entropy

MDPI

Article

Information Length Analysis of Linear Autonomous Stochastic Processes

Adrian-Josue Guel-Cortez * and Eun-jin Kim

Centre for Fluid and Complex Systems, Coventry University, Priory St, Coventry CV1 5FB, UK; ad3116@coventry.ac.uk

* Correspondence: guelcortea@uni.coventry.ac.uk

Received: 7 October 2020; Accepted: 5 November 2020; Published: 7 November 2020

Abstract: When studying the behaviour of complex dynamical systems, a statistical formulation can provide useful insights. In particular, information geometry is a promising tool for this purpose. In this paper, we investigate the information length for n-dimensional linear autonomous stochastic processes, providing a basic theoretical framework that can be applied to a large set of problems in engineering and physics. A specific application is made to a harmonically bound particle system with the natural oscillation frequency ω, subject to a damping γ and a Gaussian white-noise. We explore how the information length depends on ω and γ, elucidating the role of critical damping $\gamma = 2\omega$ in information geometry. Furthermore, in the long time limit, we show that the information length reflects the linear geometry associated with the Gaussian statistics in a linear stochastic process.

Keywords: non-equilibrium; stochastic processes; time-dependent PDF; information length; information geometry; entropy; fluctuations

1. Introduction

Stochastic processes are common in nature or laboratories, and play a major role across traditional disciplinary boundaries (e.g., see [1,2]). These stochastic processes often exhibit complex temporal behaviour and even the emergence of order (self-organization). The latter can also be artificially designed to complete an orderly task (guided self-organization) [3–6]. In order to study and compare the dynamics of different stochastic processes and self-organization, it is valuable to utilize a measurement which is independent of any specifics of a system [7–11] (e.g., physical variables, units, dimensions, etc.). This can be achieved by using information theory based on probability density functions (PDFs) and working in terms of information content or information change, e.g., by quantifying the statistical difference between two states [12–14]. Mathematically, we do this by assigning a metric to probability and by using the notion of 'length' or 'distance' in the statistical space.

One method of measuring the information content in a system is utilizing the Fisher information, which represents the degree of certainty, or order. The opposite is entropy, which is a popular concept for the uncertainty or amount of disorder. Comparing entropy at different times then gives a measure of the difference in information content between the two states, which is known as relative entropy (e.g., see [15]). Another example is the Wasserstein metric [16,17], which provides an exact solution to the Fokker-Planck equation for a gradient flow subject to the minimization of the energy functional defined as the sum of the entropy and potential energy [18–20]. This metric has units of a physical length in comparison with other metrics, for instance the dimensionless statistical distance based on the Fisher information metric [21–23]. Interestingly, there is a link between the Fisher information and the Wasserstein distance [24]. Furthermore, the relative entropy can be expressed by the integral of the Fisher information along the same path [25].

Although quite useful, the relative entropy lacks the locality of a metric as it concerns only about the difference between given two PDFs. For instance, when these two PDFs represent the two states at different times, the relative entropy between them tells us nothing about how one PDF evolves to the other PDF over time or what intermediate states a system passes through between the two PDFs. As a result, it can only inform us of the changes that affect the overall system evolution [26]. To overcome this limitation, the information length $\mathcal{L}(t)$ was proposed in recent works, which quantifies the total number of different states that the system evolves through in time [27,28]. This means that the information length is a measure that depends on the evolution path between two states (PDFs). Its formulation allows us to measure local changes in the evolution of the system as well as providing an intriguing link between stochastic processes and geometry [26].

For instance, the relation between the information length $\mathcal{L}_\infty = \mathcal{L}(t \to \infty)$ and the mean value of the initial PDF for the fixed values of all other parameters was invoked as a new way of mapping out an attractor structure in a relaxation problem where any initial PDF relaxes into its equilibrium PDF in the long time limit. Specifically, for the Ornstein-Uhlenbeck (O-U) process driven by a Gaussian white-noise (which is a linearly damped, relaxation problem), $\mathcal{L}_\infty$ increases linearly with the distance between the mean position of an initial PDF and the stable equilibrium point for further details, see [28,29], with its minimum value zero at the stable equilibrium point. This linear dependence manifests that the information length preserves the linear geometry of the underlying Gaussian process, which is lost in other metrics [26]. For a nonlinear stochastic process with nonlinear damping, $\mathcal{L}_\infty$ still takes its minimum value at the stable equilibrium point but exhibits a power-law dependence on the distance between the mean value of an initial PDF and the stable equilibrium point. In contrast, for a chaotic attractor, $\mathcal{L}_\infty$ changes abruptly under an infinitesimal change of the mean value of an initial PDF, reminiscent of the sensitive dependence on initial conditions of the Lyapunov exponent [30]. These results suggest that $\mathcal{L}_\infty$ elucidates how different (non)linear forces affect (information) geometry.

With the above background in mind, this paper aims to extend the analysis of the information length of the O-U process to an arbitrary n-th order linear autonomous stochastic processes, providing a basic theoretical framework to be utilized in a large set of problems in both engineering and physics. In particular, we provide a useful analytical result that defines the information diagnostics as a function of the covariance matrix and the mean vector of the system, which enormously reduces the computational cost of numerical simulations of high-order systems.

This is followed by a specific application to a harmonically bound particle system (Kramers equation) for the position x and velocity $v = \frac{dx}{dt}$, with the natural oscillation frequency ω, subject to a damping constant γ and a Gaussian white-noise (short-correlated). We find an exact time-dependent joint PDF $p(x, v, t)$ starting from an initial Gaussian PDF which has a finite-width. Note that as far as we are aware of, our result $p(x, v, t)$ is original since in literature, the calculation was done only for the case of a delta-function initial PDF. Since this process is governed by the two variables, x and v, we investigate how $\mathcal{L}_\infty$ depends on their initial mean values $\langle x_0 \rangle$ and $\langle v_0 \rangle$. Here, the angular brackets denote the average. Furthermore, the two characteristic time scales associated with ω and γ raise the interesting question as to their role in $\mathcal{L}_\infty$. Thus, we explore how the information length depends on ω and γ. Our principle results are as follows: (i) $\mathcal{L}_\infty$ tends to increase linearly with either the deviation of initial mean position $\langle x_0 \rangle$ or the initial mean velocity $\langle v_0 \rangle$ from their equilibrium values $\langle x(0) \rangle = \langle v(0) \rangle = 0$; (ii) a linear geometry is thus preserved for our linearly coupled stochastic processes driven by a Gaussian noise; (iii) $\mathcal{L}_\infty$ tends to take its minimum value near the critical damping $\gamma = 2\omega$ for the same initial conditions and other parameters.

The remainder of this paper is organized as follows: Section 2 presents the basic concept of information length and the formulation of our problem. In Section 3, our main theoretical results are provided (see also Appendix A). In Section 4, we apply the results in Section 3 to analyze a harmonically bound particle system with the natural oscillation frequency ω subject to a damping constant γ and a Gaussian white-noise. Finally, Section 5 contains our concluding remarks.

To help readers, we here provide a summary of our notations: $\mathbb{R}$ and $\mathbb{C}$ are the sets of real and complex numbers, respectively. $\mathbf{x} \in \mathbb{R}^n$ represents a column vector $\mathbf{x}$ of real numbers of dimension n, $\mathbf{A} \in \mathbb{R}^{n\times n}$ represents a real matrix of dimension $n \times n$, $\mathbf{tr}(\mathbf{A})$ corresponds to the trace of the matrix $\mathbf{A}$, $\mathbf{A}^T$ and $\mathbf{A}^{-1}$ are the transpose and inverse of matrix $\mathbf{A}$, respectively. (Bold-face letters are used to represent vectors and matrices.) In some places, ∂_t or the prime both are used for the partial derivative with respect to time. Besides, $\boldsymbol{i} = \sqrt{-1}$ and for $s \in \mathbb{C}$, $\mathscr{L}^{-1}\left[F(s)\right] = \frac{1}{2\pi \boldsymbol{i}} \lim_{b\to\infty} \int_{a-b\boldsymbol{i}}^{a+b\boldsymbol{i}} e^{st} F(s)\, \mathrm{d}s$ corresponds to the inverse Laplace transform of the complex function $F(s)$. Finally, the average of a random vector $\mathbf{x}$ is denoted by $\langle \mathbf{x} \rangle$.

2. Preliminaries

2.1. Information Length

As noted in Section 1, the information length [26,27,31] is a dimensionless measurement of the total number of statistically different states that a system passes through in time in non-equilibrium processes. We cannot overemphasize that it is a measure that depends on the evolution of the system, being a useful index for understanding the information geometry underlying non-equilibrium processes. For example, for a time-dependent PDF $p(x,t)$ of one stochastic variable x, the information length $\mathcal{L}(t)$ is the total information change between time 0 and t, and is defined by

$$\mathcal{L}(t) = \int_0^t \frac{\mathrm{d}t_1}{\tau(t_1)} = \int_0^t \mathrm{d}t_1 \sqrt{\mathcal{E}(t_1)} = \int_0^t \mathrm{d}t_1 \sqrt{\int_{-\infty}^{\infty} \mathrm{d}x \frac{1}{p(x,t_1)} \left[\frac{\partial p(x,t_1)}{\partial t_1}\right]^2}. \tag{1}$$

Here, $\mathcal{E}(t_1) = \int_{-\infty}^{\infty} \mathrm{d}x \frac{1}{p(x,t_1)} \left[\frac{\partial p(x,t_1)}{\partial t_1}\right]^2$ is the square of the information velocity (recalling we are working with the unit where the distance given by the information length has no dimension). As we can see, to define the information length, we compute the dynamic time unit $\tau(t) = \frac{1}{\sqrt{\mathcal{E}}}$, which quantifies the correlation time over which the PDF $p(x,t)$ changes. Besides, τ serves as the time unit in the statistical space. Alternatively, the information velocity $\frac{1}{\tau(t_1)}$ quantifies the (average) rate of change of information in time.

2.2. Problem Formulation

We consider the following linear autonomous process

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \Gamma(t). \tag{2}$$

Here, $\mathbf{A}$ is an $n \times n$ constant real matrix; $\Gamma \in \mathbb{R}^n$ is a stochastic driving given by a n dimensional vector of δ-correlated Gaussian noises Γ_i ($i = 1, 2, ...n$), with the following statistical property

$$\langle \Gamma_i(t) \rangle = 0, \langle \Gamma_i(t)\Gamma_j(t_1) \rangle = 2D_{ij}\delta(t - t_1), D_{ij} = D_{ji}, \forall i, j = 1, \ldots, n. \tag{3}$$

Note that D_{ii} represents the strength of the i-th stochastic noise while D_{ij} for $i \neq j$ denotes the correlation between i-th and j-th noises (i.e., random fluctuations). Then, from the joint PDF $p(\mathbf{x}, t)$, we define the information length $\mathcal{L}$ of system (2) by the following integral

$$\mathcal{L}(t) = \int_0^t \mathrm{d}t_1 \sqrt{\int_{-\infty}^{\infty} \mathrm{d}\mathbf{x} \frac{1}{p(\mathbf{x},t_1)} \left[\frac{\partial p(\mathbf{x},t_1)}{\partial t_1}\right]^2} = \int_0^t \mathrm{d}t_1 \sqrt{\mathcal{E}}, \tag{4}$$

where $\mathcal{E} = \int_{-\infty}^{\infty} \mathrm{d}\mathbf{x} \frac{1}{p(\mathbf{x},t_1)} \left[\frac{\partial p(\mathbf{x},t_1)}{\partial t_1}\right]^2$ is the square of the information velocity.

The first goal of this paper is to provide theoretical results for the information length (4) for the system (2) and (3). This is done in the following Section 3.

3. General Analytical Results

In the section, we provide the analytical results for Problem 2.1, summarizing the main steps required to calculate information length (4). To this end, we assume that an initial PDF is Gaussian and then take the advantage of the fact that a linear process driven by a Gaussian noise with an initial Gaussian PDF is always Gaussian. The joint PDF for (2) and (3) is thus Gaussian, whose form is provided below.

Proposition 1 (Joint probability)**.** *The system* (2) *and* (3) *for a Gaussian random variable* $\mathbf{x}$ *at any time* t *has the following joint PDF*

$$p(\mathbf{x},t) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} e^{-\frac{1}{2}\left(\mathbf{x}-\langle\mathbf{x}(t)\rangle\right)^T \Sigma^{-1}\left(\mathbf{x}-\langle\mathbf{x}(t)\rangle\right)}, \tag{5}$$

where

$$\langle\mathbf{x}(t)\rangle = e^{\mathbf{A}t}\langle\mathbf{x}(0)\rangle, \tag{6}$$

$$\Sigma(t) = e^{\mathbf{A}t}\left\langle\delta\mathbf{x}(0)\delta\mathbf{x}(0)^T\right\rangle e^{\mathbf{A}^T t} + 2\int_0^t e^{\mathbf{A}(t-t_1)}\mathbf{D}e^{\mathbf{A}^T(t-t_1)}\,\mathrm{d}t_1, \tag{7}$$

and $\mathbf{D} \in \mathbb{R}^{n\times n}$ *is the matrix of elements* D_{ij}*. Here,* $\langle\mathbf{x}(t)\rangle$ *is the mean value of* $\mathbf{x}(t)$ *while* Σ *is the covariance matrix.*

Proof. For a Gaussian PDF of $\mathbf{x}$, all we need to calculate are the mean and covariance of $\mathbf{x}$ and substitute them in the general expression for multi-variable Gaussian distribution (5). To this end, we first write down the solution of Equation (2) as follows

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-t_1)}\Gamma(t_1)\,\mathrm{d}t_1. \tag{8}$$

By taking the average of Equation (8), we find the mean value of $\mathbf{x}(t)$ of (8) as follows

$$\langle\mathbf{x}(t)\rangle = \langle e^{\mathbf{A}t}\mathbf{x}(0)\rangle + \int_0^t e^{\mathbf{A}(t-t_1)}\langle\Gamma(t_1)\rangle\,\mathrm{d}t_1 = e^{\mathbf{A}t}\langle\mathbf{x}(0)\rangle, \tag{9}$$

which is Equation (6). On the other hand, to find covariance $\Sigma(t)$, we let $\mathbf{x} = \langle\mathbf{x}\rangle + \delta\mathbf{x}$, and use the property $\langle\delta\mathbf{x}(0)\Gamma(t)\rangle = 0$ to find

$$\begin{aligned}
\Sigma(t) &= \left\langle\delta\mathbf{x}\delta\mathbf{x}^T\right\rangle \\
&= \left\langle\left(e^{\mathbf{A}t}\delta\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-t_2)}\langle\Gamma(t_2)\rangle\,\mathrm{d}t_2\right)\left(e^{\mathbf{A}t}\delta\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-t_1)}\langle\Gamma(t_1)\rangle\,\mathrm{d}t_1\right)^T\right\rangle \\
&= \left\langle\left(e^{\mathbf{A}t}\delta\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-t_2)}\Gamma(t_2)\,\mathrm{d}t_2\right)\left(\delta\mathbf{x}(0)^T e^{\mathbf{A}^T t} + \int_0^t \Gamma(t_1)^T\left(e^{\mathbf{A}(t-t_1)}\right)^T\mathrm{d}t_1\right)\right\rangle \\
&= e^{\mathbf{A}t}\left\langle\delta\mathbf{x}(0)\delta\mathbf{x}(0)^T\right\rangle e^{\mathbf{A}^T t} + \left\langle\left(\int_0^t e^{\mathbf{A}(t-t_2)}\Gamma(t_2)\,\mathrm{d}t_2\right)\left(\int_0^t \Gamma(t_1)^T e^{\mathbf{A}^T(t-t_1)}\,\mathrm{d}t_1\right)\right\rangle \\
&= e^{\mathbf{A}t}\left\langle\delta\mathbf{x}(0)\delta\mathbf{x}(0)^T\right\rangle e^{\mathbf{A}^T t} + \int_0^t\int_0^t e^{\mathbf{A}(t-t_2)}\langle\Gamma(t_2)\Gamma(t_1)^T\rangle e^{\mathbf{A}^T(t-t_1)}\,\mathrm{d}t_2\,\mathrm{d}t_1 \\
&= e^{\mathbf{A}t}\left\langle\delta\mathbf{x}(0)\delta\mathbf{x}(0)^T\right\rangle e^{\mathbf{A}^T t} + 2\int_0^t e^{\mathbf{A}(t-t_1)}\mathbf{D}e^{\mathbf{A}^T(t-t_1)}\,\mathrm{d}t_1.
\end{aligned} \tag{10}$$

Here $\delta\mathbf{x}(0) = \delta\mathbf{x}(t=0)$ is the initial fluctuation at $t = 0$. Equation (10) thus proves Equation (7). Substitution of Equations (6) and (7) in Equation (5) thus gives us a joint PDF $p(\mathbf{x}, t)$ □

Next, in order to calculate the information length from the joint PDF $p(\mathbf{x}, t)$ in Equation (5), we now use the following Theorem:

Theorem 1 (Information Length)**.** *The information length of the joint PDF of system* (2) *and* (3) *is given by the following integral*

$$\mathcal{L}(t) = \int_0^t \mathrm{d}t_1 \sqrt{\mathcal{E}(t_1)} = \frac{1}{\sqrt{2}} \int_0^t \mathrm{d}t_1 \sqrt{\partial_{t_1}\left[\boldsymbol{tr}(\mathbf{Q}\partial_{t_1}\Sigma)\right] + 2\langle \mathbf{x}'(t_1)\rangle^T \mathbf{Q} \langle \mathbf{x}'(t_1)\rangle + \boldsymbol{tr}\left(\mathbf{Q}''\Sigma\right)}, \tag{11}$$

where $\mathbf{Q} = \Sigma^{-1}$ *(recall, a prime denotes* $\frac{\partial}{\partial t}$*).*

Proof. To prove this theorem, we use the PDF (5) in (4). To simplify the expression, we let

$$\mathbf{w} \equiv \delta\mathbf{x} = \mathbf{x} - \langle \mathbf{x}(t)\rangle, \ \mathbf{Q} = \Sigma^{-1}.$$

We then compute step by step $\frac{\left[\partial_{t_1} p(\mathbf{x},t_1)\right]^2}{p(\mathbf{x},t_1)}$ as follows:

$$\begin{aligned}\partial_{t_1} p(\mathbf{x}, t_1) &= \frac{\partial}{\partial t_1}\left[\left(\det(2\pi\Sigma)\right)^{-\frac{1}{2}} e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}}\right] \\ &= -\frac{1}{2} e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}} \left(\det(2\pi\Sigma)\right)^{-\frac{3}{2}} \partial_{t_1}\left(\det(2\pi\Sigma)\right) - \frac{1}{2}\left(\det(2\pi\Sigma)\right)^{-\frac{1}{2}} e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}} \partial_{t_1}\left(\mathbf{w}^T\mathbf{Q}\mathbf{w}\right),\end{aligned} \tag{12}$$

$$\begin{aligned}\left[\partial_{t_1} p(\mathbf{x}, t_1)\right]^2 = &\frac{1}{4} e^{-\mathbf{w}^T\mathbf{Q}\mathbf{w}} \left(\det(2\pi\Sigma)\right)^{-3} \left[\partial_{t_1} \det(2\pi\Sigma)\right]^2 + \frac{1}{4}\left(\det(2\pi\Sigma)\right)^{-1} e^{-\mathbf{w}^T\mathbf{Q}\mathbf{w}} \left(\partial_{t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right]\right)^2 \\ &+ \frac{1}{2}\left(\det(2\pi\Sigma)\right)^{-2} \partial_{t_1}\left[\det(2\pi\Sigma)\right] \partial_{t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right] e^{-\mathbf{w}^T\mathbf{Q}\mathbf{w}},\end{aligned} \tag{13}$$

$$\begin{aligned}\frac{\left[\partial_{t_1} p(\mathbf{x}, t_1)\right]^2}{p(\mathbf{x}, t_1)} = &\frac{1}{4}\left(\det(2\pi\Sigma)\right)^{-\frac{5}{2}} \left[\partial_{t_1} \det(2\pi\Sigma)\right]^2 e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}} + \frac{1}{4}\left(\det(2\pi\Sigma)\right)^{-\frac{1}{2}} e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}} \left(\partial_{t_1}\left(\mathbf{w}^T\mathbf{Q}\mathbf{w}\right)\right)^2 \\ &+ \frac{1}{2}\left(\det(2\pi\Sigma)\right)^{-\frac{3}{2}} \partial_{t_1}\left(\det(2\pi\Sigma)\right) \partial_{t_1}\left(\mathbf{w}^T\mathbf{Q}\mathbf{w}\right) e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}}.\end{aligned} \tag{14}$$

Now, using Equation (12) in Equation (14), we compute the integral $\mathcal{E}(t_1) = \int_{-\infty}^{\infty} \mathrm{d}\mathbf{x} \left(\frac{\left[\partial_{t_1} p(\mathbf{x},t_1)\right]^2}{p(\mathbf{x},t_1)}\right)$ as follows

$$\begin{aligned}\mathcal{E}(t_1) &= \int_{-\infty}^{\infty} p(\mathbf{x}, t_1) \left(\frac{\left(\det(2\pi\Sigma)\right)^{-2}}{4}\left[\partial_{t_1}\left(\det(2\pi\Sigma)\right)\right]^2 + \frac{\left(\partial_{t_1}\left(\mathbf{w}^T\mathbf{Q}\mathbf{w}\right)\right)^2}{4} + \frac{\partial_{t_1}\left[\det(2\pi\Sigma)\right]\partial_{t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right]}{2\det(2\pi\Sigma)}\right) \mathrm{d}\mathbf{x} \\ &= \left\langle\left(\frac{\partial_{t_1}\left[\det(2\pi\Sigma)\right]}{2\det(2\pi\Sigma)}\right)^2\right\rangle + \left\langle\left(\frac{\partial_{t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right]}{2}\right)^2\right\rangle + \left\langle\frac{\partial_{t_1}\left[\det(2\pi\Sigma)\right]\partial_{t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right]}{2\det(2\pi\Sigma)}\right\rangle.\end{aligned} \tag{15}$$

To calculate the three averages in (15), we use the properties $\int_{-\infty}^{\infty} e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}}\,\mathrm{d}\mathbf{w} = \sqrt{\det(2\pi\Sigma)}$ [32], $\partial_{t_1}\left[e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}}\right] = -\frac{1}{2}e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}}\partial_{t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right]$ and $\partial_{t_1t_1}\left[e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}}\right] = -\frac{1}{2}\partial_{t_1t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right]e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}} + \frac{1}{4}e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}}\left(\partial_{t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right]\right)^2$. We then have

$$\begin{aligned}
\mathcal{E}(t_1) &= \left(\frac{\partial_{t_1}[\det(2\pi\Sigma)]}{2\det(2\pi\Sigma)}\right)^2 + \int_{-\infty}^{\infty} p(\mathbf{x},t_1)\left(\frac{\partial_{t_1}[\mathbf{w}^T\mathbf{Q}\mathbf{w}]}{2}\right)^2 \mathrm{d}\mathbf{x} + \frac{\partial_{t_1}[\det(2\pi\Sigma)]}{2\det(2\pi\Sigma)}\int_{-\infty}^{\infty} p(\mathbf{x},t_1)\partial_{t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right]\mathrm{d}\mathbf{x} \\
&= \frac{1}{4}\left(\frac{\partial_{t_1}[\det(2\pi\Sigma)]}{\det(2\pi\Sigma)}\right)^2 + \frac{1}{4\left(\det(2\pi\Sigma)\right)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\left[4\partial_{t_1t_1}e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}} + 2\partial_{t_1t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right]e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}}\right]\mathrm{d}\mathbf{x} \\
&\quad -2\frac{\partial_{t_1}[\det(2\pi\Sigma)]}{2\left(\det(2\pi\Sigma)\right)^{\frac{3}{2}}}\int_{-\infty}^{\infty}\partial_{t_1}\left[e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}}\right]\mathrm{d}\mathbf{x} \\
&= \frac{1}{4}\left(\frac{\partial_{t_1}[\det(2\pi\Sigma)]}{\det(2\pi\Sigma)}\right)^2 + \frac{\partial_{t_1t_1}\left[\sqrt{\det(2\pi\Sigma)}\right]}{\sqrt{\det(2\pi\Sigma)}} + \frac{1}{2\sqrt{\det(2\pi\Sigma)}}\int_{-\infty}^{\infty}\partial_{t_1t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right]e^{-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}}\,\mathrm{d}\mathbf{x} \\
&\quad -\frac{\partial_{t_1}[\det(2\pi\Sigma)]}{\left(\det(2\pi\Sigma)\right)^{\frac{3}{2}}}\partial_{t_1}\sqrt{\det(2\pi\Sigma)}.
\end{aligned} \tag{16}$$

Here

$$\partial_{t_1t_1}\left[\mathbf{w}^T\mathbf{Q}\mathbf{w}\right] = \sum_{i,j=1}^{n}\left[\partial_{t_1t_1}\left(q_{ij}w_iw_j\right)\right] = \sum_{i,j=1}^{n}\left[4q'_{ij}w'_iw_j + \underbrace{2q_{ij}w'_iw'_j}_{\text{independent of } x} + 2q_{ij}w''_iw_j + \underbrace{q''_{ij}w_iw_j}_{\mathbf{w}^T\mathbf{Q}''\mathbf{w}}\right]. \tag{17}$$

We recall that ω'_i, q'_{ij} and ω''_i, q''_{ij} denote the first and second derivative over time of the elements ω_i and q_{ij}. By substituting (17) in (16) and making some arrangements, we obtain

$$\begin{aligned}
\mathcal{E}(t_1) &= \frac{1}{4}\left(\frac{\partial_{t_1}[\det(2\pi\Sigma)]}{\det(2\pi\Sigma)}\right)^2 + \frac{\partial_{t_1t_1}\left[\sqrt{\det(2\pi\Sigma)}\right]}{\sqrt{\det(2\pi\Sigma)}} + \frac{1}{2}\cancelto{0}{\left\langle 4\sum_{i,j=1}^{n} q'_{ij}w'_iw_j\right\rangle} + \frac{1}{2}\cancelto{0}{\left\langle \sum_{i,j=1}^{n} 2q_{ij}w''_iw_j\right\rangle} \\
&\quad + \frac{1}{2}\left\langle \sum_{i,j=1}^{n} 2q_{ij}w'_iw'_j\right\rangle + \frac{1}{2}\left\langle \mathbf{w}^T\mathbf{Q}''\mathbf{w}\right\rangle - \frac{1}{2}\left(\frac{\partial_{t_1}[\det(2\pi\Sigma)]}{\det(2\pi\Sigma)}\right)^2.
\end{aligned} \tag{18}$$

Now with the help of the following relations $\left\langle \mathbf{w}^T\mathbf{Q}''\mathbf{w}\right\rangle = \mathbf{tr}\left(\mathbf{Q}''\Sigma\right)$ [33], $\partial_{t_1}\det(\Sigma) = \det(\Sigma)\mathbf{tr}(\mathbf{Q}\partial_{t_1}\Sigma)$ [34], and $\partial_{t_1t_1}\sqrt{\det(2\pi\Sigma)} = \frac{1}{4}\sqrt{\det(2\pi\Sigma)}\left(\mathbf{tr}(\mathbf{Q}\partial_{t_1}\Sigma)\right)^2 + \frac{1}{2}\sqrt{\det(2\pi\Sigma)}\partial_{t_1}(\mathbf{tr}(\mathbf{Q}\partial_{t_1}\Sigma))$, we then have

$$\begin{aligned}
\mathcal{E}(t_1) &= -\frac{1}{4}\left(\mathbf{tr}(\mathbf{Q}\partial_{t_1}\Sigma)\right)^2 + \frac{1}{2}\partial_{t_1}\left[\mathbf{tr}(\mathbf{Q}\partial_{t_1}\Sigma)\right] + \frac{1}{4}\left(\mathbf{tr}(\mathbf{Q}\partial_{t_1}\Sigma)\right)^2 \\
&\quad + \langle\mathbf{x}'(t_1)\rangle^T\mathbf{Q}\langle\mathbf{x}'(t_1)\rangle + \frac{1}{2}\mathbf{tr}\left(\mathbf{Q}''\Sigma\right) \\
&= \frac{1}{2}\partial_{t_1}\left[\mathbf{tr}(\mathbf{Q}\partial_{t_1}\Sigma)\right] + \langle\mathbf{x}'(t_1)\rangle^T\mathbf{Q}\langle\mathbf{x}'(t_1)\rangle + \frac{1}{2}\mathbf{tr}\left(\mathbf{Q}''\Sigma\right).
\end{aligned} \tag{19}$$

Equation (19) thus proves Equation (11). □

Given important properties of the covariance matrix eigenvalues (see, e.g., [35]), it is useful to express Equation (19) and the information length as a function of these covariance matrix eigenvalues. This is done in the following Corollary.

Corollary 1. *Let $\varphi_i(t)$'s ($i = 1, ...n$) be the eigenvalues of the covariance matrix Σ, and $\bar{\mathbf{x}} = \langle \mathbf{x}'(t) \rangle^T \mathbf{P}$ where $\mathbf{P}$ is an orthonormal matrix whose column vectors are linearly independent eigenvectors of $Q = \Sigma^{-1}$. We can rewrite the information length (11) as*

$$\mathcal{L}(t) = \int_0^t \mathrm{d}t_1 \sqrt{\mathcal{E}(t_1)} = \int_0^t \mathrm{d}t_1 \sqrt{\sum_{i=1}^{n} \left(\frac{\varphi_i'^2(t_1) + 2\varphi_i(t_1)\bar{x}_i^2}{2\varphi_i^2(t_1)} \right)}. \tag{20}$$

Proof. The proof follows straightforwardly from the fact that Σ is a symmetric matrix which can be diagonalised by finding the orthonormal matrix $\mathbf{P}$ such that $\mathbf{P}^T \Sigma^{-1} \mathbf{P} = \Phi$. Here Φ is the diagonal matrix whose entries are the eigenvalues $\frac{1}{\varphi_i(t)}$ $\forall i = 1, 2, \ldots, n$ (recall that $\varphi_i(t)$ is i-th the eigenvalue of Σ^{-1}). This gives us

$$\mathcal{E}(t) = \frac{1}{2} \sum_{i=1}^{n} \left(\partial_t \left[\frac{\varphi_i'(t)}{\varphi_i(t)} \right] + \varphi_i(t) \partial_{tt} \left[\frac{1}{\varphi(t)} \right] + 2 \frac{\bar{x}_i^2}{\varphi_i(t)} \right) = \sum_{i=1}^{n} \left(\frac{\varphi_i'^2(t) + 2\varphi_i(t)\bar{x}_i^2}{2\varphi_i^2(t)} \right). \tag{21}$$

This finishes the proof. □

It is useful to check that Equation (20) reproduces the previous result for the O-U process [36]

$$\mathcal{E} = \frac{\beta'^2}{2\beta^2} + 2\beta \langle x' \rangle^2, \tag{22}$$

where $\beta = \frac{1}{2\langle (x - \langle x \rangle)^2 \rangle}$ is the inverse temperature. Here, β' denotes the time derivative of β. To show this, we note that for the O-U process, the covariance matrix is a scalar ($n = 1$) with the value $\Sigma = \frac{1}{2\beta} = \varphi(t)$ and thus $\mathbf{Q} = \frac{1}{\varphi(t)} = 2\beta$ while $\langle \mathbf{x}'(t) \rangle = \langle x' \rangle$. Thus,

$$\mathcal{E}(t) = \frac{1}{2} \left(\frac{-\frac{1}{2}\beta^{-2}\beta'}{\frac{1}{2\beta}} \right)^2 + 2\beta \langle x' \rangle^2 = \frac{\beta'^2}{2\beta^2} + 2\beta \langle x' \rangle^2.$$

In sum, for the O-U process, the square of the information velocity (shown in expression (22)) increases with the 'roughness' of the process, as quantified by the squared ratio of the rate of change of the inverse temperature (or precision) and the precision – plus a term that depends upon this precision times the variance of the state velocity.

4. Kramers Equation

In this section we apply our results in Section 3 to the Kramers equation for a harmonically bound particle [19,37]. As noted in Introduction, we investigate the behaviour of the information length when varying various parameters and initial conditions to elucidate how the information geometry is affected by the damping, oscillations, strength of the stochastic noises and initial mean values.

Consider the Kramers equation

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= v \\ \frac{\mathrm{d}v}{\mathrm{d}t} &= -\gamma v - \omega^2 x + \xi(t). \end{aligned} \tag{23}$$

Here, ω is a natural frequency and γ is the damping constant, both positive real numbers. ξ is a Gaussian white-noise acting on v with the zero mean value $\langle \xi(t) \rangle = 0$, with the statistical property

$$\langle \xi(t) \xi(t_1) \rangle = 2D\delta(t - t_1). \tag{24}$$

Comparing Equations (23) and (24) with Equations (2) and (3), we note that $x_1 = x$, $x_2 = v$, $\xi_1 = 0$, $\xi_2 = \xi$, $D_{11} = 0$, $D_{12} = 0$, and $D_{22} = D$ while the matrix $\mathbf{A}$ for (23) has the element $A_{11} = 0, A_{12} = 1, A_{21} = -\omega^2, A_{22} = -\gamma$. Thus, the eigenvalues of $\mathbf{A}$ are $\lambda_{1,2} = -\frac{1}{2}\left(\gamma \pm \sqrt{\gamma^2 - 4\omega^2}\right)$.

To find the information length for the system (23), we use Proposition 1 and Theorem 1. First, Proposition 1 requires the computation of the exponential matrix $e^{\mathbf{A}t}$ involving a rather long algebra with the help of [38]. The result is:

$$e^{\mathbf{A}t} = \mathscr{L}^{-1}\left[(s\mathbf{I} - \mathbf{A})^{-1}\right] = \mathscr{L}^{-1}\left[\begin{bmatrix} \frac{s+\gamma}{(s-\lambda_1)(s-\lambda_2)} & \frac{1}{(s-\lambda_1)(s-\lambda_2)} \\ -\frac{\omega^2}{(s-\lambda_1)(s-\lambda_2)} & \frac{s}{(s-\lambda_1)(s-\lambda_2)} \end{bmatrix}\right] = \begin{bmatrix} \frac{e^{\lambda_1 t}(\gamma+\lambda_1) - e^{\lambda_2 t}(\gamma+\lambda_2)}{\lambda_1-\lambda_2} & \frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{\lambda_1-\lambda_2} \\ -\frac{\left(e^{\lambda_1 t} - e^{\lambda_2 t}\right)\omega^2}{\lambda_1-\lambda_2} & \frac{e^{\lambda_1 t}\lambda_1 - e^{\lambda_2 t}\lambda_2}{\lambda_1-\lambda_2} \end{bmatrix}. \quad (25)$$

Here, $\mathbf{I} \in \mathbb{R}^{n\times n}$ is the identity matrix. Similarly, we can show

$$2\int_0^t e^{\mathbf{A}(t-t_1)}\mathbf{D}e^{\mathbf{A}^T(t-t_1)}\,\mathrm{d}t_1 = \begin{bmatrix} \frac{D\left(\frac{-1+e^{2\lambda_1 t}}{\lambda_1} + \frac{-\lambda_1 - 4e^{(\lambda_1+\lambda_2)t}\lambda_2 + 3\lambda_2 + e^{2\lambda_2 t}(\lambda_1+\lambda_2)}{\lambda_2(\lambda_1+\lambda_2)}\right)}{(\lambda_1-\lambda_2)^2} & \frac{D\left(e^{\lambda_1 t} - e^{\lambda_2 t}\right)^2}{(\lambda_1-\lambda_2)^2} \\ \frac{D\left(e^{\lambda_1 t} - e^{\lambda_2 t}\right)^2}{(\lambda_1-\lambda_2)^2} & \frac{D\left(\left(-1+e^{2\lambda_2 t}\right)\lambda_2 + \lambda_1\left(-\frac{4e^{(\lambda_1+\lambda_2)t}\lambda_2}{\lambda_1+\lambda_2} + \frac{4\lambda_2}{\lambda_1+\lambda_2} + e^{2\lambda_1 t} - 1\right)\right)}{(\lambda_1-\lambda_2)^2} \end{bmatrix}. \quad (26)$$

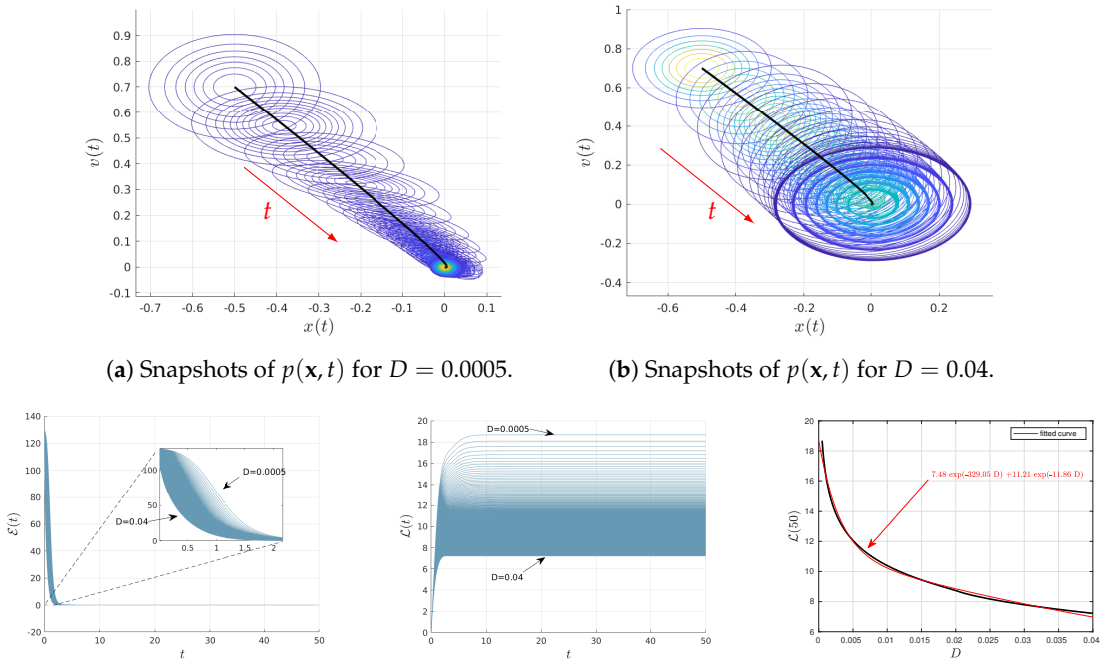
Using Equations (25) and (26) in Equations (6) and (7), we have the time-dependent (joint) PDF (5) at any time t for our system (23) and (24). To calculate Equation (11) with the help of Equations (25) and (26), we perform numerical simulations (integrations) for various parameters in Equations (23) and (24) as well as initial conditions. Note that while we have simulated many different cases, for illustration, we show some representative cases by varying D, ω, γ and $\langle x(0)\rangle$, $\langle v(0)\rangle$ in Sections 4.1–4.3 and Appendix A, respectively, for the same initial covariance matrix $\Sigma(0)$ with elements $\Sigma_{11}(0) = \Sigma_{22}(0) = 0.01$ and $\Sigma_{12}(0) = \Sigma_{21}(0) = 0$. Note that the initial marginal distributions of $p(x(0))$ and $p(v(0))$ are Gaussian with the same variance 0.01. Results in the limit $\omega \to 0$ are presented in Section 4.4.

4.1. Varying D

Figure 1 shows the results when varying D as $D \in (0.0005, 0.04)$ for the fixed parameters $\gamma = 2$ and $\omega = 1$. The initial joint PDFs are Gaussian with the fixed mean values $\langle x(0)\rangle = -0.5$, $\langle v(0)\rangle = 0.7$; as noted above, the covariance matrix $\Sigma(0)$ with elements $\Sigma_{11}(0) = \Sigma_{22}(0) = 0.01$ and $\Sigma_{12}(0) = \Sigma_{21}(0) = 0$. Consequently, at $t = 0$, the marginal distributions of $p(x(0))$ and $p(v(0))$ are Gaussian PDFs with the same variance 0.01 and the mean values $\langle x(0)\rangle = -0.5$ and $\langle v(0)\rangle = 0.7$, respectively.

Figure 1a,b show the snapshots of time-dependent joint PDF $p(\mathbf{x}, t)$ (in contour plots) for the two different values of $D = 0.0005$ and $D = 0.04$, respectively. The black solid represents the phase portrait of the mean value of $\langle x(t)\rangle$ and $\langle v(t)\rangle$ while the red arrows display the direction of time increase. Note that in Figure 1b, only some of the initial snapshots of the PDFs are shown for clarity, given the great amount of overlapping between different PDFs. Figure 1c,d show the time-evolution of the information velocity $\mathcal{E}(t)$ and information length $\mathcal{L}(t)$, respectively, for different values of $D \in (0.0005, 0.04)$. It can be seen that the system approaches a stationary (equilibrium) state for $t \gtrsim 20$ for all values of D, $\mathcal{L}(t)$ approaching constant values (recall $\mathcal{L}(t)$ does not change in a stationary state). Therefore, we approximate the total information length as $\mathcal{L}_\infty = \mathcal{L}(t = 50)$, for instance. Finally, the total information length $\mathcal{L}_\infty = \mathcal{L}(t = 50)$ is shown in Figure 1e. We determine the

dependence of $\mathcal{L}_\infty$ on D by fitting an exponential function as $\mathcal{L}_\infty(D) = 7.84e^{-329.05D} + 11.21e^{-11.86D}$ (shown in red solid line).



(**a**) Snapshots of $p(\mathbf{x}, t)$ for $D = 0.0005$.

(**b**) Snapshots of $p(\mathbf{x}, t)$ for $D = 0.04$.

(**c**) Time-evolution of $\mathcal{E}(t)$.

(**d**) Time-evolution of $\mathcal{L}(t)$.

(**e**) $\mathcal{L}_\infty = \mathcal{L}(t = 50)$ against D. A fitted curve is shown in the red solid line.

Figure 1. Results of Equations (23) and (24) for $\langle x(0) \rangle = -0.5$, $\langle v(0) \rangle = 0.7$, $\gamma = 2$, $\omega = 1$, $D \in (0.0005, 0.04)$ and the initial covariance matrix $\Sigma(0)$ with elements $\Sigma_{11}(0) = \Sigma_{22}(0) = 0.01$, $\Sigma_{12}(0) = \Sigma_{21}(0) = 0$.

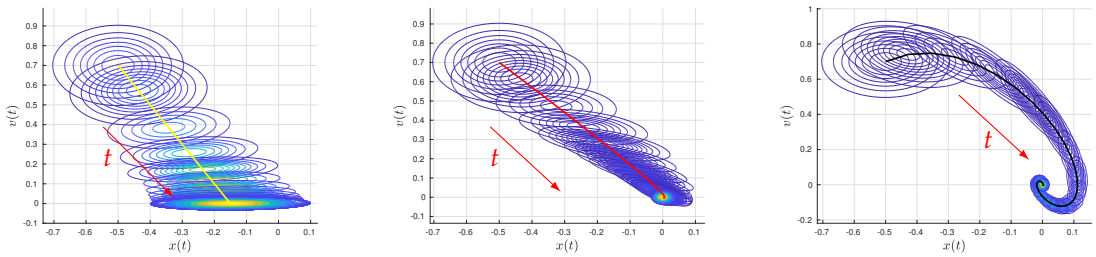
4.2. Varying ω or γ

We now explore how results depend on the two parameters ω and γ, associated with oscillation and damping, respectively. To this end, we use $D = 0.0005$ and the same initial conditions as in Figure 1 but vary $\omega \in (0, 2)$ and $\gamma \in (0, 6)$ in Figures 2 and 3, respectively. Specifically, in different panels of these figures, we show the snapshots of the joint PDF $p(\mathbf{x}, t)$, the time-evolutions of $\mathcal{E}(t)$ and $\mathcal{L}(t)$ for different values of $\omega \in (0, 2)$ and $\gamma \in (0, 6)$, and $\mathcal{L}_\infty$ against either ω or γ. From Figures 2e and 3e, we can see that the system is in a stationary state for sufficiently large $t = 10$ and $t = 100$, respectively. Thus, we use $\mathcal{L}_\infty = \mathcal{L}(t = 10) = \mathcal{L}(10)$ in Figure 2f,g and $\mathcal{L}_\infty = \mathcal{L}(t = 100) = \mathcal{L}(10)$ in Figure 3f,g.

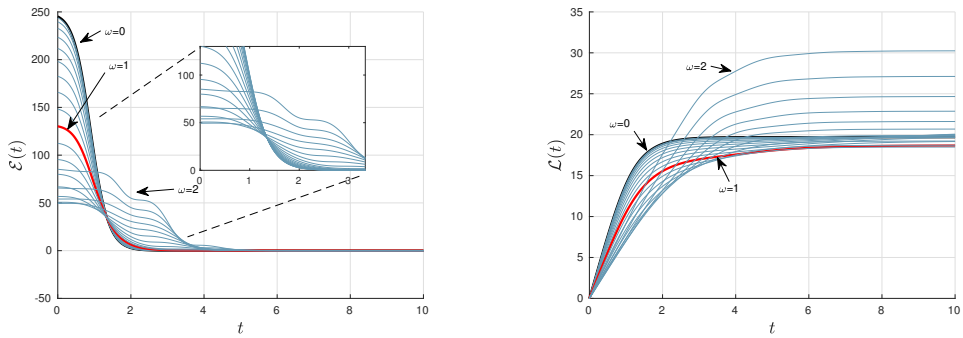
Notably, Figure 2f,g (shown on linear-linear and log-linear scales on $x - y$ axes, respectively) exhibit an interesting a non-monotonic dependence of $\mathcal{L}_\infty$ on ω for the fixed $\gamma = 2$, with the presence of a distinct minimum in $\mathcal{L}_\infty$ at certain ω. Similarly, Figure 3f,g (shown in linear-linear and log-log scales on $x - y$ axes, respectively) also shows a non-monotonic dependence of $\mathcal{L}_\infty$ on γ for the fixed $\omega = 1$. These non-monotonic dependences are more clearly seen in Figures 2g and 3g. A close inspection of these figures then reveals that the minimum value of $\mathcal{L}_\infty$ occurs close to the critical damping (CD) $\gamma \sim 2\omega$; specifically, this happens at $\omega \sim 1$ for $\gamma = 2$ in Figure 2f,g while at $\gamma \sim 2$ for $\omega = 1$ in Figure 3f,g. We thus fit $\mathcal{L}_\infty$ against ω or γ depending on whether ω or γ is smaller/larger than its critical value as follows:

$$\mathcal{L}_{10}(\omega) = -0.03e^{4.34\omega} + 19.63e^{0.06\omega} \quad \forall \quad \omega \in (0,1), \tag{27}$$
$$\mathcal{L}_{10}(\omega) = 19.52e^{-0.12\omega} + 0.11e^{2.48\omega} \quad \forall \quad \omega \in (1,2), \tag{28}$$
$$\mathcal{L}_{100}(\gamma) = 413.22e^{-12.4\gamma} + 95.39e^{-1.02\gamma} \quad \forall \quad \gamma \in (0,2), \tag{29}$$
$$\mathcal{L}_{100}(\gamma) = 3.23\gamma \quad \forall \quad \gamma \in (2,6). \tag{30}$$

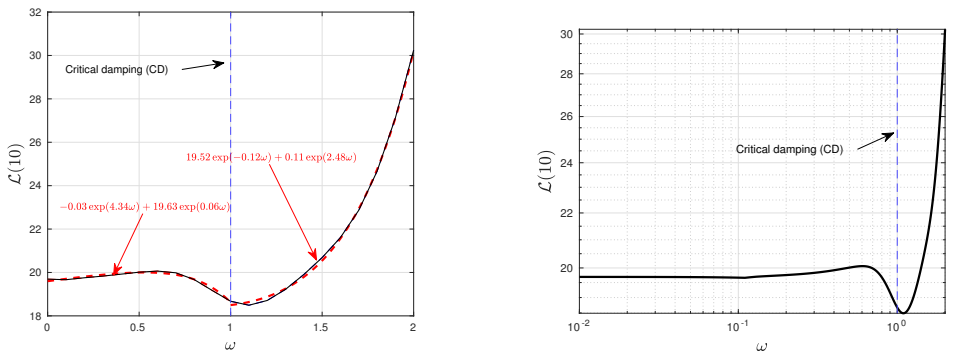
The fitted curves in Equations (27)–(30) are superimposed in Figures 2f and 3f, respectively. It is important to notice from Equations (27)–(30) that $\mathcal{L}_\infty$ tends to increase as either $\omega \to \infty$ for a finite, fixed γ ($< \infty$) or $\gamma \to \infty$ for a finite, fixed ω ($< \infty$).



(**a**) Snapshots of $p(\mathbf{x},t)$ for $\omega = 0$. (**b**) Snapshots of $p(\mathbf{x},t)$ for $\omega = 1$. (**c**) Snapshots of $p(\mathbf{x},t)$ for $\omega = 2$.

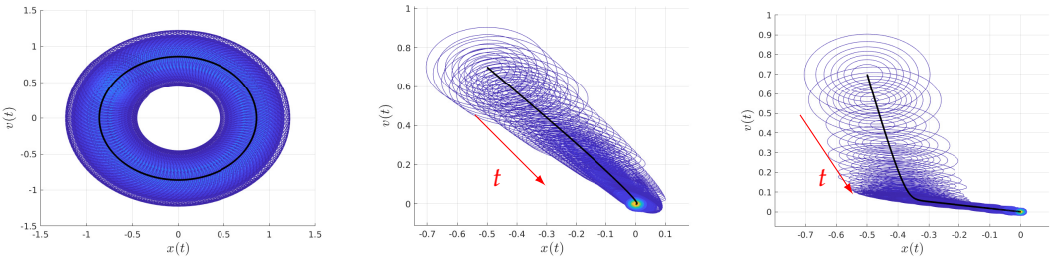


(**d**) Time-evolution of $\mathcal{E}(t)$. (**e**) Information Length $\mathcal{L}(t)$.

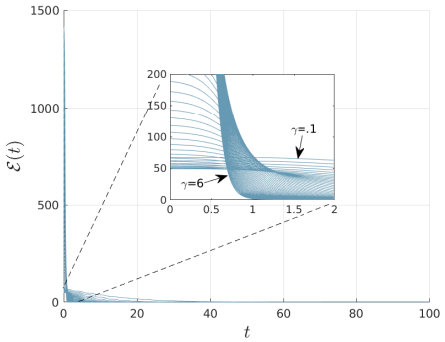


(**f**) $\mathcal{L}_\infty = \mathcal{L}(t = 10)$ against ω; fitted curves are shown in the red dashed lines while a vertical line represents $\omega = 1$ at the critical damping. (**g**) The same as panel (f) but shown on log-linear scales on $x - y$ axes.

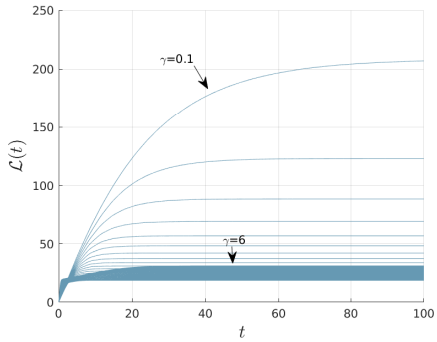
Figure 2. Results of Equations (23) and (24) for $\langle x(0) \rangle = -0.5$, $\langle v(0) \rangle = 0.7$, $\gamma = 2$, $\omega \in (0,2)$, $D = 0.0005$, and the initial covariance matrix $\Sigma(0)$ with elements $\Sigma_{11}(0) = \Sigma_{22}(0) = 0.01$, $\Sigma_{12}(0) = \Sigma_{21}(0) = 0$.



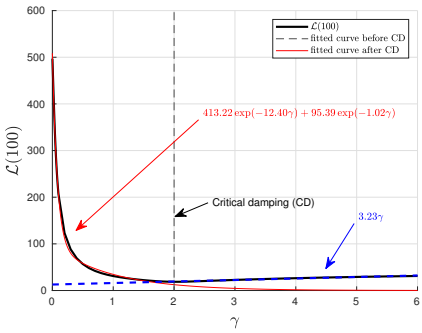
(**a**) Snapshots of $p(\mathbf{x}, t)$ for $\gamma = 0$. (**b**) Snapshots of $p(\mathbf{x}, t)$ for $\gamma = 2$. (**c**) Snapshots of $p(\mathbf{x}, t)$ for $\gamma = 6$.



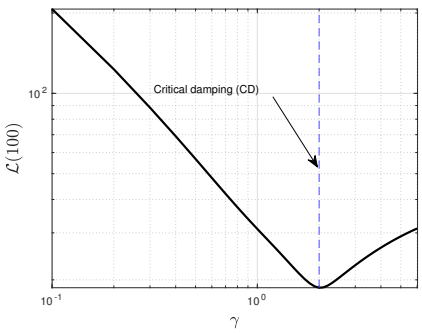
(**d**) Time-evolution of $\mathcal{E}(t)$.



(**e**) Time-evolution of $\mathcal{L}(t)$.



(**f**) $\mathcal{L}_\infty = \mathcal{L}(t = 100)$ against γ; fitted curves are shown in the red solid and blue dashed lines while a vertical line represents $\gamma = 2$ at the critical damping.



(**g**) The same as panel (f) but shown in log-log scales on x and y axes.

Figure 3. Results of Equations (23) and (24) for $\langle x(0) \rangle = -0.5$, $\langle v(0) \rangle = 0.7$, $\gamma \in (0, 6)$, $\omega = 1$, $D = 0.0005$, and the initial covariance matrix $\Sigma(0)$ with elements $\Sigma_{11}(0) = \Sigma_{22}(0) = 0.01$, $\Sigma_{12}(0) = \Sigma_{21}(0) = 0$.

Finally, we note that for the critical damping $\gamma = 2\omega$, the eigenvalue becomes a real double root with the value $\lambda_{1,2} \to -\omega$. Thus, in this limit, we have that

$$\langle \mathbf{x}(t) \rangle = \begin{bmatrix} e^{-t\omega}(x(0) + t(v(0) + (\gamma - \omega)x(0))) \\ e^{-t\omega}\left(-tx(0)\omega^2 - tv(0)\omega + v(0)\right) \end{bmatrix}, \tag{31}$$

and $\Sigma(t)$ is composed by the following elements

$$\begin{aligned}
\Sigma_{11}(t)&=\frac{e^{-2t\omega}\Big(2\omega^3\big(\Sigma_{11}(\gamma t-t\omega+1)^2+t^2((\Sigma_{12}+\Sigma_{21})(\gamma-\omega)+\Sigma_{22})+t(\Sigma_{12}+\Sigma_{21})\big)+D\big(-2t\omega(t\omega+1)+e^{2t\omega}-1\big)\Big)}{2\omega^3},\\
\Sigma_{12}(t)&=e^{-2t\omega}\Big(t\big(-\omega^2(\Sigma_{11}\gamma t+\Sigma_{11}+\Sigma_{21}t)+\Sigma_{11}t\omega^3-\Sigma_{22}t\omega+\Sigma_{22}+Dt\big)-\Sigma_{12}(t\omega-1)(\gamma t-t\omega+1)\Big),\\
\Sigma_{21}(t)&=e^{-2t\omega}\Big(t\big(-\omega^2(\Sigma_{11}\gamma t+\Sigma_{11}+\Sigma_{12}t)+\Sigma_{11}t\omega^3-\Sigma_{22}t\omega+\Sigma_{22}+Dt\big)-\Sigma_{21}(t\omega-1)(\gamma t-t\omega+1)\Big),\\
\Sigma_{22}(t)&=\frac{e^{-2t\omega}\Big(2t\omega^2\big(t\omega^2(\Sigma_{11}\omega+\Sigma_{12}+\Sigma_{21})-\omega(\Sigma_{12}+\Sigma_{21})+\Sigma_{22}(t\omega-2)\big)+2\Sigma_{22}\omega+D\big(-2t\omega(t\omega-1)+e^{2t\omega}-1\big)\Big)}{2\omega}.
\end{aligned} \tag{32}$$

Equations (31) and (32) are used in Section 4.1 (Figure 1).

4.3. Varying $\langle x(0)\rangle$ or $\langle v(0)\rangle$

To elucidate the information geometry associated with the Kramer equation (Equations (23) and (24)), we now investigate how $\mathcal{L}_\infty$ behaves near the equilibrium point $\langle x(0)\rangle = \langle v(0)\rangle = 0$. To this end, we scan over $\langle x(0)\rangle$ for $\langle v(0)\rangle = 0$ in Figure 4a–e while scanning over $\langle v(0)\rangle$ for $\langle x0)\rangle = 0$ in Figure 4f–i. For our illustrations in Figure 4, we use the same initial covariance matrix $\Sigma(0)$ as in Figures 1–3, $D = 0.0005$ and $\omega = 1$ and a few different values of γ (above/below/at the critical value $\gamma = 2$). We note that the information geometry near a non-equilibrium point is studied in Appendix A.

Specifically, snapshots of $p(\mathbf{x}, t)$ are shown in Figure 4a–f for $\gamma = 2.5$ (above its critical value $\gamma = 2 = 2\omega$) while those in Figure 4c–g are for $\gamma = 0.1$ below the critical value 2. By approximating $\mathcal{L}_\infty = \mathcal{L}(t = 100)$, we then show how $\mathcal{L}_\infty$ depends on $\langle x(0)\rangle$ and $\langle v(0)\rangle$ for different values of γ in Figure 4d,e and Figure 4h,i, respectively.

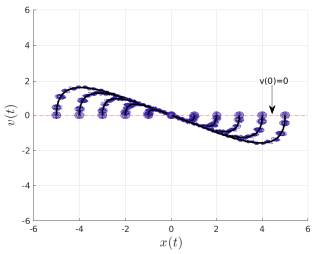
Figure 4d,e show the presence of a minimum in $\mathcal{L}_\infty$ at the equilibrium $\langle x(0)\rangle = 0$ (recall $\langle v(0)\rangle = 0$); $\mathcal{L}_\infty$ is a linear function of $\langle x(0)\rangle$ for $\langle x(0)\rangle \gg 0.1$, which can be described as $\mathcal{L}_\infty(x(0), \gamma) = h(\gamma)|\langle x(0)\rangle| + f(\gamma)$. Here, $h(\gamma)$ and $f(\gamma)$ are constant functions depending on γ for a fixed ω which represent the slope and the y-axis intercept, respectively. A non-zero value of $\mathcal{L}_\infty$ at $\langle x(0)\rangle = 0$ is caused by the adjustment (oscillation and damping) of the width of the PDFs in time due to the disparity between the width of the initial and equilibrium PDFs (see Figure 4b). In other words, even though the mean values remain in equilibrium for all time $[\langle x(0)\rangle, \langle v(0)\rangle]^T = \lim_{t\to\infty}\langle \mathbf{x}(t)\rangle = [0, 0]^T$, the information length (11) depends on the covariance matrix Σ which changes from its initial value to the final equilibrium value as follows

$$\Sigma(0) = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix} \text{ to } \lim_{t\to\infty}\Sigma(t) = \begin{bmatrix} \frac{D}{\gamma\omega^2} & 0 \\ 0 & \frac{D}{\gamma} \end{bmatrix}.$$

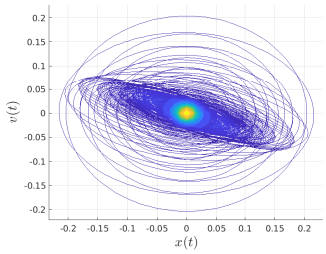
On the other hand, $\mathcal{L}_\infty$ against $\langle x(0)\rangle$ shows parabolic behaviour for small $\langle x(0)\rangle < 0.1$ in Figure 4e. This is caused by the finite width $0.1 = \sqrt{\Sigma_{11}(0)} = \sqrt{\Sigma_{22}(0)}$ of the initial $p(x, 0)$; we see that $\langle x(0)\rangle < 0.1$ is within the uncertainty of the initial $p(x, 0)$.

Similarly, Figure 4h,i exhibit a minimum in $\mathcal{L}_\infty$ at the equilibrium $\langle v(0)\rangle = 0$ (recall $\langle x(0)\rangle = 0$ in this case); $\mathcal{L}_\infty$ is a linear function of $\langle v(0)\rangle$ for $\langle v(0)\rangle \gg 0.1$ described by $\mathcal{L}_\infty(v(0), \gamma) = H(\gamma)|\langle v(0)| + F(\gamma)$ (again parabolic for $\langle v(0)\rangle < 0.1$, see Figure 4i). Here again, $H(\gamma)$ and $F(\gamma)$ are constant functions depending on γ for a fixed ω which represent the slope and the y-axis intercept, respectively.

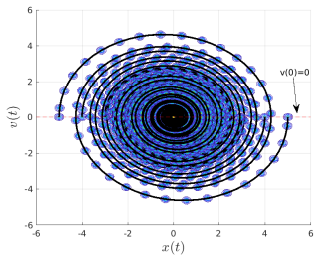
Finally, Figure 4j shows in logarithmic scale that the minimum value of $\mathcal{L}_\infty$ at $\langle x(0)\rangle = \langle v(0)\rangle$ monotonically increases with γ.



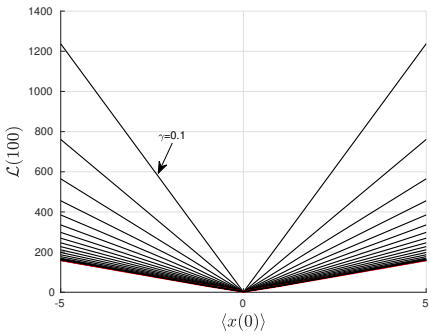
(a) Snapshots of $p(\mathbf{x}, t)$ for $\langle v(0) \rangle = 0$, $\gamma = 2.5$ and various $\langle x(0) \rangle \in (-5, 5)$.



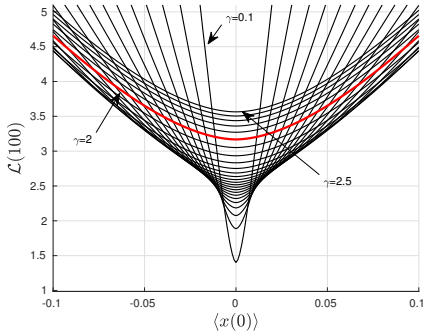
(b) Zoom-in of panel (a) showing $p(\mathbf{x}, t)$ for $\langle v(0) \rangle = 0$, $\gamma = 2.5$ and $\langle x(0) \rangle = 0$.



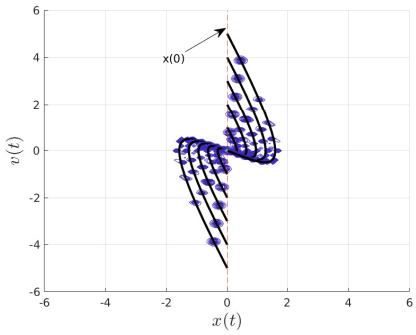
(c) Snapshots of $p(\mathbf{x}, t)$ for $\langle v(0) \rangle = 0$, $\gamma = 0.1$ and various $\langle x(0) \rangle \in (-5, 5)$.



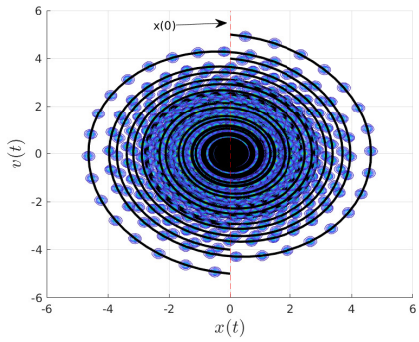
(d) $\mathcal{L}_\infty = \mathcal{L}(t = 100)$ against $\langle x(0) \rangle \in (-5, 5)$ for $\langle v(0) \rangle = 0$ and $\gamma \in (0, 2.5)$.



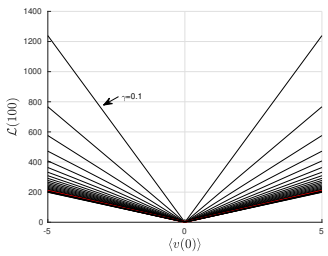
(e) Zoom-in of panel (d).



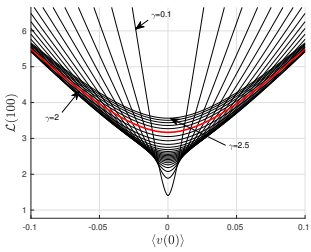
(f) Snapshots of $p(\mathbf{x}, t)$ for $\langle x(0) \rangle = 0$, $\gamma = 2.5$ and various $\langle v(0) \rangle \in (-5, 5)$.



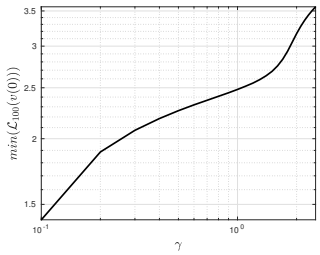
(g) Snapshots of $p(\mathbf{x}, t)$ for $\langle x(0) \rangle = 0$, $\gamma = 0.1$ and various $\langle v(0) \rangle \in (-5, 5)$.



(h) $\mathcal{L}_\infty = \mathcal{L}(t = 100)$ against $\langle v(0) \rangle \in (-5, 5)$ for $\langle x(0) \rangle = 0$ and $\gamma \in (0, 2.5)$.



(i) Zoom-in of panel (h).



(j) The minimum value of $\mathcal{L}_\infty$ at $\langle x(0) \rangle = \langle v(0) \rangle = 0$ against γ on log-log scales; $\langle x(0) \rangle = 0$.

Figure 4. Results of Equations (23) and (24) scanned over $\langle x(0) \rangle \in (-5, 5)$ for $\langle v(0) \rangle = 0$ [Figure 4a–e] and $\langle v(0) \rangle \in (-5, 5)$ for $\langle x(0) \rangle = 0$ [Figure 4f–i]. The parameter values $\omega = 1$, $D = 0.0005$, and $\gamma \in (0, 2.5)$ while the initial covariance matrix $\Sigma(0)$ has the elements $\Sigma_{11}(0) = \Sigma_{22}(0) = 0.01$, $\Sigma_{12}(0) = \Sigma_{21}(0) = 0$.

4.4. The Limit Where $\omega \to 0$

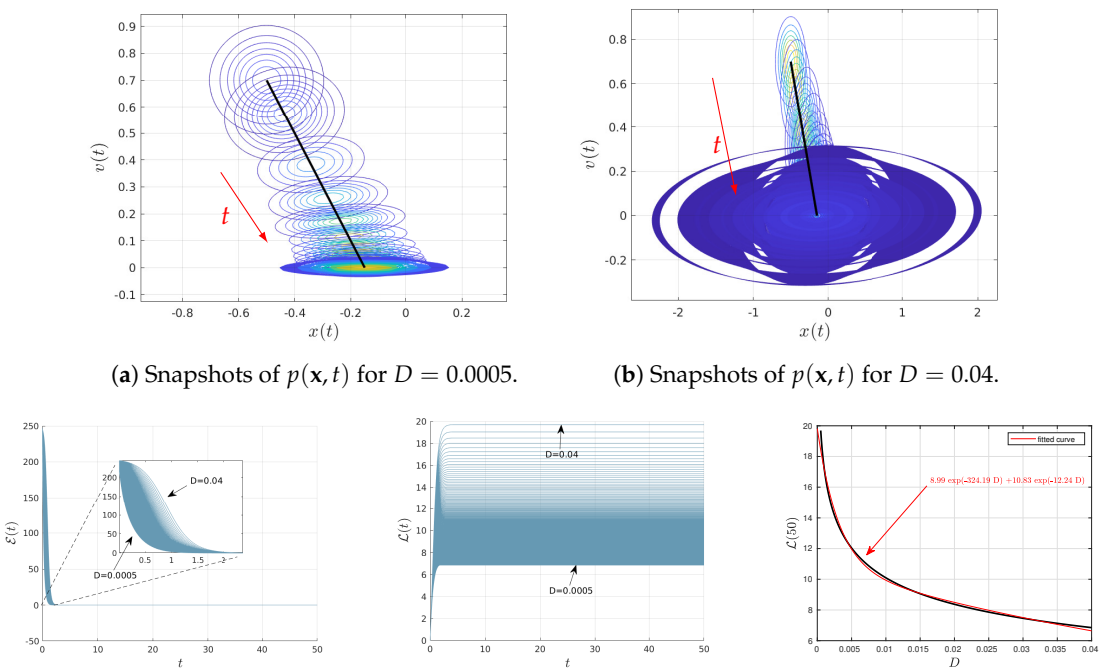
When the natural frequency $\omega = 0$ (i.e., damped-driven system like the O-U process [36]) in Equation (23), the two eigenvalues of the matrix $\mathbf{A}$ become $\lambda_1 \to -\gamma$ and $\lambda_2 \to 0$. It then easily follows that

$$\langle \mathbf{x}(t) \rangle = \begin{bmatrix} \frac{v(0) - e^{-\gamma t} v(0)}{\gamma} + x(0) \\ e^{-\gamma t} v(0) \end{bmatrix}, \tag{33}$$

and $\Sigma(t)$ is composed by the elements

$$\begin{aligned}
\Sigma_{11}(t) &= \frac{e^{-2\gamma t}\left(-D+\Sigma_{22}(0)\gamma+e^{\gamma t}(4D-\gamma(2\Sigma_{22}(0)+(\Sigma_{12}(0)+\Sigma_{21}(0))\gamma))+e^{2\gamma t}(\gamma(\Sigma_{22}(0)+\gamma(\Sigma_{12}(0)+\Sigma_{21}(0)+\Sigma_{11}(0)\gamma))+D(2\gamma t-3))\right)}{\gamma^3}, \\
\Sigma_{12}(t) &= \frac{e^{-2\gamma t}\left(D\left(-1+e^{\gamma t}\right)^2-\Sigma_{22}(0)\gamma+e^{\gamma t}\gamma(\Sigma_{22}(0)+\Sigma_{12}(0)\gamma)\right)}{\gamma^2}, \\
\Sigma_{21}(t) &= \frac{e^{-2\gamma t}\left(D\left(-1+e^{\gamma t}\right)^2-\Sigma_{22}(0)\gamma+e^{\gamma t}\gamma(\Sigma_{22}(0)+\Sigma_{21}(0)\gamma)\right)}{\gamma^2}, \\
\Sigma_{22}(t) &= \frac{e^{-2\gamma t}\left(D\left(-1+e^{2\gamma t}\right)+\Sigma_{22}(0)\gamma\right)}{\gamma}.
\end{aligned} \tag{34}$$

To investigate the case of $\omega \to 0$, we consider the scan over $D \in (0.0005, 0.04)$ for the same parameter value $\gamma = 2$, and the initial conditions as in Figure 1, apart from using $\omega = 0$ instead of $\omega = 1$. Figure 5 presents the results – snapshots of $p(\mathbf{x}, t)$, time evolutions of $\mathcal{E}(t)$, $\mathcal{L}(t)$, and $\mathcal{L}_\infty = \mathcal{L}(t = 50)$ against D in Figure 5a–e. In particular, in Figure 5e, we identify the dependence of $\mathcal{L}_\infty$ on D by fitting the results to the curve $\mathcal{L}_= 8.99e^{-324.19D} + 10.83e^{-12.24D}$.



(**a**) Snapshots of $p(\mathbf{x}, t)$ for $D = 0.0005$.

(**b**) Snapshots of $p(\mathbf{x}, t)$ for $D = 0.04$.

(**c**) Time-evolution of $\mathcal{E}(t)$.

(**d**) Time-evolution of $\mathcal{L}(t)$.

(**e**) $\mathcal{L}_\infty = \mathcal{L}(t = 50)$ against D; a fitted curve is shown in the red solid line.

Figure 5. Results of Equations (23) and (24) for $\langle x(0) \rangle = -0.5$, $\langle v(0) \rangle = 0.7$, $\gamma = 2$, $\omega = 0$, $D \in (0.0005, 0.04)$ and the initial covariance matrix $\Sigma(0)$ with elements $\Sigma_{11}(0) = \Sigma_{22}(0) = 0.01$, $\Sigma_{12}(0) = \Sigma_{21}(0) = 0$.

5. Concluding Remarks

We have presented theoretical results of time-dependent PDFs and the information length for n-th order linear autonomous stochastic processes, which can be applied to a variety of practical problems. In particular, the information length diagnostics was found as a function of the mean and covariance matrices; the latter was further expressed in terms of the covariance matrix eigenvalues. A Specific application was made to a harmonically bound particle system with the natural oscillation frequency ω, subject to a damping γ and a Gaussian white-noise (Kramer equation). We investigated how the information length depends on ω and γ, elucidating the role of critical damping $\gamma = 2\omega$ in information geometry. The fact that the information length tends to take its minimum value near the critical damping can be viewed as the simplification of dynamics and thus the decrease in information change due to the reduction of the two characteristic time scales associated with ω and γ to the one value. On the other hand, the information length in the long time limit was shown to preserve the linear geometry associated with the Gaussian statistics in a linear stochastic process, as in the case of the O-U process.

Future works would include the exploration of our results when applied to high-dimensional processes and the extension of our work to a more general (e.g., finite-correlated) stochastic noise, non-autonomous systems or non-linearly coupled systems. In particular, it will be of interest to look for a geodesic solution in non-autonomous systems [9] with the help of an external force, optimization or guiding self-organization (multi-agent systems) as well as elucidating the role of critical damping and resonances in self-organization. In addition, it would also be interesting to utilize the results introduced in [39] to predict the bound on the evolution of any observable for the Kramers problem (23), and compare it with a natural observable in such a system, the energy, for instance.

Author Contributions: Conceptualization, A.-J.G.-C. and E.-j.K.; Formal analysis, A.-J.G.-C.; Investigation, A.-J.G.-C. and E.-j.K.; Methodology, E.-j.K.; Project administration, E.-j.K.; Resources, A.-J.G.-C.; Software, A.-J.G.-C.; Supervision, E.-j.K.; Validation, A.-J.G.-C. and E.-j.K.; Visualization, A.-J.G.-C. and E.-j.K.; Writing—original draft, A.-J.G.-C. and E.-j.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no funding.

Acknowledgments: E.-j.K. acknowledges the Leverhulme Trust Research Fellowship (RF-2018-142-9).

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

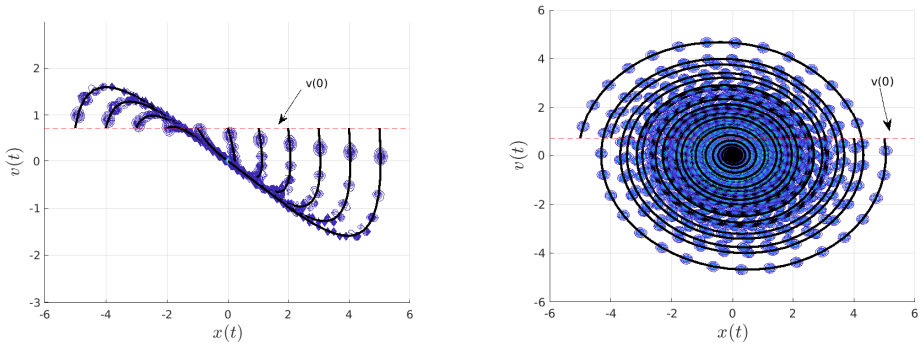
Appendix A. Analysis for Non-Zero Fixed Initial Conditions

In Section 4.3 we analysed the behaviour of the information geometry associated with the Kramer equation (Equations (23) and (24)) for different $\gamma \in (0, 2.5)$ near the equilibrium point $\langle x(0)\rangle = \langle v(0)\rangle = 0$. To this end, we plotted $\mathcal{L}_\infty$ when varying $\langle x(0)\rangle$ and $\langle v(0)\rangle$ for a fixed $\langle v(0)\rangle = 0$ and $\langle x(0)\rangle = 0$, respectively. In this Appendix, we want to show how such information geometry changes near a non-equilibrium point by scanning over $\langle x(0)\rangle$ and $\langle v(0)\rangle$ for a fixed non-zero $\langle v(0)\rangle = 0.7$ and $\langle x(0)\rangle = -0.5$, respectively. We show that the use of non-zero fixed initial conditions changes the location of the minimum $\mathcal{L}_\infty$ depending on γ. Here, we use the same parameter values $D = 0.0005$, $\omega = 1$, $\Sigma_{12}(0) = \Sigma_{21}(0) = 0$ and $\Sigma_{11}(0) = \Sigma_{22}(0) = 0.01$.

First, snapshots of $p(\mathbf{x}, t)$ are shown in Figure A1a,f for $\gamma = 2.5$ (above its critical value $\gamma = 2 = 2\omega$) while those in Figure A1b,g are for $\gamma = 0.1$ below the critical value 2. It is important to notice that there is a non-symmetric behaviour of the trajectories of the system for $\gamma \gg 0$. This is shown at Figure A1a,f whose trajectories asymmetrically vary over the initial conditions in comparison with the results shown in Figure 4a,f. By approximating $\mathcal{L}_\infty = \mathcal{L}(t = 100)$, we then show how $\mathcal{L}_\infty$ depends on $\langle x(0)\rangle$ and $\langle v(0)\rangle$ for different values of γ in Figure A1c,d and Figure A1h,i, respectively. Of prominence in Figure A1c,d is the presence of a distinct minimum in $\mathcal{L}_\infty$ for a particular value of $\langle x(0)\rangle = x_c$, $\mathcal{L}_\infty$ linearly increasing with $|\langle x(0)\rangle - x_c|$ for a sufficiently large $|\langle x(0)\rangle - x_c|$;

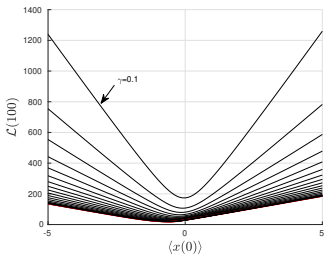
similarly, Figure A1h,i shows a distinct minimum in $\mathcal{L}_\infty$ for a particular value of $\langle v(0)\rangle = v_c$, $\mathcal{L}_\infty$ linearly increasing with $|\langle v(0)\rangle - v_c|$ for a sufficiently large $|\langle v(0)\rangle - v_c|$.

Finally, we scan over $\langle x(0)\rangle$ and $\langle v(0)\rangle$ and identify the minimum value of $\mathcal{L}_\infty$ for a given γ and plot this minimum value of $\mathcal{L}_\infty$ (at x_c and v_c) against γ in Figure A1e,j. In Figure A1e,j, $\mathcal{L}_\infty$ against γ takes its minimum near the critical damping $\gamma = 2\omega = 2$ (shown in a vertical line), as observed previously in Sections 4.1 and 4.2. This is clearly different from the behaviour of the minimum value of $\mathcal{L}_\infty$ against γ (for the equilibrium point $\langle x(0)\rangle = 0$ and $\langle v(0)\rangle = 0$) in Figure 4j where $\mathcal{L}_\infty$ monotonically increases with γ. This is because for $\langle x(0)\rangle = 0$ and $\langle v(0)\rangle = 0$, mean values does not change over time, with less effect of oscillations (ω) and thus the critical damping $\gamma = 2\omega$.

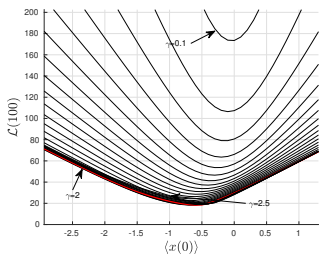


(**a**) Snapshots of $p(\mathbf{x}, t)$ for $\langle v(0)\rangle = 0.7$, $\gamma = 2.5$ and various $\langle x(0)\rangle \in (-5, 5)$.

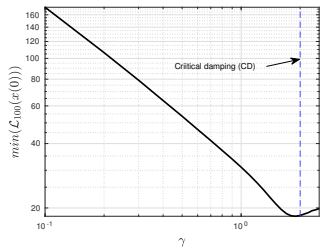
(**b**) Snapshots of $p(\mathbf{x}, t)$ for $\langle v(0)\rangle = 0.7$, $\gamma = 0.1$ and various $\langle x(0)\rangle \in (-5, 5)$.



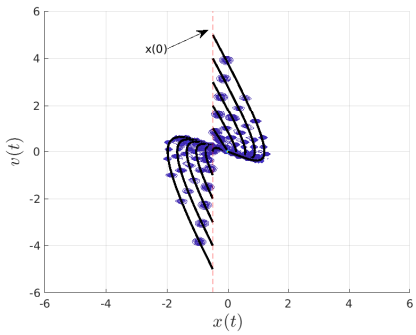
(**c**) $\mathcal{L}_\infty = \mathcal{L}(t = 100)$ against $\langle x(0)\rangle \in (-5, 5)$ for $\langle v(0)\rangle = 0.7$ and $\gamma \in (0, 2.5)$.



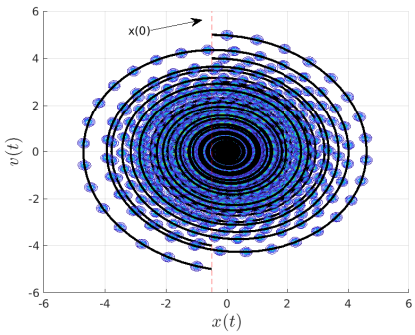
(**d**) Zoom-in of panel (c).



(**e**) The minimum value of $\mathcal{L}_\infty$ over $\langle x(0)\rangle \in (-5, 5)$ against γ on log-log scales; $\langle v(0)\rangle = 0.7$.

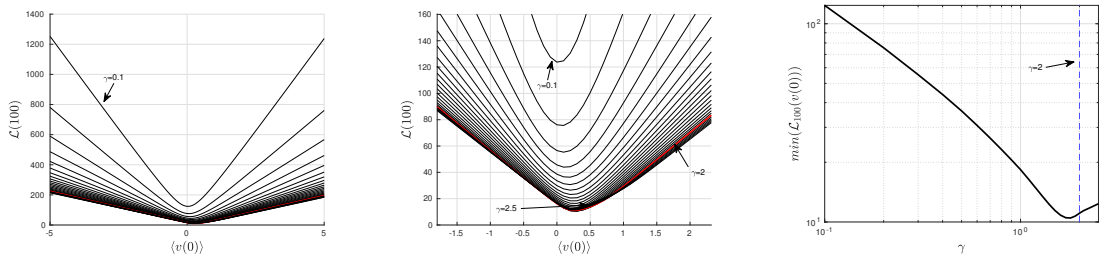


(**f**) Snapshots of $p(\mathbf{x}, t)$ for $\langle x(0)\rangle = -0.5$, $\gamma = 2.5$ and various $\langle v(0)\rangle \in (-5, 5)$.



(**g**) Snapshots of $p(\mathbf{x}, t)$ for $\langle x(0)\rangle = -0.5$, $\gamma = 0.1$ and various $\langle v(0)\rangle \in (-5, 5)$.

Figure A1. *Cont.*



(h) $\mathcal{L}_\infty = \mathcal{L}(t = 100)$ against $\langle v(0)\rangle \in (-5, 5)$ for $\langle x(0)\rangle = -0.5$ and $\gamma \in (0, 2.5)$.

(i) Zoom-in of panel (h).

(j) The minimum value of $\mathcal{L}_\infty$ over $\langle v(0)\rangle \in (-5, 5)$ against γ on log-log scales; $\langle x(0)\rangle = -0.5$.

Figure A1. Results of Equations (23) and (24) scanned over $\langle x(0)\rangle \in (-5, 5)$ for $\langle v(0)\rangle = 0.7$ [Figure A1a–e] and $\langle v(0)\rangle \in (-5, 5)$ for $\langle x(0)\rangle = -0.5$ [Figure A1f–j]. The parameter values $\omega = 1$, $D = 0.0005$, and $\gamma \in (0, 2.5)$ while the initial covariance matrix $\Sigma(0)$ has the elements $\Sigma_{11}(0) = \Sigma_{22}(0) = 0.01$, $\Sigma_{12}(0) = \Sigma_{21}(0) = 0$.

References

1. Klebaner, F. *Introduction to Stochastic Calculus with Applications*; Imperial College Press: London, UK, 2012.
2. Ramstead, M.; Friston, K.; Hipólito, I. Is the Free-Energy Principle a Formal Theory of Semantics? From Variational Density Dynamics to Neural and Phenotypic Representations. *Entropy* **2020**, *22*, 889. [CrossRef]
3. Gershenson, C. Guiding the Self-organization of Cyber-Physical Systems. *Front. Robot. AI* **2020**, *7*, 41. [CrossRef]
4. Prokopenko, M.; Boschett, F.; Ryan, A. An information-theoretic primer on complexity, self-organization. *Complexity* **2009**, *15*, 11–28. [CrossRef]
5. Trianni, V.; Nolfi, S.; Dorigo, M. *Evolution, Self-Organization and Swarm Robotics*; Springer: Berlin/Heidelberg, Germangy, 2008; pp. 163–191.
6. Wilson, A.; Schultz, J.; Murphey, T. Trajectory synthesis for Fisher information maximization. *IEEE Trans. Robot.* **2014**, *30*, 1358–1370. [CrossRef] [PubMed]
7. Correa, J. Metrics of emergence, self-organization, and complexity for EWOM research. *Front. Phys.* **2020**, *8*, 5232–5241. [CrossRef]
8. Gros, C. Generating functionals for guided self-organization. In *Guided Self-Organization: Inception*; Springer: Berlin/Heidelberg, Germangy, 2014; pp. 53–66.
9. 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]
10. Nicholson, S.; Kim, E. Investigation of the statistical distance to reach stationary distributions. *Phys. Lett. A* **2015**, *379*, 83–88. [CrossRef]
11. Li, W. Transport information geometry I: Riemannian calculus on probability simplex. *arXiv* **2020**, arXiv:math.DG/1803.06360.
12. Kowalski, A.; Martín, M.; Plastino, A.; Rosso, O.; Casas, M. Distances in Probability Space and the Statistical Complexity Setup. *Entropy* **2011**, *13*, 1055–1075. [CrossRef]
13. Prokopenko, M.; Gershenson, C. Entropy Methods in Guided Self-Organisation. *Entropy* **2014**, *16*, 5232–5241. [CrossRef]
14. Parr, T.; Da-Costa, L.; Friston, K. Markov blankets, information geometry and stochastic thermodynamics. *Philos. Trans. R. Soc. A* **2020**, *378*, 20190159. [CrossRef]
15. Erven, T.; Harremoës, P. Rényi divergence and Kullback-Leibler divergence. *IEEE Trans. Inf. Theory* **2014**, *60*, 3797–3820. [CrossRef]
16. Takatsu, A. Wasserstein geometry of Gaussian measures. *Osaka J. Math.* **2011**, *48*, 1005–1026.
17. Li, W.; Zhao, J. Wasserstein information matrix. *arXiv* **2020**, arXiv:math.ST/1910.11248.
18. Lott, J. Some Geometric Calculations on Wasserstein Space. *Commun. Math. Phys.* **2008**, *277*, 423–437. [CrossRef]

19. Risken, H. *The Fokker-Planck Equation: Methods of Solution and Applications*; Springer: Berlin, Germany, 1996.
20. Gangbo, W.; McCann, R. The geometry of optimal transportation. *Acta Math* **1996**, *177*, 113. [CrossRef]
21. Frieden, B. *Science from Fisher Information: A Unification*; Cambridge University Press: Cambridge, UK, 2004.
22. Costa, S.; Santos, S.; Strapasson, J. Fisher information distance: A geometrical reading. *Discret. Math.* **2015**, *197*, 59–69. [CrossRef]
23. Martin, M.; Plastron, A.; Ross, O. Statistical complexity and disequilibrium. *Phys. Lett. A* **2003**, *311*, 126. [CrossRef]
24. 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]
25. Zamir, R. A proof of the Fisher information inequality via a data processing argument. *IEEE Trans. Inf. Theory* **1998**, *44*, 1246–1250. [CrossRef]
26. Heseltine, J.; Kim, E. Comparing information metrics for a coupled Ornstein-Uhlenbeck process. *Entropy* **2019**, *21*, 775. [CrossRef]
27. Kim, E. Investigating Information Geometry in Classical and Quantum Systems through Information Length. *Entropy* **2018**, *20*, 574–585. [CrossRef]
28. Kim, E.; Hollerbach, R. Signature of nonlinear damping in geometric structure of a nonequilibrium process. *Phys. Rev. E* **2017**, *95*, 022137. [CrossRef]
29. Hollerbach, R.; Dimanche, D.; Kim, E. Information geometry of nonlinear stochastic systems. *Entropy* **2018**, *20*, 550. [CrossRef]
30. Nicholson, S.; Kim, E. Geometric method for forming periodic orbits in the Lorenz system. *Phys. Scr.* **2016**, *91*, 044006. [CrossRef]
31. Kim, E.; Lewis, P. Information length in quantum systems. *J. Stat. Mech. Theory Exp.* **2018**, *2018*, 043106. [CrossRef]
32. Zee, A. *Quantum Field Theory in A Nutshell*; Princeton University Press: Princeton, NJ, USA, 2010; Volume 7.
33. Tracy, D.; Jinadasa, K. Expectations of products of random quadratic forms. *Stoch. Anal. Appl.* **1986**, *4*, 111–116. [CrossRef]
34. Magnus, J.R.; Neudecker, H. *Matrix Differential Calculus with Applications in Statistics and Econometrics*; John Wiley & Sons: Chichester, UK, 2019.
35. Jolliffe, I.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef]
36. Heseltine, J.; Kim, E. Novel mapping in non-equilibrium stochastic processes. *J. Phys. A* **2016**, *49*, 175002. [CrossRef]
37. Velasco, S. On the Brownian motion of a harmonically bound particle and the theory of a Wiener process. *Eur. J. Phys.* **1985**, *6*, 259–265. [CrossRef]
38. Chen, C.T. *Linear System Theory and Design*; Oxford University Press: New York, NY, USA, 2013; Volume 7.
39. Ito, S.; Dechant, A. Stochastic Time Evolution, Information Geometry, and the Cramér-Rao Bound. *Phys. Rev. X* **2020**, *10*, 021056. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).