the fluctuation, with $\langle x'(t)\rangle = 0$. x' in Equation (18) satisfies the Ornstein–Uhlenbeck process with the effective damping constant

$$\gamma_e(t) = n\gamma\langle x(t)\rangle^{n-1}. \tag{19}$$

Thus, in this stage, the PDFs remain essentially Gaussian and evolve as

$$p(x,t) = \sqrt{\frac{\beta}{\pi}}e^{-\beta(x-\langle x\rangle)^2}, \tag{20}$$

where

$$\langle x \rangle = \frac{\mu}{[1 + \gamma(n-1)\mu^{n-1}t]^{\frac{1}{n-1}}}, \tag{21}$$

$$\frac{1}{2\beta} = \frac{e^{-2G(t)}}{2\beta_0} + \frac{D(1 - e^{-2G(t)})}{\gamma_e}, \tag{22}$$

$$G(t) = \int_0^t dt'\gamma_e(t') = \frac{n}{n-1}\ln\left[1 + \gamma(n-1)\mu^{n-1}t\right]. \tag{23}$$

The term $e^{-2G}/2\beta_0$ in Equation (22) represents the narrowing of the initial PDF width as discussed in the previous section. However, note that, when $1/2\beta_0 \ll D/\gamma_e$, a PDF can maintain the same width set by the constant value D/γ_e at the initial stage; that is, there is no nondiffusive evolution phase. See [33] for such a case with $n = 3$. For the $1/2\beta_0 > D/\gamma_e$ case, we consider here, the transition from the nondiffusive evolution phase to the quasi-linear Gaussian evolution phase occurs when $e^{-2G}/2\beta_0$ becomes comparable to D/γ_e, leading to the following criterion for the first transition timescale t_1:

$$2G \sim \ln\frac{\gamma_e}{2\beta_0 D} \sim \ln\frac{n\gamma\langle x\rangle^{n-1}}{2\beta_0 D}, \tag{24}$$

using $\gamma_e = n\gamma\langle x\rangle^{n-1}$. By using Equation (23) and $\langle x \rangle = \mu/[1 + \gamma(n-1)\mu^{n-1}t_1]^{\frac{1}{n-1}}$, we obtain

$$\left[1 + \gamma(n-1)\mu^{n-1}t_1\right]^{\frac{3n-1}{n-1}} \sim \frac{\gamma n}{2\beta_0 D}\mu^{n-1}. \tag{25}$$

For large time $\gamma(n-1)\mu^{n-1}t_1 \gg 1$, Equation (25) thus yields

$$t_1 \propto D^{-\frac{n-1}{3n-1}}\mu^{-\frac{2n(n-1)}{3n-1}}. \tag{26}$$

Note in particular how both exponents are negative for $n = 3, 5, 7$, so that smaller D and/or μ yield a larger t_1. Next, for $D \neq 0$ and $\gamma(n-1)\mu^{n-1}t \gg 1$, we approximate Equation (22) as

$$\frac{1}{2\beta} \sim \frac{D}{\gamma_e} \sim \frac{n-1}{n}Dt, \tag{27}$$

using also $\gamma_e \propto t^{-1}$. Thus, the PDF width increases as $(Dt)^{1/2}$, while the maximum amplitude decreases as

$$p_{\max}(t) \propto \beta^{1/2} \propto (Dt)^{-1/2}, \tag{28}$$

typical of Brownian motion [34]. Using Equation (26) in either Equation (15) or (28)—t_1 after all is precisely the time when one scaling transitions to the other, so they should agree at that time—we obtain the scalings of the overall maximum amplitude p_{Max} reached throughout the entire evolution as

$$p_{\mathrm{Max}} \propto D^{-\frac{n}{3n-1}} \mu^{\frac{n(n-1)}{3n-1}}. \tag{29}$$

3.3. Final Stationary Distribution

For even greater times, fluctuations become stronger while the mean values decrease. The quasi-linear analysis in the previous section is therefore eventually no longer applicable, and numerical solutions are crucial. The final stationary solution of Equation (6) can however be computed analytically from $D\partial_x p = -\gamma x^n p$ and becomes

$$p(x) = \left(\frac{\gamma}{D(n+1)}\right)^{1/(n+1)} \frac{n+1}{\Gamma(1/(n+1))} \exp\left(-\frac{\gamma}{D(n+1)}x^{n+1}\right). \tag{30}$$

Thus, the maximum amplitude

$$p_{\max} \propto D^{-\frac{1}{n+1}}, \tag{31}$$

while the width of the PDF is proportional to $D^{\frac{1}{n+1}}$. Finally, the transition from the intermediate stage in the previous section to this final stage occurs when the two formulas for the widths $(Dt)^{1/2}$ and $D^{\frac{1}{n+1}}$ become comparable, yielding the second transition timescale

$$t_2 \propto D^{-\frac{n-1}{n+1}}. \tag{32}$$

In the following section, we see how numerical solutions compare with some of these predictions such as the two timescales t_1 and t_2, as well as explore other aspects of the solutions for which no analytic predictions are possible.

4. Numerical Solutions

For $D \neq 0$, an exact analytic solution to the Fokker–Planck Equation (6) only exists for the linear case $n = 1$ [33]. For the nonlinear cases $n = 3, 5, 7$ that are the focus of this paper, we must resort to numerical solutions. Without loss of generality, we set $\gamma = 1$. The interval in x is fixed to be $[-1, 1]$, rather than the original $[-\infty, \infty]$. This may seem drastic, but actually involves no real loss of generality either, since any finite interval can always be mapped to $[-1, 1]$ by suitably rescaling x, t and D. As long as the initial condition (and D) are chosen such that p would be negligible outside $[-1, 1]$ anyway, solving the FP equation only on $[-1, 1]$, and with $p = 0$ boundary conditions is then equivalent in all essentials to the original infinite interval.

The numerical solution is done by second-order finite-differencing in x, with up to $M = 2 \times 10^6$ grid points. The time-stepping is also second-order accurate, with increments as small as $\Delta t = 10^{-6}$. Both M and Δt were varied to check the accuracy of the solutions. In the later stages, when the PDFs are evolving to increasingly broad profiles, M can also be decreased, and Δt increased, while still preserving accuracy. Regridding the solutions in this way is indeed crucial, since the final adjustment

timescale t_2 is extremely long for small D, and could not be reached if M and Δt were kept fixed at their initial values.

In [33], we considered the initial condition

$$p = \frac{1}{\sqrt{\pi \, 10^{-8}}} \exp\left[-\frac{(x - 0.7)^2}{10^{-8}}\right], \tag{33}$$

and then compared numerical solutions for $n = 3$ and $D = 10^{-3}$ to 10^{-7} with the corresponding analytical solutions for $n = 1$. That is, the initial peak was extremely narrow, and diffusion was greater than what we consider here. As a result of these choices, the peaks only became broader but never narrower; the nondiffusive regime in Section 3.1 simply does not exist in this case.

In contrast, in this work, we take the initial condition

$$p = \frac{1}{\sqrt{2\pi \, 10^{-6}}} \exp\left[-\frac{(x - \mu)^2}{2 \times 10^{-6}}\right], \tag{34}$$

and $D = 10^{-6}$ to 10^{-9}. By starting with broader peaks and having smaller D, we do have an initial nondiffusive regime here, and are able to observe the narrowing of the peaks predicted by Equations (13) and (14). We begin by fixing the initial peak position $\mu = 0.65$; below, we also consider the range $\mu = [0.01, 0.75]$.

Figure 1 shows how the peak amplitudes evolve in time. We can clearly see the three regimes deduced in Section 3: the peaks initially increase, in excellent agreement with Equation (15), then they decrease in agreement with Equation (28), and finally they equilibrate to Equation (31). To more quantitatively compare the overall peak amplitudes p_{Max} and the corresponding times t_{Max} at which they occur with the analytic predictions given by Equations (26) and (29), we note first that $D = 10^{-6}$ is clearly not yet sufficiently small for there to be an initial nondiffusive regime at all (for this particular width of the initial condition). Even $D = 10^{-7}$ only follows the nondiffusive Equation (15) for a very brief time, not yet long enough to be in the $\gamma(n - 1)\mu^{n-1}t \gg 1$ regime where Equations (26) and (29) are expected to apply. Table 1 therefore compares the $D = 10^{-8}$ and 10^{-9} cases, and uses them to extract scaling exponents of the form $D^{-\alpha}$. We see that the agreement of p_{Max} with Equation (29) is virtually perfect. The corresponding times t_{Max} are close to the expected scaling t_1, but are not fully in the asymptotic limit yet. This is hardly surprising, since even for $D = 10^{-9}$ the t_{Max} values in Table 1 only have $\gamma(n - 1)\mu^{n-1}t_{\text{Max}} \sim 8$ (for all three n values).

To similarly test the scalings that are predicted for μ, further runs were done with fixed $D = 10^{-9}$, and $\mu = 0.55$ and 0.75. As Table 2 shows, the extracted exponents are again in very good (t_{Max}) and perfect (p_{Max}) agreement with the predicted asymptotic scalings. The interesting feature that larger initial positions μ have smaller times to reach the maximum peak amplitude is certainly very well reproduced. Note finally that the two μ values used to extract these exponents in Table 2 differ by less than a factor of two even, and would thus certainly not be enough to extract reliable scaling exponents if we did not already have robust analytic predictions. As we also show in more detail below, in principle, it would be possible to have the analytic predictions extend over an arbitrarily large range in μ, but that would require increasingly small D as well, which becomes numerically too time-consuming.

Returning to the $\mu = 0.65$ runs in Figures 1 and 2 shows the expectation values $\langle x \rangle = \int xp \, dx$, and how they compare with the nondiffusive result (Equation (5)). The agreement is close to perfect even after the first timescale t_1 is reached. It is only once the very long second timescale t_2 is reached that $\langle x \rangle$ approaches 0 exponentially, far more rapidly than the $t^{-\frac{1}{n-1}}$ power law scaling in (5).
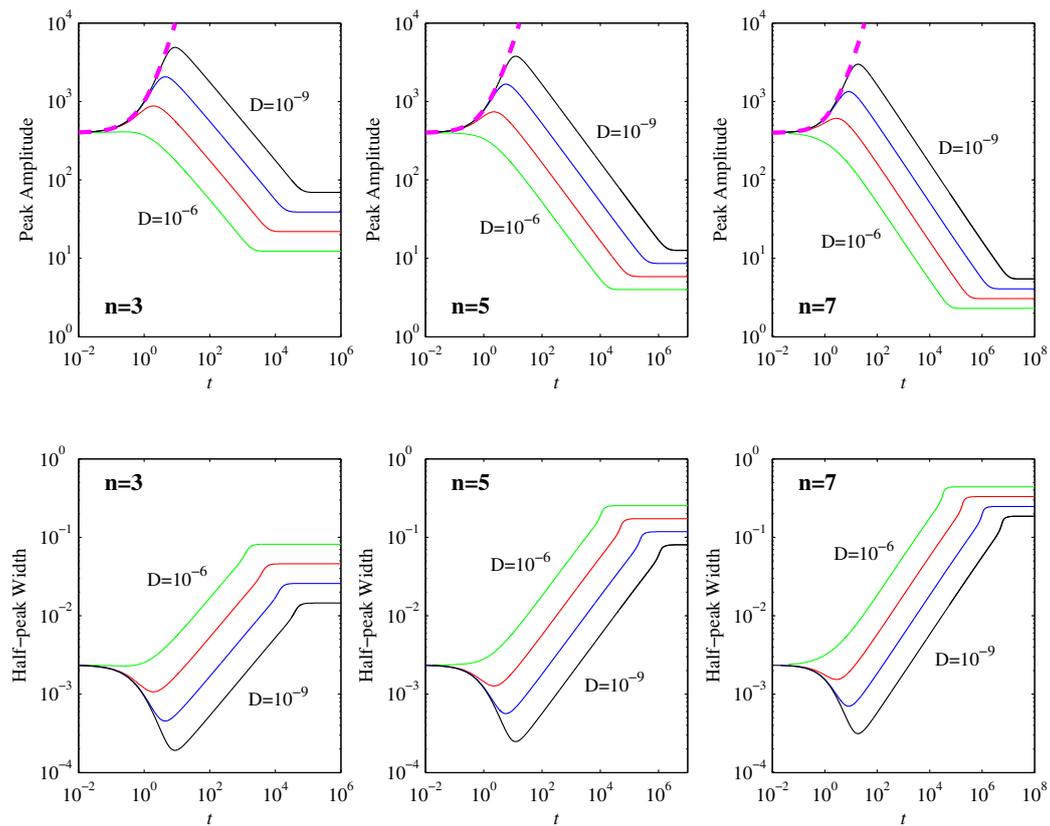
Figure 1. The top row shows the peak amplitudes as functions of time, for the initial condition in Equation (34) with $\mu = 0.65$, and $D = 10^{-6}$ to 10^{-9} as indicated. The three panels show $n = 3, 5, 7$, as labeled. The thick dashed (magenta) lines correspond to the analytic result (Equation (15)) that applies in the nondiffusive phase. The bottom row shows the equivalent widths at half-peak, which are inversely proportional to the peak amplitudes. The standard deviation $\langle (x - \langle x \rangle)^2 \rangle^{1/2}$ follows exactly the same pattern as the half-peak widths.
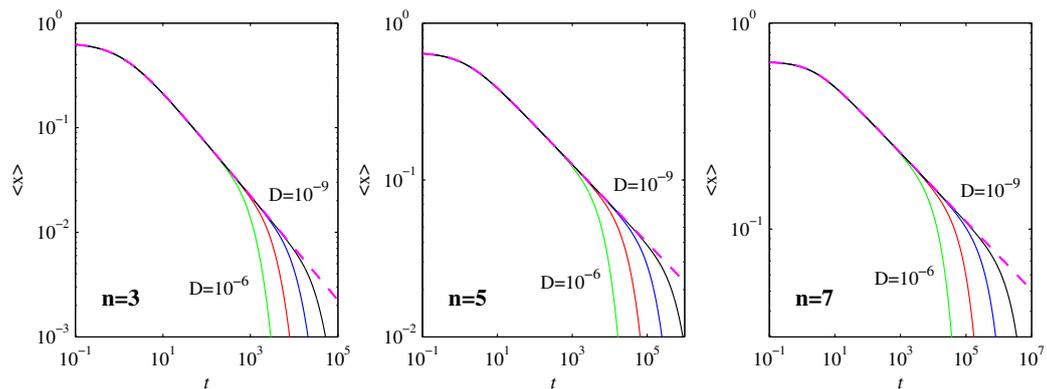

Figure 2. The mean values $\langle x \rangle$ as functions of time, for the solutions in Figure 1. The dashed (magenta) lines show Equation (5) that applies in the nondiffusive case.

Table 1. The first two rows show the overall peak amplitudes p_{Max} and the times t_{Max} at which they occur, for the results in Figure 1. $D = 10^{-8}$ and 10^{-9}, and $n = 3, 5, 7$ as indicated, and all results at $\mu = 0.65$. The row labeled "Exponent" uses the ratios of the two D values to extract scaling exponents of the form $D^{-\alpha}$. The final row compares these numerically deduced exponents with the analytic predictions from Equations (26) and (29). That is, the exponent α should equal $\frac{n-1}{3n-1}$ for t_{Max}, and $\frac{n}{3n-1}$ for p_{Max}.

	$n = 3$		$n = 5$		$n = 7$	
	t_{Max}	p_{Max}	t_{Max}	p_{Max}	t_{Max}	p_{Max}
$D = 10^{-8}$	4.42	2063	5.82	1661	8.17	1337
$D = 10^{-9}$	8.80	4888	12.55	3776	18.58	2998
Exponent	0.30	0.375	0.33	0.357	0.36	0.349
(26), (29)	0.25	0.375	0.29	0.357	0.30	0.350

Table 2. The first two rows are as in Table 1, but now for fixed $D = 10^{-9}$, $\mu = 0.55$ and 0.75, and $n = 3, 5, 7$ as indicated. The row labeled "Exponent" uses the ratios of the two μ values to extract scaling exponents of the form μ^{δ}. The final row compares these numerically deduced exponents with the analytic predictions from Equations (26) and (29). That is, the exponent δ should equal $-\frac{2n(n-1)}{3n-1}$ for t_{Max}, and $\frac{n(n-1)}{3n-1}$ for p_{Max}.

	$n = 3$		$n = 5$		$n = 7$	
	t_{Max}	p_{Max}	t_{Max}	p_{Max}	t_{Max}	p_{Max}
$\mu = 0.55$	11.17	4314	19.74	2975	35.87	2105
$\mu = 0.75$	7.18	5439	8.48	4631	10.46	4035
Exponent	-1.4	0.75	-2.7	1.43	-4.0	2.10
(26), (29)	-1.5	0.75	-2.9	1.43	-4.2	2.10

Figure 3 shows two commonly used diagnostic quantities, the skewness $\int [(x - \langle x \rangle)/\sigma]^3 \, p \, dx$ and the kurtosis $\int [(x - \langle x \rangle)/\sigma]^4 \, p \, dx$, where $\sigma = [\int (x - \langle x \rangle)^2 \, p \, dx]^{1/2}$ is the standard deviation. Skewness is a measure of how asymmetric a PDF is about its peak, with a value of 0 indicating a symmetric peak. Kurtosis measures the flatness of a PDF, especially in comparison with a Gaussian, which has kurtosis $= 3$. We see that skewness ≈ 0 and kurtosis ≈ 3 is maintained all the way until the final equilibration timescale t_2 is reached, indicating that the PDFs remain largely Gaussian up to this time, as predicted in Section 3.2. For the final equilibrated profiles in Equation (30), the skewness is again 0, whereas the kurtosis has values that depend on n (the precise values can be evaluated analytically in terms of Γ functions, but are not particularly insightful). It is interesting though to note in Figure 3 that both skewness and kurtosis exhibit non-monotonic behavior during the final adjustment process, where the PDFs transition from being largely Gaussian to their final form. Note also how the maxima reached during this transition process are independent of D, and even broadly similar for the different values of n.

Figure 4 compares the PDFs at the times when the skewness reaches its maximum (negative) value with the final equilibrated profiles in Equation (30). The entire evolution is then clear: as long as the PDFs are well outside their final profiles, they remain essentially Gaussian, with widths as in Figure 1 and positions as in Figure 2. The skewness and kurtosis start to deviate significantly from their Gaussian values when the PDFs start to reach their final positions, with the maximum skewness as in Figure 4. The final rearrangement for $t > t_2$ then merely adjusts to the final profiles (30), but with relatively little further movement of the PDFs.
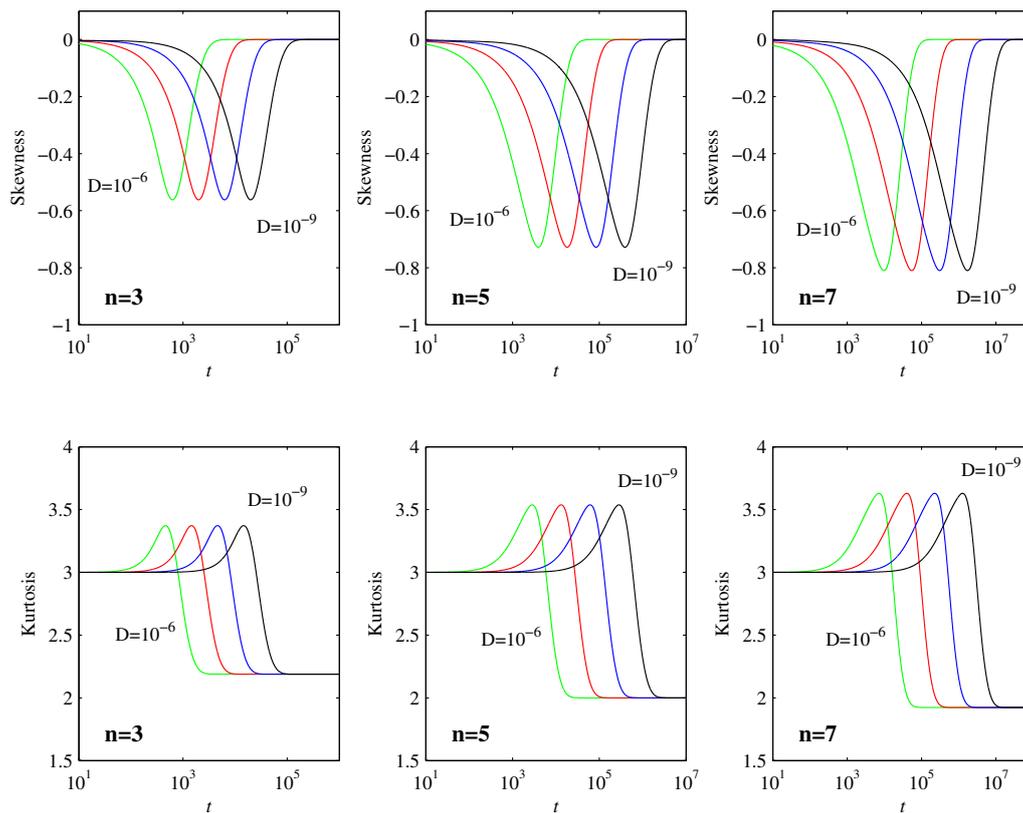
Figure 3. The top row shows the skewness $\int [(x - \langle x \rangle)/\sigma]^3 \, p \, dx$, and the bottom row shows the kurtosis $\int [(x - \langle x \rangle)/\sigma]^4 \, p \, dx$, as functions of time. The labeling of D and n is as in Figures 1 and 2. The peaks in both quantities occur at times in essentially perfect agreement with t_2 in (32).
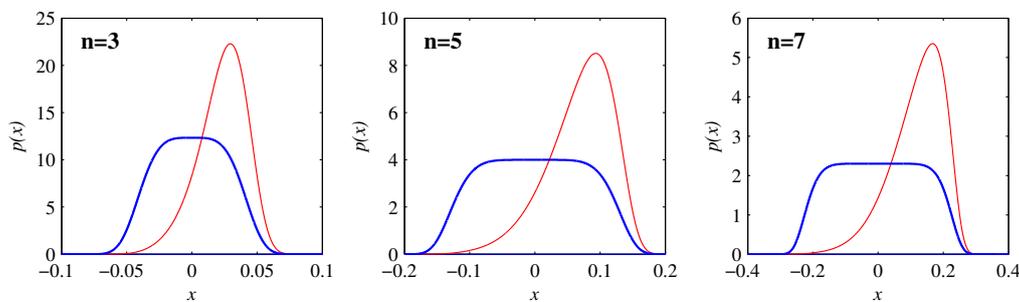

Figure 4. The heavy (blue) lines show the equilibrium profiles in Equation (30), for $D = 10^{-6}$, and $n = 3, 5, 7$ as indicated. The lighter (red) lines show $p(x)$ at the $t = 629, 3930, 9794$, for $n = 3, 5, 7$, respectively. As shown in Figure 3, these are the times when the skewness reaches its maximum negative values. Results for other values of D are identical, once x and $p(x)$ are rescaled as in Equation (31), and t is shifted as in Figure 3 to consistently have the correct skewness values.

Figure 5 shows the diagnostic quantities $\mathcal{E}(t)$ and $\mathcal{L}(t)$. As predicted by Equation (17), $\mathcal{E}$ remains constant in the initial nondiffusive phase. Once the first timescale t_1 is reached, $\mathcal{E}$ decreases as $t^{-\frac{3n-1}{n-1}} D^{-1}$. Once the second timescale t_2 is reached, $\mathcal{E}$ decreases exponentially. (This is not included in Figure 5, however, as $\mathcal{E} \propto D^{\frac{2(n-1)}{n+1}}$ at that point, and is thus already negligibly small.) The behavior of $\mathcal{L}$ is as expected: while $\mathcal{E}$ is constant, $\mathcal{L}$ increases linearly in time, and, once $\mathcal{E}$ starts decreasing, $\mathcal{L}$ levels off to its final value $\mathcal{L}_\infty$.
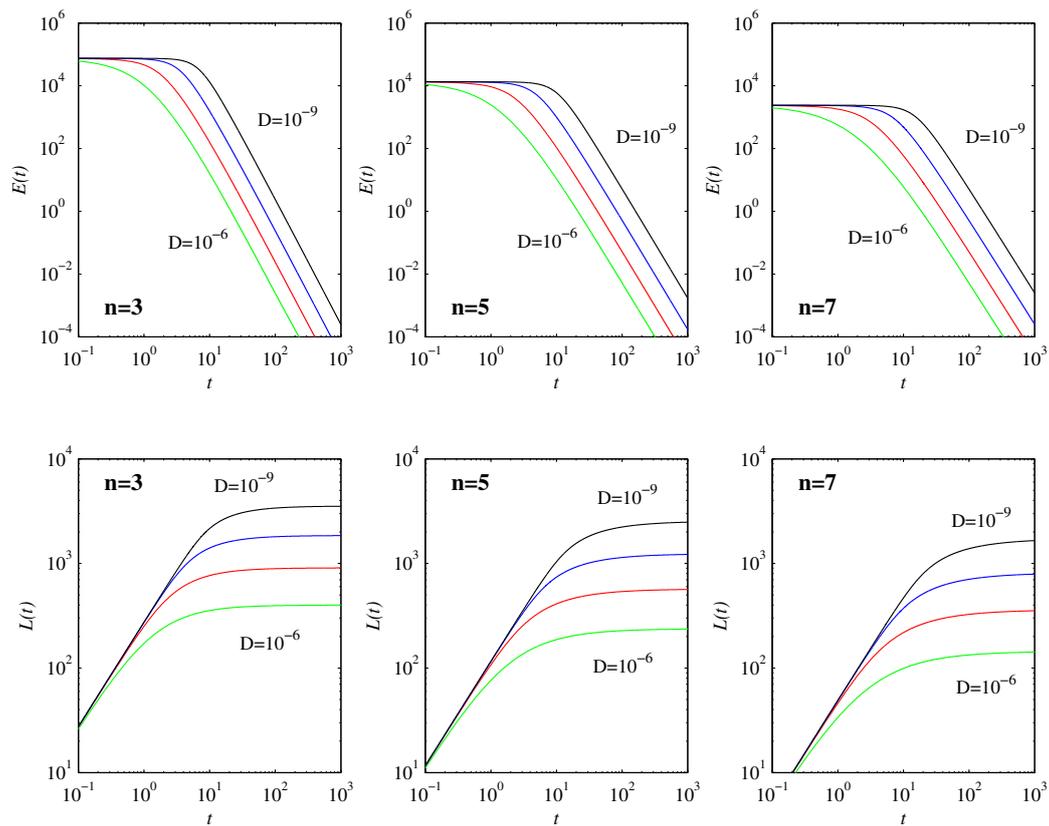
Figure 5. The top row shows $\mathcal{E}(t)$ and the bottom row shows $\mathcal{L}(t)$. The labeling of D and n is as in Figures 1–3. The peaks are initially located at $\mu = 0.65$, consistent with the initial plateau in $\mathcal{E}$ being reduced by a factor of $0.65^4 = 0.1785$ when comparing $n = 3$ and 5, and similarly $n = 5$ and 7.

Based on the scaling for $\mathcal{E}$, namely that $\mathcal{E} \propto \mu^{2n}$ initially, and the range t_1 for which this applies, we expect at least a lowest-order estimate for $\mathcal{L}_\infty$ to be given simply by

$$\mathcal{L}_\infty \propto \sqrt{\mathcal{E}}\, t_1 \propto D^{-\frac{n-1}{3n-1}}\, \mu^{\frac{n(n+1)}{3n-1}}. \tag{35}$$

Figure 6 shows numerically computed results, for the same $D = 10^{-6}$ to 10^{-9} range considered throughout, and $\mu = [0.01, 0.75]$. At least for sufficiently large μ, we do indeed see clear evidence of power-law behavior, with a negative exponent for D and a positive one for μ. Concentrating on the variation with μ, for $n = 3$, the slopes vary between 1.8 for $D = 10^{-6}$ and 1.6 for $D = 10^{-9}$; for $n = 5$, they vary between 2.7 and 2.3; and for $n = 7$, they vary between 3.5 and 2.9. By comparison, according to Equation (35), the exponent should be $\frac{n(n+1)}{3n-1}$, which yields 1.5, 2.1, 2.8 for $n = 3, 5, 7$, respectively. The agreement is thus quite good, especially when we recall (Tables 1 and 2) that even $D = 10^{-9}$ was not quite small enough yet to obtain precise agreement with the asymptotic formula for t_1, which in turn affects Equation (35) as well.

$\mathcal{L}_\infty$ as a function of μ provides a mapping from the physical distance μ (the distance between the peak position μ of an initial PDF and the final equilibrium point 0) to the information length (the total number of statistically different states of a system reaching a final stationary PDF from an initial PDF). From Equation (35), this mapping is linear for a linear force; that is, $\mathcal{L}_\infty$ is linearly proportional to the physical distance. This linear relation breaks down for nonlinear forces as $\mathcal{L}_\infty$ depends nonlinearly on the physical distance μ. Specifically, for $n = 3$, $\mathcal{L}_\infty \propto \mu^{1.5}$; for $n = 5$, $\mathcal{L}_\infty \propto \mu^{2.1}$; and, for $n = 7$, $\mathcal{L}_\infty \propto \mu^{2.8}$. Thus, we can envision that nonlinear forces affect the information geometry, changing a linear (flat) coordinate to a power-law (curved) coordinate. This is reminiscent of gravity changing a

flat to a curved space-time. Furthermore, interestingly, Equation (35) shows that $\mathcal{L}_\infty$ is independent of D for the linear force, manifesting the information geometry as independent of the resolution (set by D). In contrast, $\mathcal{L}_\infty$ decreases with $n = 3, 5, 7$ as $D^{-1/4}$, $D^{-2/7}$, $D^{-3/10}$, suggesting that the information geometry is fractal, depending on the resolution. This would be equivalent to the I theorem of [14].

Finally, what about the small μ limit in Figure 6, which clearly does not follow the expected scaling (35)? The explanation is that the initial position μ is already within the $O(D^{\frac{1}{n+1}})$ width of the final distribution in Equation (30). In this limit, the behavior is different, and the peak merely spreads out, resulting in a small and relatively uniform $\mathcal{L}_\infty$. For any given μ, D must therefore satisfy $D \ll \mu^{n+1}$ to be in the regime where Equation (35) is expected to apply. That is, in principle, it would be possible to have Equation (35) apply over several orders of magnitude in μ, but D would have to be far smaller than is numerically feasible.
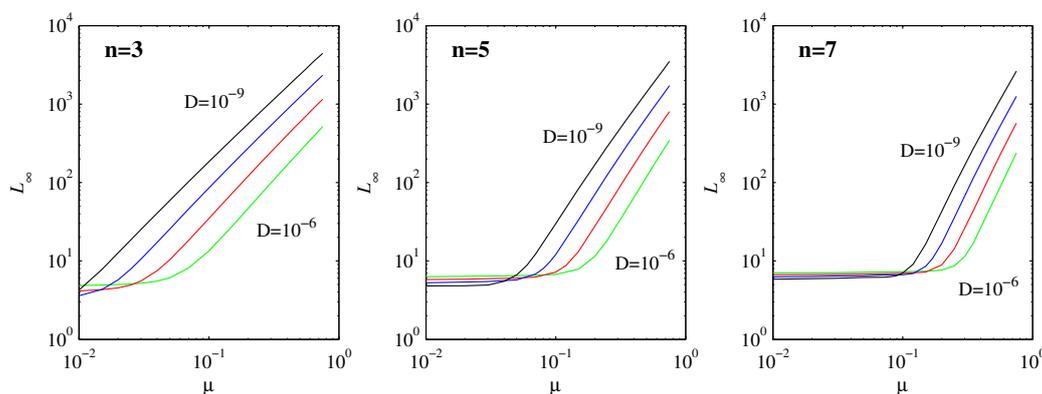

Figure 6. $\mathcal{L}_\infty$ as a function of the initial position μ, for $D = 10^{-6}$ to 10^{-9} and $n = 3, 5, 7$ as labeled.

5. Conclusions

The motivation of this research was to elucidate the link between force and geometry in a stochastic system. To this end, we investigated the transient relaxation of initial PDFs of a stochastic variable x under different nonlinear forces proportional to $-x^n$ ($n = 3, 5, 7$) and different strength D of δ-correlated stochastic noise. We identified the three main stages consisting of nondiffusive evolution, quasi-linear Gaussian evolution and settling into stationary PDFs. The strength of the stochastic noise is shown to play a crucial role in determining these timescales $t_1 \propto D^{-\frac{n-1}{3n-1}}$ and $t_2 \propto D^{-\frac{n-1}{n+1}}$, as well as the peak amplitude and width of PDFs. Note in particular how both timescales have (n-dependent) power-law scalings with D, in sharp contrast with the linear $n = 1$ case, where the entire evolution occurs on $t = O(1)$ timescales, with no dependence on D.

We further computed the rate of information change, and the information length $\mathcal{L}(t)$ representing the number of distinguishable statistical states that the system evolves through in time. We identified a robust geodesic (where the information changes at a constant rate) in the initial nondiffusive stage, and mapped out a geometric structure of an attractor as $\mathcal{L}_\infty \propto \mu^m$, where μ is the peak position of the initial Gaussian PDF. For sufficiently small D the exponent m was shown to be $\frac{n(n+1)}{3n-1}$, but still exhibits (moderate) variation with D even for values as small as 10^{-9}. This suggests that the geometry is curved by the nonlinear interaction, in contrast with the linear geometry of the Ornstein-Uhlenbeck process which has $m = 1$ and no variation with D. Our results thus highlight ubiquitous power-laws and multi-scalings of information geometry.

This work will form a basis for future investigation of a more general case of the superposition of nonlinear forces with different ns. Multiplicative noise would also be interesting, but the situation becomes somewhat more complicated as a deterministic force could appear as $-x^n$ as here, but the noise ξ could appear as ξx^m, with m not necessarily the same as n. Fractional noise would also be worth pursuing, although this would be challenging both analytically and numerically. Finally, it will

also be interesting to extend our work to analyse a heterogeneous, linear system [35–37], and to explore applications of our work to estimators.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gangbo, W.; McCann, R.J. Optimal maps in Monge's mass transport problem. *Comptes Rendus Acad. Sci. Paris* **1995**, *321*, 1653–1658.
2. Jordan, R.; Kinderlehrer, D.; Otto, F. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **1998**, *29*, 1–17. [CrossRef]
3. Otto, F. The geometry of dissipative evolution equations: The porous medium equation. *Comm. Part. Diff. Equ.* **2001**, *26*, 101–174. [CrossRef]
4. Gibbs, A.L.; Su, F.E. On choosing and bounding probability metrics. *Int. Stat. Rev.* **2002**, *70*, 419–435. [CrossRef]
5. Frieden, B.R. *Science from Fisher Information*; Cambridge University Press: Cambridge, UK, 2004.
6. Lott, J. Some geometric calculations on Wasserstein space. *Commun. Math. Phys.* **2008**, *277*, 423–437. [CrossRef]
7. Takatsu, A. Wasserstein geometry of Gaussian measures. *Osaka J. Math.* **2011**, *48*, 1005–1026.
8. Ferradans, S.; Xia, G.-S.; Peyré, G.; Aujol, J.-F. Static and dynamic texture mixing using optimal transport. In *Lecture Notes in Computer Science*; Kuijper, A., Bredies, K., Pock, T., Bischof, H., Eds.; Springer: Berlin, Germany, 2013; Volume 7893.
9. Costa, S.; Santos, S.; Strapasson, J. Fisher information distance. *Discret. Appl. Math.* **2015**, *197*, 59–69. [CrossRef]
10. Chevallier, E.; Kalunga, E.; Angulo, J. Kernel density estimation on spaces of Gaussian distributions and symmetric positive definite matrices. *SIAM J. Imaging Sci.* **2017**, *10*, 191–215. [CrossRef]
11. Feng, E.H.; Crooks, G.E. Far-from-equilibrium measurements of thermodynamic length. *Phys. Rev. E* **2009**, *79*, 012104. [CrossRef] [PubMed]
12. Wilson, A.D.; Schultz, J.A.; Murphey, T.D. Trajectory synthesis for Fisher information maximization. *IEEE Trans. Robot.* **2014**, *30*, 1358–1370. [CrossRef] [PubMed]
13. Kim, E.; Lee, U.; Heseltine, J.; Hollerbach, R. Geometric structure and geodesic in a solvable model of nonequilibrium process. *Phys. Rev. E* **2016**, *93*, 062127. [CrossRef] [PubMed]
14. Plastino, A.R.; Casas, M.; Plastino, A. Fisher's information, Kullback's measure, and H-theorems. *Phys. Lett. A* **1998**, *246*, 498–504. [CrossRef]
15. Polettini, M.; Esposito, M. Nonconvexity of the relative entropy for Markov dynamics: A Fisher information approach. *Phys. Rev. E.* **2013**, *88*, 012112. [CrossRef] [PubMed]
16. Wootters, W.K. Statistical distance and Hilbert space. *Phys. Rev. D* **1981**, *23*, 357–362. [CrossRef]
17. Ruppeiner, G. Thermodynamics: A Riemannian geometric model. *Phys. Rev. A* **1979**, *20*, 1608–1613. [CrossRef]
18. Schlögl, F. Thermodynamic metric and stochastic measures. *Z. Phys. B Cond. Matt.* **1985**, *59*, 449–454. [CrossRef]
19. Braunstein, S.L.; Caves, C.M. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.* **1994**, *72*, 3439–3443. [CrossRef] [PubMed]
20. Nulton, J.; Salamon, P.; Andresen, B.; Anmin, Q. Quasistatic processes as step equilibrations. *J. Chem. Phys.* **1985**, *83*, 334–338. [CrossRef]
21. Sivak, D.A.; Crooks, G.E. Thermodynamic metrics and optimal paths. *Phys. Rev. Lett.* **2012**, *8*, 190602. [CrossRef] [PubMed]
22. Salamon, P.; Nulton, J.D.; Siragusa, G.; Limon, A.; Bedeaus, D.; Kjelstrup, S. A simple example of control to minimize entropy production. *J. Non-Equilib. Thermodyn.* **2002**, *27*, 45–55. [CrossRef]

23. Nicholson, S.B.; Kim, E. Investigation of the statistical distance to reach stationary distributions. *Phys. Lett. A* **2015**, *379*, 83–88. [CrossRef]

24. Nicholson, S.B.; Kim, E. Structures in sound: Analysis of classical music using the information length. *Entropy* **2016**, *18*, 258. [CrossRef]

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

26. Kim, E.; Hollerbach, R. Signature of nonlinear damping in geometric structure of a nonequilibrium process. *Phys. Rev. E* **2017**, *95*, 022137. [CrossRef] [PubMed]

27. Hollerbach, R.; Kim, E. Information geometry of non-equilibrium processes in a bistable system with a cubic damping. *Entropy* **2017**, *19*, 268. [CrossRef]

28. Kim, E.; Tenkès, L.-M.; Hollerbach, R.; Radulescu, O. Far-from-equilibrium time evolution between two gamma distributions. *Entropy* **2017**, *19*, 511. [CrossRef]

29. Tenkès, L.-M.; Hollerbach, R.; Kim, E. Time-dependent probability density functions and information geometry in stochastic logistic and Gompertz models. *J. Stat. Mech.* **2017**, *2017*, 123201. [CrossRef]

30. Kim, E.; Lewis, P. Information length in quantum system. *J. Stat. Mech.* **2018**, *2018*, 043106. [CrossRef]

31. Wilde, M.M. *Quantum Information Theory*; Cambridge University Press: Cambridge, UK, 2017; Volume 345, pp. 424–427.

32. Ott, E. *Chaos in Dynamical Systems*; Cambridge University Press: Cambridge, UK, 2002.

33. Kim, E.; Hollerbach, R. Time-dependent probability density function in cubic stochastic processes. *Phys. Rev. E* **2016**, *94*, 052118. [CrossRef] [PubMed]

34. Risken, H. *The Fokker-Planck Equation: Methods of Solution & Applications*; Springer: Berlin, Germany, 1996.

35. Cherstvy, A.G.; Chechkin, A.V.; Metzler, R., Anomalous diffusion and ergodicity breaking in heterogeneous diffusion processes. *New J. Phys.* **2010**, *15*, 083039. [CrossRef]

36. Andrey G.; Cherstvy, A.G.; Metzler, R. Population splitting, trapping, and non-ergodicity in heterogeneous diffusion processes. *Phys. Chem. Chem. Phys.* **2013**, *15*, 20220.

37. Sandev, T.; Schulz, A.; Kantz, H.; Iomin, A. Heterogeneous diffusion in comb and fractal grid structures. *Chaos Solitons Fractals* **2017**. [CrossRef]