

Information Geometry Control under the Laplace Assumption [†]

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

Centre for Fluid and Complex Systems, Coventry University, Coventry CV1 2TT, UK

* Correspondence: adrianjguelc@gmail.com

[†] Presented at the 41st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Paris, France, 18–22 July 2022.

Abstract: By combining information science and differential geometry, information geometry provides a geometric method to measure the differences in the time evolution of the statistical states in a stochastic process. Specifically, the so-called information length (the time integral of the information rate) describes the total amount of statistical changes that a time-varying probability distribution takes through time. In this work, we outline how the application of information geometry may permit us to create energetically efficient and organised behaviour artificially. Specifically, we demonstrate how nonlinear stochastic systems can be analysed by utilising the Laplace assumption to speed up the numerical computation of the information rate of stochastic dynamics. Then, we explore a modern control engineering protocol to obtain the minimum statistical variability while analysing its effects on the closed-loop system's stochastic thermodynamics.

Keywords: information geometry; non-linear stochastic systems; information length; stochastic thermodynamics



Citation: Guel-Cortez, A.-J.; Kim, E.-j. Information Geometry Control under the Laplace Assumption. *Phys. Sci. Forum* **2022**, *5*, 25. <https://doi.org/10.3390/psf2022005025>

Academic Editors: Frédéric Barbaresco, Ali Mohammad-Djafari, Frank Nielsen and Martino Trassinelli

Published: 12 December 2022

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



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Stochastic systems are ubiquitous and include a large set of complex systems, such as the time evolution of molecular motors [1], the stock market [2], decision making [3], population dynamics [4] or engineering systems with parameter uncertainties [5]. The description of stochastic dynamics commonly involves the calculation of time-varying probability density functions (PDFs) governed by a Fokker–Planck equation and its corresponding stochastic differential equation [1]. Since a time-varying PDF describes all the possible trajectories the stochastic system can take in time, such a formalism has been advantageously applied in emergent fields as stochastic thermodynamics [6] or inference control [7].

As time-varying PDFs contain enormous dynamical information, defining a metric of the path's length that a stochastic system takes through time can bring benefits when, for example, designing “efficient” systems (for instance, see [8]). In this regard, the field of information geometry has brought to light a true metric of the differences in the time evolution of the statistical states in a stochastic process [9,10]. Specifically, the concept of information length (IL) [11,12], given by the time integral of the stochastic dynamics information rate, describes the total amount of statistical changes that a time-varying probability distribution takes through time. The previous works showed that IL provides a link between stochastic processes, complexity and geometry [12]. Additionally, IL has been applied to the quantification of hysteresis in forward–backward processes [8,13], correlation and self-regulation among different players [13], phase transitions [14], and prediction of sudden events [15]. It is worth noting that in nonlinear stochastic systems, IL is generally difficult to obtain because the analytical/numerical solution of the Fokker–Planck equation and the execution of stochastic simulations are usually complicated and computationally costly. Hence, analytical simplifications are advantageous to ease IL's calculation while increasing the possibility of applying IL to broader practical scenarios [16].

Even though the information rate may seem purely like a statistical quantity, its meaning can be understood in relation to the thermodynamics [11,12]. This result is of great advantage as we may use it to quantify the effects, for example, that a minimum statistical variability (constant information rate) control could have on the system's energetic behaviour. Note that the idea of thermodynamic informed control systems is not something new. For instance, since the beginning of the 21st century, various works have proposed the consideration of entropy-informed control protocols to generate “intelligent/efficient” systems (for further details, see [17] and the references therein). Yet, describing the effects of an information geometry-informed control protocol over the system's stochastic thermodynamics is to be carried out.

In this regard, we consider the application of the so-called Laplace assumption (Gaussian approximation of the system's time-varying PDF) to the computation of IL and the information rate for a set of nonlinear stochastic differential equations. By using this assumption, we derive the values of the entropy rate, entropy production and entropy flow and their relation to the information rate. Then, we formulate an optimisation problem for the minimum information variability control and study the closed-loop stochastic dynamics and thermodynamics in a numerical example. Thus, creating a connection between information geometry, stochastic thermodynamics and control engineering (Figure 1).

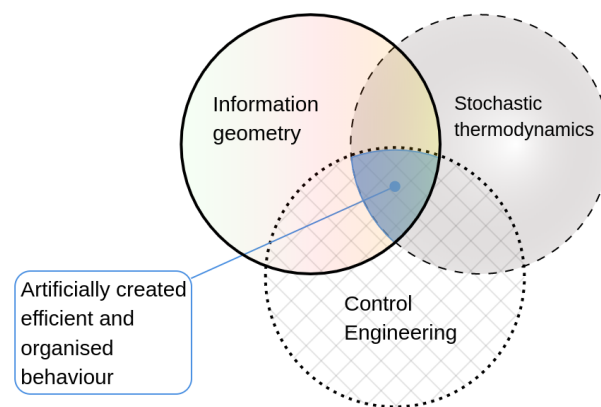


Figure 1. The combination of methods from information geometry, stochastic thermodynamics and control engineering may lead to the creation of energetically efficient and organised behaviour.

To help readers, in the following, we summarise our notations. $\mathbb{R}$ is the set of real numbers; $\mathbf{x} \in \mathbb{R}^n$ 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 $n \times n$ (bold-face letters are used to represent vectors and matrices); $\text{Tr}(\mathbf{A})$ corresponds to the trace of the matrix $\mathbf{A}$; $|\mathbf{A}|$, $\text{vec}(\mathbf{A})$, $\mathbf{A}^\top$ and $\mathbf{A}^{-1}$ are the determinant, vectorisation, transpose and inverse of matrix $\mathbf{A}$, respectively. The value $\mathbf{I}_n$ denotes the identity matrix of order n . Newton's notation is used for the partial derivative with respect to the variable t (i.e. $\frac{\partial y}{\partial t} = \dot{y}$). Finally, the average of a random vector $\boldsymbol{\zeta}$ is denoted by $\boldsymbol{\mu} := \langle \mathbf{x} \rangle$, the angular brackets representing the average.

2. Model

Throughout this work, the following set of nonlinear Langevin equations is considered

$$\dot{\mathbf{x}} = f(\mathbf{x}, u) + \boldsymbol{\zeta}. \quad (1)$$

Here, $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a function taking as input a vector $\mathbf{x} \in \mathbb{R}^n$ and a bounded smooth time-dependent deterministic function $u(t) \in \mathbb{R}$ to output a vector $f(\mathbf{x}) \in \mathbb{R}^n$ with elements f_i ($i = 1, 2, \dots, n$); $\boldsymbol{\zeta} \in \mathbb{R}^n$ is a Gaussian stochastic noise given by an n dimensional vector of δ -correlated Gaussian noises ζ_i ($i = 1, 2, \dots, n$), with the following statistical property

$$\langle \zeta_i(t) \rangle = 0, \langle \zeta_i(t) \zeta_j(t_1) \rangle = 2D_{ij}(t)\delta(t - t_1), D_{ij}(t) = D_{ji}(t), \forall i, j = 1, \dots, n. \quad (2)$$

The Fokker–Planck equation of (1) is

$$\dot{p} = -\nabla \cdot J = -\nabla \cdot fp + \nabla \cdot \mathbf{D}\nabla p = -\sum_{i=1}^n \partial_{x_i}(f_i p) + \sum_{i,j=1}^n (\partial_{x_i} D_{ij} \partial_{x_j}) p, \quad (3)$$

where $J = [f_1 p - \mathbf{D}_1 \nabla p, \dots, f_n p - \mathbf{D}_n \nabla p]$.

2.1. The Laplace Assumption

The Laplace assumption allows us to describe the solution of (3) through a fixed multivariable Gaussian distribution given by [4]

$$p(\mathbf{x}; t) = \frac{1}{\sqrt{|2\pi\mathbf{\Sigma}|}} e^{\frac{1}{2}Q(\mathbf{x}; t)}, \quad (4)$$

where $Q(\mathbf{x}; t) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}(t))^T \mathbf{\Sigma}^{-1}(t)(\mathbf{x} - \boldsymbol{\mu}(t))$, and $\boldsymbol{\mu}(t) \in \mathbb{R}^n$ and $\mathbf{\Sigma}(t) \in \mathbb{R}^{n \times n}$ are the mean and covariance value of the random variable $\mathbf{x}$. The value of the mean $\boldsymbol{\mu}(t)$ and covariance matrix $\mathbf{\Sigma}(t)$ can be obtained from the following result.

Proposition 1 (The Laplace assumption). *Under the Laplace assumption, the dynamics of the mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$ at any time t of a nonlinear stochastic differential system (1) are governed by the following differential equations*

$$\dot{\boldsymbol{\mu}} = \left[f_1(\boldsymbol{\mu}, u) + \frac{1}{2} \text{Tr}(\mathbf{\Sigma} \mathbf{H}_{f_1}), f_2(\boldsymbol{\mu}, u) + \frac{1}{2} \text{Tr}(\mathbf{\Sigma} \mathbf{H}_{f_2}), \dots, f_n(\boldsymbol{\mu}, u) + \frac{1}{2} \text{Tr}(\mathbf{\Sigma} \mathbf{H}_{f_n}) \right]^T \quad (5)$$

$$\dot{\mathbf{\Sigma}} = \mathbf{J}_f \mathbf{\Sigma} + \mathbf{\Sigma} \mathbf{J}_f^T + \mathbf{D} + \mathbf{D}^T \quad (6)$$

where $\mathbf{H}_{f_i}$ is the Hessian matrix of the function $f_i(\mathbf{x}, u)$, and $\mathbf{J}_f$ is the Jacobian of the function $f(\mathbf{x}, u)$.

Proof. See Appendix A. $\square$

Note that when $f_i(\mathbf{x}; t)$ in (1) is a linear function defined as

$$f_i(\mathbf{x}; t) := \mathbf{A}_i \mathbf{x}(t) + \mathbf{B}_i \mathbf{u}(t) = \sum_{j=1}^n a_{ij} x_j(t) + \sum_j^p b_{ij} u_j(t), \quad (7)$$

i.e., we consider a set of particles driven by a harmonic potential ($\mathbf{A}_i \mathbf{x}(t)$) and a deterministic force ($\mathbf{B}_i \mathbf{u}(t)$), the value $\mathbf{H}_{f_i} = 0$, meaning that the mean value $\boldsymbol{\mu}$ is not affected by the covariance matrix $\mathbf{\Sigma}$.

Limits of the Laplace Assumption

Since the Laplace assumption does not always hold, we first check on its limitation by considering the following cubic stochastic differential equation

$$\dot{x}(t) = -\gamma x(t)^3 + u(t) + \zeta(t), \quad (8)$$

where $\langle \zeta \rangle = 0$, $\langle \zeta(t) \zeta(t') \rangle = 2D\delta(t - t')$ and $\gamma \in \mathbb{R}^+$. Then, we denote $q(x, t)$ that is based on the Laplace Assumption, which take the following Gaussian form

$$q(x, t) = \frac{1}{\sqrt{2\pi\Sigma}} e^{\frac{1}{2} \frac{(x-\mu)^2}{\Sigma}}, \quad (9)$$

where μ and Σ are determined by the solution of

$$\dot{\mu} = -\gamma\mu^3 + u - 3\gamma\mu\Sigma, \quad (10)$$

$$\dot{\Sigma} = -6\gamma\Sigma\mu^2 + 2D. \quad (11)$$

To obtain the real system PDF $\tilde{p}(x, t)$ of system (8), we use stochastic simulations and kernel density estimators (for further details see [18]). Now, to highlight the limits of the Gaussian approximation $q(x, t)$, we apply the Kullback divergence (KL) D_{KL} or relative entropy between the estimated $\tilde{p}$ and the Gaussian approximation q of the time-varying system (8) PDFs defined as

$$D_{KL}(\tilde{p}||q) = \int_{\mathbb{R}} p(x; t) \log\left(\frac{p(x; t)}{q(x; t)}\right) dx. \quad (12)$$

Figure 2 shows the KL divergence trough time between $\tilde{p}$ and q when changing the parameters γ and D in Equation (8). The result shows that a valid LA requires a small damping (slow behaviour) and a wider noise amplitude in comparison with the initial value of Σ .

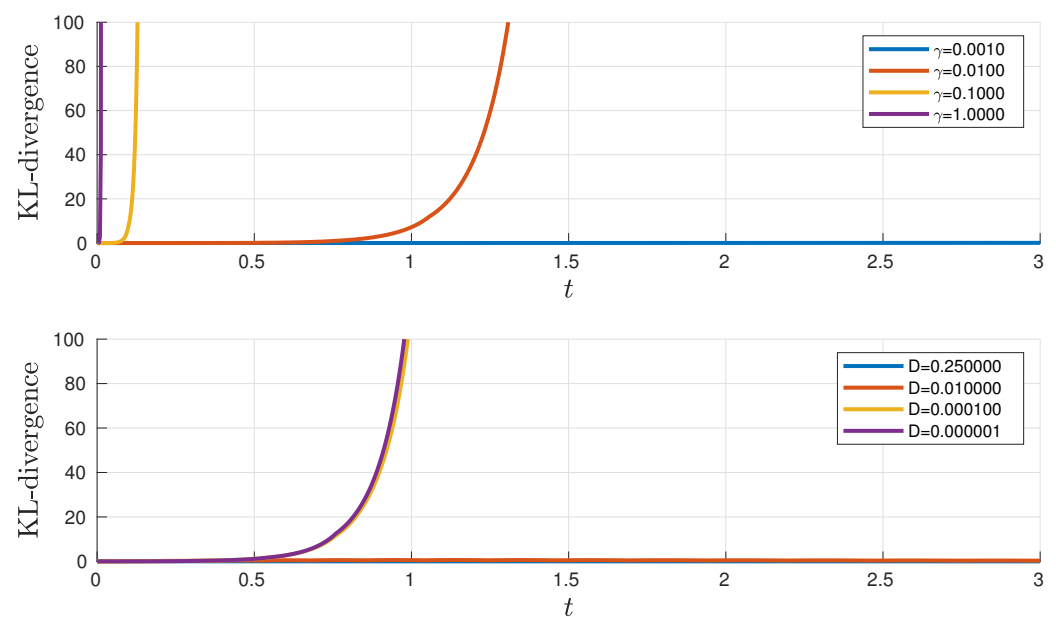


Figure 2. KL divergence between the value $\tilde{p}(x; t)$ and the value $q(x, y)$ varying the values γ and D of Equation (8). When γ changes, $D = 0.01$; when D changes, $\gamma = 0.01$. The initial condition is a Gaussian distribution defined by $\mu(0) = 5$ and $\Sigma(0) = 0.01$. See code at [19].

3. Stochastic Thermodynamics

Stochastic thermodynamics uses stochastic calculus to draw a connection between the “micro/mesoscopic stochastic dynamics” and the “macroscopic thermodynamics” [1,6]. In physical terms, this means that stochastic thermodynamics describes the interaction of a micro/mesoscopic system with one or multiple reservoirs (for instance, the dynamics of a Brownian particle suspended in a fluid in thermodynamic equilibrium described by a Langevin/Fokker Planck equation).

3.1. Entropy Rate

Given a time-varying multivariable PDF $p(\mathbf{x}; t)$, we can calculate the entropy rate, a fundamental concept of stochastic thermodynamics, as follows [20]

$$\dot{S}(t) = \frac{d}{dt} S(t) = - \int_{\mathbb{R}^n} \dot{p}(\mathbf{x}; t) \ln(p(\mathbf{x}; t)) d\mathbf{x} = \Pi - \Phi. \quad (13)$$

Proposition 2. Under the Laplace assumption, the value of entropy rate $\dot{S}$, entropy production Π and entropy flow Φ is given by

$$\dot{S} = \text{Tr}(\Sigma^{-1} \mathbf{D}) + \text{Tr}(\mathbf{J}_f) = \frac{1}{2} \text{Tr}(\Sigma^{-1} \dot{\Sigma}), \quad (14)$$

$$\Pi = \text{Tr}(\Sigma^{-1} \mathbf{D}) + \text{Tr}(f(\mu, u)^\top \mathbf{D}^{-1} f(\mu, u)) + \text{Tr}(\mathbf{J}_f \mathbf{D}^{-1} \mathbf{J}_f^\top \Sigma) + 2\text{Tr}(\mathbf{J}_f), \quad (15)$$

$$\Phi = \text{Tr}(f(\mu, u)^\top \mathbf{D}^{-1} f(\mu, u)) + \text{Tr}(\mathbf{J}_f \mathbf{D}^{-1} \mathbf{J}_f^\top \Sigma) + \text{Tr}(\mathbf{J}_f). \quad (16)$$

Proof. See Appendix B. $\square$

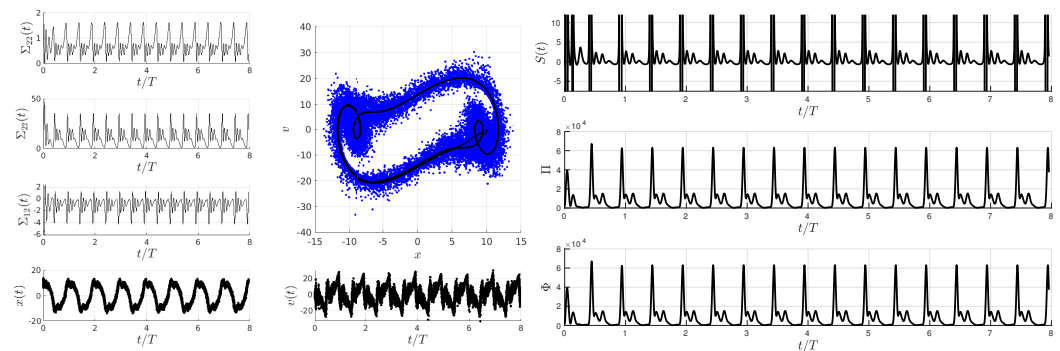
3.2. Example

To illustrate the application of Proposition 2, consider the following Langevin form of the Duffing equation

$$\begin{aligned} \dot{x}(t) &= v(t) + \zeta_1(t) \\ \dot{v}(t) &= -\delta v(t) - \alpha x(t) - \beta x(t)^3 + \gamma \cos(\omega t) + \zeta_2(t) \end{aligned} \quad (17)$$

where $x(t)$ is the displacement at time t , $v(t) = \dot{x}(t)$ is the first derivative of x with respect to time, i.e., velocity, ζ is a delta correlated noise, and the values $\delta, \alpha, \beta, \gamma$ and ω are given constants.

Figure 3 shows a simulation of (17) using the deterministic equations of the mean vector $\mu = [\langle x \rangle, \langle v \rangle]^\top$ and covariance matrix Σ as described by Proposition 1. Specifically, Figure 3a includes the time evolution of the random variables x and v with its phase portrait, and the time evolution of Σ_{11}, Σ_{12} and Σ_{22} . Figure 3b shows the time evolution of the system's stochastic thermodynamics including the entropy rate $\dot{S}$, the entropy production Π and the entropy flow Φ . In all subplots, time is scaled by the factor $T = 2\pi/\omega$.



(a) Time evolution.

(b) Stochastic thermodynamics.

Figure 3. Simulation of dynamics and stochastic thermodynamics of the Duffing equation under the Laplace assumption.

More importantly, Figure 3 shows that via Propositions 1 and 2 it is possible to describe the thermodynamics of any nonlinear stochastic system at every instant of time. Hence, as will be discussed in Section 5, Propositions 1 and 2 allow us to perceive the effects of a control protocol on the closed-loop system thermodynamics.

4. Information Length and Information Rate

For a time-varying multivariable PDF $p(\mathbf{x}; t)$, we define its IL $\mathcal{L}$ as [15,16]

$$\mathcal{L}(t) = \int_0^t \left(\sqrt{\int_{\mathbb{R}^n} p(\mathbf{x}; \tau) [\partial_\tau \ln p(\mathbf{x}; \tau)]^2 d\mathbf{x}} \right) d\tau = \int_0^t \Gamma(\tau) d\tau, \quad (18)$$

where Γ is called the information rate. The value of Γ^2 can be understood as the Fisher information where the time is the control parameter [12]. Since Γ gives the rate of change of $p(\mathbf{x}; t)$, its time integral $\mathcal{L}$ quantifies the amount of statistical changes that the system goes through in time from the initial PDF $p(\mathbf{x}; 0)$ to a final PDF $p(\mathbf{x}; t)$ [16].

Under the Laplace assumption, i.e., when $p(\mathbf{x}; t)$ is a Gaussian PDF, the value of information rate Γ of the joint PDF takes the compact form [11,16]

$$\Gamma^2 = \dot{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma}^{-1} \dot{\boldsymbol{\mu}} + \frac{1}{2} \text{Tr}((\boldsymbol{\Sigma}^{-1} \dot{\boldsymbol{\Sigma}})^2), \quad (19)$$

where the time derivatives of $\dot{\boldsymbol{\mu}}$ and $\dot{\boldsymbol{\Sigma}}$ are given by Equations (5) and (6), respectively.

Relation with Stochastic Thermodynamics

Considering a fully decoupled nonlinear stochastic system and using the Laplace assumption, the value of the information rate Γ^2 is related to the entropy production Π and the entropy flow $\dot{S}$ as follows

$$\Gamma^2 = \sum_{i=1}^n \frac{D_{ii}}{\Sigma_{ii}} \Pi_i + \sum_{i=1}^n \dot{S}_i^2 + \frac{1}{2} \sum_{i=1}^n H_{f_i} (\dot{\mu}_i + f_i(\mu_i, u)), \quad (20)$$

where Π_i and $\dot{S}_i$ are the entropy production and entropy rate from the marginal PDF $p(x_i, t)$ of x_i . $H_{f_i} = \frac{\partial^2}{\partial x_i^2} f_i(\mu_i, t)$. If f_i describes a harmonic potential (7), then $H_{f_i} = 0$ and (20) lead to the expression

$$\Gamma^2 = \sum_i^n \frac{D_{ii}}{\Sigma_{ii}} \Pi_i + \sum_i^n \dot{S}_i^2. \quad (21)$$

Note that Equation (21) gives us a case where a minimum information length $\mathcal{L}$ would produce both a minimum entropy production/rate and a minimum statistical variability behaviour.

5. Minimum Variability Control

To impose a minimum statistical variability when going from an initial to a desired state (for instance, see [21]), we propose the optimisation problem with the following cost function

$$\mathbf{c} = \arg \min_{\mathbf{c}} \left(J = \int_0^{t_f} ((\Gamma(t) - \Gamma(0))^2 + (\mathbf{Y}(t) - \mathbf{Y}_d)^\top \mathbf{Q} (\mathbf{Y}(t) - \mathbf{Y}_d) + \tilde{\mathbf{c}}(t)^\top \mathbf{R} \tilde{\mathbf{c}}(t)) dt \right) \quad (22)$$

where $\mathbf{Y}(t) := [\boldsymbol{\mu}(t), \text{vec}(\boldsymbol{\Sigma}(t))]^\top$, $\mathbf{Y}_d := [\boldsymbol{\mu}_d, \text{vec}(\boldsymbol{\Sigma}_d)]^\top$, $\tilde{\mathbf{c}}(t) := [u(t), \text{vec}(\mathbf{D}(t))]^\top$, $\mathbf{Q} \in \mathbb{R}^{n+n \times n}$, and $\mathbf{R} \in \mathbb{R}^{1+n \times n}$. The solution $\mathbf{c}(t)$ corresponds to the control vector that allows us to obtain the minimum statistical variability. In Equation (22), the term $(\Gamma(t) - \Gamma(0))^2$ keeps the information variability constant. The term involving $\mathbf{Q}$ drives the system to reach a given PDF defined by $\mathbf{Y}_d$. The term containing $\mathbf{R}$ in the right-hand side of (22) regularises the control action $\mathbf{c}$ to avoid abrupt changes in the inputs. Note that the values of $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$ and Γ can be easily computed for any nonlinear stochastic process through Proposition (1) or the Laplace assumption. A control that comes after solving (22) would be called an information length quadratic regulator (IL-QR).

5.1. Model Predictive Control

A solution to the proposed optimisation problem (22) can be obtained by one of the most popular optimisation-based control techniques currently available—the so-called model-predictive-control (MPC) scheme [22]. Generally, MPC is an online optimisation algorithm for constrained control problems whose benefits have been recognized in applications to robotics [23], solar energy [24] or bioengineering [25]. Furthermore, MPC has the advantage of being easily implemented owing to packages such as CasADi [26] or the Hybrid Toolbox [27].

Figure 4 briefly details the working principle of the MPC's optimiser in the form of a block diagram. The MPC method consists of utilising a prediction model to solve the optimisation problem in a finite horizon. Then, the optimal solution is applied to the system

in real-time. Finally, the system's output is fed back to the MPC algorithm to start the optimisation procedure again. In this work, we ease the prediction and simulation of the stochastic process by employing the Laplace assumption.

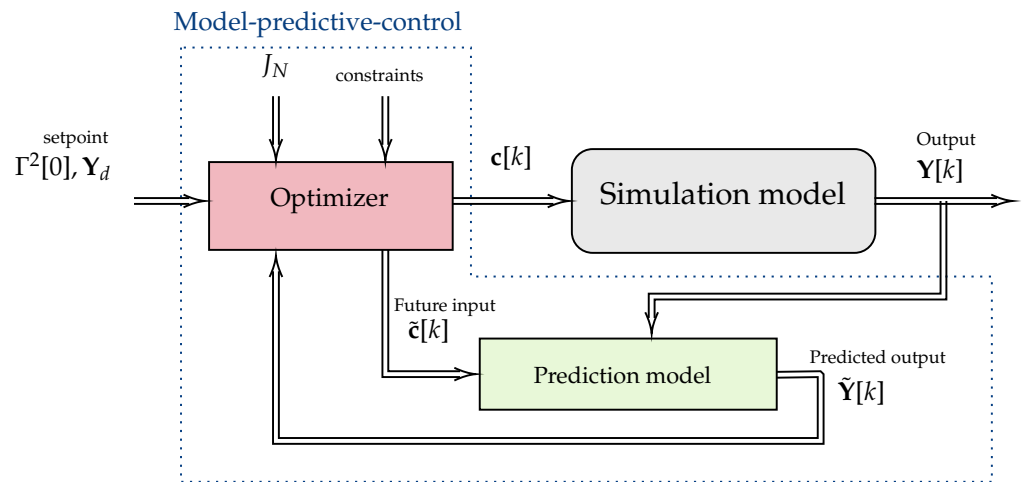


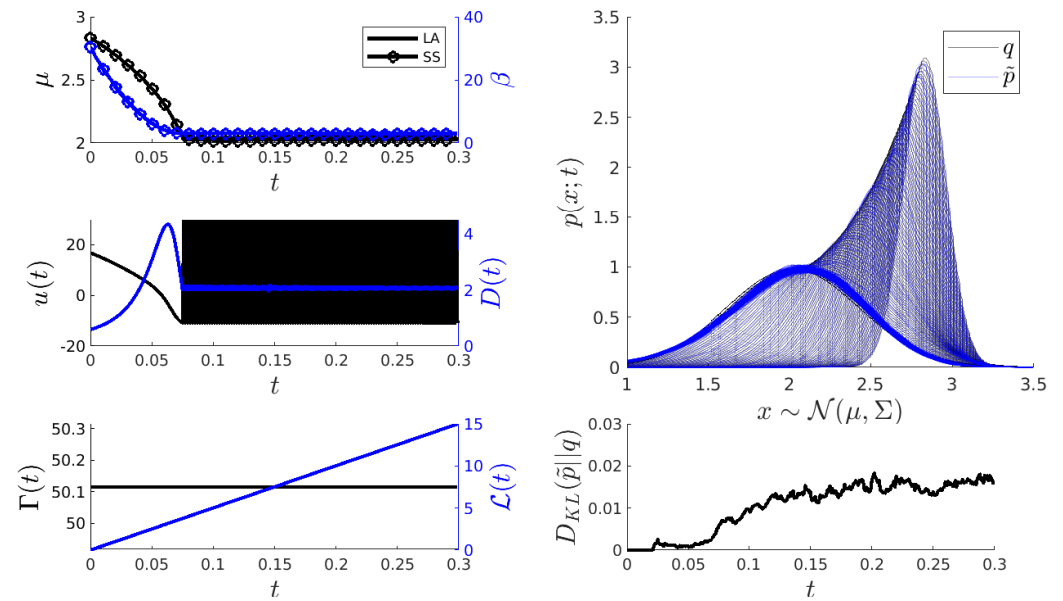
Figure 4. Model predictive control block diagram.

5.2. Example

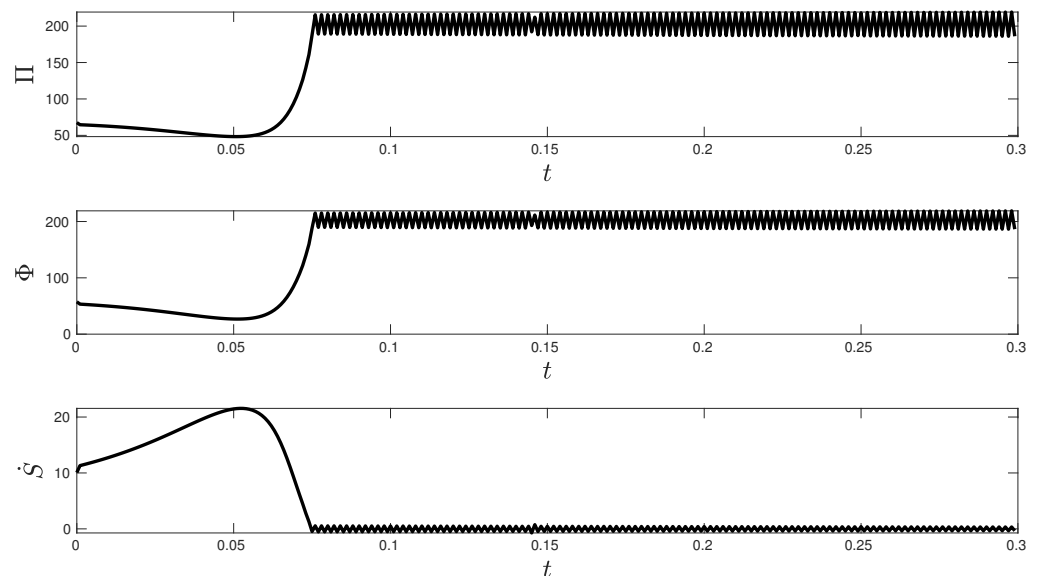
We now present an example of the application of the MPC method to obtain the minimum variability behaviour of a stochastic system. Figure 5 shows the IL-QR applied to the cubic stochastic process given by Equation (8), where the control vector and the state vector are given by $\mathbf{c} = [u, D]^T$ and $\mathbf{Y} = [\mu, \Sigma]^T$, respectively. In the simulation, the initial state is $\mathbf{Y}(0) = [2 + 5/6, 1/(2 \times 30)]^T$, while the desired state is $\mathbf{Y}_d(t) = [2 + 1/30, 1/(2 \times 3)]^T$. Additionally, we consider the parameters $\gamma = 0.1$, $T_s = 1 \times 10^{-3}$, $N = 5$, $I_L = 1 \times 10^3$, $\mathbf{R} = 1 \times 10^{-5} \mathbf{I}_2$, $Q_{12} = Q_{21} = 0$, $Q_{11} = 1 \times 10^2$ and $Q_{22} = 8 \times 10^2$. Here, T_s is the integration time step, and N is the number of future time steps considered in the prediction model. The value of $\Gamma(0)$ is imposed via the initial conditions and Equation (19).

In Figure 5a, we show the time evolution of the mean μ , the inverse temperature $\beta = \frac{1}{2\Sigma}$, the input force u , the noise amplitude D , the information rate Γ^2 and the information length $\mathcal{L}$. We also show the PDF time evolution of the simulation model computed via the Laplace approximation (q) or via stochastic simulations ($\tilde{p}$) and the corresponding KL -divergence (12) between them. In the subplot of μ and β , the legend LA and SS stand for the Laplace assumption and stochastic simulations, respectively. Interestingly, we can see from this that the Laplace approximation works fine when used as a prediction model in the MPC method. The controls have a chattering effect (oscillations having a finite frequency and amplitude) similar to the one encountered when implementing other control methods, such as the sliding mode control [28], when trying to keep the system in the desired state $\mathbf{Y}_d$.

Figure 5b demonstrates the effects of controls (22) on the closed-loop system stochastic thermodynamics. The results show that at the desired state $\mathbf{Y}_d$ the value of $\dot{S}$ oscillates around zero with a small amplitude. This means $\Phi = -\Pi$ holds at some instants of time when $\mathbf{Y}$ reaches $\mathbf{Y}_d$. In other words, all the energy is exchanged with the system's environment when the control keeps $\mathbf{Y}$ on the nonequilibrium state $\mathbf{Y}_d$.



(a) PDF, D , u , μ , $\beta = \frac{1}{2\Sigma}$, Γ and $\mathcal{L}$ time evolution.



(b) Stochastic thermodynamics time evolution.

Figure 5. IL-QR under LA applied to system (8) with $\mathbf{Y}(0) = [2 + 5/6, 1/(2 \times 30)]^\top$ and $\mathbf{Y}_d(t) = [2 + 1/30, 1/(2 \times 3)]^\top$. The control is applied in $u(t)$ and $D(t)$. Moreover, $\gamma = 0.1$, $T_s = 1 \times 10^{-3}$, $N = 5$, $I_L = 1 \times 10^3$, $\mathbf{R} = 1 \times 10^{-5} \mathbf{I}_2$, $Q_{12} = Q_{21} = 0$, $Q_{11} = 1 \times 10^2$ and $Q_{22} = 8 \times 10^2$.

6. Conclusions

In this work, we developed a new control MPC method to derive the evolution of a system with a minimum information variability in systems governed by nonlinear stochastic differential equations. Specifically, we identified the limitations of the Laplace assumption and utilised it to reduce the computational cost of calculating the time-varying PDFs and to develop a prediction model in the MPC algorithm. We also derived the relations that permit us to analyse the controller effects on the close-loop system's thermodynamics.

In future work, we aim to apply our results for maximising the free-energy (minimum entropy production) [12] and for the analysis of the closed-loop stochastic thermodynamics in higher-order systems.

Author Contributions: Conceptualization, A.-J.G.-C. and E.-j.K.; methodology, A.-J.G.-C.; software, A.-J.G.-C.; validation, A.-J.G.-C. and E.-j.K.; formal analysis, A.-J.G.-C.; investigation, A.-J.G.-C.; resources, A.-J.G.-C. and E.-j.K.; data curation, A.-J.G.-C.; writing—original draft preparation, A.-J.G.-C.; writing—review and editing, A.-J.G.-C. and E.-j.K.; visualization, A.-J.G.-C.; supervision, E.-j.K.; project administration, E.-j.K.; funding acquisition, E.-j.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Appendix A. Proof of Proposition 1

To prove Proposition 1, we start by defining the first two moments of the ensemble density $p(\mathbf{x})$. This is given as follows

$$\dot{\mu}_i = \int_{\mathbb{R}^n} x_i \dot{p}(\mathbf{x}; t) d^n x, \quad (\text{A1})$$

$$\dot{\Sigma}_{ij} = \int_{\mathbb{R}^n} \bar{x}_i \bar{x}_j \dot{p}(\mathbf{x}; t) d^n x. \quad (\text{A2})$$

Here, $\bar{x}_i = x_i - \mu_i$. Using (A1)–(A2) and (3), while avoiding the arguments for simplicity, we have

$$\begin{aligned} \dot{\mu}_i &= \int_{\mathbb{R}^n} x_i \left[- \sum_{i=1}^n \partial_{x_i} (f_i p) + \sum_{i,j=1}^n (\partial_{x_i} D_{ij} \partial_{x_j}) p \right] d^n x, \\ &= - \int_{\mathbb{R}^n} x_i \partial_{x_i} (f_i p) d^n x + \int_{\mathbb{R}^n} x_i \partial_{x_i} \left(\sum_{j=1}^n D_{ij} \partial_{x_j} p \right) d^n x, \\ &= \int_{\mathbb{R}^n} f_i p d^n x = \langle f_i \rangle. \end{aligned} \quad (\text{A3})$$

$$\begin{aligned} \dot{\Sigma}_{ij} &= \int_{\mathbb{R}^n} \bar{x}_i \bar{x}_j \left[- \sum_{i=1}^n \partial_{x_i} (f_i p) + \sum_{i,j=1}^n (\partial_{x_i} D_{ij} \partial_{x_j}) p \right] d^n x, \\ &= - \int_{\mathbb{R}^n} \bar{x}_i \bar{x}_j \partial_{x_i} (f_i p) d^n x - \int_{\mathbb{R}^n} \bar{x}_i \bar{x}_j \partial_{x_j} (f_j p) d^n x + \int_{\mathbb{R}^n} \bar{x}_i \left(\sum_{j=1}^n D_{ij} \partial_{x_j} p \right) d^n x + \int_{\mathbb{R}^n} \bar{x}_j \left(\sum_{i=1}^n D_{ji} \partial_{x_i} p \right) d^n x, \\ &= \langle \bar{x}_j f_i + \bar{x}_i f_j \rangle + D_{ij} + D_{ji}. \end{aligned} \quad (\text{A4})$$

A closed-form solution to (A4)–(A5) can be obtained by exploiting the Laplace assumption; i.e., we recover the sufficient statistics (A1)–(A2) of system (1) through the first three terms of the nonlinear flow $f_i(\mathbf{x}, u)$ Taylor expansion around the expected state $\boldsymbol{\mu}$. This is given as follows

$$f_i(\mathbf{x}, u) = f_i(\boldsymbol{\mu}, u) + \sum_{j=1}^n \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_j} \bar{x}_j + \frac{1}{2} \sum_{j,k=1}^n \frac{\partial^2 f_i(\boldsymbol{\mu}, u)}{\partial x_j \partial x_k} \bar{x}_j \bar{x}_k + \dots \quad (\text{A5})$$

Under Gaussian assumptions $\langle \bar{x}_i \rangle = 0$ and $\langle \bar{x}_i \bar{x}_j \rangle = \Sigma_{ij}$ and applying (A5) to (A4)–(A5), we have

$$\dot{\boldsymbol{\mu}} = \left\langle f_i(\boldsymbol{\mu}, u) + \sum_{j=1}^n \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_j} \bar{x}_j + \frac{1}{2} \sum_{j,k=1}^n \frac{\partial^2 f_i(\boldsymbol{\mu}, u)}{\partial x_j \partial x_k} \bar{x}_j \bar{x}_k \right\rangle,$$

$$= f_i(\boldsymbol{\mu}, u) + \frac{1}{2} \sum_{j,k=1}^n \frac{\partial^2 f_i(\boldsymbol{\mu}, u)}{\partial x_j \partial x_k} \Sigma_{jk}. \quad (\text{A6})$$

$$\begin{aligned} \dot{\Sigma}_{ij} &= \left\langle (\bar{x}_j + \bar{x}_i) \left(f_i(\boldsymbol{\mu}, u) + \sum_{k=1}^n \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_k} \bar{x}_k \right) \right\rangle + D_{ij} + D_{ji}, \\ &= \sum_{k=1}^n \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_k} \Sigma_{jk} + \sum_{k=1}^n \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_k} \Sigma_{ik} + D_{ij} + D_{ji}. \end{aligned} \quad (\text{A7})$$

Equations (A6)–(A7) are the expansion of the equations shown in Proposition 1. This finishes the proof.

Appendix B. Proof of Proposition 2

By substituting (3) in (13), we obtain

$$\frac{d}{dt} S(t) = \int_{\mathbb{R}^n} \left(\sum_i \frac{\partial}{\partial x_i} J_i(\mathbf{x}; t) \right) \ln(p(\mathbf{x}; t)) d^n x = - \int_{\mathbb{R}^n} \sum_i J_i(\mathbf{x}; t) \left(\frac{\partial}{\partial x_i} \ln(p(\mathbf{x}; t)) \right) d^n x. \quad (\text{A8})$$

Now, after substituting the i – th term of J in (A8), we have

$$\frac{d}{dt} S(t) = - \int_{\mathbb{R}^n} \sum_i J_i(\mathbf{x}; t) \left(\frac{f_i(\mathbf{x}; t)}{D_{ii}} - \frac{J_i(\mathbf{x}; t)}{D_{ii} p(\mathbf{x}; t)} - \frac{\sum_{j \neq i} D_{ij} \frac{\partial}{\partial x_j} p(\mathbf{x}; t)}{D_{ii} p(\mathbf{x}; t)} \right) d^n x. \quad (\text{A9})$$

From (A9), the entropy production rate of the system corresponds to the positive definite part

$$\Pi = \sum_i \Pi_i = \int_{\mathbb{R}^n} \sum_i \frac{J_i(\mathbf{x}; t)^2}{D_{ii} p(\mathbf{x}; t)} d^n x, \quad (\text{A10})$$

while the entropy flux (entropy from the system to the environment) is

$$\Phi = \int_{\mathbb{R}^n} \sum_i \left(\frac{J_i(\mathbf{x}; t) f_i(\mathbf{x}; t)}{D_{ii}} - \frac{\sum_{j \neq i} D_{ij} J_i(\mathbf{x}; t) \frac{\partial}{\partial x_j} p(\mathbf{x}; t)}{D_{ii} p(\mathbf{x}; t)} \right) d^n x. \quad (\text{A11})$$

In this paper, we focus on the case when $D_{ij} = 0$ if $i \neq j$ to simplify (A11) as

$$\Phi = \sum_i \Phi_i = \int_{\mathbb{R}^n} \sum_i \left(\frac{J_i(\mathbf{x}; t) f_i(\mathbf{x}; t)}{D_{ii}} \right) d^n x. \quad (\text{A12})$$

Notice that (A10)–(A11) require that $D_{ii} > 0$. If $D_{ii} = 0$, we have $\Pi_i = 0$ and

$$\Phi_i = \left\langle \frac{\partial f_i(\mathbf{x}, t)}{\partial x_i} \right\rangle. \quad (\text{A13})$$

We start by applying the definition of entropy (A10) production and entropy flux (A12), giving us

$$\Pi_i = \frac{1}{D_{ii}} \langle f_i(\mathbf{x}, t)^2 \rangle + D_{ii} \left\langle \left(\frac{\partial Q(\mathbf{x})}{\partial x_i} \right)^2 \right\rangle + 2 \left\langle \frac{\partial f_i(\mathbf{x}, t)}{\partial x_i} \right\rangle, \quad (\text{A14})$$

$$\Phi_i = \frac{1}{D_{ii}} \langle f_i(\mathbf{x}, t)^2 \rangle + \left\langle \frac{\partial f_i(\mathbf{x}, t)}{\partial x_i} \right\rangle. \quad (\text{A15})$$

Before continuing, it is useful to note that [29]

$$\frac{\partial Q}{\partial x_k} = -\frac{1}{2} \left[\sum_i \bar{x}_i \Sigma_{ki}^{-1} + \sum_j \bar{x}_j \Sigma_{jk}^{-1} \right] = -\sum_i \bar{x}_i \Sigma_{ki}^{-1} = -\bar{\mathbf{x}}^\top \boldsymbol{\Sigma}_k^{-1} \quad (\text{A16})$$

where $\bar{x}_i = x_i - \mu_i$, $\bar{\mathbf{x}} := \mathbf{x} - \boldsymbol{\mu} = [\bar{x}_1, \dots, \bar{x}_n]^\top$ and $\boldsymbol{\Sigma}_k^{-1}$ is the k -th column of the inverse matrix $\boldsymbol{\Sigma}^{-1}$ of $\boldsymbol{\Sigma}$. Therefore, [29]

$$\left\langle D_{ii} \left(\frac{\partial Q(\mathbf{x})}{\partial x_i} \right)^2 \right\rangle = D_{ii} \langle \bar{\mathbf{x}}^\top \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\Sigma}_i^{-1})^\top \bar{\mathbf{x}} \rangle = D_{ii} \text{Tr}(\boldsymbol{\Delta}_i \boldsymbol{\Sigma}), \quad (\text{A17})$$

and

$$\begin{aligned} \frac{\langle f_i(\mathbf{x})^2 \rangle}{D_{ii}} &= \frac{1}{D_{ii}} \left\langle \left(f_i(\boldsymbol{\mu}, u) + \sum_{j=1}^n \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_j} \bar{x}_j \right) \left(f_i(\boldsymbol{\mu}, u) + \sum_{k=1}^n \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_k} \bar{x}_k \right) \right\rangle \\ &= \frac{1}{D_{ii}} \left(f_i(\boldsymbol{\mu}, u)^2 + \sum_{j,k=1}^n \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_j} \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_k} \Sigma_{jk} \right) = \frac{1}{D_{ii}} (f_i(\boldsymbol{\mu}, u)^2 + \nabla^\top f_i(\boldsymbol{\mu}, u) \boldsymbol{\Sigma} \nabla f_i(\boldsymbol{\mu}, u)), \end{aligned} \quad (\text{A18})$$

$$\left\langle \frac{\partial f_i(\mathbf{x}, t)}{\partial x_i} \right\rangle = \left\langle \frac{\partial}{\partial x_i} \left(f_i(\boldsymbol{\mu}, u) + \sum_{j=1}^n \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_j} \bar{x}_j + \frac{1}{2} \sum_{j,k=1}^n \frac{\partial^2 f_i(\boldsymbol{\mu}, u)}{\partial x_j \partial x_k} \bar{x}_j \bar{x}_k \right) \right\rangle = \frac{\partial f_i(\boldsymbol{\mu}, u)}{\partial x_i}. \quad (\text{A19})$$

Hence,

$$\Pi = \sum_{i=1}^n \Pi_i = \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{D}) + \text{Tr}(\mathbf{D}^{-1} f(\boldsymbol{\mu}, u) f(\boldsymbol{\mu}, u)^\top) + \text{Tr}(\mathbf{J}_f \mathbf{D}^{-1} \mathbf{J}_f^\top \boldsymbol{\Sigma}) + 2\text{Tr}(\mathbf{J}_f) \quad (\text{A20})$$

$$\Phi = \sum_{i=1}^n \Phi_i = \text{Tr}(\mathbf{D}^{-1} f(\boldsymbol{\mu}, u) f(\boldsymbol{\mu}, u)^\top) + \text{Tr}(\mathbf{J}_f \mathbf{D}^{-1} \mathbf{J}_f^\top \boldsymbol{\Sigma}) + \text{Tr}(\mathbf{J}_f) \quad (\text{A21})$$

$$\dot{\mathbf{S}} = \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{D}) + \text{Tr}(\mathbf{J}_f) \quad (\text{A22})$$

References

- Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.* **2012**, *75*, 126001. [CrossRef] [PubMed]
- Gontis, V.; Kononovicius, A. Consentaneous agent-based and stochastic model of the financial markets. *PLoS ONE* **2014**, *9*, e102201. [CrossRef] [PubMed]
- de Freitas, R.A.; Vogel, E.P.; Korzenowski, A.L.; Rocha, L.A.O. Stochastic model to aid decision making on investments in renewable energy generation: Portfolio diffusion and investor risk aversion. *Renew. Energy* **2020**, *162*, 1161–1176. [CrossRef]
- Marreiros, A.C.; Kiebel, S.J.; Daunizeau, J.; Harrison, L.M.; Friston, K.J. Population dynamics under the Laplace assumption. *Neuroimage* **2009**, *44*, 701–714. [CrossRef] [PubMed]
- Maybeck, P.S. *Stochastic Models, Estimation, and Control*; Academic Press: Cambridge, MA, USA, 1982.
- Peliti, L.; Pigolotti, S. *Stochastic Thermodynamics: An Introduction*; Princeton University Press: Princeton, NJ, USA, 2021.
- Baltieri, M.; Buckley, C.L. PID control as a process of active inference with linear generative models. *Entropy* **2019**, *21*, 257. [CrossRef] [PubMed]
- Kim, E.; Hollerbach, R. Geometric structure and information change in phase transitions. *Phys. Rev. E* **2017**, *95*, 062107. [CrossRef]
- Nielsen, F. An elementary introduction to information geometry. *Entropy* **2020**, *22*, 1100. [CrossRef]
- Amari, S.I. *Information Geometry and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 194.
- Kim, E. Information geometry and nonequilibrium thermodynamic relations in the over-damped stochastic processes. *J. Stat. Mech. Theory Exp.* **2021**, *2021*, 093406. [CrossRef]
- Kim, E. Information Geometry, Fluctuations, Non-Equilibrium Thermodynamics, and Geodesics in Complex Systems. *Entropy* **2021**, *23*, 1393. [CrossRef]
- Hollerbach, R.; Kim, E.; Schmitz, L. Time-dependent probability density functions and information diagnostics in forward and backward processes in a stochastic prey–predator model of fusion plasmas. *Phys. Plasmas* **2020**, *27*, 102301. [CrossRef]
- Kim, E.; Heseltine, J.; Liu, H. Information length as a useful index to understand variability in the global circulation. *Mathematics* **2020**, *8*, 299. [CrossRef]
- Guel-Cortez, A.J.; Kim, E. Information Geometric Theory in the Prediction of Abrupt Changes in System Dynamics. *Entropy* **2021**, *23*, 694. [CrossRef] [PubMed]
- Guel-Cortez, A.J.; Kim, E. Information length analysis of linear autonomous stochastic processes. *Entropy* **2020**, *22*, 1265. [CrossRef] [PubMed]
- Saridis, G.N. *Entropy in Control Engineering*; World Scientific: Singapore, 2001; Volume 12.
- Fan, J.; Marron, J.S. Fast implementations of nonparametric curve estimators. *J. Comput. Graph. Stat.* **1994**, *3*, 35–56.

19. Stochastic Simulation Versus Laplace Assumption in a Cubic System. Available online: <https://github.com/AdrianGuel/StochasticProcesses/blob/main/CubicvsLA.ipynb> (accessed on 6 June 2022).
20. Tomé, T. Entropy production in nonequilibrium systems described by a Fokker-Planck equation. *Braz. J. Phys.* **2006**, *36*, 1285–1289. [CrossRef]
21. Soto, F.; Wang, J.; Ahmed, R.; Demirci, U. Medical micro/nanorobots in precision medicine. *Adv. Sci.* **2020**, *7*, 2002203. [CrossRef] [PubMed]
22. Lee, J.H. Model predictive control: Review of the three decades of development. *Int. J. Control. Autom. Syst.* **2011**, *9*, 415–424. [CrossRef]
23. Mehrez, M.W.; Worthmann, K.; Cenerini, J.P.; Osman, M.; Melek, W.W.; Jeon, S. Model predictive control without terminal constraints or costs for holonomic mobile robots. *Robot. Auton. Syst.* **2020**, *127*, 103468. [CrossRef]
24. Kristiansen, B.A.; Gravdahl, J.T.; Johansen, T.A. Energy optimal attitude control for a solar-powered spacecraft. *Eur. J. Control* **2021**, *62*, 192–197. [CrossRef]
25. Salesch, T.; Gesenhues, J.; Habigt, M.; Mechelinck, M.; Hein, M.; Abel, D. Model based optimization of a novel ventricular assist device. *at-Automatisierungstechnik* **2021**, *69*, 619–631. [CrossRef]
26. Andersson, J.A.E.; Gillis, J.; Horn, G.; Rawlings, J.B.; Diehl, M. CasADi—A software framework for nonlinear optimization and optimal control. *Math. Program. Comput.* **2019**, *11*, 1–36. [CrossRef]
27. Bemporad, A. Hybrid Toolbox—User’s Guide. 2004. Available online: <http://cse.lab.imtlucca.it/~bemporad/hybrid/toolbox> (accessed on 1 June 2022).
28. Utkin, V.; Lee, H. Chattering problem in sliding mode control systems. In Proceedings of the International Workshop on Variable Structure Systems, Alghero, Sardinia, 5–7 June 2006; pp. 346–350.
29. Petersen, K.B.; Pedersen, M.S. The matrix cookbook. *Tech. Univ. Den.* **2008**, *7*, 510.