Relations between entropy rate, entropy production and information geometry in linear stochastic systems

Guel-Cortez, A-J. & Kim, E.

Original citation & hyperlink:
Guel-Cortez, A-J & Kim, E 2023, 'Relations between entropy rate, entropy production and information geometry in linear stochastic systems', Journal of Statistical Mechanics: Theory and Experiment, vol. 2023, no. 3, 033204.
https://dx.doi.org/10.1088/1742-5468/acbc24

DOI 10.1088/1742-5468/acbc24
ISSN 1742-5468

Publisher: IOP Publishing

Relations between entropy rate, entropy production and information geometry in linear stochastic systems

Adrian-Josue Guel-Cortez[12],**Eun-jin Kim**[1]

[1]Centre for fluid and complex systems, Coventry University, Coventry, UK.

E-mail: [2]adrianjguelc@gmail.com

August 2017

Abstract. In this work, we investigate the relation between the concept of "information rate", an information geometric method for measuring the speed of the time evolution of the statistical states of a stochastic process, and stochastic thermodynamics quantities like entropy rate and entropy production. Then, we propose the application of entropy rate and entropy production to different practical applications such as abrupt event detection, correlation analysis, and control engineering. Specifically, by utilising the Fokker-Planck equation of multi-variable linear stochastic processes described by Langevin equations, we calculate the exact value for information rate, entropy rate, and entropy production and derive various inequalities among them. Inspired by classical correlation coefficients and control techniques, we create entropic-informed correlation coefficients as abrupt event detection methods and information geometric cost functions as optimal thermodynamic control policies, respectively. The methods are analysed via the numerical simulations of common prototypical systems.

Keywords: Information Geometry, Stochastic Thermodynamics, Linear Stochastic Systems Submitted to: *J. Stat. Mech.*

1. Introduction

Information geometry or the application of differential geometry to the information sciences has brought to light new tools for the analysis of dynamical systems [1, 2]. For instance, the information length (IL) [3, 4], given by the time integral of information rate, describes the total amount of statistical changes that a time-varying probability distribution takes through in time. IL appears as an important tool for the analysis of the non-equilibrium processes by providing a possible link between stochastic processes, complexity and geometry. In fact, IL has already been applied with success to different scenarios such as the quantification of hysteresis in forward-backward processes

[5, 6], correlation and self-regulation among different players [6], phase transitions [7], and prediction of sudden events [8]. Yet, the thermodynamic significance of IL and information rate seem to be less understood.

To find a connection between IL and thermodynamics, we can use stochastic thermodynamics (ST) [9, 10]. As it has already been popularised, ST makes use of stochastic calculus to draw a correspondence between micro/mesoscopic stochastic dynamics and macroscopic thermodynamics. In other words, we use ST to describe 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 is described by a Langevin/Fokker Planck equation. ST introduces time information uncertainty relations in thermodynamics [11], information theory to causality, modelling and control [12], optimal protocols [16], or fluctuation relations [14, 9]. In addition, it has inspired works in neuroscience [17], system dynamics [18] and control theory [19].

The main aim of this paper is to elucidate the link between thermodynamic quantities (e.g., entropy rate [20, 21]) and information rate (detailed in §IV) while exploring their application to abrupt event detection, correlation analysis, and control engineering. Even though some relations between information geometry and stochastic thermodynamics have been derived already (for instance, see [22, 23, 24]), this work focuses on generalising a relation initially proposed in [3, 4] for the Ornstein–Uhlenbeck (OU) process governed by a first-order linear stochastic differential equation. Note that, in comparison to [24] which determines inequalities between information geometry and thermodynamic observables. The relation reported in [3] connects information geometry to the "entropy balance equation" by explicitly using the values of entropy rate and entropy production. We generalise this relation to a set of linear Langevin equations using the corresponding Fokker-Planck equation [25] for multivariable linear processes. By utilising the exact time-varying probability density function (PDF), we calculate the values of the entropy rate, entropy production rate, and information rate. Then, we establish an inequality and an equality between information rate, entropy production and entropy rate for general linear stochastic systems and fully decoupled systems, respectively. Our results suggest that a weighted value of entropy production plus the square of the entropy rate is generally an upper bound of the information rate and that such a relation is fully determined by the structure of the system's harmonic potential.

Furthermore, after establishing an information-thermodynamic relation, we propose normalised correlation coefficients based on information rate and entropy production [27], and compare them in the analysis of abrupt events. Specifically, we perform the numerical simulation of a second-order stochastic process where the abrupt events are stimuli in the form of impulse-like functions applied to the system's mean value and covariance matrix. The results suggest that the information rate coefficient performs better when detecting/predicting abrupt events. Moreover, we conduct abrupt events analysis for systems with more than two random variables via the Euclidean norm of the information rate and entropy production as approximate correlation coefficients.

Finally, we explore the design of optimal "static" feedback control algorithms and the corresponding effects on entropy rate and entropy production for the minimisation of the number of statistical fluctuations via the solution of cost functions using IL and entropy production.

The rest of the paper is organised as follows. Section 2 describes the mathematical model and preliminary results to be used throughout the paper. Section 3 details the calculation of the entropy rate, entropy production and entropy flow in multivariable stochastic systems. Section 4 introduces the concept of information rate and IL and presents their relation to the thermodynamic quantities of entropy production and entropy rate. In Section 5, new normalised correlation coefficients in terms of entropy rate and entropy production are introduced. Section 6 demonstrates the application of the main results through the numerical simulation of different toy models. Section 7 states the conclusions of the paper and future work. Finally, from Appendix A to Appendix D, we present the detailed steps for the derivation of the main results in the paper.

2. Model

Consider the following set of Langevin equations

$$\frac{dx_i}{dt} = f_i(\mathbf{x};t) + \xi_i(t). \tag{1}$$

Here, $f_i : \mathbb{R}^n \to \mathbb{R}$ is a function that maps the variables of the vector $\mathbf{x} \in \mathbb{R}^n := [x_1, x_2, \ldots, x_n]^\top$ to a real value at a given time $t \in \mathbb{R}$; ξ_i is a short-correlated random noise satisfying $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_j(t') \rangle = 2D_{ij}\delta(t-t')$, where D_{ij} is the amplitude of the correlation function $\langle \xi_i(t)\xi_j(t') \rangle$. In this work, $f_i(\mathbf{x};t)$ is a linear function defined as

$$f_i(\mathbf{x};t) := \mathbf{A}_i\mathbf{x}(t) + \mathbf{B}_i\mathbf{u}(t) = \sum_{j=1}^{n} a_{ij}x_j(t) + \sum_{j}^{p} b_{ij}u_j(t), \tag{2}$$

where it becomes clear that $\mathbf{u}(t) \in \mathbb{R}^p$ is a vector of continuous time dependant functions such that $\mathbf{u}(t) := [u_1(t), u_2(t), \ldots, u_p(t)]^\top$; $\mathbf{A}_i$ and $\mathbf{B}_i$ are the i-th row vectors of the constant matrices $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$ used to couple the different system states $x_j(t)$ and deterministic forces $u_j(t)$, respectively. Hence, we consider that the set of particles described by (1) are driven by a harmonic potential ($\mathbf{A}_i\mathbf{x}(t)$), deterministic force ($\mathbf{B}_i\mathbf{u}(t)$ due to x-independent but possibly time-dependent $u(t)$), and a stochastic forcing term $\xi_i(t)$. Note that we have used a state space representation [28] of the term f_i as we are concerned with applications of stochastic thermodynamics/information geometry to dynamical control systems, causality and abrupt events analysis.

The Fokker-Planck equation corresponding to the Langevin equation (1) is given by

$$\frac{\partial p(\mathbf{x};t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}\left(f_i(\mathbf{x};t)p(\mathbf{x};t)\right) + \sum_i\sum_j D_{ij}\frac{\partial^2 p(\mathbf{x};t)}{\partial x_i \partial x_j}. \tag{3}$$

Equation (3) describes the time evolution of the PDF $p(\mathbf{x};t) : \mathbb{R}^n \to \mathbb{R}$ for $\mathbf{x} := \{x_1, x_2, \ldots, x_n\}$ in (1). Equation (3) can be written in terms of a probability current as follows

$$\frac{\partial p(\mathbf{x};t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i} J_i(\mathbf{x};t), \tag{4}$$

where J_i is the i-th component of the probability current J defined by

$$J_i(\mathbf{x};t) = f_i(\mathbf{x};t)p(\mathbf{x};t) - \sum_j D_{ij}\frac{\partial}{\partial x_j} p(\mathbf{x};t). \tag{5}$$

In general, the solutions of Equation (4) require numerical simulations. However, thanks to the linearity of $f_i(\mathbf{x};t)$ in (1), we have an analytical time-dependent solution of Equation (4) given by the following Proposition.

Proposition 1 *Given an initial multivariable Gaussian PDF $p(\mathbf{x};t_0)$, the solution of $p(\mathbf{x};t)$ in (3) is Gaussian and described by [29]*

$$p(\mathbf{x};t) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}} e^{Q(\mathbf{x};t)}, \tag{6}$$

where $Q(\mathbf{x};t) = -\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}(t)\right)^\top \boldsymbol{\Sigma}^{-1}(t)\left(\mathbf{x}-\boldsymbol{\mu}(t)\right)$; $\boldsymbol{\mu}(t) \in \mathbb{R}^n$ and $\boldsymbol{\Sigma}(t) \in \mathbb{R}^{n\times n}$ are the mean and covariance value of the random variable $\mathbf{x}$.

According to Proposition 1, the dynamics of the PDF at every instant of time are governed solely by the value of $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t)$. In the case of linear stochastic systems, the value of $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t)$ can be obtained by solving the following set of differential equations [30]

$$\dot{\mu}_i(t) = \sum_j^n a_{ij}\mu_j(t) + \sum_j^p b_{ij}u_j(t), \tag{7}$$

$$\dot{\Sigma}_{ij}(t) = \sum_k^n a_{ik}\Sigma_{kj}(t) + \sum_k^n a_{jk}\Sigma_{ki}(t) + 2D_{ij}(t). \tag{8}$$

Equations (7) and (8) can be rewritten in the following matricial form

$$\dot{\boldsymbol{\mu}}(t) = \mathbf{A}\boldsymbol{\mu}(t) + \mathbf{B}\mathbf{u}(t), \tag{9}$$

$$\dot{\boldsymbol{\Sigma}}(t) = \mathbf{A}\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\mathbf{A}^\top + 2\mathbf{D}(t). \tag{10}$$

In Equation (10), $\mathbf{D}$ is an square matrix whose elements are $D_{ij}(t)$.

3. Entropy rate

Given a time-varying multivariable PDF $p(\mathbf{x};t)$, its entropy rate is defined as [20]

$$\dot{S}(t) = \frac{d}{dt}S(t) = -\int_{\mathbb{R}^n} \dot{p}(\mathbf{x};t)\ln\left(p(\mathbf{x};t)\right) \mathrm{d}^n x. \tag{11}$$

By substituting (3) in (11), we obtain

$$\frac{d}{dt}S(t) = \int_{\mathbb{R}^n} \left(\sum_i \frac{\partial}{\partial x_i} J_i(\mathbf{x};t) \right) \ln\left(p(\mathbf{x};t)\right) \mathrm{d}^n x = -\int_{\mathbb{R}^n} \sum_i J_i(\mathbf{x};t) \left(\frac{\partial}{\partial x_i} \ln\left(p(\mathbf{x};t)\right) \right) \mathrm{d}^n x. \tag{12}$$

Now, after substituting (5) in (12), we have

$$\frac{d}{dt}S(t) = -\int_{\mathbb{R}^n} \sum_i J_i(\mathbf{x};t) \left(\frac{f_i(\mathbf{x};t)}{D_{ii}} - \frac{J_i(\mathbf{x};t)}{D_{ii}p(\mathbf{x};t)} - \frac{\sum_{j\neq i} D_{ij} \frac{\partial}{\partial x_j} p(\mathbf{x};t)}{D_{ii}p(\mathbf{x};t)} \right) \mathrm{d}^n x. \tag{13}$$

From (13), the entropy production rate of the system corresponds to the positive definite part

$$\Pi = \int_{\mathbb{R}^n} \sum_i \frac{J_i(\mathbf{x};t)^2}{D_{ii}p(\mathbf{x};t)} \mathrm{d}^n x, \tag{14}$$

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

$$\Phi = \int_{\mathbb{R}^n} \sum_i \left(\frac{J_i(\mathbf{x};t) f_i(\mathbf{x};t)}{D_{ii}} - \frac{\sum_{j\neq i} D_{ij} J_i(\mathbf{x};t) \frac{\partial}{\partial x_j} p(\mathbf{x};t)}{D_{ii}p(\mathbf{x};t)} \right) \mathrm{d}^n x. \tag{15}$$

In this paper, we focus on the case when $D_{ij} = 0$ if $i \neq j$ to simplify (15) as

$$\Phi = \int_{\mathbb{R}^n} \sum_i \left(\frac{J_i(\mathbf{x};t) f_i(\mathbf{x},t)}{D_{ii}} \right) \mathrm{d}^n x. \tag{16}$$

From Equations (14)-(16), we can define the entropy production Π_{J_i} and the entropy flow Φ_{J_i} from the J_i probability current as

$$\begin{aligned} \Pi_{J_i} &= \int_{\mathbb{R}^n} \frac{J_i(\mathbf{x};t)^2}{D_{ii}p(\mathbf{x};t)} \mathrm{d}^n x, \\ \Phi_{J_i} &= \int_{\mathbb{R}^n} \frac{J_i(\mathbf{x};t) f_i(\mathbf{x};t)}{D_{ii}} \mathrm{d}^n x. \end{aligned} \tag{17}$$

Finally, entropy rate can be expressed as follows

$$\frac{dS}{dt} = \Pi - \Phi, \tag{18}$$

where, from (17), clearly $\Pi = \sum_i \Pi_{J_i}$ and $\Phi = \sum_i \Phi_{J_i}$. Equation (18) is a well known expression for irreversible processes [31]. Notice that (14)-(15) require that $D_{ii} > 0$. If $D_{ii} = 0$, we have $\Pi_{J_i} = 0$ and

$$\Phi_{J_i} = \left\langle \frac{\partial f_i(\mathbf{x},t)}{\partial x_i} \right\rangle = a_{ii}. \tag{19}$$

3.1. Entropy rate for Gaussian dynamics

By using $p(\mathbf{x};t)$ given in equation (6), we derive the expressions for entropy production (14) and entropy flux (16) in terms of $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{A}$ and $\mathbf{D}$ as follows.

Relation 1 *The value of entropy production* Π *and entropy flow* Φ *in a Gaussian process whose mean* $\boldsymbol{\mu}$ *and covariance* $\boldsymbol{\Sigma}$ *are governed by equations* (9)-(10) *are given by*

$$\Pi = \dot{\boldsymbol{\mu}}^\top \mathbf{D}^{-1}\dot{\boldsymbol{\mu}} + \mathrm{Tr}\left(\mathbf{A}^\top \mathbf{D}^{-1}\mathbf{A}\boldsymbol{\Sigma}\right) + \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{D}\right) + 2\mathrm{Tr}(\mathbf{A}), \tag{20}$$

$$\Phi = \dot{\boldsymbol{\mu}}^\top \mathbf{D}^{-1}\dot{\boldsymbol{\mu}} + \mathrm{Tr}\left(\mathbf{A}^\top \mathbf{D}^{-1}\mathbf{A}\boldsymbol{\Sigma}\right) + \mathrm{Tr}(\mathbf{A}). \tag{21}$$

Proof 1 *See Appendix A.*

From Relation 1, we readily obtain the entropy rate (for further details, see (A.12) in Appendix A)

$$\frac{dS}{dt} = \frac{1}{2}\frac{d}{dt}\ln|\boldsymbol{\Sigma}| = \frac{1}{2}\,\mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}\right). \tag{22}$$

Clearly, equation (22) can also be obtained directly from the time derivative of the entropy S in a Gaussian Process

$$S = \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln|\boldsymbol{\Sigma}| + \frac{1}{2}n. \tag{23}$$

Again, if $D_{ii} = 0\ \forall i = 1, 2, \ldots, n$, we have $\Pi = 0$ and

$$\Phi = \mathrm{Tr}(\mathbf{A}). \tag{24}$$

Expressions similar to (20)-(22) but with an emphasis in distinguishing between variables that are even and odd under time reversal can be found in [31].

4. Relation between Entropy rate and information rate

As noted previously in Section 1, we aim to elucidate the link between entropy production Π, entropy rate $\dot{S}$, and information rate. We recall that for a time-varying multivariable PDF $p(\mathbf{x};t)$, we define its IL $\mathcal{L}$ as

$$\mathcal{L}(t) = \int_0^t \left(\sqrt{\int_{\mathbb{R}^n} p(\mathbf{x};\tau)\left[\partial_\tau \ln p(\mathbf{x};\tau)\right]^2 \mathrm{d}^n x}\right)\mathrm{d}\tau = \int_0^t \Gamma(\tau)\mathrm{d}\tau, \tag{25}$$

where $\Gamma(\tau) = \sqrt{\int_{\mathbb{R}^n} p(\mathbf{x};\tau)\left[\partial_\tau \ln p(\mathbf{x};\tau)\right]^2 \mathrm{d}^n x}$ is called the information rate and $\Gamma^2(\tau)$ the information energy. The value of $\Gamma^2(\tau)$ can also be understood as the Fisher information where the time is the control parameter [4]. 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 $p(\mathbf{x};t)$ [25].

When $p(\mathbf{x};t)$ is a Gaussian PDF, the information rate Γ of the joint PDF takes the compact form [3, 25, 26]

$$\Gamma^2 = \dot{\boldsymbol{\mu}}^\top \boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\mu}} + \frac{1}{2}\,\mathrm{Tr}\left((\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}})^2\right). \tag{26}$$

As it is mentioned in [3], if the PDF of Equation (1) is described by an univariate Gaussian PDF (i.e. the Ornstein–Uhlenbeck (OU) process) the information rate Γ is related to the entropy rate $\dot{S}$ and the entropy production Π via

$$\Gamma^2 = \frac{D}{\Sigma}\Pi + \dot{S}^2. \tag{27}$$

Equation (27) can easily be confirmed after some algebra with the following expressions

$$\begin{aligned} \Pi &= \frac{\dot{\mu}^2}{D} + \frac{a^2\Sigma}{D} + \frac{D}{\Sigma} + 2a = \frac{\dot{\mu}^2}{D} + \frac{\dot{\Sigma}^2}{4\Sigma D}, \\ \dot{S} &= \frac{1}{2}\frac{\dot{\Sigma}}{\Sigma}, \quad \Gamma^2 = \frac{\dot{\mu}^2}{\Sigma} + \frac{1}{2}\left(\frac{\dot{\Sigma}}{\Sigma}\right)^2, \end{aligned} \tag{28}$$

where μ, Σ, D and a are the scalar version of $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}$ and $\mathbf{A}$, respectively, in Proposition 1. Extending (27) to the case of a n-th order Gaussian process is one of the main contributions of this work. Such a result is given by the following relations.

Relation 2 *Given an n-th order Gaussian process whose mean and covariance are described by Equations* (9)-(10)*, a relationship between entropy production Π, entropy rate $\dot{S}$, and information rate Γ is given by*

$$0 \leq \Gamma^2 \leq \mathcal{E}_u := \mathrm{Tr}(\boldsymbol{\Sigma}^{-1})\Pi\,\mathrm{Tr}(\mathbf{D}) + \dot{S}^2 - 2g(\mathbf{s}), \tag{29}$$

where $\mathbf{s} = [\dot{S}_{J_1}, \dot{S}_{J_2}, \ldots, \dot{S}_{J_n}]^\top$, $g(\mathbf{s}) := \sum_{i<j}^{n} \dot{S}_{J_i}\dot{S}_{J_j}$ and $\dot{S}_{J_i}$ is the contribution to entropy rate by the current flow J_i, i.e.

$$\dot{S}_{J_i} = -\int_{\mathbb{R}^n} \frac{\partial}{\partial x_i} J_i(\mathbf{x};t)\ln\left(p(\mathbf{x};t)\right)\mathrm{d}^n x = \Pi_{J_i} - \Phi_{J_i}. \tag{30}$$

Proof 2 *See Appendix B.*

Relation 2 provides an inequality between information rate Γ (an information metric), entropy rate $\dot{S}$ and entropy production Π where the entropy rate $\dot{S}_{J_i}$ of each current flow J_i is explicitly taken into account. In the relation, Π is normalised by the product between the amplitude of the environmental fluctuations and the inverse of the system dynamics covariance due to its nature as an extensive quantity [3]. If $\boldsymbol{\Sigma}$ is constant, the normalised entropy production dominates the square of the information rate. On the other hand, when the system fluctuations change, the flow of information between the system and its environment is considered by the term $\dot{S}^2 - 2g(\mathbf{s})$ in (29). In addition, from Relation 2 we have

$$\mathcal{L}(t) \leq \mathcal{L}_u(t) := \int_0^t \sqrt{\mathcal{E}_u(\tau)}\mathrm{d}\tau. \tag{31}$$

Since minimising $\mathcal{L}_u$ will minimise $\mathcal{L}$, we can obtain both a minimum entropy production and a minimum statistical variability behaviour through $\mathcal{L}_u$.

For unstable systems, we can avoid the computation of the term $g(\mathbf{s})$ involving the contribution to entropy rate by each current flow J_i by using the following relation.

Relation 3 *Given the same conditions as in Relation 2, but considering that the eigenvalues $\varphi_i \in \mathbb{C}$ of the matrix $\mathbf{A}$ satisfy the following inequality*

$$\Re\{\varphi_i\} > 0 \quad \forall i = 1, 2, \ldots, n. \tag{32}$$

Then, the following result holds

$$0 \leq \Gamma^2 \leq \mathcal{I}_u := \frac{1}{n}\mathrm{Tr}(\boldsymbol{\Sigma}^{-1})\Pi_u \mathrm{Tr}(\mathbf{D}) + \dot{S}^2, \tag{33}$$

where $\Pi_u \geq \Pi$ (an upper bound of entropy production) defined by

$$\Pi_u := \mathrm{Tr}(\dot{\boldsymbol{\mu}}\dot{\boldsymbol{\mu}}^\top)\mathrm{Tr}(\mathbf{D}^{-1}) + \frac{1}{4}\mathrm{Tr}(\boldsymbol{\Sigma}^{-1})\mathrm{Tr}(\dot{\boldsymbol{\Sigma}})^2\mathrm{Tr}(\mathbf{D}^{-1}). \tag{34}$$

Proof 3 *See Appendix C.*

Now, we investigate the case when a relation between Γ, $\dot{S}$ and Π can be expressed in the form of equality. If and only if $\mathbf{A}$ in (9)-(10) is a diagonal matrix, i.e. we have a set of linearly independent stochastic differential equations (this can be after applying decoupling transformations [32]), the following result holds.

Relation 4 *Given a n-th order Gaussian process where all its random variables are independent, we have*

$$\Gamma^2 := \sum_i \frac{D_{ii}}{\Sigma_{ii}}\Pi_i + \sum_i \dot{S}_i^2, \tag{35}$$

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, respectively.

Proof 4 *See Appendix D .*

Again, the value of Π_i in (35) is given by Equation (17) while $\dot{S}_i$ is given by (30). From (35), it is inferred that a first order system is a special case of Relation 4. In addition, Equation (35) tells us that the geodesic (length-minimising curve between the initial and final PDF) of $\mathcal{L}(t)$ can be computed utilising the entropy rate and entropy production values (for further details on the geodesic problem, see [33]). More importantly, since Relation 4 permits us to equate the effects of IL geodesic dynamics on the system stochastic thermodynamics, Equation (35) can be used as part of a cost function employed to design controls that lead to both high energetic efficiency and minimum variability system closed-loop responses. (for further details, see [3, 4]).

5. On correlation coefficients

In this section, we define different types of correlations coefficients in terms of entropy production and information rate utilising mutual information and Pearson correlation coefficient, which will be used for abrupt event detection and causality analysis as discussed in Section 6.

5.1. Mutual information

The mutual information between two continuous random variables x_i and x_j with a joint PDF $p(\mathbf{x};t)$ at time t is defined as

$$I_{ij}(t) := \int_{\mathbb{R}^2} p(\mathbf{x};t) \ln\left(\frac{p(\mathbf{x};t)}{p(x_i,t)p(x_j,t)}\right) \mathrm{d}^2x = S_i(t) + S_j(t) - S(t). \tag{36}$$

Here, $p(x_i,t)$ and $p(x_j,t)$ are the marginal PDFs of the random variables x_i and x_j, respectively. Hence, the sub-index i in the entropy S refers to the entropy from the marginal PDF of x_i and its value is simply

$$S_i(t) = \frac{1}{2} + \ln\left(\sqrt{2\pi\Sigma_{ii}(t)}\right). \tag{37}$$

Mutual information represents the amount of information of a random variable that can be obtained by observing another random variable. Hence, it is a measure of the mutual dependence between the two variables [27]. To measure correlations between two random variables in a process, we can utilise common normalised variants of the mutual information, for instance, the total correlation formula [34, 35]

$$\rho_I(t) := 2\frac{I_{ij}(t)}{S_i(t) + S_j(t)}, \tag{38}$$

where x_i and x_j are treated symmetrically. Equation (38) is a weighted average of the asymmetrical uncertainty coefficients $\mathcal{C}_{ij}(t)$ and $\mathcal{C}_{ji}(t)$, defined as

$$\mathcal{C}_{ij}(t) := \frac{I_{ij}(t)}{S_i(t)}, \quad \mathcal{C}_{ji}(t) := \frac{I_{ij}(t)}{S_j(t)}, \tag{39}$$

weighted by the entropy of each variable separately [36]. The uncertainty coefficient (39) gives a value between 0 and 1, indicating no association or complete predictability of x_i from x_j (given x_j, what fraction of x_i we can predict), respectively. Thus, (38) gives an average of the predictability between x_i and x_j. The total correlation formula (38) is as an alternative to the well-known Pearson correlation coefficient

$$\rho := \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}, \tag{40}$$

when dealing with non-linear relationships between the random variables [37, 38, 39]. Pearson correlation coefficient ρ and mutual information I are related to each other through the following expression

$$I_{ij} = -\frac{1}{2}\log\left(1-\rho^2\right). \tag{41}$$

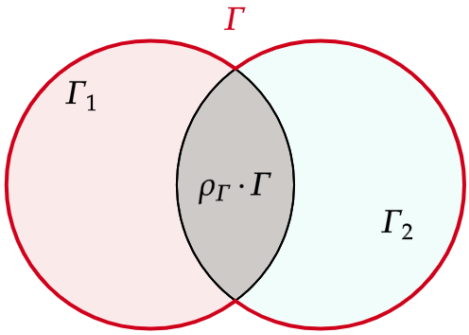


Figure 1: Venn diagram describing the meaning of (42). A similar diagram can be made for (43).

5.2. Information rate and entropy production correlation coefficients

In analogy to (38) to (40), we define new normalised correlation coefficients between two variables x_i and x_j in terms of information rate and entropy production as follows

$$\rho_\Gamma(t) := \frac{\Gamma_i(t) + \Gamma_j(t) - \Gamma(t)}{\Gamma(t)}, \tag{42}$$

$$\rho_\Pi(t) := \frac{\Pi_i(t) + \Pi_j(t) - \Pi(t)}{\Pi(t)}. \tag{43}$$

Here, Π_i and Π_j are the contributions from the variable x_i and x_j to the entropy production Π (see Equation (D.2)). The values of Γ_i and Γ_j are the information rates from the marginal PDFs of x_i and x_j, respectively. For instance, given the marginal PDF $p(x_i, t)$ of the random variable x_i the value of Γ_i is defined as follows

$$\Gamma_i := \sqrt{\int_{\mathbb{R}} p(x_i; t) \left[\partial_\tau \ln p(x_i, t)\right]^2 \mathrm{d}x_i}. \tag{44}$$

Equations (42)-(43) are not defined exactly as the Pearson correlation coefficient (40) or the normalised correlation coefficient of the mutual information (38). Instead, they are expressed analogously to the information quality ratio, a quantity of the amount of information of a variable based on another variable against total uncertainty [40]. Hence, ρ_Γ/ρ_Π is said to quantify the predictability of information rate/entropy production of a variable based on another variable. In other words, these coefficients give a weight of how much information one variable can add to the information rate or entropy production of another variable. Also, importantly, these quantities can work for any nonlinear relation. A graphical description of Equation (42) in the form of Venn diagram is shown in Figure 1.

6. Applications

So far, we have introduced a complete set of tools that establish a connection between information geometry and thermodynamics for linear systems. In this section, we apply

these to the different linear stochastic toy models and scenarios. Firstly, we start by exploring a simple second order stochastic model §6.1 representing a harmonically bound particle. Secondly, we will explore a simple decoupled third order stochastic model §6.1.3 commonly used to represent the simplified dynamics of an optical trap in the three-dimensional space. Thirdly, we will investigate higher order models to elucidate the effects of system's dimension on entropy production §6.2. Fourthly, we will provide the analysis of abrupt events/perturbations in the dynamics of a second and fourth order stochastic system §6.3. Finally, we will explore two optimisation problems for the design of thermodynamic control algorithms §6.4.

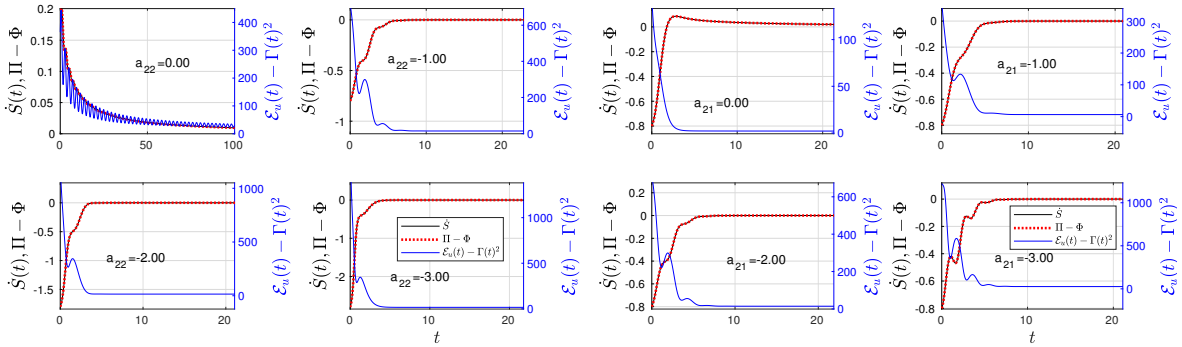
6.1. Second and third order process

6.1.1. Harmonically bound particle Consider the following Langevin equation describing a harmonically bound particle

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -\gamma \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix}, \tag{45}$$

where the parameters ω and γ are related to the system's natural frequency and damping, respectively.

First, to explore Relations 1 and 2, in Figure 2 we plot the changes on entropy rate $\dot{S}$ computed by using equation (22) and compared them with the value of $\Pi - \Phi$ obtained from equations (20)-(21), confirming the expected relation $\dot{S} = \Pi - \Phi$. Second, to briefly demonstrate that $\mathcal{E}_u > \Gamma^2$, we also show the difference between Γ^2 (using equation (26)) and $\mathcal{E}_u$ (from Relation 2). Our simulations were done for fixed value of ω by varying the value of γ (Figure 2a) and vice-versa (Figure 2b).



(a) $a_{11} = 0, a_{12} = 1, a_{21} = -2, a_{22} = \{0, -1, -2, -3\}$.

(b) $a_{11} = 0, a_{12} = 1, a_{21} = \{0, -1, -2, -3\}, a_{22} = -1$.

Figure 2: Computational experiment of a second order Langevin equation using $D_{11} = D_{22} = 0.01$, $x(0) = 1$, $y(0) = 1$, $\Sigma^0_{11} = \Sigma^0_{22} = 0.1$ and $\Sigma^0_{12} = \Sigma^0_{21} = 0$.

As can be concluded from Figure 2a, for an undamped harmonic oscillator with $\gamma = 0$, the value of $\mathcal{E}_u - \Gamma^2$ tends to decrease with time, meaning that they become

equal over time. Here, $\Gamma, \dot{S} > 0\ \forall t \geq 0$ because the system is permanently oscillating. Once we increase γ, the system goes to the equilibrium giving $\Gamma, \dot{S} \to 0$ and $\mathcal{E}_u \geq 0$ due to $\Pi = \Phi$. In general, for any $\mathbf{A}$ whose eigenvalues have negative real part and $u(t) = 0$, entropy production Π and entropy flow Φ in the long-time limit take the following values

$$\begin{aligned}\lim_{t\to\infty} \Pi(t) &= \mathrm{Tr}\left(\mathbf{A}^\top \mathbf{D}^{-1}\mathbf{A}\boldsymbol{\Sigma}(\infty) + \boldsymbol{\Sigma}^{-1}(\infty)\mathbf{D} + 2\mathbf{A}\right), \\ \lim_{t\to\infty} \Phi(t) &= \mathrm{Tr}\left(\mathbf{A}^\top \mathbf{D}^{-1}\mathbf{A}\boldsymbol{\Sigma}(\infty) + \mathbf{A}\right), \end{aligned} \tag{46}$$

where $\boldsymbol{\Sigma}(\infty) = 2\lim_{t\to\infty}\{\int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{D}e^{\mathbf{A}^\top(t-\tau)}\mathrm{d}\tau\}$. As we see, the transient behaviour and longtime limit of entropy production and entropy flow are fully determined by the value of $e^{\mathbf{A}t}$, which in turn (obviously) depends on the eigenvalues of the matrix $\mathbf{A}$. In §6.4, we discuss how such eigenvalues can be modified through a control algorithm (for example, using a full-state feedback control method [41]). In system (45), the bigger the value of γ the quicker we arrive to equilibrium. On the other hand, increasing the value of ω with $\gamma > 0$ increments the oscillations on the transitory response (see Figures 2a and 2b) [8, 25].

6.1.2. Correlation coefficients Using a bivariate PDF that satisfies (45), we will compare the different correlation coefficients defined in Section 5. To this end, Figure 3 shows the time evolution of $\rho_\Gamma, \rho_\Pi, \rho_I$ and ρ; the phase portraits of x_1 vs x_2, Γ_1 vs Γ_2 and Π_1 vs Π_2. The parameters used in this simulation are $\omega = 1, \gamma = 2$, $\mu_1(0) = 0.5, \mu_2(0) = 0.7$, $\boldsymbol{\Sigma}(0) = 0.01\mathbf{I}$ and $\mathbf{D} = 0.001\mathbf{I}$ where $\mathbf{I}$ is the identity matrix. To facilitate the analysis, the phase portrait shows time snapshots with the complete contour of the PDF where times are marked by vertical dashed lines.

The time evolution of the Pearson correlation coefficient ρ and the total correlation coefficient ρ_I may be the easiest coefficients to interpret since their values are related to the shape of the bivariate PDF. For instance, Figure 3a shows that ρ has negative sign because of the negative slope on the "linear" correlation between x_1 and x_2. Similarly, the total correlation coefficient ρ_I is negative with a maximum of -0.34 at some $0.68135 < t < 1.9323$. Recall that an advantage of the mutual information coefficient ρ_I over the Pearson coefficient ρ is that ρ_I does not assume whether the association of the random variables is linear or not while ρ does. This causes ρ to be zero even when the variables are still stochastically dependent (for further details, see [37, 38, 39]).

Understanding the meaning of ρ_Γ and ρ_Π is a bit more complicated as its behaviour cannot be easily related to the shape of the PDF through time. This is because even though the values of ρ_Γ and ρ_Π clearly take into account the correlation effects through $\boldsymbol{\Sigma}$, they are also sensitive to the mean values of x_1 and x_2. Hence, they should be interpreted by comparing the time evolution of the pair of quantities whose correlation is studied. For instance, Figure 3b shows that $\rho_\Gamma > 0$ in the time intervals where the time evolution of Γ_1 and Γ_2 have the same tendency or similar behaviour. ρ_Γ reaches its maximum value when Γ_1 and Γ_2 are almost similar in amplitude and rate of change. The different behaviour in the time evolution of Γ_1 and Γ_2 at $0.26 \lessapprox t \lessapprox 2$

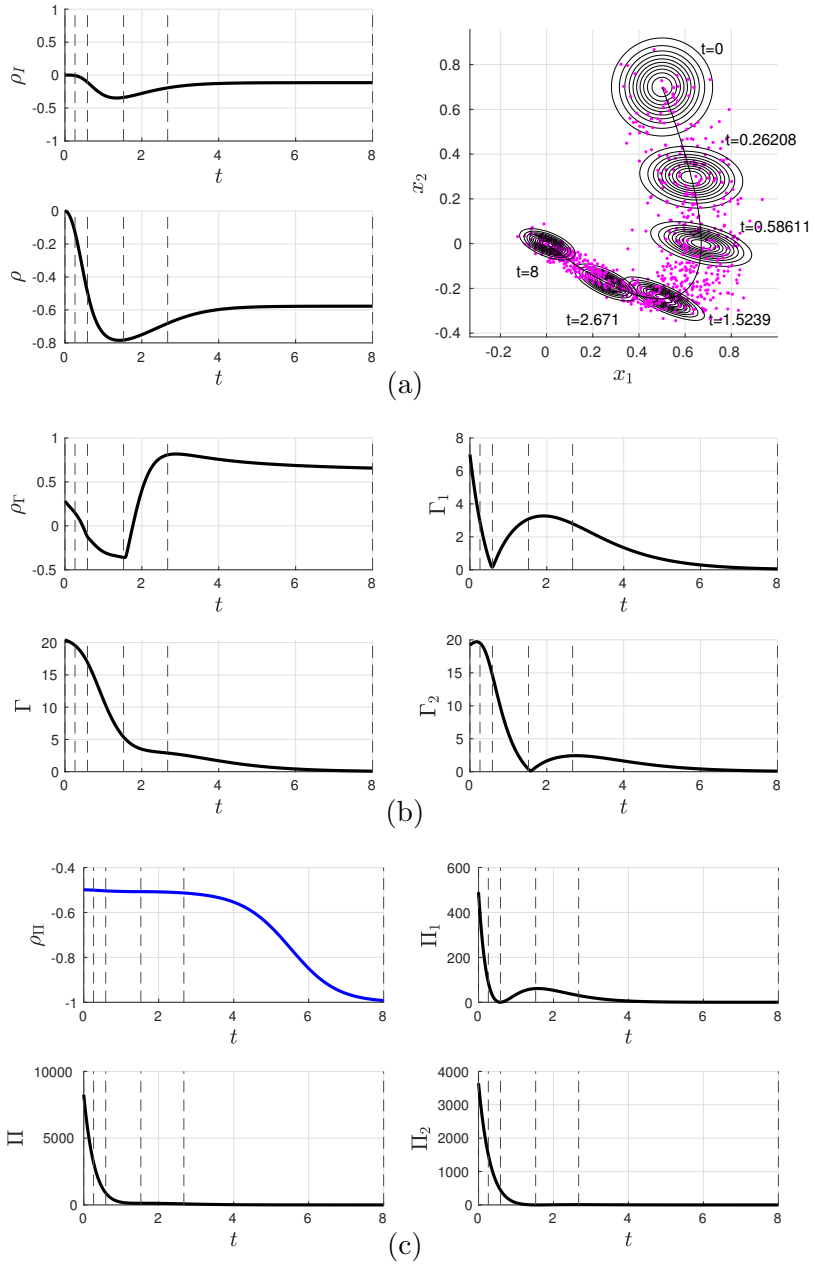


Figure 3: Time evolution of ρ_Γ, ρ_Π, ρ_I, ρ, Γ_1, Γ_2, Π_1 and Π_2 in comparison with the evolution of the snapshots of the phase portraits of x_1 vs x_2. The parameters used in this simulation are $\omega = 1, \gamma = 2$, $\mu_1(0) = 0.5, \mu_2(0) = 0.7$, $\mathbf{\Sigma}(0) = 0.01\mathbf{I}$ and $\mathbf{D} = 0.001\mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

depicts an anticorrelation which is corroborated by a value of $\rho_\Gamma < 0$. Regarding the thermodynamic effects, Figure 3 shows that Π_1 and Π_2 are always anticorrelated as $\rho_\Pi < 0$ at all t. The monotonically decrement of ρ_Π represents a continously increasing loss of predictability of Π_2 by Π_1.

6.1.3. Three-dimensional decoupled process We now consider fully decoupled linear stochastic systems (i.e. where $\mathbf{A}$ is a diagonal matrix). A practical example of a three-dimensional linear decoupled process corresponds to the simplified version of the mathematical description of an optical trap shown in Figure 4. The model consists of a set of three independent overdamped Langevin equations [42] given by equation (47). Here, x and y represent the position of the particle in the plane perpendicular to the beam propagation direction and z represents the position of the particle along the propagation direction. The stiffnesses of the trap in each of these directions are κ_x, κ_y and κ_z, respectively. γ is the particle friction coefficient. ξ_1, ξ_2 and ξ_3 are independent delta-correlated noises, i.e. $\langle \xi_i(t) \rangle = 0$, $\langle \xi_i(t)\xi_i(t') \rangle = 2D_{ii}\delta(t-t')$ and $\langle \xi_i(t)\xi_j(t') \rangle = 0 \quad \forall i \neq j$ with $i = 1, 2, 3$.

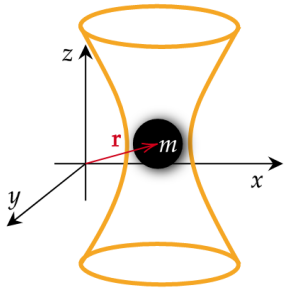


Figure 4: Particle of mass m in a three dimensional optical trap.

$$\begin{bmatrix} \dot{x}(t) \\ \dot{y}(t) \\ \dot{z}(t) \end{bmatrix} = \begin{bmatrix} -\frac{\kappa_x}{\gamma} & 0 & 0 \\ 0 & -\frac{\kappa_y}{\gamma} & 0 \\ 0 & 0 & -\frac{\kappa_z}{\gamma} \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} + \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \\ \xi_3(t) \end{bmatrix}. \tag{47}$$

Since (47) is a fully decoupled linear stochastic model, it permits us to show the applicability of Relation 4. To this end, in the left plot of Figure 5, we show $\dot{S}$, $\Pi - \Phi$, and Γ computed from equations (22), (20) minus (21), and (26)-(35), respectively. In Figure 5 we can see that both entropy rate $\dot{S} \to 0$ and information rate $\Gamma \to 0$ as the system goes to equilibrium. The right plot of Figure 5, a plot depicting the value of the information rate upper bound $\mathcal{E}_u$ (Relation 2) minus Γ^2 showing $\mathcal{E}_u \to 0$ at equilibrium. The exact value of $\mathcal{E}_u(t)$ in equilibrium is

$$\lim_{t\to\infty} \mathcal{E}_u = \mathrm{Tr}(\mathbf{\Sigma}^{-1}(\infty))\Pi(\infty)\,\mathrm{Tr}(\mathbf{D}) - 2g(\mathbf{H}(\infty)) = 0, \tag{48}$$

where

$$\mathbf{\Sigma}^{-1}(\infty) = -\mathbf{D}^{-1}\mathbf{A}, \quad \Pi(\infty) = 0. \quad \mathbf{H}(\infty) = 0. \tag{49}$$

Equation (49) applies only to systems with diagonal and stable $\mathbf{A}$ (proof at Appendix E). Since $\Pi \to 0$, any decoupled linear system is reversible at equilibrium.

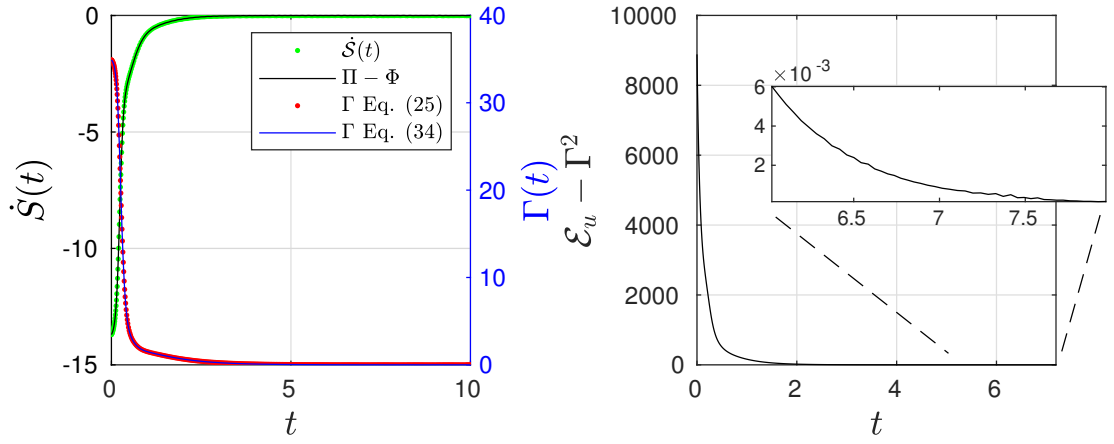


Figure 5: Computational experiment of a three-dimensional optical trap using $\kappa_x = 10, \kappa_y = 3$, $\kappa_z = 1, \gamma = 1$, $x(0) = 1$, $y(0) = 0.1$, $z(0) = 0.5$, $\mathbf{\Sigma} = 0.1\mathbf{I}$ and $\mathbf{D} = 0.01\mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

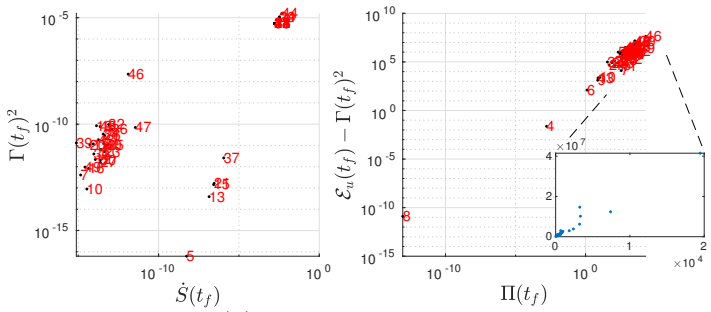
6.2. Higher order systems and the upper bounds of information rate Γ

As discussed in Section 4, it is possible to relate $\dot{S}$, Π and Γ using the inequalities from Relations 2 and 3. These inequalities become highly relevant when the order of the stochastic models increase, for instance, when using toy models in control engineering scenarios [19]. In this section, we take the case when $\mathbf{A}$ is a randomly chosen Hurwitz matrix whose size varies from $2 - 50$, i.e. we choose linear stochastic systems that contain from 2 to 50 random variables. Figure 6 shows the phase portrait of $\Gamma(t_f)^2$ vs $\dot{S}(t_f)$, and the phase portrait of $\mathcal{E}_u(t_f) - \Gamma(t_f)^2$ vs $\Pi(t_f)$. Figure 6 is also separated in sub-figures 6a and 6b showing the cases when matrix $\mathbf{A}$ is diagonal and non diagonal, respectively. Note, t_f refers to the time close enough the system's equilibrium; in our simulations $t_f = 300$. The phase portraits contain numbers to indicate the value at $t = t_f$; the number also indicates the order of the stochastic system.

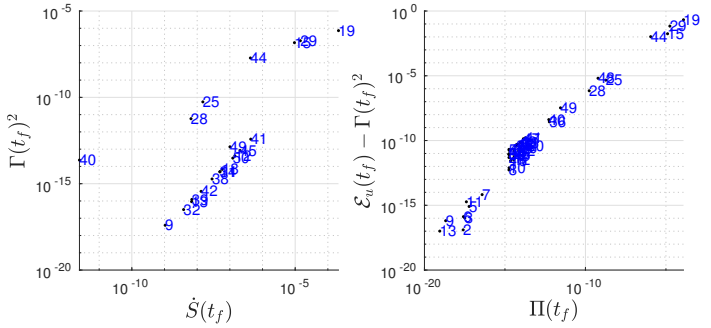
Regarding the portraits of $\Gamma(t_f)^2$ vs $\dot{S}(t_f)$, for every Hurwitz $\mathbf{A}$ (i.e. diagonal and non diagonal matrix with negative real part eigenvalues) $\lim_{t\to\infty} \mathcal{E}(t) = \lim_{t\to\infty} \dot{S}(t) = 0$ as expected. Meanwhile, for the same processes $\Pi > 0$ at equilibrium (see equation (46)). When looking at the phase portraits of $\mathcal{E}_u(t_f) - \Gamma(t_f)^2$ vs $\Pi(t_f)$, we see that $\mathcal{E}_u(t_f) > 0$ when $\mathbf{A}$ is non diagonal due to $\Pi(t_f) > 0$ for some t (see Figure 6a). On the other hand, as demonstrated in Equations (48) and (49) $\mathcal{E}_u(t_f) = \Pi(t_f) = 0$. Again, meaning that every fully decoupled linear system is reversible at equilibrium.

6.3. Abrupt events analysis

When the temperature changes abruptly in a system like (45), the value of D (noise amplitude) is affected, contributing to the uncertainty in the control of the particle's position. To bring light to the analysis and study of abrupt events, we use our toy models and simulate an abrupt change in the system's temperature by using the following



(a) **A** is non diagonal.



(b) **A** is diagonal.

Figure 6: Entropy rate $\dot{S}(t_f)$ vs the square of information rate $\Gamma(t_f)^2$ and the values of $\mathcal{E}_u(t_f) - \Gamma(t_f)^2$ vs $\Pi(t_f)$. The simulations use randomly chosen stable linear systems from order $n = 2$ to $n = 50$. The red numbers indicate the order of the system and its position the value at $t_f = 300$. In all simulations the system starts out of the equilibirum with initial conditions $\boldsymbol{\mu} = [1, 1, \ldots, 1]^\top$, $\boldsymbol{\Sigma} = 0.1\mathbf{I}$ and $\mathbf{D} = 0.01\mathbf{I}$ where $\mathbf{I}$ is the identity matrix varying from size 2 to 50.

impulse like function for the ii-element of the noise amplitude matrix $\mathbf{D}$ and on the input function $u(t)$

$$D_{ii}(t) = D_0 + \frac{1}{|a|\sqrt{\pi}} e^{-\left(\frac{t-t_p}{a}\right)^2}, \tag{50}$$

$$u(t) = \frac{1}{|a|\sqrt{\pi}} e^{-\left(\frac{t-t_p}{a}\right)^2}. \tag{51}$$

Here, the second term on RHS of (50) and (51) takes a non-zero value for a short time interval around t_p and a changes the amplitude of the impulses.

6.3.1. Harmonically bound particle Again, we start our analysis by considering system (45). Figures 7 and 8 show the computer simulation results and time evolution of abrupt events in $\mathbf{D}$ and u, respectively. The noise amplitude is perturbed via the element D_{22} of the matrix $\mathbf{D}$ and the input force u only affects the state x_2. Figures 7 and 8 are divided in three panels, 7a/8a which includes the time evolution of the correlation coefficients ρ and ρ_I (Figure 8a also includes the phase portrait of x_1 vs x_2); 7b/8b shows the time evolution of $\rho_\Gamma, \Gamma, \Gamma_1$ and Γ_2; 7c/8c the time evolution of ρ_Π, Π, Π_1 and Π_2. From Figure

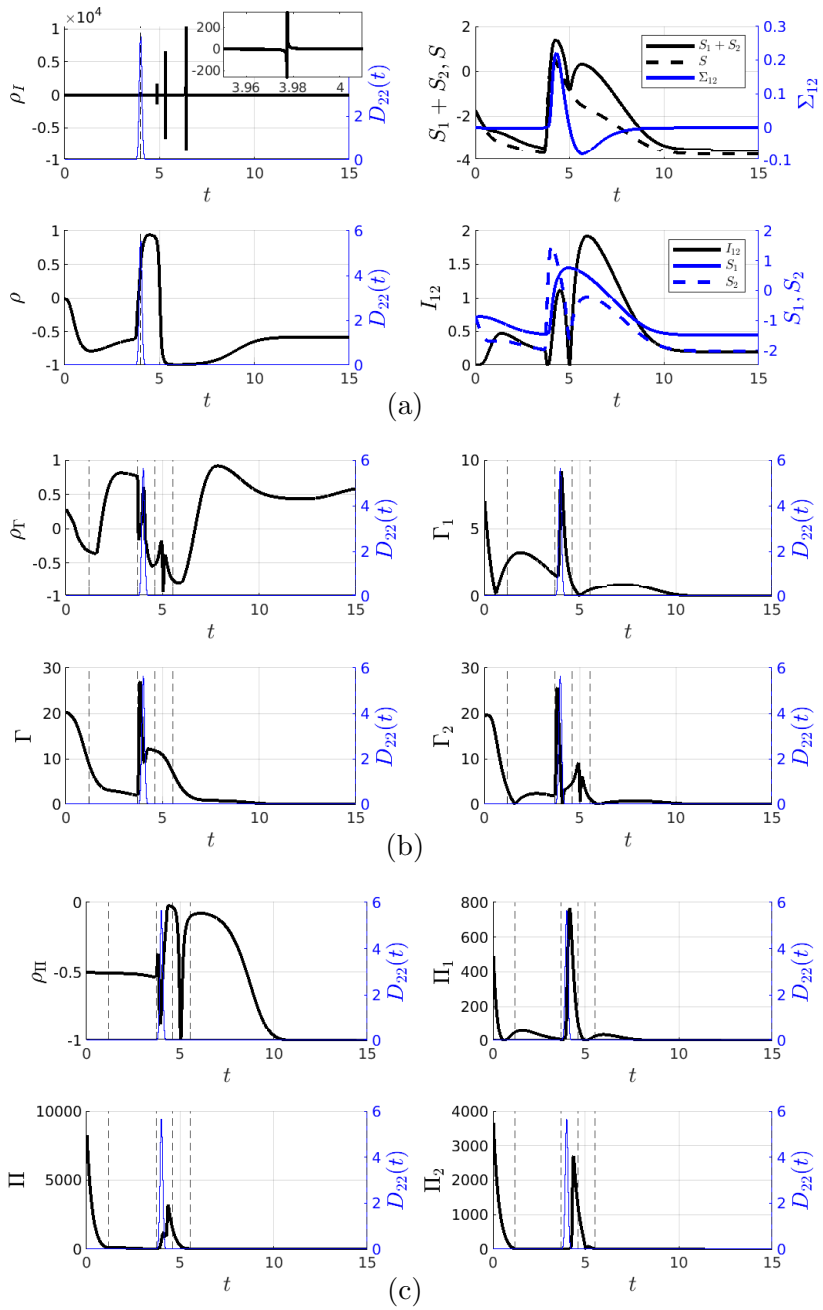


Figure 7: Abrupt event numerical experiment $u(t) = 0, D_{11}(t) = 0.001$ and $D_{22}(t) = 0.001 + \frac{1}{|0.1|\sqrt{\pi}} e^{-\left(\frac{t-4}{0.1}\right)^2}$. The rest of the simulation parameters are $\omega = 1, \gamma = 2$, $\mu_1(0) = 0.5, \mu_2(0) = 0.7$, and $\mathbf{\Sigma}(0) = 0.01\mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

7, the coefficient ρ_I is the most sensitive to noise amplitude perturbations, as it shows an abrupt change before the peak of the perturbation at $t = 4$ followed by a couple of extra abrupt changes at $t \geq 5$. To explain the sensitiveness of ρ_I, in 7a, we have included time evolution of $S_1 + S_2, S, \Sigma_{22}, I_{12}, S_1$ and S_2 instead of a phase portrait as in 8a. The divergences in ρ_I can be explained if we notice that the values of entropy (specially S_2) go from negative to positive values after the perturbation occurs. Let us recall that

differential entropy S can be negative and it is understood as a relative privation of information. When negative, its value means we have less disorder/uncertainty or more information/order. The increment in temperature given by the perturbation increases the level of uncertainty in the system and thus entropy, then when it decreases entropy decreases as well. Such changes lead to S, S_1+S_2 and I_{12} to be zero provoking divergences in ρ_I at different instants of time. Meanwhile, since it precedes the aforementioned perturbation (see Figure 7b), the value of Γ predicts the abrupt event (corroborating the previous results shown in [8]). Regarding the perturbation in $u(t)$ shown in Figure 8, the coefficients ρ and ρ_I are no longer useful because they are not sensitive to changes in the mean value of the PDF (see Figure 8a). In contrast, an abrupt event in the mean value is well captured by ρ_Γ and ρ_Π.

Figures 8b and 8c show that the values of ρ_Γ and ρ_Π change abruptly at the time $t = 4$ when perturbation occurs. Figure 8b presents negative ρ_Γ at $t \approx 4$ due to the large difference between Γ_2 and Γ_1. For the similar reasons, ρ_Π also presents a high decrement at $t \approx 4$. Here, the coefficients are able to detect the perturbation over the mean value but they are no longer able to predicted it.

6.3.2. Controllable canonical form The analysis of abrupt events in high order systems can be done by an offline method, such as the application of the Euclidean norm to each marginal or joint information rate/entropy production of the random variables in the system. Recall that the Euclidean norm of any time dependant function $\vartheta(t)$ is defined as follows

$$\|\vartheta(t)\| := \left(\int_0^{t_f} \vartheta(\tau)^2 d\tau\right)^{\frac{1}{2}}. \tag{52}$$

As a demonstration, using Equation (52), here we study the effects of abrupt events in the noise amplitude matrix $\mathbf{D}(t)$ and the force input $u(t)$ of the popular controllable canonical form of the state-space realization of a linear system given by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -d_n & -d_{n-1} & -d_{n-2} & \cdots & -d_1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} u(t) + \boldsymbol{\xi}(t). \tag{53}$$

Here, $\mathbf{x} := [x_1, x_2, \ldots, x_n]^\top \in \mathbb{R}^n$, $u \in \mathbb{R}$ and $\boldsymbol{\xi} := [\xi_1, \xi_2, \ldots, \xi_n]^\top \in \mathbb{R}^n$ is a vector of random variables with $\langle \xi_i(t) \rangle = 0$, and $\langle \xi_i(t)\xi_j(t') \rangle = 2D_{ij}\delta(t - t')$. Model (53) provides us with a structure where the input enters a chain of integrators making it to move every state in the Langevin equation (i.e. they are fully controllable). We consider the case when (53) is of 4th order. The values of the parameters are $\mathbf{d}_4 = [d_1, \ldots, d_n]^\top = [-1.5165, -5.2614, -6.7985, -4.2206]^\top$. In Figures 9 to 11, we use the notation $x_i \ \forall i = 1, 2, 3, 4$ and $\mathbf{x}$ to refer to the values of $\Pi, \dot{S}$ and Γ computed from marginal PDF $p(x_i, t)$ and from the joint PDF $p(\mathbf{x}; t)$, respectively.

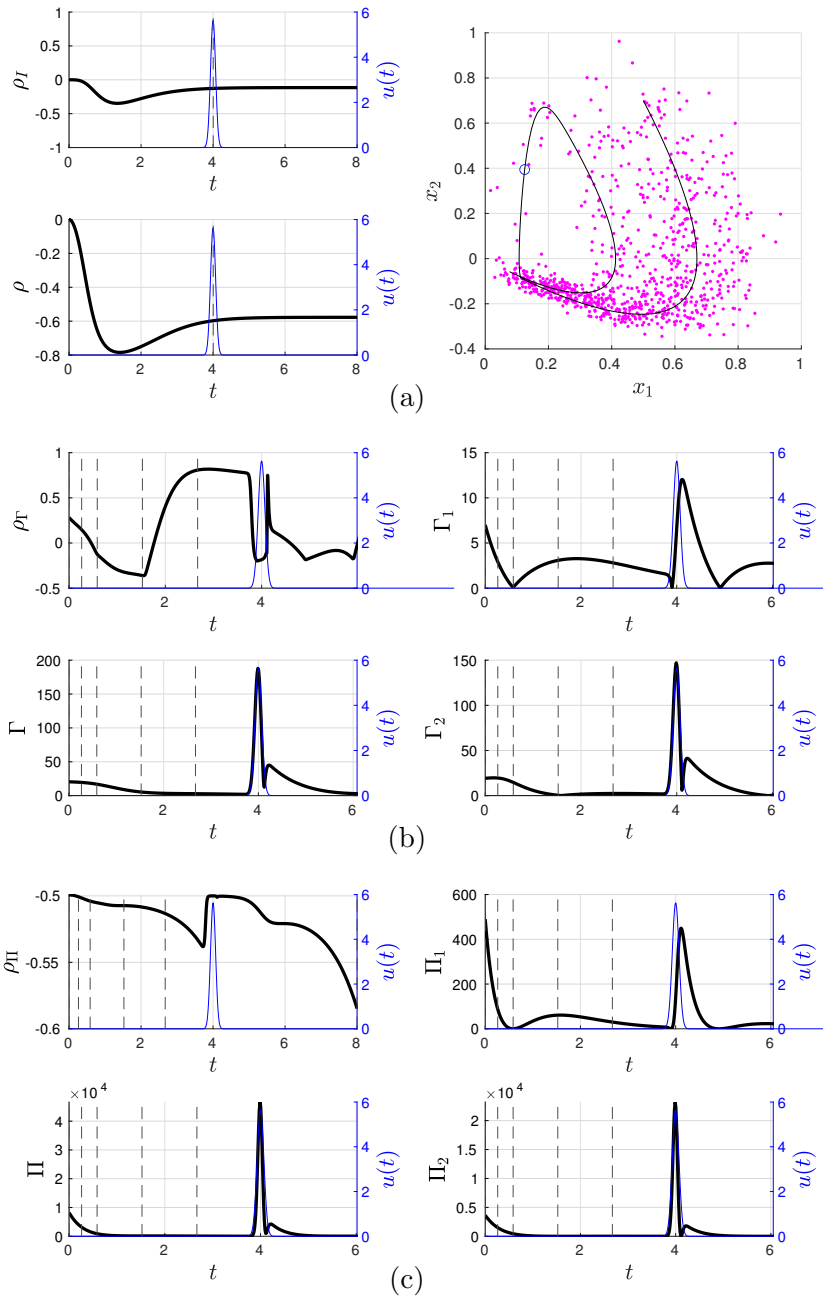
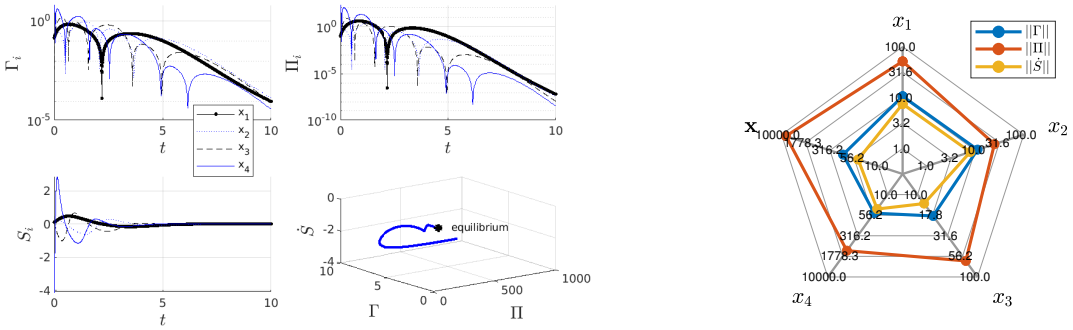


Figure 8: Abrupt event numerical experiment $D_{22}(t) = D_{11}(t) = 0.001$ and $u(t) = \frac{1}{|0.1|\sqrt{\pi}} e^{-\left(\frac{t-4}{0.1}\right)^2}$. The rest of the simulation parameters are $\omega = 1, \gamma = 2$, $\mu_1(0) = 0.5, \mu_2(0) = 0.7$, and $\mathbf{\Sigma}(0) = 0.01\mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

Figure 9a depicts the time evolution of Γ_i, Π_i and $\dot{S}_i$ $\forall i = 1, 2, 3, 4$. It also includes the time evolution of the three dimensional space $(\Gamma, \Pi, \dot{S})$. Figure 9b shows the norms of $\Gamma_i, \Pi_i, \dot{S}_i$ $\forall i = 1, 2, 3, 4$ and $\Gamma, \Pi, \dot{S}$ in the form of a spider plot. For instance, the value of the norm of the information rate Γ computed from the joint PDF $p(\mathbf{x}; t)$ and from the marginal PDF $p(x_2, t)$ is $||\Gamma|| \approx 10000$ and $||\Gamma_2|| \approx 31.6$, respectively. As we

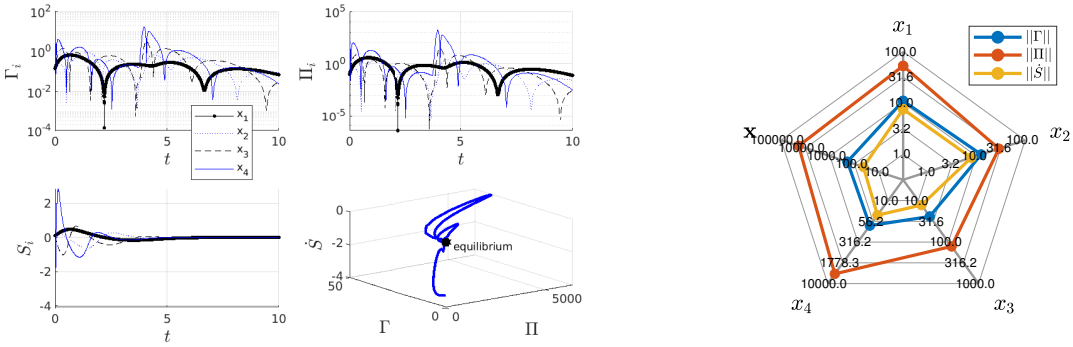
can see, the effects on Γ_i, Π_i and $\dot{S}_i$ $\forall i = 1, 2, 3, 4$ by the random variables is hierarchical with regards to their amplitude (for example $|\dot{S}_4| > |\dot{S}_3| > |\dot{S}_2| > |\dot{S}_1|$ at almost all the time) and the equilibrium of $(\Gamma, \Pi, \dot{S})$ is $(0, 0, 0)$. When we add a perturbation in u



(a) (b)

Figure 9: Behaviour with no abrupt event $D_{ii}(t) = D_0 = 0.01, u(t) = 0$. The rest of the simulation parameters are $\mu(0) = [0, 0, 0, 1]^\top$ and $\mathbf{\Sigma}(0) = 0.1\mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

which affects directly to x_4 (see Equation (53)), we obtain the results shown in Figure 10. Such an abrupt event causes an notable increment in the norms of the states (See Figure 10b) which still maintains the same hierarchical order in the states (x_4 is the most affected in comparison with x_1) due to the system's structure as expected. The direct effect of the abrupt event on each variable's time evolution is shown in Figure 10a. Recall that u is applied directly to x_4. Again, $\dot{S}$ is unperturbed since the event affects only the mean value of the PDF. On the other hand, if we separately include a



(a) (b)

Figure 10: Behaviour with abrupt event $D_{ii}(t) = D_0 = 0.01, u(t) = \frac{1}{|0.1|\sqrt{\pi}} e^{-\left(\frac{t-t_p}{0.1}\right)^2}$ with $t_p = 4$. The rest of the simulation parameters are $\mu(0) = [0, 0, 0, 1]^\top$ and $\mathbf{\Sigma}(0) = 0.1\mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

perturbation in each element D_{ii} of the noise matrix $\mathbf{D}$, similar results occur. Figure

11 illustrates the norms of $\Gamma, \Pi, \dot{S}$ in the form of bar plots (after applying perturbations to each D_{ii}). The plots indicate a domino effect in the marginal PDFs as follow. When only D_{11} is perturbed no clear effects can be seen in the rest of the variables but when D_{33} is perturbed x_3, x_2 and x_1 increase their values. Same happens after perturbing D_{44}, again this is due to the structure interaction of the system we are studying. This implies that norms provide an approximate value of the dependence between the variables of the random process.

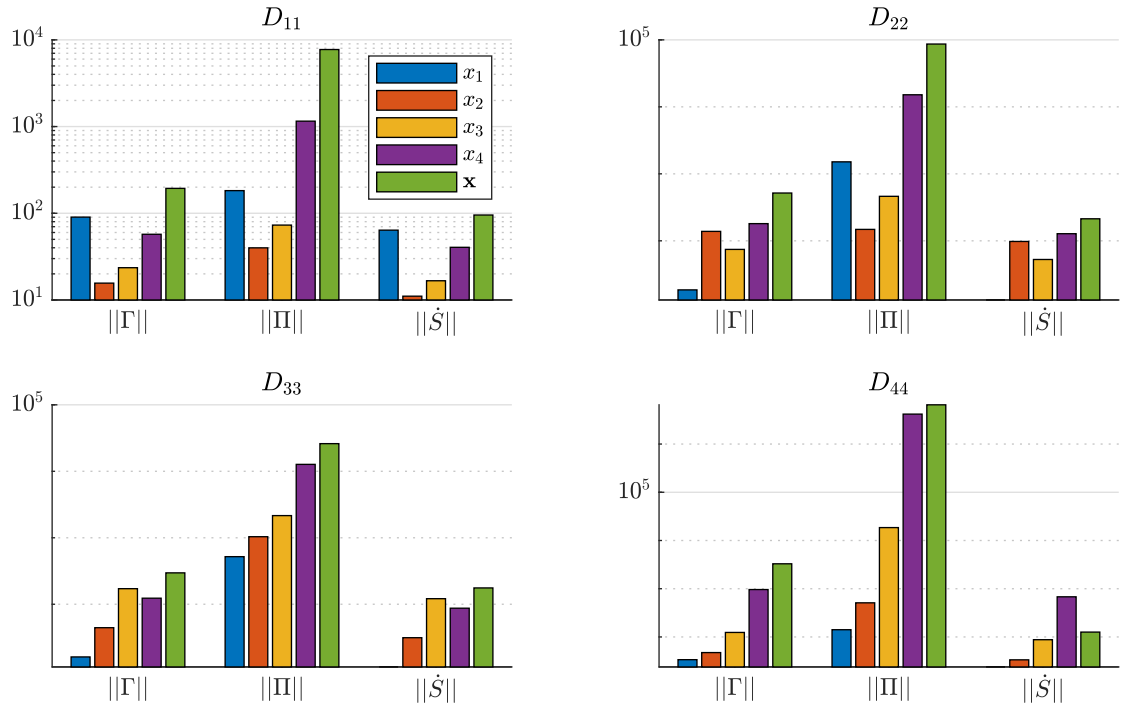


Figure 11: Abrupt event analysis using the norms of Π, $\dot{S}$, Γ at the marginal PDF $p(x_i, t)$ and the joint PDF $p(\mathbf{x}; t)$. Each plot depicts a perturbation on a given D_{ii} according to (50) at $t_p = 6$ with $a = 0.1$ and $D_0 = 0.01$ in system (53). The rest of the simulation parameters are $\mu(0) = [0, 0, 0, 1]^\top$ and $\mathbf{\Sigma}(0) = 0.1\mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

6.4. Minimum variability control design

The design of "efficient" machines remains still a common practical problem in different research areas. Note that here, efficiency can be related to the existence of minimum variability or entropy production as it was highlighted in previous works [5, 33, 43, 44]. The generation of efficient processes, as we describe here, can be accomplished by designing optimal protocols through work and heat minimisation [45], the Wasserstein distance [46] or inverse engineering (for a complete review, see [47]) to drive our system while being subject to thermodynamic constraints. In this context, as a final application, here we explore the use of IL to design a classical control technique in a given linear stochastic process. From the field of control engineering, we will take the full-state

feedback controller given by

$$u(t) = -\mathbf{k}\mathbf{x}(t), \tag{54}$$

where $\mathbf{k} \in \mathbb{R}^{1\times n}$. Through this control, we obtain the following closed-loop system

$$\dot{\mathbf{x}}(t) = \mathbf{A}_{cl}\mathbf{x}(t) + \boldsymbol{\xi}(t), \tag{55}$$

where $\mathbf{A}_{cl} = \mathbf{A} - \mathbf{B}\mathbf{k}$. The full-state feedback control permits us to manipulate the system's mean value via changing the eigenvalues of $\mathbf{A}$. As discussed previously, such eigenvalues also modify the time evolution of $\boldsymbol{\Sigma}$. In systems like (45), the value of $\boldsymbol{\Sigma}$ can as well be manipulated by the temperature of the environment whose value is related to the elements D_{11} and D_{22} of the noise amplitude matrix $\mathbf{D}$.

Taking the aforementioned details into consideration, we propose the following optimisation problems for the design of minimum variability controls

$$\begin{aligned}
\min_{\mathbf{k},\mathbf{D}} \quad & J_1 = \int_0^{t_f} \Gamma(\tau)\mathrm{d}\tau, \\
\text{s.t.} \quad & \dot{\boldsymbol{\mu}} = \mathbf{A}_{cl}\boldsymbol{\mu} \\
& \dot{\boldsymbol{\Sigma}} = \mathbf{A}_{cl}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}_{cl}^{\top} + 2\mathbf{D} \\
& \boldsymbol{\mu}(0) = \mathbf{m}, \boldsymbol{\Sigma}(0) = \mathbf{S} \\
& k_{l,i} \leq k_i \leq k_{u,i}, \quad 0 \leq D_{ii} \leq D_{\max} \\
& \forall i = 1, 2, \ldots, n,
\end{aligned} \tag{56}$$

and

$$\begin{aligned}
\min_{\mathbf{k},\mathbf{D}} \quad & J_2 = ||\Gamma(t)^2 - \Gamma(0)^2||, \\
\text{s.t.} \quad & \dot{\boldsymbol{\mu}} = \mathbf{A}_{cl}\boldsymbol{\mu} \\
& \dot{\boldsymbol{\Sigma}} = \mathbf{A}_{cl}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}_{cl}^{\top} + 2\mathbf{D} \\
& \boldsymbol{\mu}(0) = \mathbf{m}, \boldsymbol{\Sigma}(0) = \mathbf{S} \\
& k_{l,i} \leq k_i \leq k_{u,i}, \quad 0 < D_{ii} \leq D_{\max} \\
& \forall i = 1, 2, \ldots, n.
\end{aligned} \tag{57}$$

In Equation (56), J_1 is a cost function that considers the minimisation of IL from $t = 0$ to $t = t_f$ to obtain the "minimum" statistical changes in the given period of time. On the other hand, Equation (57) considers a cost function J_2 equal to the norm of $\Gamma(t)^2 - \Gamma(0)^2$. The objective of J_2 is to keep Γ^2 constant through time (with the least amount of fluctuations) to approximately follow the "geodesic", a problem well described in [33]. Both optimisation problems are subject to the dynamics of the mean and covariance of the PDF given certain initial conditions for them. The problems also consider upper and lower limits to k_i and D_{ii} $\forall i = 1, 2, \ldots, n$ given by $k_{l,i}, k_{u,i}$ and $0, D_{\max}$, respectively. Note $D_{ii} \geq 0$ because the temperature cannot be negative. The values of $k_{l,i}$ and $k_{u,i}$ are determined such that the following stability condition is satisfied

$$|s\mathbf{I} - \mathbf{A}_{cl}| \neq 0 \quad \forall s \in \mathbb{C} \quad \text{s.t.} \quad \Re s > 0. \tag{58}$$

Using $\omega = 1, \gamma = 2, \mu_1(0) = 0.5, \mu_2(0) = 0.7, \Sigma_{11} = \Sigma_{22} = 0.01, \Sigma_{12} = \Sigma_{21} = 0$, $D_{\max} = \infty$, $k_{l,1} = -1, k_{l,2} = -2, k_{u,1} = k_{u,2} = \infty$ and $u(t) = -[k_1 \quad k_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ in system (45), we have explored the solution of Equations (56) and (57) via the MATLAB Toolbox FMINCON [48]. The solutions give us the set values of $\mathbf{k}$ and $\mathbf{D}$ that give at least a local minimum. Note that our goal here is to see the implications of a solution to such problems instead of rigorously finding the global optimal solution. Figure 12 depicts the

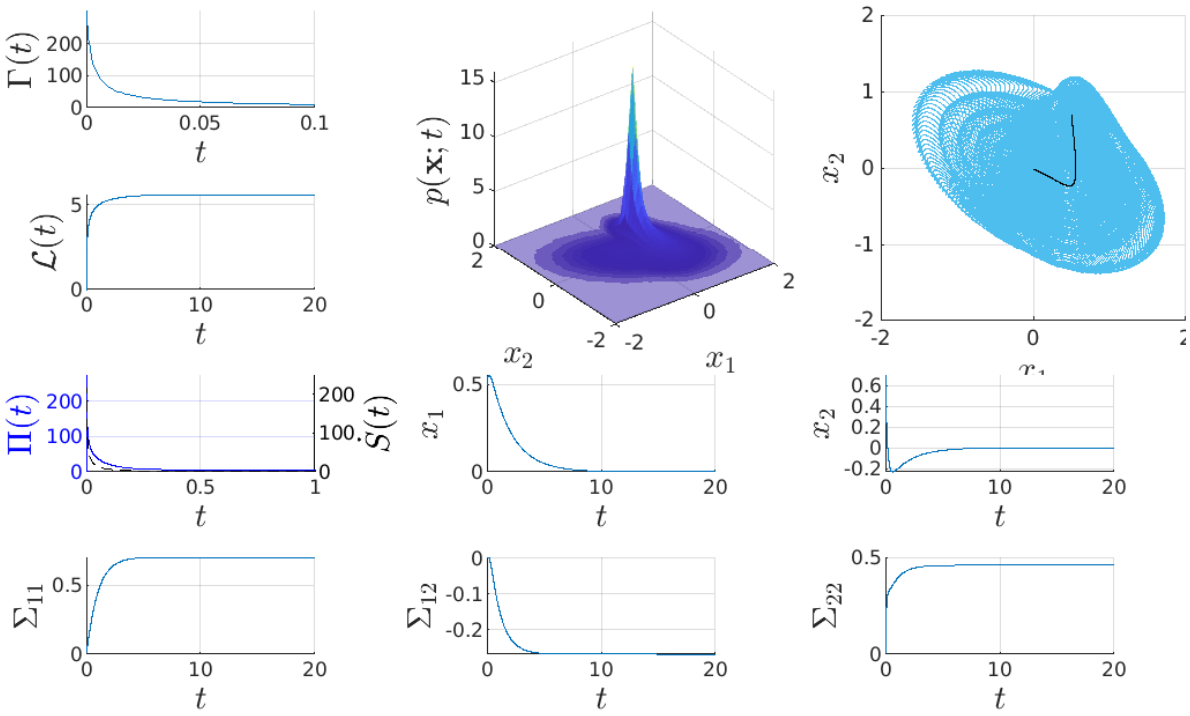


Figure 12: Full-state feedback control and temperature setting minimising J_1. A local minima is at $\mathbf{k} = [2.1229, 4.4453]^\top$ and $\mathbf{D} = [0.2684, 0; 0, 2.1181]$. The system parameters and initial conditions are $\omega = 1, \gamma = 2, \mu_1(0) = 0.5, \mu_2(0) = 0.7, \Sigma_{11} = \Sigma_{22} = 0.01, \Sigma_{12} = \Sigma_{21} = 0$.

time evolution of $\Gamma, \mathcal{L}, \Pi, x_1, x_2$ and the spaces $(p(\mathbf{x}, t), x_1, x_2)$ and (x_1, x_2) after applying the values of $\mathbf{k}$ and $\mathbf{D}$ that give a solution to the optimisation problem (56). As a result, the value of $\mathbf{k}$ contains $k_1 = 2.1229$ and $k_2 = 4.4453$ and $\mathbf{D}$, contains $D_{11} = 0.2684$ and $D_{22} = 2.1181$. This values produce an abrupt change in Γ, a quasi-logarithmic change in $\mathcal{L}$ with a maximum value slightly over 5 and a slow almost critically damped change in the system dynamics towards the equilibrium. In addition, the control quickly drives Π and $\dot{S}$ to zero. Even though the control action imposes a slow evolution of the mean value, the information rate quickly decreases. Such behaviour is desirable for systems where minimum information variability is more important than the speed under which we reach the equilibrium.

The solution of the optimisation problem (57) depicted in Figure 13 shows that a geodesic solution is obtained when entropy production Π and entropy rate $\dot{S}$ are

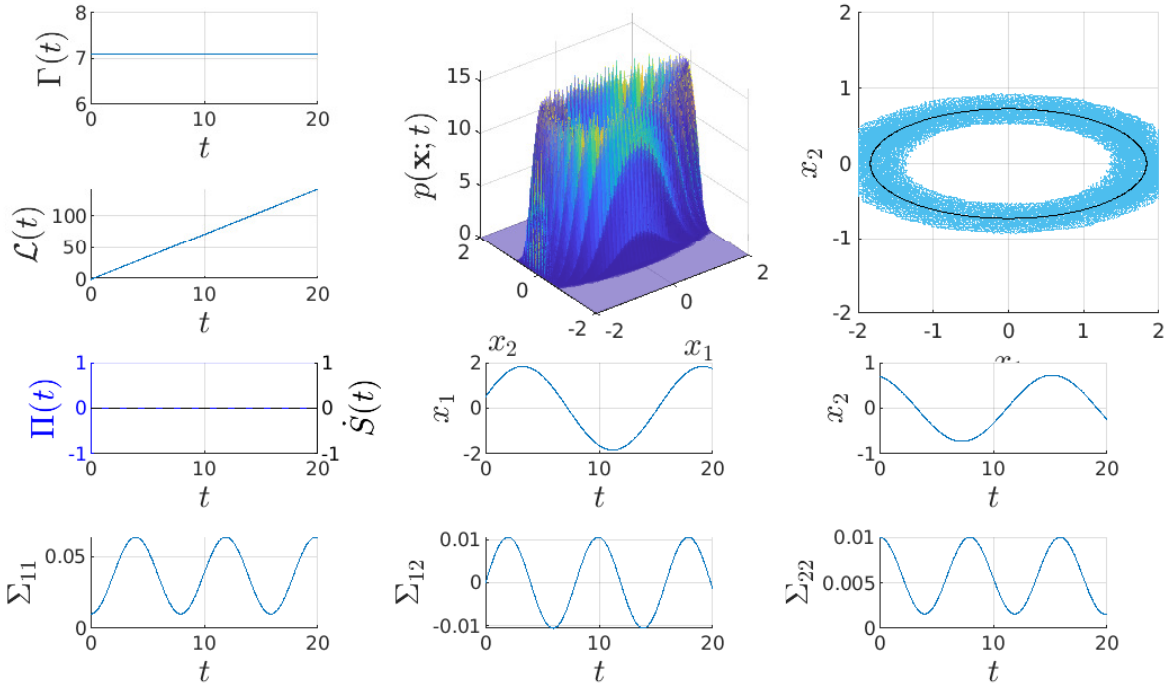


Figure 13: Full-state feedback control and temperature setting minimising J_2. A local minima is at $\mathbf{k} = [-0.8431, -2]^\top$ and $\mathbf{D} = [0, 0; 0, 0]$. The system parameters and initial conditions are $\omega = 1, \gamma = 2, \mu_1(0) = 0.5, \mu_2(0) = 0.7, \Sigma_{11} = \Sigma_{22} = 0.01, \Sigma_{12} = \Sigma_{21} = 0$.

zero. This is imposed by the resultant control parameters $k_1 = -0.8431$, $k_2 = -2$, $D_{11} = 0$ and $D_{22} = 0$ which generate a harmonic oscillatory behaviour of the mean value $\boldsymbol{\mu} = [\mu_1, \mu_2]^\top$ and small changes in the time evolution of the covariance matrix elements Σ_{11}, Σ_{12} and Σ_{22} to keep Γ constant at all t. Note that, $D_{11} = D_{22} = 0$ just means the absence of any external stochastic noise. However, there is stochasticity in the system due to the stochasticity in the initial condition (that is, our initial PDFs have a finite width). This optimal result is similar to the one shown in [49] where authors find that the optimal energy landscape corresponds to an harmonic oscillator at all times.

7. Conclusions

We have derived relations between information rate, entropy production and entropy rate for linear stochastic Gaussian processes changing according to the structure of the harmonic potential. Given such information-thermodynamic connection, we explore the applicability of these quantities to the design of thermodynamic optimal control algorithms and the detection of abrupt events in models of different nature. The results demonstrate that information rate ρ_Γ and entropy production ρ_Π correlation coefficients predict and detect abrupt events in the first and second moments of the stochastic dynamics, respectively. For higher-order systems, the norm of the information/thermodynamic quantities represents a fair approximation of the

correlation between all the system random variables. Besides, the control applications show that it is possible to obtain the geodesic of the information length via a simple full-state feedback control algorithm.

Since the proposed results permit us to establish a clear connection between information geometry and thermodynamic quantities, we can undoubtedly create cost functions that lead to energetically efficient (minimum entropy production) and organised (with minimum information variability) behaviours through IL. In the future, we plan to explore further connections with the area of control theory for applications in more complex scenarios. For instance, we can extend our results to non-linear systems by employing approximation methods such as the Laplacian assumption [52, 53, 50]. Then, we can utilise modern control techniques such as the model-predictive-control [54, 51] to find the solution to the proposed optimisation problems J_1 and J_2. This could bring potential benefits in the research areas of population dynamics [55] or inference control [56, 57].

References

[1] Nielsen, F. An elementary introduction to information geometry. *Entropy.* **22**, 1100 (2020)

[2] Amari, S. Information geometry and its applications. (Springer,2016)

[3] Kim, E. Information geometry and non-equilibrium thermodynamic relations in the over-damped stochastic processes. *Journal Of Statistical Mechanics: Theory And Experiment.* **2021**, 093406 (2021)

[4] Kim, E. Information Geometry, Fluctuations, Non-Equilibrium Thermodynamics, and Geodesics in Complex Systems. *Entropy.* **23**, 1393 (2021)

[5] Kim, E. & Hollerbach, R. Geometric structure and information change in phase transitions. *Physical Review E.* **95**, 062107 (2017)

[6] 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. *Physics Of Plasmas.* **27**, 102301 (2020)

[7] Kim, E., Heseltine, J. & Liu, H. Information length as a useful index to understand variability in the global circulation. *Mathematics.* **8**, 299 (2020)

[8] Guel-Cortez, A. & Kim, E. Information Geometric Theory in the Prediction of Abrupt Changes in System Dynamics. *Entropy.* **23**, 694 (2021)

[9] Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports On Progress In Physics.* **75**, 126001 (2012)

[10] Peliti, L. & Pigolotti, S. Stochastic Thermodynamics: An Introduction. (Princeton University Press,2021)

[11] Nicholson, S., Garcia-Pintos, L., Campo, A. & Green, J. Time–information uncertainty relations in thermodynamics. *Nature Physics.* **16**, 1211-1215 (2020)

[12] Lozano-Durán, A. & Arranz, G. Information-theoretic formulation of dynamical systems: causality, modeling, and control. *ArXiv Preprint ArXiv:2111.09484.* (2021)

[13] Dewar, R. Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. *Journal Of Physics A: Mathematical And General.* **36**, 631 (2003)

[14] Jarzynski, C. Fluctuation relations and strong inequalities for thermally isolated systems. *Physica A: Statistical Mechanics And Its Applications.* **552** pp. 122077 (2020)

[15] Prigogine, I. Time, structure, and fluctuations. *Science.* **201**, 777-785 (1978)

[16] Deffner, S. & Bonança, M.V. Thermodynamic control—An old paradigm with new applications. *Europhysics Letters.* **2**, 20001 (2020)
[17] Friston, K., Heins, C., Ueltzhöffer, K., Da Costa, L. & Parr, T. Stochastic chaos and markov blankets. *Entropy.* **23**, 1220 (2021)
[18] Haddad, W. A dynamical systems theory of thermodynamics. (Princeton University Press,2019)
[19] Bechhoefer, J. Control Theory for Physicists. (Cambridge University Press,2021)
[20] Tomé, T. Entropy production in nonequilibrium systems described by a Fokker-Planck equation. *Brazilian Journal Of Physics.* **36**, 1285-1289 (2006)
[21] Cocconi, L., Garcia-Millan, R., Zhen, Z., Buturca, B. & Pruessner, G. Entropy production in exactly solvable systems. *Entropy.* **22**, 1252 (2020)
[22] Ito, S., & Dechant, A. Stochastic time evolution, information geometry, and the Cramér-Rao bound. *Physical Review X.* **2**, 021056 (2020)
[23] Nicholson, S.B., del Campo, A., & Green, J.R. Nonequilibrium uncertainty principle from information geometry. *Physical Review E.* **3**, 032106 (2018)
[24] Ito, S. Stochastic thermodynamic interpretation of information geometry. *Physical review letters.* **3**, 030605 (2018)
[25] Guel-Cortez, A. & Kim, E. Information length analysis of linear autonomous stochastic processes. *Entropy.* **22**, 1265 (2020)
[26] Malagò, L. and Giovanni P. Information geometry of the Gaussian distribution in view of stochastic optimization. *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII.* (2015)
[27] Duncan, T. On the calculation of mutual information. *SIAM Journal On Applied Mathematics.* **19**, 215-220 (1970)
[28] Chen, C.T. Linear System Theory and Design. (Oxford University Press: New York, NY, USA, Volume 7, 2013)
[29] Tomé, T. & De Oliveira, M. Stochastic dynamics and irreversibility. (Springer,2015)
[30] Maybeck, P. Stochastic models, estimation, and control. (Academic press,1982)
[31] Landi, G., Tomé, T. & De Oliveira, M. Entropy production in linear Langevin systems. *Journal Of Physics A: Mathematical And Theoretical.* **46**, 395001 (2013)
[32] Falb, P. & Wolovich, W. Decoupling in the design and synthesis of multivariable control systems. *IEEE transactions on automatic control.* **6**, 651-659 (1967)
[33] Kim, E., Lee, U., Heseltine, J. & Hollerbach, R. Geometric structure and geodesic in a solvable model of nonequilibrium process. *Physical Review E.* **93**, 062127 (2016)
[34] Nguyen Thi Thanh, N., Nguyen Kim, K., Ngo Hong, S. & Ngo Lam, T. Entropy correlation and its impacts on data aggregation in a wireless sensor network. *Sensors.* **18**, 3118 (2018)
[35] Cahill, N. Normalized measures of mutual information with general definitions of entropy for multimodal image registration. *International Workshop On Biomedical Image Registration.* pp. 258-268 (2010)
[36] Press, W., Teukolsky, S., Vetterling, W. & Flannery, B. Numerical recipes 3rd edition: The art of scientific computing. (Cambridge university press,2007)
[37] Veyrat-Charvillon, N. & Standaert, F. Mutual information analysis: how, when and why?. *International Workshop On Cryptographic Hardware And Embedded Systems.* pp. 429-443 (2009)
[38] Li, W. Mutual information functions versus correlation functions. *Journal Of Statistical Physics.* **60**, 823-837 (1990)
[39] Dionisio, A., Menezes, R. & Mendes, D. Mutual information: a measure of dependency for nonlinear time series. *Physica A: Statistical Mechanics And Its Applications.* **344**, 326-329 (2004)
[40] Wijaya, D., Sarno, R. & Zulaika, E. Information Quality Ratio as a novel metric for mother wavelet selection. *Chemometrics And Intelligent Laboratory Systems.* **160** pp. 59-71 (2017)
[41] Sontag, E.D. Mathematical control theory: deterministic finite dimensional systems (Vol. 6) (Springer Science & Business Media, 2013)
[42] Pesce, G., Jones, P., Maragò, O. & Volpe, G. Optical tweezers: theory and practice. *The European*

Physical Journal Plus. **135**, 1-38 (2020)

[43] Salamon, P., Nitzan, A., Andresen, B. & Berry, R. Minimum entropy production and the optimization of heat engines. *Physical Review A.* **21**, 2115 (1980)

[44] Martyushev, L., Nazarova, A. & Seleznev, V. On the problem of the minimum entropy production in the nonequilibrium stationary state. *Journal Of Physics A: Mathematical And Theoretical.* **40**, 371 (2006)

[45] Aurell, E., Mejía-Monasterio, C., & Muratore-Ginanneschi, P. Optimal protocols and optimal transport in stochastic thermodynamics. *Physical review letters.* **25**, 250601 (2011)

[46] Dechant, A., & Sakurai, Y. Thermodynamic interpretation of Wasserstein distance. *ArXiv Preprint arXiv:1912.08405* (2019)

[47] Guéry-Odelin, D., Jarzynski, C., Plata, C. A., Prados, A., & Trizac, E. Driving rapidly while remaining in control: classical shortcuts from Hamiltonian to stochastic dynamics. *Reports on Progress in Physics.* (2022)

[48] The MathWorks, I. Nonlinear programming solver. (2022), https://www.mathworks.com/help/optim/ug/fmincon.html

[49] Proesmans, K. Precision-dissipation trade-off for driven stochastic systems. *ArXiv Preprint arXiv:2203.00428.* (2022)

[50] Guel-Cortez, A.J., & Kim, E.J. Information Geometry Control under the Laplace Assumption. *In Physical Sciences Forum. Multidisciplinary Digital Publishing Institute* **1** p. 25 (2022)

[51] Guel-Cortez, A.J., Kim, E.J., & Mehrez, M.W. Minimum Information Variability in Linear Langevin Systems Via Model Predictive Control. *Preprint available at SSRN 4214108* (2022)

[52] Marreiros, A., Kiebel, S., Daunizeau, J., Harrison, L. & Friston, K. Population dynamics under the Laplace assumption. *Neuroimage.* **44**, 701-714 (2009)

[53] Da Costa, L., Friston, K., Heins, C. & Pavliotis, G. Bayesian mechanics for stationary processes. *Proceedings Of The Royal Society A.* **477**, 20210518 (2021)

[54] Camacho, E. & Alba, C. Model predictive control. (Springer science & business media,2013)

[55] Schwartenbeck, P., FitzGerald, T., Dolan, R. & Friston, K. Exploration, novelty, surprise, and free energy minimization. *Frontiers In Psychology.* pp. 710 (2013)

[56] Lanillos, P., Meo, C., Pezzato, C., Meera, A., Baioumy, M., Ohata, W., Tschantz, A., Millidge, B., Wisse, M., Buckley, C. & Others Active Inference in Robotics and Artificial Agents: Survey and Challenges. *ArXiv Preprint ArXiv:2112.01871.* (2021)

[57] Baltieri, M. & Buckley, C. PID control as a process of active inference with linear generative models. *Entropy.* **21**, 257 (2019)

[58] Petersen, K., Pedersen, M. & Others The matrix cookbook. *Technical University Of Denmark.* **7**, 510 (2008)

[59] Thiruthummal, A.A., & Kim, E.J. Monte Carlo simulation of stochastic differential equation to study information geometry. *Entropy.* **8** 1113 (2022)

[60] Yang, X. A matrix trace inequality. *Journal Of Mathematical Analysis And Applications.* **250**, 372-374 (2000)

[61] Shebrawi, K. & Albadawi, H. Trace inequalities for matrices. *Bulletin Of The Australian Mathematical Society.* **87**, 139-148 (2013)

[62] Patel, R. & Toda, M. Trace inequalities involving Hermitian matrices. *Linear Algebra And Its Applications.* **23** pp. 13-20 (1979)

Appendix A. Derivation of Entropy rate for Gaussian Process

We start by applying the definition of entropy (14) production and entropy flux (16) giving us

$$\Pi_{J_i}=\frac{1}{D_{ii}}\left\langle f_i(\mathbf{x},t)^2\right\rangle+D_{ii}\left\langle\left(\frac{\partial Q(\mathbf{x})}{\partial x_i}\right)^2\right\rangle+2\left\langle\frac{\partial f_i(\mathbf{x},t)}{\partial x_i}\right\rangle, \tag{A.1}$$

$$\Phi_{J_i}=\frac{1}{D_{ii}}\langle f_i(\mathbf{x},t)^2\rangle+\left\langle\frac{\partial f_i(\mathbf{x},t)}{\partial x_i}\right\rangle. \tag{A.2}$$

Before continuing, it is useful to note that [58]

$$\frac{\partial Q}{\partial x_k}=-\frac{1}{2}\left[\sum_i \delta x_i\Sigma^{-1}_{ki}+\sum_j \delta x_j\Sigma^{-1}_{jk}\right]=-\sum_i \delta x_i\Sigma^{-1}_{ki}=-\delta\mathbf{x}^\top\boldsymbol{\Sigma}^{-1}_k \tag{A.3}$$

where $\delta x_i = x_i-\mu_i$, $\delta\mathbf{x}:=\mathbf{x}-\boldsymbol{\mu}=[\delta x_1,\dots,\delta x_n]^\top$ and $\boldsymbol{\Sigma}^{-1}_k$ is the k-th column of the inverse matrix $\boldsymbol{\Sigma}^{-1}$ of $\boldsymbol{\Sigma}$. Besides,

$$f_i(\mathbf{x})^2=\mathbf{x}^\top\mathbf{A}_i^\top\mathbf{A}_i\mathbf{x}+\mathbf{u}^\top\mathbf{B}_i^\top\mathbf{B}_i\mathbf{u}+2\mathbf{u}^\top\mathbf{B}_i^\top\mathbf{A}_i\mathbf{x}, \tag{A.4}$$

where we recall that $\mathbf{A}_i$ is the i-th arrow of the matrix $\mathbf{A}$. Therefore [58]

$$\left\langle D_{ii}\left(\frac{\partial Q(\mathbf{x})}{\partial x_i}\right)^2\right\rangle=D_{ii}\left\langle\delta\mathbf{x}^\top\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\Sigma}_i^{-1})^\top\delta\mathbf{x}\right\rangle=D_{ii}\,\mathrm{Tr}(\boldsymbol{\Delta}_i\boldsymbol{\Sigma}), \tag{A.5}$$

and

$$\frac{\langle f_i(\mathbf{x})^2\rangle}{D_{ii}}=\frac{1}{D_{ii}}\left(\mathrm{Tr}(\boldsymbol{\Gamma}_i\boldsymbol{\Sigma})+\boldsymbol{\mu}^\top\boldsymbol{\Gamma}_i\boldsymbol{\mu}+\mathbf{u}^\top\boldsymbol{\Omega}_i\mathbf{u}+2\mathbf{u}^\top\boldsymbol{\varphi}_i\boldsymbol{\mu}\right) \tag{A.6}$$

where $\boldsymbol{\Delta}_i=\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\Sigma}_i^{-1})^\top$, $\boldsymbol{\Gamma}_i=\mathbf{A}_i^\top\mathbf{A}_i$, $\boldsymbol{\Omega}_i=\mathbf{B}_i^\top\mathbf{B}_i$, and $\boldsymbol{\varphi}_i=\mathbf{B}_i^\top\mathbf{A}_i$. Furthermore, we have that

$$\frac{\partial f_i(\mathbf{x})}{\partial x_i}=a_{ii}.$$

Then

$$\begin{aligned}\Pi_{J_i}&=\frac{1}{D_{ii}}\left(\mathrm{Tr}(\boldsymbol{\Gamma}_i\boldsymbol{\Sigma})+\boldsymbol{\mu}^\top\boldsymbol{\Gamma}_i\boldsymbol{\mu}+\mathbf{u}^\top\boldsymbol{\Omega}_i\mathbf{u}+2\mathbf{u}^\top\boldsymbol{\varphi}_i\boldsymbol{\mu}\right)+D_{ii}\,\mathrm{Tr}(\boldsymbol{\Delta}_i\boldsymbol{\Sigma})+2a_{ii},\\ \Phi_{J_i}&=\frac{1}{D_{ii}}\left(\mathrm{Tr}(\boldsymbol{\Gamma}_i\boldsymbol{\Sigma})+\boldsymbol{\mu}^\top\boldsymbol{\Gamma}_i\boldsymbol{\mu}+\mathbf{u}^\top\boldsymbol{\Omega}_i\mathbf{u}+2\mathbf{u}^\top\boldsymbol{\varphi}_i\boldsymbol{\mu}\right)+a_{ii}.\end{aligned} \tag{A.7}$$

Finally, since

$$\sum_i^n\frac{\boldsymbol{\mu}^\top\boldsymbol{\Gamma}_i\boldsymbol{\mu}}{D_{ii}}=\boldsymbol{\mu}^\top\left(\frac{\mathbf{A}_1^\top\mathbf{A}_1}{D_{11}}+\cdots+\frac{\mathbf{A}_n^\top\mathbf{A}_n}{D_{nn}}\right)\boldsymbol{\mu}=\boldsymbol{\mu}^\top\mathbf{A}^\top\mathbf{D}^{-1}\mathbf{A}\boldsymbol{\mu}, \tag{A.8}$$

$$\begin{aligned}\sum_i^n\frac{\mathrm{Tr}(\boldsymbol{\Delta}_i\boldsymbol{\Sigma})}{D_{ii}}&=\mathrm{Tr}\left(\left(\frac{\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\Sigma}_1^{-1})^\top}{D_{11}}+\cdots+\frac{\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\Sigma}_n^{-1})^\top}{D_{nn}}\right)\boldsymbol{\Sigma}\right)\\ &=\mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{D}^{-1}(\boldsymbol{\Sigma}^{-1})^\top\boldsymbol{\Sigma}\right).\end{aligned} \tag{A.9}$$

Now, after applying the same reasoning to all the terms in the right hand side of (A.7) and simplifying, we get

$$\Pi = \dot{\boldsymbol{\mu}}^\top \mathbf{D}^{-1}\dot{\boldsymbol{\mu}} + \operatorname{Tr}\left(\mathbf{A}^\top \mathbf{D}^{-1}\mathbf{A}\boldsymbol{\Sigma}\right) + \operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{D}\right) + 2\operatorname{Tr}(\mathbf{A}), \tag{A.10}$$

$$\Phi = \dot{\boldsymbol{\mu}}^\top \mathbf{D}^{-1}\dot{\boldsymbol{\mu}} + \operatorname{Tr}\left(\mathbf{A}^\top \mathbf{D}^{-1}\mathbf{A}\boldsymbol{\Sigma}\right) + \operatorname{Tr}(\mathbf{A}), \tag{A.11}$$

$$\dot{S} = \operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{D}\right) + \operatorname{Tr}(\mathbf{A}) = \frac{1}{2}\operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}\right), \tag{A.12}$$

which corresponds to the result given in Relation 1.

To provide extra confidence in the validity of expressions (A.10)-(A.12). Consider as an example, the stochastic simulation [59], i.e. the multiple numerical solution of the stochastic differential equation, of (45) using the parameters $\boldsymbol{\mu}(0) = [1, 0.5]^\top, a_{11} = 0, a_{12} = 1, a_{21} = -2, a_{22} = -3$, $\mathbf{D} = 0.01\mathbf{I}$ and $\boldsymbol{\Sigma}(0) = 0.1\mathbf{I}$ to compute the value of entropy production Π via (A.1). Then, let us compare it with the "analytical" solution obtained using (A.10) and the solution of equations (9)-(10).

Since the stochastic simulation uses its solutions to compute the averages $\langle f_i\rangle$ and $\langle(\partial_{x_i}Q)^2\rangle$ in (A.1). Figure A1 includes a comparison of the analytical vs numerical values of $\langle f_1\rangle, \langle f_2\rangle, \langle(\partial_{x_1}Q)^2\rangle, \langle(\partial_{x_2}Q)^2\rangle$ and Π showing a perfect agreement between both approaches as expected.

Appendix B. Proof of Relation 2

For any real matrix $\mathbf{A}$ in system (9)-(10), we can rewrite the second term in the right hand side of (26) as follows

$$\begin{aligned}
\operatorname{Tr}\left((\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}})^2\right) &= \operatorname{Tr}\left(2\mathbf{A}^2 + 2\boldsymbol{\Sigma}^{-1}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\right. \\
&\quad \left. + 8\boldsymbol{\Sigma}^{-1}\mathbf{A}\mathbf{D} + 4\boldsymbol{\Sigma}^{-1}\mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}\right) \\
&= \operatorname{Tr}(2\mathbf{A}^2 + 4\boldsymbol{\Sigma}^{-1}\mathbf{A}\mathbf{D} + 2\boldsymbol{\Sigma}^{-1}\mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}) \\
&\quad + \operatorname{Tr}(2\boldsymbol{\Sigma}^{-1}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A} + 4\boldsymbol{\Sigma}^{-1}\mathbf{A}\mathbf{D} + 2\boldsymbol{\Sigma}^{-1}\mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}) \\
&= 2\operatorname{Tr}\left((\boldsymbol{\Sigma}^{-1}\mathbf{D} + A)^2\right) \\
&\quad + 2\operatorname{Tr}(\boldsymbol{\Sigma}^{-1}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top\mathbf{D}^{-1} + 2\mathbf{A} + \mathbf{D}\boldsymbol{\Sigma}^{-1})\mathbf{D}).
\end{aligned} \tag{B.1}$$

Equation (B.1) can be written in terms of Entropy production Π and entropy rate $\dot{S}$ using the following results. First, from the fact that $\Pi \geq 0$ and $\Sigma^{-1}, \mathbf{D} \succeq 0$ we get

$$\begin{aligned}
\operatorname{Tr}(\boldsymbol{\Sigma}^{-1})\Pi\operatorname{Tr}(\mathbf{D}) &= \operatorname{Tr}(\boldsymbol{\Sigma}^{-1})\operatorname{Tr}\left(\dot{\boldsymbol{\mu}}\dot{\boldsymbol{\mu}}^\top\mathbf{D}^{-1} + \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top\mathbf{D}^{-1} + 2\mathbf{A} + \mathbf{D}\boldsymbol{\Sigma}^{-1}\right)\operatorname{Tr}(\mathbf{D}) \\
&\geq \operatorname{Tr}(\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\mu}}\dot{\boldsymbol{\mu}}^\top) + \operatorname{Tr}(\boldsymbol{\Sigma}^{-1}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top\mathbf{D}^{-1} + 2\mathbf{A} + \mathbf{D}\boldsymbol{\Sigma}^{-1})\mathbf{D}).
\end{aligned} \tag{B.2}$$

Now, taking λ_i as the eigenvalues of the matrix $\mathbf{H} := \boldsymbol{\Sigma}^{-1}\mathbf{D} + \mathbf{A}$, we have

$$\dot{S}^2 = \operatorname{Tr}(\mathbf{H})^2 = (\sum_{i=1}^{n}\lambda_i)^2 = \sum_{i=1}^{n}\lambda_i^2 + 2\sum_{j=1}^{n}\sum_{i=1}^{j-1}\lambda_i\lambda_j = \operatorname{Tr}(\mathbf{H}^2) + 2\sum_{i<j}^{n}\lambda_i\lambda_j = \operatorname{Tr}(\mathbf{H}^2) + 2g(\mathbf{H}). \tag{B.3}$$

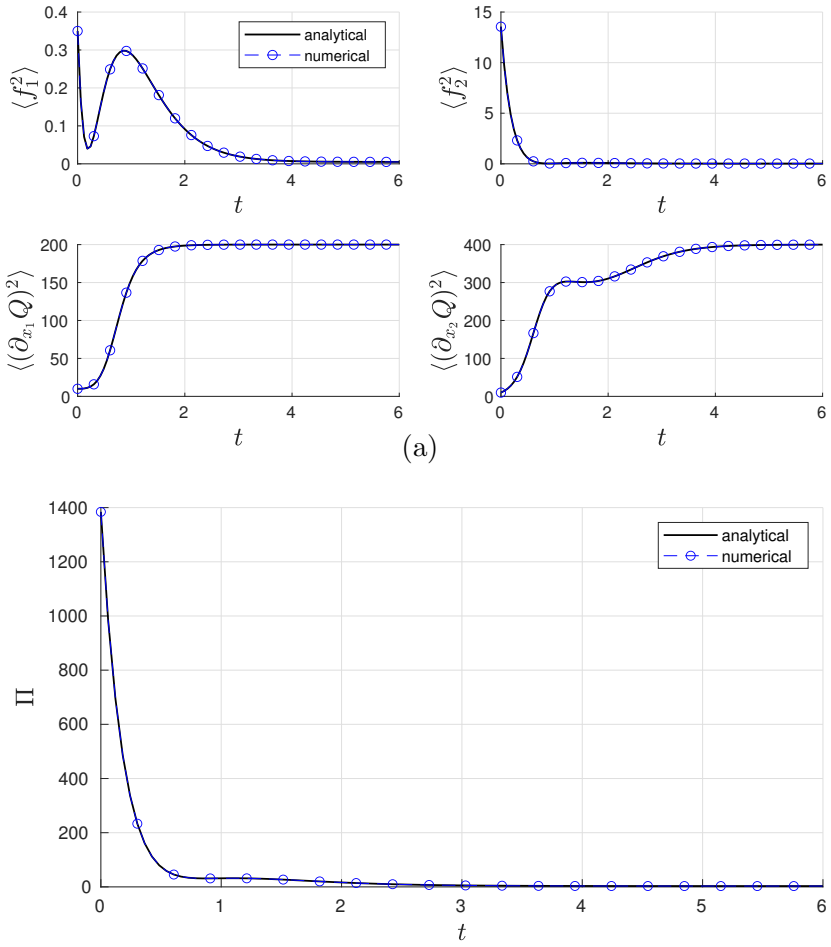


Figure A1: Stochastic simulation (numerical) vs analytical solution of the entropy production Π for (45) using the parameters $\boldsymbol{\mu}(0) = [1, 0.5]^\top, a_{11} = 0, a_{12} = 1, a_{21} = -2, a_{22} = -3$, $\mathbf{D} = 0.01\mathbf{I}$ to produce $\boldsymbol{\Sigma}(0) = 0.1\mathbf{I}$. The results show an agreement between the analytical and numerical results as expected.

Finally, using (B.1)-(B.3) in (26) we get

$$\begin{aligned} 2g(\mathbf{H}) \leq \Gamma^2 + 2g(\mathbf{H}) &\leq \operatorname{Tr}(\boldsymbol{\Sigma}^{-1})\Pi \operatorname{Tr}(\mathbf{D}) + \operatorname{Tr}(\mathbf{H}^2) + 2g(\mathbf{H}), \\ 2g(\mathbf{H}) \leq \Gamma^2 + 2g(\mathbf{H}) &\leq \operatorname{Tr}(\boldsymbol{\Sigma}^{-1})\Pi \operatorname{Tr}(\mathbf{D}) + \dot{S}^2, \\ 0 \leq \Gamma^2 &\leq \operatorname{Tr}(\boldsymbol{\Sigma}^{-1})\Pi \operatorname{Tr}(\mathbf{D}) + \dot{S}^2 - 2g(\mathbf{H}). \end{aligned} \quad \text{(B.4)}$$

Now, since $\dot{S} = \sum_i^n \dot{S}_{J_i}$ where again $\dot{S}_{J_i}$ is the contribution of the current flow J_i to the total entropy rate $\dot{S}$, we see that each eigenvalue $\lambda_i = \dot{S}_{J_i}$. Note that, for simplicity, we have changed the argument of the function g from $\mathbf{H}$ to a new variable vector $\mathbf{s} = [\dot{S}_{J_1}, \dot{S}_{J_2}, \ldots, \dot{S}_{J_n}]^\top$, i.e. now $g : \mathbf{R}^n \to \mathbf{R}$. This ends our proof.

Appendix C. Proof Relation 3

To derive the result shown in Relation 2, we first consider the following preliminary results [60, 61, 62]

$$\begin{aligned}
\operatorname{Tr}(\mathbf{XY}) &\leq \operatorname{Tr}(\mathbf{X})\operatorname{Tr}(\mathbf{Y}) \quad \forall \quad \mathbf{X},\mathbf{Y} \succeq 0, && \text{(C.1)}\\
\operatorname{Tr}(\dot{\mathbf{\Sigma}}) &= \operatorname{Tr}(\mathbf{A\Sigma}) + \operatorname{Tr}(\mathbf{\Sigma A}^\top) + \operatorname{Tr}(2\mathbf{D}) \\
&= 2\operatorname{Tr}(\mathbf{\Sigma A} + \mathbf{D}), && \text{(C.2)}\\
\operatorname{Tr}(\dot{\mathbf{\Sigma}})^2 &= 4\operatorname{Tr}(\mathbf{\Sigma A} + \mathbf{D})\operatorname{Tr}(\mathbf{\Sigma A}^\top + \mathbf{D}) \\
&\geq 4\operatorname{Tr}(\mathbf{\Sigma A \Sigma A}^\top + 2\mathbf{\Sigma A D} + \mathbf{D}^2), && \text{(C.3)}\\
\operatorname{Tr}(\mathbf{\Sigma}^{-1})^2\operatorname{Tr}(\dot{\mathbf{\Sigma}})^2 &= \operatorname{Tr}(\mathbf{\Sigma}^{-1})\operatorname{Tr}(\dot{\mathbf{\Sigma}})\operatorname{Tr}(\mathbf{\Sigma}^{-1})\operatorname{Tr}(\dot{\mathbf{\Sigma}}) \\
&\geq \operatorname{Tr}(\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}})^2 \geq \operatorname{Tr}((\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}})^2) && \text{(C.4)}
\end{aligned}$$

Then, by applying the previous results to the definition of Γ^2 in (26), we have

$$\begin{aligned}
0 \leq \Gamma^2 &= \operatorname{Tr}(\dot{\boldsymbol{\mu}}^\top \mathbf{\Sigma}^{-1}\dot{\boldsymbol{\mu}}) + \frac{1}{2}\operatorname{Tr}\left((\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}})^2\right) \\
&\leq \operatorname{Tr}(\dot{\boldsymbol{\mu}}\dot{\boldsymbol{\mu}}^\top)\operatorname{Tr}(\mathbf{\Sigma}^{-1}) + \frac{1}{2}\operatorname{Tr}(\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}})^2 && \text{(C.5)}\\
&\leq \operatorname{Tr}(\dot{\boldsymbol{\mu}}\dot{\boldsymbol{\mu}}^\top)\operatorname{Tr}(\mathbf{\Sigma}^{-1}) + \frac{1}{4}\operatorname{Tr}(\mathbf{\Sigma}^{-1})^2\operatorname{Tr}(\dot{\mathbf{\Sigma}})^2 + \dot{S}^2.
\end{aligned}$$

Now, multiplying both sides of inequality (C.5) by $\operatorname{Tr}(\mathbf{D}^{-1}\mathbf{D})$ and factorising $\operatorname{Tr}(\mathbf{\Sigma}^{-1})$ from its right hand side, we have

$$\begin{aligned}
0 \leq n\Gamma^2 &\leq \operatorname{Tr}(\mathbf{\Sigma}^{-1})\{\operatorname{Tr}(\dot{\boldsymbol{\mu}}\dot{\boldsymbol{\mu}}^\top)\operatorname{Tr}(\mathbf{D}^{-1}\mathbf{D}) \\
&\quad + \frac{1}{4}\operatorname{Tr}(\mathbf{\Sigma}^{-1})\operatorname{Tr}(\dot{\mathbf{\Sigma}})^2\operatorname{Tr}(\mathbf{D}^{-1}\mathbf{D})\} + n\dot{S}^2 \\
&\leq \operatorname{Tr}(\mathbf{\Sigma}^{-1})\{\operatorname{Tr}(\dot{\boldsymbol{\mu}}\dot{\boldsymbol{\mu}}^\top)\operatorname{Tr}(\mathbf{D}^{-1}) && \text{(C.6)}\\
&\quad + \frac{1}{4}\operatorname{Tr}(\mathbf{\Sigma}^{-1})\operatorname{Tr}(\dot{\mathbf{\Sigma}})^2\operatorname{Tr}(\mathbf{D}^{-1})\}\operatorname{Tr}(\mathbf{D}) + n\dot{S}^2.
\end{aligned}$$

From the right hand side of (C.6), we define the part inside the curly brackets as

$$\Pi_u := \operatorname{Tr}(\dot{\boldsymbol{\mu}}\dot{\boldsymbol{\mu}}^\top)\operatorname{Tr}(\mathbf{D}^{-1}) + \frac{1}{4}\operatorname{Tr}(\mathbf{\Sigma}^{-1})\operatorname{Tr}(\dot{\mathbf{\Sigma}})^2\operatorname{Tr}(\mathbf{D}^{-1}). \quad \text{(C.7)}$$

Which gives us the expression in our result (29). The value of Π_u can be proved to be an upper bound of Π from the following reasoning

$$\begin{aligned}
\Pi_u &\geq \operatorname{Tr}(\dot{\boldsymbol{\mu}}^\top\mathbf{D}^{-1}\dot{\boldsymbol{\mu}}) \\
&\quad + \operatorname{Tr}(\mathbf{\Sigma}^{-1})\operatorname{Tr}(\mathbf{\Sigma A \Sigma A}^\top + 2\mathbf{\Sigma A D} + \mathbf{D}^2)\operatorname{Tr}(\mathbf{D}^{-1}) \\
&\geq \dot{\boldsymbol{\mu}}^\top\mathbf{D}^{-1}\dot{\boldsymbol{\mu}} + \mathbf{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}(\mathbf{A\Sigma A}^\top\mathbf{D}^{-1} + 2\mathbf{A} + \mathbf{\Sigma}^{-1}\mathbf{D})\mathbf{D}\mathbf{D}^{-1}\right) \\
&= \dot{\boldsymbol{\mu}}^\top\mathbf{D}^{-1}\dot{\boldsymbol{\mu}} + \mathbf{tr}(\mathbf{A\Sigma A}^\top\mathbf{D}^{-1} + 2\mathbf{A} + \mathbf{\Sigma}^{-1}\mathbf{D}). && \text{(C.8)}
\end{aligned}$$

Note that for $\Pi_u \geq \Pi$ we need $\mathbf{A} \succeq 0$. A similar result can be found starting from the definition of Π_u in (C.7) as follows

$$\begin{aligned}
\operatorname{Tr}(\mathbf{\Sigma}^{-1})\Pi_u\operatorname{Tr}(\mathbf{D}) \geq \operatorname{Tr}(\mathbf{\Sigma}^{-1})\{&\operatorname{Tr}(\dot{\boldsymbol{\mu}}\dot{\boldsymbol{\mu}}^\top)\operatorname{Tr}(\mathbf{D}^{-1}\mathbf{D}) \\
&+ \frac{1}{4}\operatorname{Tr}(\mathbf{\Sigma}^{-1})\operatorname{Tr}(\dot{\mathbf{\Sigma}})^2\operatorname{Tr}(\mathbf{D}^{-1}\mathbf{D})\}. \quad \text{(C.9)}
\end{aligned}$$

From (C.8), the main result follows straightforwardly using (C.6) leading to our main result in Relation 2.

Appendix D. Proof of Relation 4

If $\mathbf{A}$ is an $n \times n$ diagonal matrix, then $\mathbf{\Sigma}$ and $\dot{\mathbf{\Sigma}}$ are also diagonal and the following expressions hold

$$\Gamma^2 = \sum_i \Gamma_i^2 = \sum_i \left(\frac{\dot{\mu}_i^2}{\Sigma_{ii}} + \frac{1}{2} \left(\frac{\dot{\Sigma}_{ii}}{\Sigma_{ii}} \right)^2 \right), \tag{D.1}$$

$$\begin{aligned} \Pi &= \sum_i \Pi_i = \dot{\boldsymbol{\mu}}^\top \mathbf{D}^{-1} \dot{\boldsymbol{\mu}} + \frac{1}{4} \operatorname{Tr} \left(\mathbf{\Sigma}^{-1} \dot{\mathbf{\Sigma}}^2 \mathbf{D}^{-1} \right) \\ &= \sum_i \left(\frac{\dot{\mu}_i^2}{D_{ii}} + \frac{\dot{\Sigma}_{ii}^2}{4 \Sigma_{ii} D_{ii}} \right), \end{aligned} \tag{D.2}$$

$$\dot{S} = \sum_i \dot{S}_i = \frac{1}{2} \sum_i \frac{\dot{\Sigma}_{ii}}{\Sigma_{ii}}. \tag{D.3}$$

By rearranging Equations (D.2) and (D.3) to form Γ^2, we have

$$\Gamma_i^2 = \frac{D_{ii}}{\Sigma_{ii}} \Pi_i + \dot{S}_i^2. \tag{D.4}$$

which leads to our result.

Appendix E. Proof of equation (49)

Given that $\mathbf{A}$ is a diagonal and stable matrix, we have

$$\dot{\mathbf{\Sigma}} = 2\mathbf{A}\mathbf{\Sigma} + 2\mathbf{D}, \tag{E.1}$$

$$\dot{\boldsymbol{\mu}} = \mathbf{A}\boldsymbol{\mu}, \tag{E.2}$$

which gives

$$\mathbf{\Sigma}^{-1}(\infty) = -\mathbf{D}^{-1}\mathbf{A}, \tag{E.3}$$

$$\dot{\mathbf{\Sigma}}(\infty) = 0, \tag{E.4}$$

$$\boldsymbol{\mu}(\infty) = 0, \tag{E.5}$$

$$\dot{\boldsymbol{\mu}}(\infty) = 0. \tag{E.6}$$

In this manner, the entropy production and entropy rate are

$$\Pi(\infty) = \sum_i \left(\frac{\dot{\mu}_i(\infty)^2}{D_{ii}} + \frac{\dot{\Sigma}_{ii}(\infty)^2}{4 \Sigma_{ii}(\infty) D_{ii}} \right) = 0, \tag{E.7}$$

$$\dot{S}(\infty) = \sum_i \dot{S}_{J_i} = \frac{1}{2} \sum_i \frac{\dot{\Sigma}_{ii}(\infty)}{\Sigma_{ii}(\infty)} = 0. \tag{E.8}$$

Giving us the result mentioned in equation (49).

$$\mathbf{\Sigma}^{-1}(\infty) = -\mathbf{D}^{-1}\mathbf{A}, \quad \Pi(\infty) = 0. \quad \mathbf{s}(\infty) = 0. \tag{E.9}$$