


Article

Comparing Information Metrics for a Coupled Ornstein–Uhlenbeck Process

James Heseltine * and Eun-jin Kim * 

School of Mathematics and Statistics, University of Sheffield, Sheffield S3 7RH, UK

* Correspondence: jheseltine1@sheffield.ac.uk (J.H.); ejk92122@gmail.com (E.-j.K.)

Received: 9 July 2019; Accepted: 6 August 2019; Published: 8 August 2019



Abstract: It is often the case when studying complex dynamical systems that a statistical formulation can provide the greatest insight into the underlying dynamics. When discussing the behavior of such a system which is evolving in time, it is useful to have the notion of a metric between two given states. A popular measure of information change in a system under perturbation has been the relative entropy of the states, as this notion allows us to quantify the difference between states of a system at different times. In this paper, we investigate the relaxation problem given by a single and coupled Ornstein–Uhlenbeck (O-U) process and compare the information length with entropy-based metrics (relative entropy, Jensen divergence) as well as others. By measuring the total information length in the long time limit, we show that it is only the information length that preserves the linear geometry of the O-U process. In the coupled O-U process, the information length is shown to be capable of detecting changes in both components of the system even when other metrics would detect almost nothing in one of the components. We show in detail that the information length is sensitive to the evolution of subsystems.

Keywords: stochastic processes; Langevin equation; Fokker–Planck equation; information length; Fisher information; metrics; O-U process; probability density function

1. Introduction

Describing many natural systems statistically can give great insight into the system’s dynamics, when uncertainty or degrees of freedom are too high to do otherwise. Measures of information change can be particularly useful in understanding the evolution of a system under perturbation, or comparing data (e.g., see [1]). Here, by information, we specifically refer to a measurable, statistical difference between the states of a system, defined by probability density functions (PDFs), and avoid any of the more diaphanous definitions of the term. The statistical difference can be quantified by assigning a metric to probability, which then endows a stochastic system with a geometric structure. Previously, different metrics (e.g., Refs. [2–10]) have been considered depending on the question of interest.

A popular measure of the information change in a system would be entropy, which measures the uncertainty or ‘disorder’ of the system. More specifically, it is a measure of the number of states that are accessible from the current state. Comparing entropy at different times gives a measure of the difference in information for the system, called the relative entropy. We can use this relative entropy as a metric. Another example is the Wasserstein metric, which was used to optimize transport cost in the optimal transport problem [4,6]; for Gaussian PDFs, the Wasserstein metric is defined in the product space consisting of Euclidean and positive symmetric matrices for the mean and variance, respectively (e.g., see [2]). The link between the Fisher information [8] and the Wasserstein distance was made in [6] where the integral of the Fisher information [8] along the Ornstein–Uhlenbeck semigroup was shown to be the same as the Wasserstein distance. Furthermore [1] stated that relative entropy was the integral of Fisher information along the same path.

However, the way in which the relative entropy has mostly been used in the past lacks a sense of locality to a metric of the system as it focuses on quantifying the difference between two given PDFs, for instance, PDFs at time t_1 and t_2 . As a result, they are independent of the intermediate PDFs between time t_1 and t_2 (the history/path of a system), and thus can only inform us about changes which affect the overall structure of the system. The work of [11] was, in part, a search for a disequilibrium component for a statistical complexity measure (SCM). In short, an SCM is a measure of both the ‘order’ and ‘disorder’ of a system, which can help to reveal hidden structures of a disordered system. They proposed several metrics ‘disorder’ or disequilibrium component of the SCM.

In this paper we compare several of the proposed metrics of [11] with the information length \mathcal{L} [12–20]. The information length, proportional to the time integral of the square root of the infinitesimal relative entropy, depends on the intermediate states between $t = t_1$ and $t = t_2$ and is thus a Lagrangian measure. Also, the formulation of the information length allows us to measure local change for the system in time. \mathcal{L}_∞ , the total information length over the entire evolution $t = t_1 = 0$ and $t = t_2 \rightarrow \infty$, was shown to be useful to quantify the proximity of any initial PDF to a final attractor of a dynamical system. For instance, for the Ornstein–Uhlenbeck process (O-U) [16,18], \mathcal{L}_∞ was shown to take its minimum value at the stable equilibrium point and increase linearly with the mean position of an initial PDF from the stable equilibrium point. This linear dependence manifests that the information length preserves the linear geometry of the underlying Gaussian process. In this paper, we will show that this linear relation is lost for other metrics (e.g., relative entropy, Jensen divergence). Note that for a chaotic attractor, \mathcal{L}_∞ varies sensitively with the mean position of a narrow initial PDF, taking its minimum value at the most unstable point [21]. This sensitive dependence of \mathcal{L}_∞ on the initial PDF is similar to a Lyapunov exponent.

We note that the O-U is a prototypical relaxation problem and can be particularly useful to study, as its attractor provides a natural equilibrium state. It can model many stochastic systems which relax to a stable equilibrium. The solution to this process is Gaussian, and so has ‘nice’ properties of analytical tractability, permitting us to perform detailed investigation under the change of parameters. We first compare different metrics for a single O-U process then move to a coupled O-U process. The O-U process is a well studied model, though less so for the coupled system. Our focus is to compare different metrics and to see if the information length may be more revealing the behavior of the components of the coupled system, as well as the overall system. The remainder of this paper is organized as follows. Section 2 provides the definition of different metrics. Section 3 is devoted to the discussion of a single O-U process. Section 4 provides analytical solutions to the coupled O-U process and Section 5 compares different metrics for the coupled O-U process. Conclusions are found in Section 6. In Appendix A, we present how to solve the Fokker–Planck equation(s) numerically by using a second-order accuracy method in time and compare analytical results with numerical results. Appendix B comments on the Langevin equation for our coupled O-U process.

2. Information Length and Other Metrics

We consider a PDF $P(x, t)$ for a stochastic variable x in the following.

2.1. Information Length

The information length $\mathcal{L}(t)$ between time 0 and t is given by

$$\mathcal{L}(t) = \int_0^t \frac{dt'}{\tau(t')} = \int_0^t dt' \sqrt{\int_{-\infty}^{\infty} dx \left(\frac{\partial P(x, t')}{\partial t'} \right)^2 \frac{1}{P(x, t')}} \quad (1)$$

where $\frac{1}{|\tau(t)|^2}$ is the second moment given by,

$$\frac{1}{|\tau(t)|^2} = \int_{-\infty}^{\infty} dx \left(\frac{\partial P(x, t)}{\partial t} \right)^2 \frac{1}{P(x, t)}. \quad (2)$$

Here, $\tau(t)$ has the unit of time while \mathcal{L} has no dimension. The parameter $\tau(t)$ is the characteristic timescale of the system, and quantifies the correlation time for the system [15]. Hence, $\frac{1}{\tau(t)}$ is the rate of change of the information in time. Integrating $\frac{1}{\tau(t)}$ over $[0, t]$ gives the total number of statistically different states that a system passes through in time. We note that \mathcal{L} quantifies the information change in time through the root-mean-squared fluctuating energy, using the second moment of the partial derivative with respect to time. When the parameters governing a PDF are known, the information length $\mathcal{L}(t)$ can be written in terms of the Fisher information metric [8,12,15].

As noted in Section 1, $\mathcal{L}(t)$ is a Lagrangian quantity and has the property of being a local measure, being sensitive to how $P(x, t)$ evolves at different x in time. In comparison with entropy which is independent of the spatial (x) gradient of a PDF, it is this property that may elevate $\mathcal{L}(t)$ above entropy in revealing micro-scale interactions within a system.

The discrete version of Equations (1) and (2) are as follows:

$$\mathcal{L}_n = h \sum_{i=0}^n \frac{1}{\tau_i}, \quad (3)$$

$$\frac{1}{\tau_i^2} = \frac{s}{h^2} \sum_j P_i^j \left[\ln(P_{i+1}^j) - \ln(P_i^j) \right]^2. \quad (4)$$

Here, i and j represent the discrete time and spatial point, respectively; P_i^j is the discrete version of $P(x, t)$, $h = t/n$ is the time step while s is the spatial step.

2.2. Other Metrics

Here, we list the metrics taken from [11]. We will calculate each metric relative to the initial state PDF $P(x, 0)$ at $t = 0$ in order to compare their time-evolution with that of the information length. That is, the metrics below are based on comparing two PDFs $P(x, 0)$ and $P(x, t)$. For convenience, the metrics are given both for the continuous process and discrete approximation (that is used for numerical calculation) by using i, j as an index representing time and space, respectively and s as the spatial step, as above. The reference probability $P(x, 0)$ [P_0] will be the initial PDF while $P(x, t)$ [P_i] is the PDF at time t [i].

2.2.1. Euclidean norm

$$\begin{aligned} \|P(x, t) - P(x, 0)\|^2 &= \int dx [P(x, 0) - P(x, t)]^2, \\ \|P_0 - P_i\|^2 &= s \sum_j [P_0^j - P_i^j]^2. \end{aligned} \quad (5)$$

Applying our standard notion of distance in Euclidean space seems like a natural extension. However, it quickly becomes apparent that the statistical space of stochastic systems is rarely well described by Euclidean metrics. Mostly included here as a base case, whilst this formulation seems appealing, it does not yield illuminating details about the disequilibrium of the system. We will use it as an example of a poor measure of information change.

2.2.2. Wootters' Distance

$$W[(P(x,0), (P(x,t))] = \cos^{-1} \left[\int dx [P(x,0)]^{\frac{1}{2}} [P(x,t)]^{\frac{1}{2}} \right], \quad (6)$$

$$W[P_0, P_i] = \cos^{-1} \left[s \sum_j [P_0^j]^{\frac{1}{2}} [P_i^j]^{\frac{1}{2}} \right].$$

This metric, as the notion of statistical distance itself, originates in quantum information theory [9]. However, as quantum information theory is purely statistical in formulation, it can be applied to any systems defined by a PDF. Fundamentally, this metric is based on the principle that any finite number of measurements on a stochastic system will yield results that may not be exactly the same as underlying probability distributions. It would be impossible to distinguish 2 states whose real underlying probabilities are different less than a typical fluctuation of the error of measurement. This intrinsically defines a resolution for the system. The Wootters' distance was shown to be a monotone transformation of the Hellinger distance [22].

2.2.3. Kullback-Leibler relative entropy

$$K(P(x,0)|P(x,t)) = \int dx P(x,0) \ln \left(\frac{P(x,0)}{P(x,t)} \right), \quad (7)$$

$$K[P_0|P_i] = s \sum_j P_0^j \log \left(\frac{P_0^j}{P_i^j} \right).$$

Kullback–Leibler relative entropy was first introduced by Solomon Kullback and Richard Leibler [10], and sometimes is referred to as the Kullback–Leibler divergence. It represents a measure of the difference between a probability distribution and some other reference probability distribution. Whilst a useful tool, it is not strictly a metric as it does not satisfy the triangle inequality. It is however used in the definition of some other quantities, such as the mutual information of two co-varying random variables, and the Jensen divergence.

2.2.4. Jensen Divergence

$$J(P(x,0)|P(x,t)) = \frac{1}{2} [K(P(x,0)|P(x,t)) + K(P(x,t)|P(x,0))], \quad (8)$$

$$J[P_0|P_i] = \frac{1}{2} [K[P_0|P_i] + K[P_i|P_0]].$$

The Jensen divergence is simply the symmetric version of the Kullback–Leibler divergence. Often it is referred to as the Jensen distance, and the square root of this quantity can be shown to be a metric [11], which can allow us to examine the statistical geometry of the system. The Jensen divergence is the mutual information of a random variable x , with a mixture distribution from $P(x,0)$ and $P(x,t)$, and a binary indicator variable used to build the distribution. In other words it is a measure of the mutual dependence of x on the way you construct the mixture, and thus quantifies the amount of information difference between the two distributions.

3. The O-U Process

The one-dimensional O-U process is based on the Langevin equation

$$\dot{x} = -\gamma(x - \mu) + \xi, \tag{9}$$

where x is the stochastic variable (e.g., position, velocity, etc.), γ is the damping constant, μ is the position of the attractor for the system, and ξ is a δ -correlated, Gaussian-distributed stochastic forcing, i.e.,

$$\langle \xi(t_1)\xi(t_2) \rangle = 2D\delta(t_1 - t_2), \tag{10}$$

where D is the strength of the stochastic forcing. The corresponding Fokker–Planck equation [23,24] is given by

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}[\gamma(x - \mu)P] + D\frac{\partial^2 P}{\partial x^2}, \tag{11}$$

where the solution $P = P(x, t)$ is the time-dependant PDF which describes the evolution of the system. It can be shown that the solution to Equation (11) is given by [13]

$$P(x, t) = \sqrt{\frac{\beta}{\pi}} e^{-\beta(x - \langle x \rangle)^2}, \tag{12}$$

where $\frac{1}{2\beta} = \frac{e^{-2\gamma t}}{2\beta_0} + \frac{D}{\gamma}(1 - e^{-2\gamma t})$ and $\langle x \rangle = x_0 e^{-\gamma t}$, given the initial condition

$$P(x, 0) = \sqrt{\frac{\beta_0}{\pi}} e^{-\beta_0(x - x_0)^2}, \tag{13}$$

where $\beta = 1/2\langle(x - \langle x \rangle)^2\rangle$ and $\langle x \rangle$ in Equation (12) represent the inverse temperature and the mean value of x , respectively, and β_0 and x_0 in Equation (13) are the values of β and $\langle x \rangle$ at $t = 0$, respectively.

3.1. Information Length

In [13], we showed that the information length $\mathcal{L}(t)$ for the O-U process is given by

$$\mathcal{L} = \frac{1}{\sqrt{2}} \left[\ln \left(\frac{y - r}{y + r} \right) \right]_{y_i}^{y_f} + \frac{\sqrt{2}}{r} \int_{y_i}^{y_f} \frac{qr - r^2}{y^2 + qr - r^2} dy, \tag{14}$$

where $y = \sqrt{r^2 + qT}$, $r = 2\beta_0 D - \gamma$, $q = \beta_0 \gamma x_0^2$, and $T = \beta_0 D(e^{2\gamma t} - 1) + \gamma$. y_i is y evaluated at the initial time ($t = 0$ in our case) and y_f is y evaluated at final time. x_0 is the initial mean position.

Let the integral in the last term of Equation (14) be H , and let $r \neq q \neq 0$. Then Equation (14) can be written as

$$\mathcal{L} = \frac{1}{\sqrt{2}} \left[\ln \left(\frac{y - r}{y + r} \right) \right]_{y_i}^{y_f} + \frac{\sqrt{2}}{r} H, \tag{15}$$

where

$$H = \begin{cases} \sqrt{qr - r^2} \tan^{-1} \left(\frac{Y}{\sqrt{qr - r^2}} \right) & \text{if } qr - r^2 > 0, \\ -\frac{\sqrt{r^2 - rq}}{2} \ln \left(\frac{Y - \sqrt{r^2 - rq}}{Y + \sqrt{r^2 - rq}} \right) & \text{if } qr - r^2 < 0. \end{cases} \tag{16}$$

Note that this is continuous through $q = r$. For $q = 0$, we can directly compute

$$\mathcal{L} = \frac{1}{\sqrt{2}} \frac{|r|}{r} \ln \left(\frac{T}{T + r} \right). \tag{17}$$

For $r = 0$ we have

$$\mathcal{L} = -\sqrt{2q} \left[\frac{1}{\sqrt{T}} \right]_{T_i}^{T_f}. \tag{18}$$

3.2. Wootters' Distance

By using Equations (12) and (13) in (6), we obtain

$$\begin{aligned} W[P(x,0), P(x,t)] &= \left[\frac{\beta_0\beta}{\pi^2} \right]^{\frac{1}{4}} \int dx e^{-\frac{\beta_0}{2}(x-x_0)^2 - \frac{\beta}{2}(x-\langle x \rangle)^2} \\ &= \left[\frac{2\sqrt{\beta_0\beta}}{\beta_0 + \beta} \right]^{\frac{1}{2}} \exp\left(-\frac{\beta\beta_0(x_0 - \langle x \rangle)^2}{2(\beta_0 + \beta)}\right), \end{aligned} \tag{19}$$

where $\langle x \rangle = x_0 e^{-\gamma t}$.

3.3. Kullback–Leibler Relative Entropy

By using Equations (12) and (13) in (7), we can show

$$\begin{aligned} K[P(x,0)|P(x,t)] &= \int dx P(x,0) \left[\ln \sqrt{\frac{\beta_0}{\beta}} e^{-\beta_0(x-x_0)^2 + \beta(x-\langle x \rangle)^2} \right] \\ &= \ln \sqrt{\frac{\beta_0}{\beta}} + \int dx P(x,0) \left[-\beta_0(x-x_0)^2 + \beta(x-\langle x \rangle)^2 \right] \\ &= \ln \sqrt{\frac{\beta}{\beta_0}} + \beta(x_0 - \langle x \rangle)^2 + \frac{\beta}{2\beta_0} - \frac{1}{2}, \end{aligned} \tag{20}$$

where $\langle x \rangle = x_0 e^{-\gamma t}$.

3.4. Jensen Divergence

By using Equations (12) and (13) in (8), we obtain

$$J[P(x,0)|P(x,t)] = \left(\frac{\beta + \beta_0}{2} \right) (x_0 - \langle x \rangle)^2 + \frac{\beta^2 + \beta_0^2}{4\beta_0\beta} - \frac{1}{2}. \tag{21}$$

3.5. Comparison

Figure 1 shows the final value for each metric as we vary the initial position x_0 of the system. The total information length $\mathcal{L}_\infty = \mathcal{L}(t \rightarrow \infty)$, Wootters' distance, K-L relative entropy and Jensen divergence against x_0 are shown in blue, orange, green and red, respectively, in the long time limit as $t \rightarrow \infty$. Note that the green and red lines are overlapped. It is notable in Figure 1 that the linear relation between the metric and the initial mean value x_0 is obtained only by the information length. That is, it is only the information length that preserves the linear geometry underlying a linear stochastic process. For all other metrics, this linear relation is lost. We show in the Appendix A that our analytical metrics in Figure 1 has a good agreement with those calculated directly from the numerical solutions to the Fokker–Planck Equation (11) by time-stepping (see Figure A1).

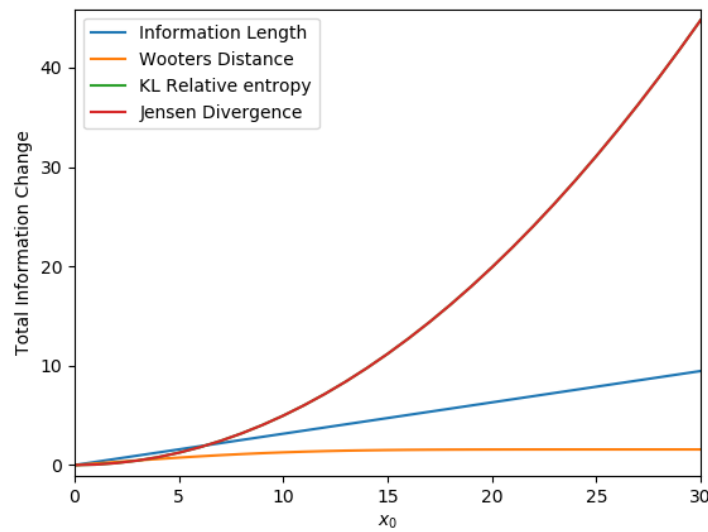


Figure 1. The metrics against x_0 in the long time limit for a single Ornstein–Uhlenbeck (O-U) process.

4. The Coupled O-U Process

We now consider the coupled system of equations

$$\frac{\partial P_1}{\partial t} = \frac{\partial}{\partial x} [\gamma_1(x - \mu)P_1] + D \frac{\partial^2 P_1}{\partial x^2} - f_0 P_1 + g_0 P_2, \tag{22}$$

$$\frac{\partial P_2}{\partial t} = \frac{\partial}{\partial x} [\gamma_2(x - \mu)P_2] + D \frac{\partial^2 P_2}{\partial x^2} + f_0 P_1 - g_0 P_2, \tag{23}$$

where D is the strength of a short-correlated Gaussian noise given by Equation (10). These equations are a pair of O-U processes, linked by coupling constants f_0 and g_0 . The coupling f_0 and g_0 are due to the Dichotomous noise [25] (see Appendix B for the Langevin equation corresponding to Equations (22) and (23)).

We choose a coupled system like this one to examine the localized dynamics of these interacting sub-processes. This system could model any process for which there are two competing components relaxing to an equilibrium, like evaporation in a closed system, or a reversible chemical reaction.

Since we are mainly interested in the relaxation process from non-equilibrium initial states, we choose the different initial conditions for P_1 and P_2 while for simplicity, considering the case where $\gamma_1 = \gamma_2 = \gamma$ and $f_0 = g_0 = \epsilon$ where the Fokker–Planck Equations (22) and (23) are reduced to

$$\frac{\partial}{\partial t} P_1(x, t) = \frac{\partial}{\partial x} [\gamma P_1(x, t)] + D \frac{\partial^2}{\partial x^2} P_1(x, t) + \epsilon(P_2 - P_1), \tag{24}$$

$$\frac{\partial}{\partial t} P_2(x, t) = \frac{\partial}{\partial x} [\gamma P_2(x, t)] + D \frac{\partial^2}{\partial x^2} P_2(x, t) + \epsilon(P_1 - P_2). \tag{25}$$

Specifically, as initial conditions, we use the following different Gaussian PDFs

$$P_1(x, 0) = \frac{1}{2} \sqrt{\frac{\beta_{10}}{\pi}} \exp[-\beta_{10}x^2], \tag{26}$$

$$P_2(x, 0) = \frac{1}{2} \sqrt{\frac{\beta_{20}}{\pi}} \exp[-\beta_{20}(x - x_0)^2]. \tag{27}$$

Note that β_{10} and β_{20} are the initial inverse temperatures for P_1 and P_2 , respectively. On the other hand, we fix the initial mean of P_1 to be zero while the initial P_2 is taken to have any arbitrary mean value x_0 . We also note that as $t \rightarrow \infty$, P_1 and P_2 approach the same PDF.

To solve Equations (24) and (25), we take the Fourier transform [$\tilde{P}_m(k, t) = \int dx e^{ikx} P_m(x, t)$ for $m = 1, 2$] and use the characteristic equation to recast Equations (24) and (25) as

$$\frac{d\tilde{P}_1}{dt} = -Dk^2\tilde{P}_1 + \epsilon(\tilde{P}_2 - \tilde{P}_1), \quad (28)$$

$$\frac{d\tilde{P}_2}{dt} = -Dk^2\tilde{P}_2 + \epsilon(\tilde{P}_1 - \tilde{P}_2). \quad (29)$$

Here $\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{dk}{dt} \frac{\partial}{\partial k}$ is the total derivative along the characteristic defined by

$$\frac{dk}{dt} = \gamma k, \quad (30)$$

which has the solution

$$k(t) = k_0 e^{\gamma t}, \quad (31)$$

where $k_0 = k(0)$ is the initial wavenumber. We solve Equations (28) and (29) in terms of new variables

$$\bar{P}_m = \tilde{P}_m \exp\left(\epsilon t + D \int_0^t dt_1 k(t_1)^2\right), \quad (32)$$

for $m = 1, 2$. The coupled Equations (28) and (29) are then simplified as

$$\frac{d\bar{P}_1}{dt} = \epsilon \bar{P}_2, \quad (33)$$

$$\frac{d\bar{P}_2}{dt} = \epsilon \bar{P}_1. \quad (34)$$

We write down the solutions to Equations (33) and (34) using the two constants a and b

$$\bar{P}_1 = ae^{\epsilon t} + be^{-\epsilon t}, \quad (35)$$

$$\bar{P}_2 = ae^{\epsilon t} - be^{-\epsilon t}. \quad (36)$$

To determine a and b , we take the Fourier transform of the initial conditions Equations (26) and (27) to obtain

$$\tilde{P}_1(0) = \frac{1}{2} e^{-\frac{k_0^2}{4\beta_2}}, \quad (37)$$

$$\tilde{P}_2(0) = \frac{1}{2} e^{ik_0 x - \frac{k_0^2}{4\beta_1}}. \quad (38)$$

Thus, evaluating Equation (32) for $m = 1, 2$ at $t = 0$ and equating them to Equations (37) and (38), we find

$$a = \frac{1}{2} \tilde{P}_1(0) + \frac{1}{2} e^{-ik_0 x} \tilde{P}_2(0), \quad (39)$$

$$b = \frac{1}{2} \tilde{P}_1(0) - \frac{1}{2} e^{-ik_0 x} \tilde{P}_2(0). \quad (40)$$

On the other hand, using Equation (31) in Equation (32), we write \tilde{P}_m ($m = 1, 2$) in terms of \bar{P}_m as

$$\tilde{P}_m = \exp\left[-\epsilon t - \frac{D}{2\gamma}(1 - e^{-2\gamma t})k^2\right] \bar{P}_m. \quad (41)$$

Finally, taking the inverse Fourier transform $[P_m(x, t) = \frac{1}{2\pi} \int dx e^{-ikx} \tilde{P}_m(x, t)$ for $m = 1, 2]$ and performing several Gaussian integrals, we obtain

$$P_1(x, t) = \frac{1}{4} \left[\sqrt{\frac{\beta_1}{\pi}} (1 + e^{-2\epsilon t}) e^{-\beta_1 x^2} + \sqrt{\frac{\beta_2}{\pi}} (1 - e^{-2\epsilon t}) e^{-\beta_2 (x - e^{-\gamma t} x_0)^2} \right], \tag{42}$$

$$P_2(x, t) = \frac{1}{4} \left[\sqrt{\frac{\beta_1}{\pi}} (1 - e^{-2\epsilon t}) e^{-\beta_1 x^2} + \sqrt{\frac{\beta_2}{\pi}} (1 + e^{-2\epsilon t}) e^{-\beta_2 (x - e^{-\gamma t} x_0)^2} \right], \tag{43}$$

where

$$\frac{1}{2\beta_m} = \frac{e^{-2\gamma t}}{2\beta_{m0}} + \frac{D}{\gamma} (1 - e^{-2\gamma t}), \tag{44}$$

for $m = 1, 2$. We can check that at $t = 0$, Equations (42) and (43) recover Equations (26) and (27). On the other hand, in the limit of $t \rightarrow \infty$, Equations (42) and (43) give us

$$P_m(x, t) = \frac{1}{2} \sqrt{\frac{\beta_m(t \rightarrow \infty)}{\pi}} e^{-\beta_m(t \rightarrow \infty) x^2}, \tag{45}$$

which is the stationary solution to a single O-U process where $\beta_m(t \rightarrow \infty) = \frac{\gamma}{2D}$. We note that the total PDF $P = P_1 + P_2$ is the solution of this single O-U process with the initial condition given by the sum of Equations (26) and (27).

Using these analytical solutions in Equations (42)–(44), we present the different metrics in Equations (1) and (5)–(8) in Section 5.

5. Results for the Coupled O-U Process

For the illustration in this section, we show results for the fixed parameter values $\gamma = 0.1$, $D = 1$ and $\epsilon = 0.5$. We recall from Section 4 that for these parameter values, in equilibrium (see Equation (44)), $\beta_1 = \beta_2 = \beta(t \rightarrow \infty) = \frac{\gamma}{2D} = 0.05$ while the mean values in Equation (45) are zero for both P_1 and P_2 . For comparing metrics, we consider the case where P_1 is initially in the final equilibrium with the zero mean value and inverse temperature $\beta_{10} = 0.05$. On the other hand, P_2 at $t = 0$ is taken to have either different mean values x_0 or different inverse temperatures β_{20} . Here, we present results obtained by using analytical solutions in Section 4 only. (See Appendix A for the numerical solutions and comparison with the analytical solutions.)

5.1. Varying β_{20}

We first examine how the system changes when varying the initial inverse temperature β_{20} of P_2 for the fixed zero mean position ($x_0 = 0$) (the equilibrium value). We investigate what changes are detected by each metric.

Figure 2 shows the total information length (in blue), Euclidean norm (in orange), Wootters' distance (in green), K-L relative entropy (in red), and Jensen divergence (in purple) against β_{20} in the long time limit as $t \rightarrow \infty$. Panels (a), (b) and (c) are for the overall system P , and the components P_1 and P_2 , respectively. Since P_1 is initially chosen to be in its equilibrium state, the behavior for the overall system P in panel (a) is more similar to P_2 in panel (c) than P_1 in panel (b). What is prominent in both panels (a) and (c) is the presence of the distinct minimum in the information length around $\beta_{20} = 0.05$. This is because the final equilibrium has the inverse temperature 0.05, demonstrating that the total information also maps out the underlying attractor structure when varying β_{20} , taking its minimum value at the equilibrium state (reminiscent of the results for the single O-U process above and previous works [13,16,18]). The minimum value around $\beta_{20} = 0.05$ is also observed for other metrics (apart from the Euclidean norm) although less pronounced than the total information length. In fact, for $\beta_{20} = \beta_{10} = \beta(t \rightarrow \infty)$, there is no temporal change in both P_1 and P_2 , so all the metrics

apart from the Wootters' one take the value of zero. Also, of interest is an almost linear increase in the total information length for P_2 in panel (c) as $|\beta_{20} - 0.05|$ increases.

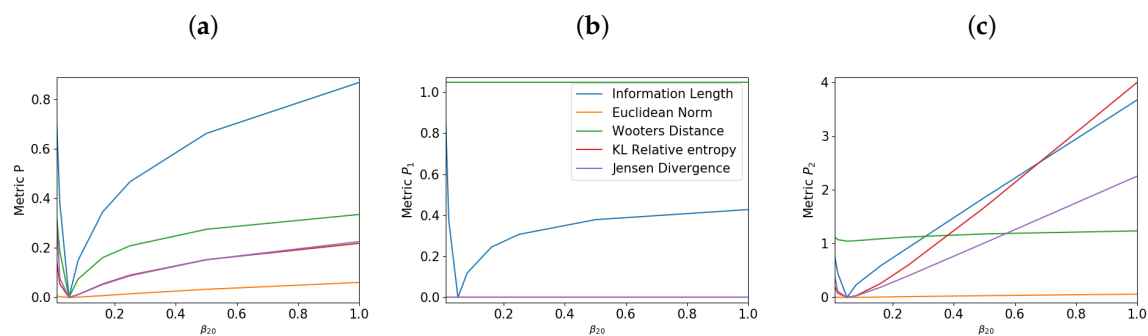


Figure 2. Behavior of the metrics for varying β_{20} for the overall system in panel (a) and components P_1 in panel (b) and P_2 in panel (c).

Now, what is happening to P_1 which starts with the equilibrium state ($\beta_{10} = 0.05$ and zero mean value)? While Equation (44) shows that $\beta_1 = \beta_{10} = \gamma/2D = 0.05$ for all time, the actual PDF P_1 in Equation (42) changes with time due to its coupling to P_2 via ϵ . That is, P_1 changes over time, initially deviating from the equilibrium state due to the interaction with P_2 and then finally relaxing back to the final equilibrium state. Associated with this time evolution of P_1 is the information change which can be measured by different metrics. However, Figure 2b shows that for P_1 , it is only information length that detects a noticeable difference in the information change as β_{20} changes. Furthermore, the information length for P_1 takes the minimum value for the equilibrium value of $\beta_{20} = 0.05$, as was the case for P_2 . This result demonstrates that the information length is sensitive to the evolution of the component P_1 (a subsystem).

To investigate the evolution of the metrics for the P_1 around β_{20} , we show in Figure 3 how the metrics evolve over time for the component P_1 for the two different values of $\beta_{20} = 0.02$ and 0.08 near the equilibrium value $\beta_{20} = 0.05$. Panel a) is for $\beta_{20} = 0.02 < 0.05$ and panel b) is for $\beta_{20} = 0.08 > 0.05$. The different metrics are denoted by using the same color as used in Figure 2. This small deviation of β_{20} from the equilibrium value induces a small change in P_1 over time due to the coupling to P_2 , as noted above. However, in panels a) and b), we see a significant change in the Euclidean norm, K-L relative entropy, and Jensen divergence in time, with a large increase before settling down to a lower value. In comparison, the information length shows no such spike. These spikes are caused by the deviation of P_1 from its initial equilibrium state due to the coupling to P_2 before settling into the equilibrium state as P_2 approaches the equilibrium. What is interesting is that the information length does not show such a spike since it measures the local information change; this would thus be the more sensible view in this instance since the change in the component P_1 is small compared to its width (uncertainly). Furthermore, such spikes do not appear for P_2 nor P (results not shown) since P_2 starting from a non-equilibrium state monotonically approaches its equilibrium.

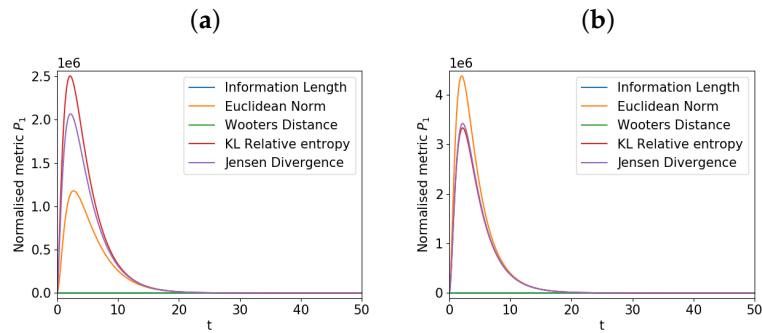


Figure 3. The metrics against time for P_1 around $\beta_{20} = 0.05$. $\beta_{20} = 0.02$ in panel (a) and $\beta_{20} = 0.08$ in panel (b). The Y-axis scaling on the panels is 10^6 .

5.2. Varying x_0

We now fix $\beta_{20} = 0.05$ and vary the initial mean position x_0 of P_2 to examine how metrics depend on x_0 , as we have done for the single O-U process in Section 3. Figure 4 shows the total information change for each metric for different values of x_0 , for P , P_1 and P_2 in panels (a), (b), and (c) respectively. Specifically, Figure 4a shows that for P , the total information length against x_0 is linear, capturing the linearity of the system as expected from the single O-U process. None of the other metrics are capable of showing the linear relationship in the same way. On the other hand, Figure 4b,c shows an interesting non-monotonic dependence of the information length on x_0 .

To understand this, we show in Figure 5 the time evolution of P , P_1 , and P_2 in panels (a), (b), and (c), respectively, by using $x_0 = 20$. Of interest is that the evolution of P_1 and P_2 in Figure 5b,c involves the formation of the two peaks from the initial one peak, followed by merging of these two peaks into one peak as a system settles into the equilibrium. This formation of the two peaks is due to the interaction between P_1 and P_2 when they are initially widely separated for a sufficiently large $x_0 \gtrsim 20$. The formation of two PDFs peaks for $x_0 \gtrsim 20$ leads to the maximum in the total information length around $x_0 = 20$ in Figure 4b,c. Specifically, the formation of the two peaks in P_1 and P_2 shown in Figure 5b,c takes place when the two peak are a full PDF width apart, and facilitates broadening of both PDFs in the relaxation process. As x_0 is further increased from 20, P_1 and P_2 form two peaks which are more widely separated, leading to P_1 and P_2 becoming effectively broader. This in turn reduces the information length (as x_0 increases further from $x_0 = 20$) since large fluctuations (uncertainty) associated with a broad PDF reduces the information length. Again, the information length is the only measure which detects the difference in the overall information change for P_1 due to its sensitive dependence on the local evolution of a system.

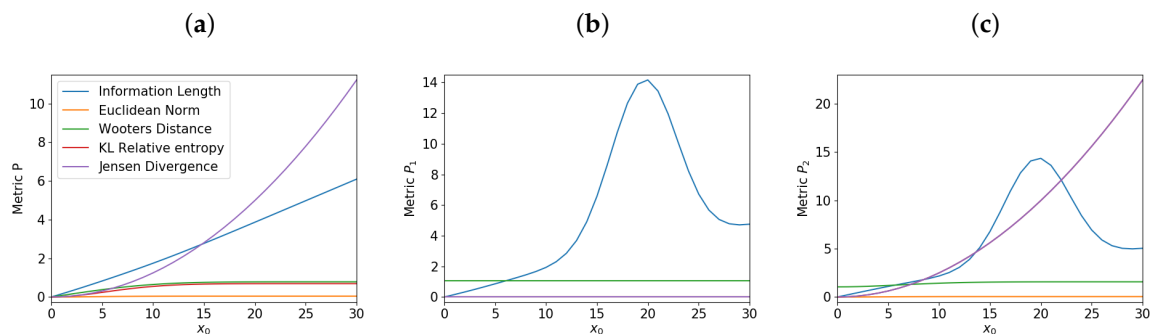


Figure 4. Behavior of the metrics for varying x_0 for the overall system P in panel (a), P_1 in (b), and P_2 in (c).

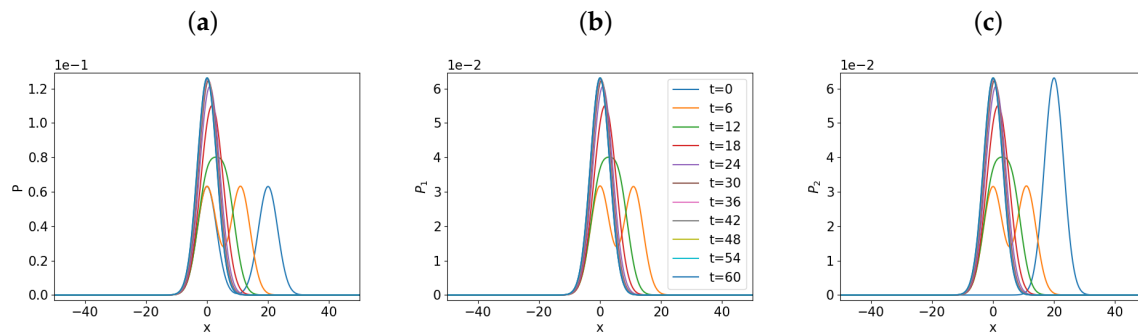


Figure 5. Time-dependent partial differential equations (PDFs) for P in panel (a), for P_1 in panel (b) and for P_2 in panel (c); $x_0 = 20$.

6. Conclusions

When searching for a way to quantify the information change in a given dynamical system, our choices are many and varied. Our aim here was to show the power of the information length \mathcal{L} when compared with some of the more popular methods of measuring information change. Utilizing the O-U, we compared several relative-entropy formulations with our information length to investigate what each could reveal about the dynamics of system.

Specifically, we investigated the relaxation problem given by a single and coupled O-U process and compared the information length \mathcal{L} with K-L relative entropy, Jensen divergence, Wootters' distance, and Euclidean norm. By measuring the total information length in the long time limit, we showed that \mathcal{L} was unique in detecting the linear spatial relationship between the total information change and the initial position of a PDF. In the coupled O-U process, the information length was shown to be the most effective in detecting changes in the components of the system even when the others would detect almost nothing in one of the components. In particular, when P_1 started with an equilibrium state with the zero mean value, the formation of the two peaks of P_1 (or P_2) from an initial one peak P_1 (or P_2) and the merging of the two peaks into one peak as a system settled into equilibrium was detected only by the information length with its intriguing non-monotonic dependence on x_0 (the mean value of P_2). This underscores the sensitivity of the information length on the evolution of subsystems.

Future work will include the study of a system with multiple attractor positions for the system or how the system behaves when changing the position of the attractor. It would also be interesting to examine the case where the coupling parameters f_0 and g_0 are not constant, but are functions of time. This could result in a periodic equilibrium where the PDF varies between 2 or more unstable states. This could represent physical systems like reversible chemical reactions, or even fluctuating financial markets. It would also be of interest to investigate implication of the information length for the deep neural network [26], in particular, to elucidate the role of geodesic along which the information length is minimized [27].

Author Contributions: Conceptualization, E.-j.K.; methodology, J.H. and E.-j.K.; software, J.H.; investigation, J.H. and E.-j.K.; writing—original draft preparation, J.H.; writing—review and editing, E.-j.K.; visualization, J.H.; supervision, E.-j.K.

Funding: This research was funded by Engineering & Physical Sciences Research Council Doctoral Training Partnership Scholarship (J.H.) and by Leverhulme Trust Research Fellowship RF-2018-142-9 (E.-j.K.).

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

Appendix A

In this appendix, we show the numerical solutions to the Fokker-Planck equation by time-stepping and metrics using these numerical solutions for the O-U process and the coupled O-U processes.

Appendix A.1. The O-U Process

We use the finite difference method to approximate the partial derivatives from Equation (11), where h is our time-step, s is our spatial step, i is the current time step, and j is the current spatial step.

$$\frac{P_{i+1}^j - P_i^j}{h} = \gamma \frac{(x_j - x_0)[P_i^{j+1} - P_i^j]}{s} + D \frac{P_i^{j+1} - 2P_i^j + P_i^{j-1}}{s^2}. \quad (\text{A1})$$

Then using a second-order Runge-Kutta Method, we create a time-stepped solution. Specifically, we first create an intermediate value \tilde{P} by

$$(\tilde{P})_{i+1}^j = P_i^j + hf(t_i, P_i^j), \quad (\text{A2})$$

where

$$f(t_i, P_i^j) = \gamma \frac{(x_j - x_0)[P_i^{j+1} - P_i^j]}{s} + D \frac{P_i^{j+1} - 2P_i^j + P_i^{j-1}}{s^2}, \quad (\text{A3})$$

which is the time-gradient for the system. Here, t_i is the current time value. Then we use the intermediate value to calculate the next time step value, P_{i+1}^j , using the formula

$$P_{i+1}^j = P_i^j + \frac{h}{2}[f(t_i, P_i^j) + f(t_{i+1}, \tilde{P})_{i+1}^j]. \quad (\text{A4})$$

For simplicity, we set $D = 1$. This is to consider the case of the one parameter system, as the O-U process can always be arranged to combine γ and D into a single parameter. For our stepping, we use $h = 0.04$ and $s = 0.1$. We check on the convergence of our solutions by reducing h and s . We use the initial condition given by Equation (13). All solutions are coded in python. We present the numerically computed metrics from the time-stepping in Figure A1, which shows the final value for each metric as we vary the initial position of the system, x_0 . Figure A1 is quite similar to the analytically calculated metrics shown in Figure 1, demonstrating a good agreement between the analytical and numerical solutions.

To test the accuracy of our time-stepped model further, Figure A2 compares the time evolution of different metrics for the largest x_0 value, i.e., the largest perturbation from equilibrium in our testing, with analytic and time-stepped solutions on the left and right respectively. The agreement between the analytical and numerical solutions is observed to be quite good.

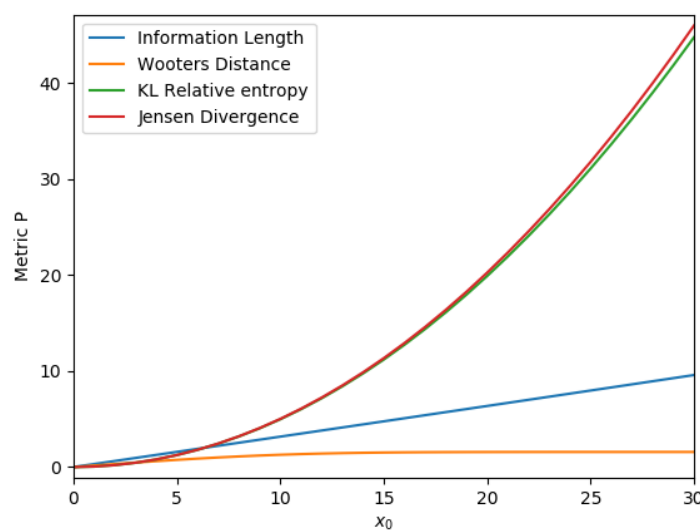


Figure A1. The metrics calculated by the numerical solutions to the Fokker–Planck equation by time-stepping.

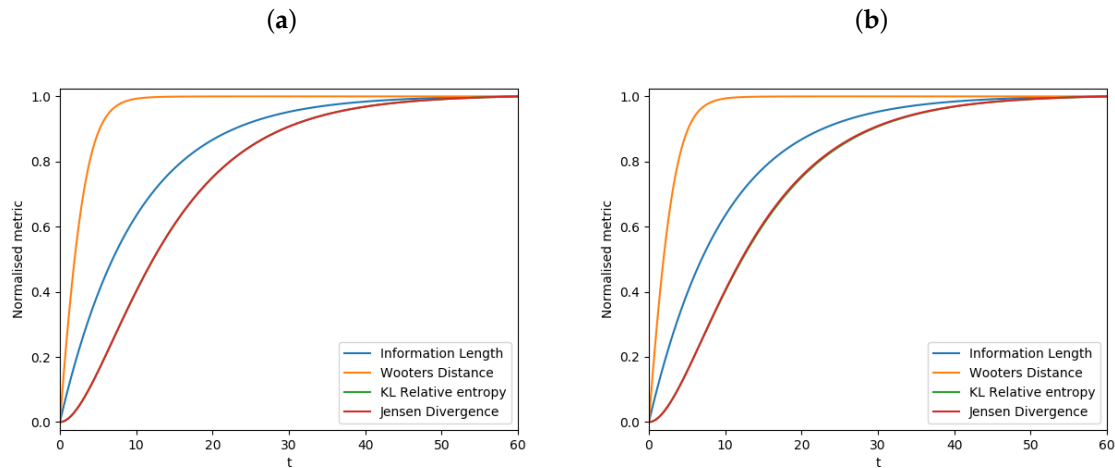


Figure A2. Comparing the evolution of the metrics with $x_0 = 30$. Panels (a) and (b), respectively, are obtained by using the analytical and numerical solutions to the Fokker–Planck equation.

Appendix A.2. Coupled O-U Process

We solve (22) and (23) numerically by time-stepping, using the same method as for the single equation case. Our discrete versions of (22) and (23) become

$$\frac{(P_1)_{i+1}^j - (P_1)_i^j}{h} = \gamma_1 \frac{(x_j - \mu_1)[(P_1)_i^{j+1} - (P_1)_i^j]}{s} + D_1 \frac{(P_1)_i^{j+1} - 2(P_1)_i^j + (P_1)_i^{j-1}}{s^2} - \epsilon(P_1)_i^j + \epsilon(P_2)_i^j \quad (A5)$$

$$\frac{(P_2)_{i+1}^j - (P_2)_i^j}{h} = \gamma_2 \frac{(x_j - \mu_2)[(P_2)_i^{j+1} - (P_2)_i^j]}{s} + D_2 \frac{(P_2)_i^{j+1} - 2(P_2)_i^j + (P_2)_i^{j-1}}{s^2} + \epsilon(P_1)_i^j - \epsilon(P_2)_i^j, \quad (A6)$$

where h is our time-step, and s is our spatial-step. We then use Huen’s method to approximate the time-evolution for the system. So our final numerical model becomes

$$(\tilde{P}_1)_{i+1}^j = (P_1)_i^j + hf_1(t_i, (P_1)_i^j, (P_2)_i^j), \quad (A7)$$

$$(\tilde{P}_2)_{i+1}^j = (P_2)_i^j + hf_2(t_i, (P_1)_i^j, (P_2)_i^j), \quad (A8)$$

$$(P_1)_{i+1}^j = (P_1)_i^j + \frac{h}{2}[f_1(t_i, (P_1)_i^j, (P_2)_i^j) + f_1(t_{i+1}, (\tilde{P}_1)_{i+1}^j, (\tilde{P}_2)_{i+1}^j)], \quad (A9)$$

$$(P_2)_{i+1}^j = (P_2)_i^j + \frac{h}{2}[f_2(t_i, (P_1)_i^j, (P_2)_i^j) + f_2(t_{i+1}, (\tilde{P}_1)_{i+1}^j, (\tilde{P}_2)_{i+1}^j)], \quad (A10)$$

where

$$f_1(t_i, (P_1)_i^j, (P_2)_i^j) = \gamma_1 \frac{(x_j - \mu_1)[(P_1)_i^{j+1} - (P_1)_i^j]}{s} + D_1 \frac{(P_1)_i^{j+1} - 2(P_1)_i^j + (P_1)_i^{j-1}}{s^2} - \epsilon(P_1)_i^j + \epsilon(P_2)_i^j \quad (A11)$$

$$f_2(t_i, (P_1)_i^j, (P_2)_i^j) = \gamma_2 \frac{(x_j - \mu_2)[(P_2)_i^{j+1} - (P_2)_i^j]}{s} + D_2 \frac{(P_2)_i^{j+1} - 2(P_2)_i^j + (P_2)_i^{j-1}}{s^2} - \epsilon(P_2)_i^j + \epsilon(P_1)_i^j, \quad (A12)$$

and $(\tilde{P}_k)_{i+1}^j$ are the intermediate values. Here we set $D_1 = D_2 = 1$ and use $h = 0.04$ and $s = 0.1$.

We give our system components a Gaussian initial condition, given by

$$(P_1)_0^j = \frac{1}{2} \sqrt{\frac{\beta_1}{\pi}} e^{-\beta_1(x_j - v_1)^2}, \quad (A13)$$

$$(P_2)_0^j = \frac{1}{2} \sqrt{\frac{\beta_2}{\pi}} e^{-\beta_2(x_j - v_2)^2}, \quad (A14)$$

where β_1 and β_2 are the inverse temperature for P_1 and P_2 respectively, and ν_1 and ν_2 are their initial mean positions.

When varying the system parameters, we fix the parameters for P_1 at equilibrium ($\nu_1 = 0$, $\beta_1 = 0.05$, $\gamma_1 = 0.1$) and vary the parameters for P_2 (ν_2 , β_2). The choice between fixing P_1 or P_2 is arbitrary. We solve the Fokker–Planck equations for the coupled O-U process (Equations (22) and (23)) by time-stepping and then numerically calculated the metrics. These are shown to have a good agreement with the results using the analytic formulation of the PDF (Equations (42) and (43)) in Section 5.

Appendix B

The Langevin equation corresponding to Equations (22) and (23) is given by

$$\partial_t x = -(x - \mu)\eta(t) + \xi. \quad (\text{A15})$$

Here, ξ is the Gaussian noise given by Equation (10), and $\eta(t)$ is the dichotomous Markov noise (DMN) [25,28] which takes the two value $\{\gamma_1, \gamma_2\}$, with constant transition rates $k_{\pm} = f_0, g_0$ between the two states. A pure DMN case is governed by

$$\partial_t x = \eta(t), \quad (\text{A16})$$

and the waiting time τ_{\pm} in the two states are exponentially-distributed stochastic variable, with the probability

$$P(\tau_{\pm}) = k_{\pm} e^{-k_{\pm} \tau_{\pm}}. \quad (\text{A17})$$

Note that $\eta(t)$ is not a white noise, but has the exponentially decaying correlation function as

$$\langle \eta(t)\eta(t') \rangle = \frac{Q}{\tau_c} \exp\left(-\frac{|t-t'|}{\tau_c}\right). \quad (\text{A18})$$

Here, $Q = k_+ k_- \tau_c^3 (\gamma_1 + \gamma_2)^2$ and $\tau_c = \frac{1}{k_+ + k_-}$ is the characteristic relaxation time to the stationary state of the DMN. For the DMN, due to a finite correlation time, an exact, analytical time-dependent PDF, in general, is unavailable. For instance, the pure DMN case given by Equation (A17) leads to a coupled linear Fokker–Planck equation (e.g., see Equations (9) and (10) in [28]), but a resulting equation for the total PDF is a differential and integral equation in time (see, e.g., Equation (11) in [28]).

References

1. Zamir, R. A proof of the Fisher information inequality via a data processing argument. *IEEE Trans. Inf. Theory* **1998**, *44*, 1246–1250. [\[CrossRef\]](#)
2. Gibbs, A.L.; Su, F.E. On choosing and bounding probability metrics. *Int. Stat. Rev.* **2002**, *70*, 419–435. [\[CrossRef\]](#)
3. Jordan, R.; Kinderlehrer, D.; Otto, F. The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* **1998**, *29*, 1–17. [\[CrossRef\]](#)
4. Lott, J. Some geometric calculations on Wasserstein space. *Commun. Math. Phys.* **2008**, *277*, 423–437. [\[CrossRef\]](#)
5. Takatsu, A. Wasserstein geometry of Gaussian measures. *Osaka J. Math.* **2011**, *48*, 1005–1026.
6. Otto, F.; Villani, C. Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality. *J. Funct. Anal.* **2000**, *173*, 361–400. [\[CrossRef\]](#)
7. Costa, S.; Santos, S.; Strapasson, J. Fisher information distance. *Discret. Appl. Math.* **2015**, *197*, 59–69. [\[CrossRef\]](#)
8. Frieden, B.R. *Science from Fisher Information*; Cambridge University Press: Cambridge, UK, 2004.
9. Wootters, K.W. Wootters, W.K. Statistical distance and Hilbert space. *Phys. Rev. D* **1981**, *23*, 357–362. [\[CrossRef\]](#)

10. Kullback, S. Letter to the Editor: The Kullback-Leibler distance. *Am. Stat.* **1951**, *41*, 340–341.
11. Kowalski, A.M.; Martin, M.T.; Plastino, A.; Rosso, O.A.; Casas, M. Distances in Probability Space and the Statistical Complexity Setup. *Entropy* **2011**, *13*, 1055–1075. [[CrossRef](#)]
12. Information Length. Available online: <https://encyclopedia.pub/238> (accessed on 29 July 2019).
13. Heseltine, J.; Kim, E. Novel mapping in non-equilibrium stochastic processes. *J. Phys. A* **2016**, *49*, 175002. [[CrossRef](#)]
14. Kim, E. Investigating Information Geometry in Classical and Quantum Systems through Information Length. *Entropy* **2018**, *20*, 574. [[CrossRef](#)]
15. Kim, E.; Lewis, P. Information length in quantum systems. *J. Stat. Mech.* **2018**, 043106. [[CrossRef](#)]
16. Kim, E.; Hollerbach, R. Signature of nonlinear damping in geometric structure of a nonequilibrium process. *Phys. Rev. E* **2017**, *95*, 022137. [[CrossRef](#)]
17. Kim, E.; Hollerbach, R. Geometric structure and information change in phase transitions. *Phys. Rev. E* **2017**, *95*, 062107. [[CrossRef](#)]
18. Hollerbach, R.; Dimanche, D.; Kim, E. Information geometry of nonlinear stochastic systems. *Entropy* **2018**, *20*, 550. [[CrossRef](#)]
19. Hollerbach, R.; Kim, E.; Mahi, Y. Information length as a new diagnostic in the periodically modulated double-well model of stochastic resonance. *Physica A* **2019**, *525*, 1313–1322. [[CrossRef](#)]
20. Kim, E.; Hollerbach, R. Time-dependent probability density function in cubic stochastic processes. *Phys. Rev. E* **2016**, *94*, 052118. [[CrossRef](#)]
21. Nicholson, S.B.; Kim, E. Investigation of the statistical distance to reach stationary distributions. *Phys. Lett. A* **2015**, *379*, 83–88. [[CrossRef](#)]
22. Matey, A.; Lamberti, P.W.; Martin, M.T.; Plastron, A. Wotters' distance resisted: A new distinguishability criterium. *Eur. Phys. J. D* **2005**, *32*, 413–419.
23. Risken, H. *The Fokker-Planck Equation: Methods of Solution and Applications*; Springer: Berlin, Germany, 1996.
24. Klebaner, F. *Introduction to Stochastic Calculus with Applications*; Imperial College Press: London, UK, 2012.
25. Bena, I. Dichotomous Markov Noise: Exact results for out-of-equilibrium systems (a brief overview). *Int. J. Mod. Phys. B* **2006**, *20*, 2825–2888. [[CrossRef](#)]
26. Shwartz-Ziv, R.; Tishby, N. Opening the Black Box of Deep Neural Networks via Information. *arXiv* **2017**, arXiv:1703.00810.
27. 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](#)]
28. Van Den Brock, C. On the relation between white shot noise, Gaussian white noise, and the dichotomic Markov process. *J. Stat. Phys.* **1983**, *31*, 467–483. [[CrossRef](#)]

