computer sciences and mathematics forum

MDPI

Proceeding Paper

Recursive Way to Get the Cross-Entropy Rate to Evaluate Data Model Dissimilarity in a State-Space Framework †

Eric Grivel

IMS UMR CNRS 5218, Bordeaux INP, University of Bordeaux, 351 Cours de la Liberation, 33400 Talence, France; eric.grivel@ims-bordeaux.fr

† Presented at the 12th International Conference on Time Series and Forecasting, Canaria, Spain, 15–17 July 2026.

Abstract

In many signal processing problems, such as GPS navigation, target tracking, and speech enhancement, the available data can be modeled *a priori* within a state-space framework. This framework describes how the state vector evolves over time and how it is related to the observations. Comparing or classifying state-space representations (SSRs) of a system can be of interest for estimation purposes, but not only. However, the number k of consecutive state vectors considered over time may significantly impact the evaluation of similarity between the models. In this paper, we derive analytical expressions for the entropy associated with the equation updating the state vector under a given model, as well as for the Kullback–Leibler divergence and the cross-entropy between two different models, using a recursive approach. Under certain conditions given in the paper, which are explicitly stated, we show that these measures, as functions of k, are either equal to or well approximated by affine functions. The corresponding slopes define quantities known as rates in information theory. These rates do not depend on k and can therefore be used as meaningful criteria for comparing SSRs and, consequently, the *a priori* models used to describe the time series to be processed. Illustrations are provided for motion models.

Keywords: times series; cross-entropy; entropy; Kulbback–Leibler divergence; rate; recursive method

1. Introduction

Many estimation problems based on time series can be formulated in a state-space framework: speech enhancement, GPS navigation, object tracking, etc. In this case, the first equation describes the system dynamics and how the state vector, which stores the state variables, evolves over time, whereas the second equation makes it possible to show how the state variables are related to the available observations (In many cases, when the state vector is estimated using H_∞ filtering, a third equation can be introduced to focus on specific components of the state vector.) State-space representations can be either linear or nonlinear, and may rely on Gaussian assumptions or not. Depending on the assumptions that are made, the state-vector estimation can be done by using Kalman filtering or smoothing techniques, including the extended Kalman filter (EKF) [1], the second-order Kalman filter [2], the unscented Kalman filter (UKF) [3], the central difference Kalman filter (CDKF) [4], the sigma point Kalman filter (SPKF) [5] or particle filtering.

To relax the *a priori* put on the representation of the system, approaches based on multiple models can be considered, such as the standard interactive multiple model (IMM) [6] or the variable-structure IMM methods [7]. In this framework, switching between models

check for updates

Academic Editors: Olga Valenzuela, Fernando Rojas, Luis Javier Herrera, Hector Pomares and Ignacio Rojas

Published: 29 September 2026

Copyright: © 2026 by the author. 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.

is represented by a discrete-time Markov chain. The corresponding transition probabilities can be fixed *a priori* or recursively estimated from the data [8]. However, the relevance of IMM approaches relies on the existence of sufficient diversity between models chosen *a priori*. To this end, working on the divergence, and more generally the cross-entropy, between models can be of interest when considering their state-space representation (SSR).

A first possible approach would consist of evaluating the Kullback–Leibler (KL) divergence between the probability density functions (pdf) associated with values of the state vector taken at k consecutive time instants, under two models. Alternative divergences could be also considered such as the Rényi divergence, etc. However, the choice of k constitutes a strong prior assumption. Models that seem to be weakly different for small values of k may exhibit significantly larger differences when k increases. To address this limitation, the divergence rate can be considered. The most standard way to deduce it is to express the difference between the divergences computed for k and $k-1$ consecutive time instants. When k increases and tends to infinity, the difference between the divergences can converge to a limit that corresponds to the rate. Therefore, the divergence rate characterizes the asymptotic evolution of the divergence with respect to k. In particular, when the divergence rate is finite, the divergence grows asymptotically as an affine function of k, whose slope corresponds to the divergence rate. This framework can be then extended by considering the entropy rate, defined as the entropy per unit time, and the cross-entropy rate.

This paper is a complementary study to previous studies dealing with entropies, divergences, and cross-entropies. They focused on 1/ finding closed-form expressions of these statistical measures for different families of pdf [9,10], 2/ ways to estimate these measures from Gaussian or non Gaussian data [11–14] or 3/ determining subsets of similar models by (i) building a matrix composed of Jeffrey's divergences between model pairs, (ii) transforming it to get the same properties as a correlation matrix, and then (iii) performing an eigenvalue decomposition [15].

The entropy rate has been investigated in several papers, including [16]. The divergence rates of stationary Gaussian processes have been studied initially in [17], but also in [18]. In these contributions, the authors derive closed-form expressions for the KL divergence rate and the Rényi divergence rate of zero-mean Gaussian processes, respectively. Their analyses rely on the theory of Toeplitz matrices and on the asymptotic distribution of the eigenvalues of Toeplitz forms. The resulting expressions are given as integrals involving functions of the power spectral densities of the processes under comparison. More recently, the rates of the Jeffreys, KL or Rényi divergence for Gaussian scalar short-memory processes such as autoregressive moving average (ARMA) processes (with parameters varying over time or not) and long-memory processes such as AR fractionally integrated MA (ARFIMA) processes were obtained in [19–22] by exploiting both the expressions of the divergence in the Gaussian case and the statistical properties of the stochastic processes. The analysis was done for the noisy and noise-free cases. The resulting divergence rates depend on the process parameters. Note that among the other results that were presented, it was shown that the KL divergence rate corresponds to the Itakura–Saito distance, up to a multiplicative factor. Divergences to compare sums of complex exponentials disturbed by additive white noises were addressed in [23]. More recently [24], we focused our attention on first-order MA or AR processes. We proposed to (i) introduce a multiscale approach such as coarse-graining to get different signals from the original ones and (ii) to derive the entropy rates of the resulting processes. Note that coarse-graining was initially used to obtain robust information of a dynamical system, it consists of applying an operator like rounding, min, max, and averaging on the signal. Combined with an average filtering, it was used by Costa [25] to define the multiscale entropy in order to evaluate the complexity of time-series. Consequently, the stochastic signal can be characterized by a set of markers

https://doi.org/10.3390/cmsf2026016015

(instead of one), making comparison between stochastic processes "easier". This concept was extended to the multiscale cross-entropy rate [26].

Although the derivation of the rates for stochastic processes such as ARMA and ARFIMA is more or less straightforward, the analysis of other classes of random processes—whose time-evolution may depend on other stochastic processes—may be more "difficult". In such cases, a state-space representation can prove to be a useful framework for addressing these difficulties. This is the reason why this paper addresses a recursive way to get the cross-entropy rate using a state-space framework.

In the following, we first recall the definition of the linear state-space representation of a system under Gaussian assumptions, with illustrations based on motion models. We then show how to recursively derive expressions for the entropy, KL divergence, and cross-entropy, and study the conditions under which finite rates can be obtained. Finally, some illustrative examples are presented.

2. Preamble: Discrete-Time State-Space Representation of Some Motion Models

When describing a system in a state-space framework, the state vector $\underline{x}$ is defined as the minimum set of variables, called state variables, that are required to predict the future evolution of the system given the inputs. The state vector evolves according to a state equation, which models the system dynamics. In the following, let us focus on the discrete-time linear Gaussian case defined by

$$\underline{x}(k) = \Phi_{k-1,k}\underline{x}(k-1) + G\underline{u}(k) \tag{1}$$

where G is the gain and $\Phi_{k-1,k}$ is the transition matrix. When its elements do not vary over time, the subscript $k-1,k$ can be removed and the notation $\Phi_{k-1,k}$ can be reduced to Φ. $\underline{u}(k)$ is the model noise, assumed to be Gaussian zero-mean with covariance matrix Q.

Various mathematical models of target motions have been developed over the past three decades. The next three subsections will give examples when dealing with coordinate-uncoupled maneuver motion-models. Then, first-order multivariate AR modeling will be considered for coupled or uncoupled motion models.

2.1. Constant-Velocity Motion Model (CV)

The CV motion model is defined by an acceleration that is a zero-mean memory-less process with variance σ^2. The equation describing the way the state vector $\underline{x}(k)$ evolves for the CV motion model along one dimension is then given by [27]

$$\underline{x}(k) = \begin{bmatrix} x(k) \\ \dot{x}(k) \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x(k-1) \\ \dot{x}(k-1) \end{bmatrix} + \underline{u}(k) \text{ with } Q = E[\underline{u}(k)\underline{u}^T(k)] = \sigma^2 \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix} \tag{2}$$

with T being the sampling period, $\dot{x}(k)$ the velocity at time kT, and $x(k)$ the position at time kT. The transition matrix does not change over time and is equal to $\Phi = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$. Note that the determinant of the model-noise covariance is given by $|Q| = \frac{T^4}{12}\sigma^4$. Moreover, when two different models are considered that differ by the model noise covariance matrix (denoted as Q_1 and Q_2), one has $\frac{1}{2}Tr(Q_2^{-1}Q_1) = \frac{\sigma_1^2}{\sigma_2^2}$, where Tr denotes the trace of a matrix.

The way the motion is modeled along the x-axis has been addressed, but the same reasoning could be applicable for the y-axis. As the coordinates are assumed to be uncoupled, the models can be used separately. As an alternative, one can store the state vectors for the x-axis and the y-axis in an extended state vector.

https://doi.org/10.3390/cmsf2026016015

2.2. Constant-Acceleration Motion Model (CA)

Let us now recall the state-vector update equation for the CA motion model, in which the continuous-time jerk (the derivative of the acceleration) is modeled as a zero-mean, memory-less process with variance σ^2. In that case, one gets

$$\begin{bmatrix} x(k) \\ \dot{x}(k) \\ \ddot{x}(k) \end{bmatrix} = \begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x(k-1) \\ \dot{x}(k-1) \\ \ddot{x}(k-1) \end{bmatrix} + \underline{u}(k) \text{ with } Q = E[\underline{u}(k)\underline{u}^T(k)] = \sigma^2 \begin{bmatrix} \frac{T^5}{20} & \frac{T^4}{8} & \frac{T^3}{6} \\ \frac{T^4}{8} & \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^3}{6} & \frac{T^2}{2} & T \end{bmatrix} \quad (3)$$

with $\ddot{x}(k)$ being the acceleration, $\dot{x}(k)$ the velocity, and $x(k)$ the position at time kT. Once again, the transition matrix remains unchanged over time and only depends on T. Note that $|Q| = \sigma_i^6 \frac{T^9}{8640}$. When dealing with two different models, one has $\frac{1}{2} Tr(Q_2^{-1} Q_1) = \frac{3}{2} \frac{\sigma_1^2}{\sigma_2^2}$.

2.3. Singer Acceleration Motion Model

Let us introduce the state update formulation of the Singer motion model, where the continuous-time acceleration is assumed to be a zero-mean stochastic process with variance σ^2. Its autocorrelation function exponentially decays with a rate equal to $-\frac{1}{\tau_m} = -\alpha$. Note that the Singer model tends to the CA model when α tends to 0. When α is very large, the Singer model tends to the CV model. The influence of α is illustrated in Figure 1, which provides the evolution of the abscissa of trajectories generated by Singer models based on the same model-noise realization with different values of α.

$$\begin{bmatrix} x(k) \\ \dot{x}(k) \\ \ddot{x}(k) \end{bmatrix} = \begin{bmatrix} 1 & T & \phi_{13} \\ 0 & 1 & \phi_{23} \\ 0 & 0 & \phi_{33} \end{bmatrix} \begin{bmatrix} x(k-1) \\ \dot{x}(k-1) \\ \ddot{x}(k-1) \end{bmatrix} + \underline{u}(k) \text{ with } Q = E[\underline{u}(k)\underline{u}^T(k)] = 2\alpha\sigma^2 \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \quad (4)$$

with $\phi_{13} = \frac{1}{\alpha^2}(-1 + \alpha T + \rho)$, $\phi_{23} = \frac{1}{\alpha}(1 - \rho)$, and $\phi_{33} = \rho$, with $\rho = \exp(-\alpha T)$. Moreover, $q_{11} = \frac{1}{2\alpha^5}\Big(1 + 2\alpha T - 2\alpha^2 T^2 + \frac{2}{3}\alpha^3 T^3 - 4\alpha T \exp(-\alpha T) - \exp(-2\alpha T)\Big)$, $q_{12} = q_{21} = \frac{1}{2\alpha^4}\Big(1 - 2\alpha T + \alpha^2 T^2 + (-2 + 2\alpha T)\exp(-\alpha T) + \exp(-2\alpha T)\Big)$, $q_{13} = q_{31} = \frac{1}{2\alpha^3}\Big(1 - 2\alpha T \exp(-\alpha T) - \exp(-2\alpha T)\Big)$, $q_{22} = \frac{1}{2\alpha^3}\Big(-3 + 2\alpha T + 4\exp(-\alpha T) - \exp(-2\alpha T)\Big)$, $q_{23} = q_{32} = \frac{1}{2\alpha^2}\Big(1 - \exp(-\alpha T)\Big)^2$ and $q_{33} = \frac{1}{2\alpha}\Big(1 - \exp(-2\alpha T)\Big)$.

Remark 1. *The transition matrix* $\Phi = \begin{bmatrix} 1 & T & \phi_{13} \\ 0 & 1 & \phi_{23} \\ 0 & 0 & \phi_{33} \end{bmatrix}$ *is unchanged and depends on* T *and* α*. Its eigenvalues are equal to the elements of the diagonal as the matrix is upper triangular. Hence, two eigenvalues are equal to 1 while the third is equal to* $exp(-\alpha T)$*.*

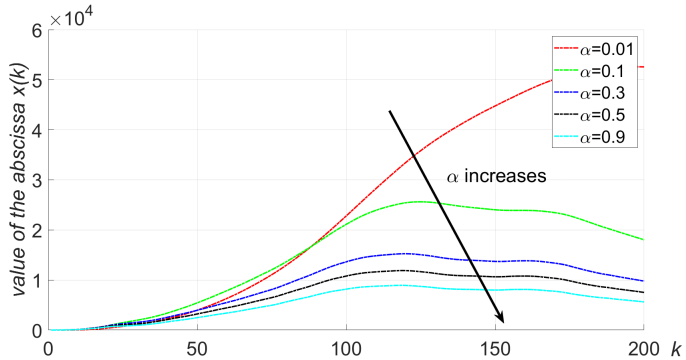


Figure 1. Examples of the evolution of the abscissa of trajectories generated by Singer models based on the same model-noise realization with different values of α.

2.4. First-Order Multivariate-Autoregressive (AR) Process for Uncoupled or Coupled Motion Model

In some cases, the coordinates of the object can be modeled without explicitly considering any assumption on the velocity, the acceleration or the jerk. In a 2D space, the coordinates x and y at time k are stored in the state vector, whose time evolution can be modeled by a first-order multivariate AR process. In that case, one has

$$\begin{bmatrix} x(k) \\ y(k) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x(k-1) \\ y(k-1) \end{bmatrix} + \begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix} \tag{5}$$

where the transition matrix is defined by the set of parameters $\{a_{ij}\}_{i=1,2;j=1,2}$. When $a_{12} = a_{21} = 0$, the coordinates are uncoupled and each process can be seen as an AR process. In that case, if $|a_{ii}|$ is equal or larger than 1, the i^{th} process diverges. When a_{12} and/or a_{21} are different from 0, the coordinates are coupled.

Remark 2. *When applying the stability criterion to the above multivariate AR process (if $\{A_i\}_{i=1,\dots,p}$ denote the AR matrices of the p^{th}-order multivariate AR process defined by $\underline{X}_k = -\sum_{i=1}^{p} A_i \underline{X}_k + \underline{U}_k$, the roots of $|I + \sum_{i=1}^{p} A_i^{-i}| = 0$ must be inside the unit circle in the z-plane), the eigenvalues of the transition matrix must have their modulus strictly smaller than 1. This remark will be useful in the rest of the paper.*

3. Cross-Entropy Based on the Recursive Computation

The entropy and the cross-entropy as well as their rates have been widely used for the classification and comparison of time series and stochastic signals, as they provide criteria to quantify differences in their underlying statistical structures.

In the following, $\underline{x}_{0:k}$ will denote the set of state vectors from time 0 to k. Moreover, $p_1(\underline{x}_{0:k})$ is the pdf of $\underline{x}_{0:k}$ for the first model described in the SSR of the system.

The cross-entropy $H(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k}))$ of the pdf $p_1(\underline{x}_{0:k})$ relative to the pdf $p_2(\underline{x}_{0:k})$ can be expressed as the sum of two terms: the entropy $H(p_1(\underline{x}_{0:k}))$ related to the pdf $p_1(\underline{x}_{0:k})$ and the KL divergence $KL(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k}))$ between $p_1(\underline{x}_{0:k})$ and $p_2(\underline{x}_{0:k})$:

$$H(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k})) = H(p_1(\underline{x}_{0:k})) + KL(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k})) \tag{6}$$

In this section, our purpose is to first derive these quantities recursively when dealing with Gaussian pdfs. Then, we will study the difference between two consecutive measures. We will see if this leads to analytical expressions of the rates.

3.1. Recursive Computation of the Entropy Leading to the Definition of the Entropy Rate

In what follows, let us recursively derive the entropy $H(p_1(\underline{x}_{0:k}))$. This will allow us to deduce the corresponding entropy rate, which corresponds to the difference between $H(p_1(\underline{x}_{0:k}))$ and $H(p_1(\underline{x}_{0:k-1}))$ when k tends to infinity. By definition, one has

$$H(p_1(\underline{x}_{0:k})) = -\int_{\underline{x}_{0:k}} p_1(\underline{x}_{0:k}) \log(p_1(\underline{x}_{0:k})) d\underline{x}_{0:k} \tag{7}$$

As $p_1(\underline{x}_{0:k}) = p_1(\underline{x}_{0:k-1}) p_1(\underline{x}_k / \underline{x}_{k-1})$, (7) can be rewritten as follows:

$$\begin{aligned} H(p_1(\underline{x}_{0:k})) &= -\int_{\underline{x}_{0:k}} p_1(\underline{x}_{0:k}) \log(p_1(\underline{x}_{0:k-1})) d\underline{x}_{0:k} - \int_{\underline{x}_{0:k}} p_1(\underline{x}_{0:k}) \log(p_1(\underline{x}_k / \underline{x}_{k-1})) d\underline{x}_{0:k} \\ &= -\int_{\underline{x}_{0:k}} p_1(\underline{x}_{0:k}) \log(p_1(\underline{x}_k / \underline{x}_{k-1})) d\underline{x}_{0:k} - \int_{\underline{x}_{0:k-1}} p_1(\underline{x}_{0:k-1}) \int_{x_k} p_1(\underline{x}_k / \underline{x}_{k-1}) dx_k \log(p_1(\underline{x}_{0:k-1})) d\underline{x}_{0:k-1} \\ &= H(p_1(\underline{x}_{0:k-1})) - \int_{\underline{x}_{0:k}} p_1(\underline{x}_{0:k}) \log(p_1(\underline{x}_k / \underline{x}_{k-1})) d\underline{x}_{0:k} \end{aligned} \tag{8}$$

At that stage, the second term of (8) can be simplified and one gets

$$\begin{aligned} H(p_1(\underline{x}_{0:k})) &= H(p_1(\underline{x}_{0:k-1})) - \int_{\underline{x}_{k-1:k}} p_1(\underline{x}_{k-1:k}) \log(p_1(\underline{x}_k/\underline{x}_{k-1}))d\underline{x}_{k-1:k} \\ &= H(p_1(\underline{x}_{0:k-1})) - \int_{x_{k-1:k}} p_1(\underline{x}_k/\underline{x}_{k-1})p_1(\underline{x}_{k-1}) \log(p_1(\underline{x}_k/\underline{x}_{k-1}))d\underline{x}_{k-1:k} \\ &= H(p_1(\underline{x}_{0:k-1})) - \int_{x_{k-1}} p_1(\underline{x}_{k-1})\Big(\int_{x_k} p_1(\underline{x}_k/\underline{x}_{k-1}) \log(p_1(\underline{x}_k/\underline{x}_{k-1}))dx_k\Big)d\underline{x}_{k-1} \end{aligned} \tag{9}$$

Based on (9), the difference between the entropies computed for k consecutive time instants and $k-1$ consecutive time instants is given by

$$\Delta H_k = H(p_1(\underline{x}_{0:k})) - H(p_1(\underline{x}_{0:k-1})) = E_{p_1(\underline{x}_{k-1})}[H(p_1(\underline{x}_k/\underline{x}_{k-1}))] \tag{10}$$

where $E_{p_1(\underline{x}_{k-1})}[.]$ denotes the statistical mean with respect to $p_1(\underline{x}_{k-1})$.

In the following, let us consider that $p_1(\underline{x}_k/\underline{x}_{k-1})$ is real and Gaussian and $\underline{x}_k$ is a column vector of dimension m. In that case, when dealing with a linear SSR of the system whose transition matrix is equal to Φ_1, $p_1(\underline{x}_k/\underline{x}_{k-1})$ is defined as follows

$$p_1(\underline{x}_k/\underline{x}_{k-1}) = \frac{1}{(2\pi)^{m/2}|Q_1|^{1/2}} \exp(-\frac{1}{2}(\underline{x}_k - \underline{\mu}_1)^T Q_1^{-1}(\underline{x}_k - \underline{\mu}_1)) \tag{11}$$

where $\underline{\mu}_1 = \Phi_1 \underline{x}_{k-1}$ and Q_1 are the mean and the full-rank covariance matrix characterizing the Gaussian pdf, respectively.

In that case, one can easily show that

$$\begin{aligned} &E_{p_1(\underline{x}_{k-1})}[H(p_1(\underline{x}_k/\underline{x}_{k-1}))] \\ &= \frac{m}{2}\log(2\pi) + \frac{1}{2}\log(|Q_1|) + \frac{1}{2}\int_{\underline{x}_{k-1}} Tr((\underline{x}_k - \Phi_1\underline{x}_{k-1})(\underline{x}_k - \Phi_1\underline{x}_{k-1})^T Q_1^{-1})p_1(\underline{x}_{k-1})d\underline{x}_{k-1} \\ &= \frac{m}{2}\log(2\pi) + \frac{1}{2}\log(|Q_1|) + \frac{1}{2}Tr(Q_1 Q_1^{-1}) \end{aligned} \tag{12}$$

Therefore, by combining (10) and (12), the entropy increment is equal to

$$\Delta H_k = H(p_1(\underline{x}_{0:k})) - H(p_1(\underline{x}_{0:k-1})) = \frac{m}{2}(1 + \log(2\pi)) + \frac{1}{2}\log(|Q_1|) \tag{13}$$

We can notice that this difference is always the same regardless of k. Hence, it is equal to the entropy rate, denoted as ΔH. It only depends on the size of the state vector and the determinant of the model-noise covariance matrix. In other words, the entropy is an affine function with respect to k whose slope is equal to the rate ΔH.

In the next subsection, let us derive the KL divergence.

3.2. Recursive Computation of the KL Divergence Leading to the Definition of the Divergence Rate When It Exists

Let us now focus our attention on the way the KL divergence evolves. To this end, let us first recall its definition:

$$KL(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k})) = \int_{\underline{x}_{0:k}} p_1(\underline{x}_{0:k}) \log \frac{p_1(\underline{x}_{0:k})}{p_2(\underline{x}_{0:k})} d\underline{x}_{0:k} \tag{14}$$

As $p_i(\underline{x}_{0:k}) = p_i(\underline{x}_{0:k-1})p_i(\underline{x}_k/\underline{x}_{k-1})$ with $i = 1, 2$, one gets

$$
\begin{aligned}
&KL(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k})) \\
&= \int_{\underline{x}_{0:k-1}} p_1(\underline{x}_{0:k}) \log \frac{p_1(\underline{x}_{0:k-1})}{p_2(\underline{x}_{0:k-1})} d\underline{x}_{0:k-1} + \int_{\underline{x}_{0:k}} p_1(\underline{x}_{0:k}) \log \frac{p_1(\underline{x}_k/\underline{x}_{k-1})}{p_2(\underline{x}_k/\underline{x}_{k-1})} d\underline{x}_{0:k} \\
&= \int_{\underline{x}_{0:k-1}} p_1(\underline{x}_{0:k-1}) \log \frac{p_1(\underline{x}_{0:k-1})}{p_2(\underline{x}_{0:k-1})} \Big(\int_{x_k} p_1(\underline{x}_k/\underline{x}_{k-1}) dx_k \Big) d\underline{x}_{0:k-1} + \int_{\underline{x}_{0:k}} p_1(\underline{x}_{0:k}) \log \frac{p_1(\underline{x}_k/\underline{x}_{k-1})}{p_2(\underline{x}_k/\underline{x}_{k-1})} d\underline{x}_{0:k} \\
&= \int_{\underline{x}_{0:k-1}} p_1(\underline{x}_{0:k-1}) \log \frac{p_1(\underline{x}_{0:k-1})}{p_2(\underline{x}_{0:k-1})} d\underline{x}_{0:k-1} + \int_{x_{k-1:k}} p_1(\underline{x}_{k-1:k}) \log \frac{p_1(\underline{x}_k/\underline{x}_{k-1})}{p_2(\underline{x}_k/\underline{x}_{k-1})} d\underline{x}_{k-1:k} \\
&= KL(p_1(\underline{x}_{0:k-1}), p_2(\underline{x}_{0:k-1})) + \int_{x_{k-1:k}} p_1(\underline{x}_k/\underline{x}_{k-1}) p_1(\underline{x}_{k-1}) \log \frac{p_1(\underline{x}_k/\underline{x}_{k-1})}{p_2(\underline{x}_k/\underline{x}_{k-1})} d\underline{x}_{k-1:k}
\end{aligned} \tag{15}
$$

In other words, one has

$$
KL(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k})) = KL(p_1(\underline{x}_{0:k-1}), p_2(\underline{x}_{0:k-1})) + E_{p_1(\underline{x}_{k-1})}\Big(KL(p_1(\underline{x}_k/\underline{x}_{k-1}), p_2(\underline{x}_k/\underline{x}_{k-1})) \Big) \tag{16}
$$

At that stage, let us assume that $p_1(\underline{x}_k/\underline{x}_{k-1})$ and $p_2(\underline{x}_k/\underline{x}_{k-1})$ are real and Gaussian and characterized by the means and covariance matrices, namely, $\underline{\mu}_1$ and Q_1 and $\underline{\mu}_2$ and Q_2 and let us express $E_{p_1(\underline{x}_{k-1})}\Big(KL(p_1(\underline{x}_k/\underline{x}_{k-1}), p_2(\underline{x}_k/\underline{x}_{k-1})) \Big)$. One gets

$$
KL(p_1(\underline{x}_k/\underline{x}_{k-1}), p_2(\underline{x}_k/\underline{x}_{k-1}) = \frac{1}{2}\Big(Tr(Q_2^{-1}Q_1) - m - \log \frac{|Q_1|}{|Q_2|} + (\underline{\mu}_2 - \underline{\mu}_1)^T Q_2^{-1} (\underline{\mu}_2 - \underline{\mu}_1) \Big) \tag{17}
$$

where $\underline{\mu}_i = \Phi_i \underline{x}_{k-1}$ when dealing with a linear SSR of the system. Therefore, by introducing $\Delta\Phi = \Phi_2 - \Phi_1$, one gets

$$
\begin{aligned}
KL(p_1(\underline{x}_k/\underline{x}_{k-1}), p_2(\underline{x}_k/\underline{x}_{k-1}) &= \frac{1}{2}\Big(Tr(Q_2^{-1}Q_1) - m - \log \frac{|Q_1|}{|Q_2|} + (\Delta\Phi \underline{x}_{k-1})^T Q_2^{-1} (\Delta\Phi \underline{x}_{k-1}) \Big) \\
&= \frac{1}{2}\Big(Tr(Q_2^{-1}Q_1) - m - \log \frac{|Q_1|}{|Q_2|} + Tr(\underline{x}_{k-1}\underline{x}_{k-1}^T \Delta\Phi^T Q_2^{-1} \Delta\Phi) \Big)
\end{aligned} \tag{18}
$$

This leads to

$$
\begin{aligned}
&E_{p_1(\underline{x}_{k-1})}\Big(KL(p_1(\underline{x}_k/\underline{x}_{k-1}), p_2(\underline{x}_k/\underline{x}_{k-1})) \Big) \\
&= \frac{1}{2}\Big(Tr(Q_2^{-1}Q_1) - m - \log \frac{|Q_1|}{|Q_2|} + \int_{\underline{x}_{k-1}} Tr(\underline{x}_{k-1}\underline{x}_{k-1}^T \Delta\Phi^T Q_2^{-1} \Delta\Phi) p(\underline{x}_{k-1}) d\underline{x}_{k-1} \Big) \\
&= \frac{1}{2}\Big(Tr(Q_2^{-1}Q_1) - m - \log \frac{|Q_1|}{|Q_2|} + Tr(P_{1,k-1} \Delta\Phi^T Q_2^{-1} \Delta\Phi) \Big)
\end{aligned} \tag{19}
$$

where

$$
P_{1,k} = \Phi_1 P_{1,k-1} \Phi_1^T + Q_1 \tag{20}
$$

Let us now combine (16) and (19) to express the difference between the KL divergence from $k-1$ to k time instants:

$$
\begin{aligned}
\Delta KL_k &= KL(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k})) - KL(p_1(\underline{x}_{0:k-1}), p_2(\underline{x}_{0:k-1})) \\
&= \frac{1}{2}\Big(Tr(Q_2^{-1}Q_1) - m - \log \frac{|Q_1|}{|Q_2|} + Tr(P_{1,k-1} \Delta\Phi^T Q_2^{-1} \Delta\Phi) \Big)
\end{aligned} \tag{21}
$$

It should be noted (21) was initially presented in [19]. In this paper, let us now examine this expression in more detail. One of its components, $\frac{1}{2}\Big(Tr(Q_2^{-1}Q_1) - m - \log \frac{|Q_1|}{|Q_2|} \Big)$, is constant and depends on the model-noise covariance matrices and the size of the state vector. The second one varies over time and depends on $P_{1,k-1}$.

There exist cases where ΔKL_k is a constant or tends to a constant when k increases.

- First case ($\Delta\Phi$ is null): The difference ΔKL_k simply reduces to

$$
\Delta KL_k = \frac{1}{2}\Big(Tr(Q_2^{-1}Q_1) - m - \log \frac{|Q_1|}{|Q_2|} \Big) = C^{te} \tag{22}
$$

- Second case (Convergence of $P_{1,k}$): If the sequence $\{P_{1,k}\}$, which is defined in (20), admits a limit denoted by P_1, P_1 satisfies the following discrete-time Lyapunov equation:

$$P_1 = \lim_{k \to +\infty} P_{1,k} = \Phi_1 P_1 \Phi_1^T + Q_1 \tag{23}$$

If the state transition matrix Φ_1 is stable (a sufficient condition is that all its eigenvalues lie strictly inside the unit circle, and it ensures that Φ^k tends to the zero matrix exponentially as k increases), (23) admits a unique symmetric positive-definite solution. It can be expressed in the following analytical form:

$$P_1 = \sum_{k=0}^{+\infty} \Phi_1^k Q_1 (\Phi_1^T)^k \tag{24}$$

The reader can check it is true by substituting (24) in (23). In that case, one gets

$$\Delta KL = \lim_{k \to +\infty} \Delta KL_k = \frac{1}{2}\left(Tr(Q_2^{-1}Q_1) - m - \log \frac{|Q_1|}{|Q_2|} + Tr(P_1 \Delta\Phi^T Q_2^{-1} \Delta\Phi) \right) \tag{25}$$

Remark 3. *For the Singer acceleration model, we show in the Appendix A that the sequence $P_{1,k}$ does not admit a limit. As a consequence, if a Singer model is compared with another Singer model or a CA model, the difference ΔKL_k will not converge to a limit.*

Remark 4. *If one compares two first-order multivariate AR processes, the transition matrices will be potentially different. If they are stable, the Lyapunov equation will admit a solution and a rate exists.*

In the above two cases, the difference is directly equal to or tends to the divergence rate ΔKL, which is finite. The divergence is or tends to an affine function whose slope is ΔKL.

3.3. Deducing the Cross-Entropy and Its Rate When It Exists

Let us now combine (13) and (21) to get the recursive way the cross-entropy evolves:

$$\begin{aligned} &H(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k})) = H(p_1(\underline{x}_{0:k-1})) + KL(p_1(\underline{x}_{0:k-1}), p_2(\underline{x}_{0:k-1})) + \frac{m}{2}(1 + \log(2\pi)) \\ &+ \frac{1}{2}\log(|Q_1|) + \frac{1}{2}\left(Tr(Q_2^{-1}Q_1) - m - \log \frac{|Q_1|}{|Q_2|} + Tr(P_{1,k-1} \Delta\Phi^T Q_2^{-1} \Delta\Phi) \right) \\ &= H(p_1(\underline{x}_{0:k-1}), p_2(\underline{x}_{0:k-1})) + \frac{m}{2}\log(2\pi) + \frac{1}{2}\log(|Q_2|) + \frac{1}{2}\left(Tr(Q_2^{-1}Q_1) + Tr(P_{1,k-1} \Delta\Phi^T Q_2^{-1} \Delta\Phi) \right) \end{aligned} \tag{26}$$

When considering again the particular cases introduced in Section 3.2, one gets:

- First case ($\Delta\Phi$ is null): The following cross-entropy difference can be obtained:

$$\begin{aligned} \Delta H_k(1,2) &= H(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k})) - H(p_1(\underline{x}_{0:k-1}), p_2(\underline{x}_{0:k-1})) \\ &= \frac{m}{2}\log(2\pi) + \frac{1}{2}\log(|Q_2|) + \frac{1}{2} Tr(Q_2^{-1}Q_1) = \Delta H(1,2) \end{aligned} \tag{27}$$

This increment does not depend on k. This means that the cross-entropy, as a function of k, is an affine function whose slope is equal to $\Delta H(1,2)$. Hence, this rate is enough to characterize the cross-entropy behavior.

- Second case (Convergence of $P_{1,k}$): One gets the following cross-entropy difference:

$$\begin{aligned} \Delta H(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k})) &= \lim_{k \to +\infty} \Delta H_k(p_1(\underline{x}_{0:k}), p_2(\underline{x}_{0:k})) \\ &= \frac{m}{2}\log(2\pi) + \frac{1}{2}\log(|Q_2|) + \frac{1}{2}\left(Tr(Q_2^{-1}Q_1) + Tr(P_1 \Delta\Phi^T Q_2^{-1} \Delta\Phi) \right) \end{aligned} \tag{28}$$

Given the above mathematical development, let us now look at some illustrations by considering motion models briefly recalled in Section 2.

4. Illustrations

4.1. Comparing CV Models

This case corresponds to case 1 ($\Delta\Phi$ is null). Therefore, using (27), the cross-entropy rate denoted as $\Delta H_{CV}(1,2)$ exists and is equal to

$$\Delta H_{CV}(1,2) = \log(2\pi) + \log(\frac{T^2}{2\sqrt{3}}) + \log(\sigma_2^2) + \frac{\sigma_1^2}{\sigma_2^2} \tag{29}$$

It only depends on the variances of the continuous-time accelerations of the two models and the sampling period T.

Remark 5. *The rate of the symmetric cross-entropy could be considered and is equal to the sum of the cross-entropy rates $\Delta H(1,2) + \Delta H(2,1)$. Hence, it depends on the logarithms of the continuous-time acceleration variances and on the variance ratios.*

Remark 6. *If one plans to compare a Singer model with a CV model, the problem stands in the dimension of the state vectors, which are not the same. Starting from the largest-size state vector, one idea would be to pre-multiply this vector by a selection matrix S in order to consider only the position and the velocity. $S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. The covariance matrix of the model noise would become SQ_iS^T . The transition matrix would become $S\Phi_i(S^TS)^{-1}S^T$, but $(S^TS)^{-1}$ is not invertible. For this reason, one option is to consider a Singer model, having in mind that when α becomes very large, the model tends to a CV model whereas when α is small and tends to zero, the Singer model tends to a CA model.*

4.2. Comparing CA Models

This also corresponds to case 1. The cross-entropy rate exists and is equal to

$$\Delta H_{CA}(1,2) = \frac{3}{2}\Big(\log(2\pi) + \log(\frac{T^3}{12\sqrt[3]{5}})\Big) + \frac{3}{2}\Big(\log(\sigma_2^2) + \frac{\sigma_1^2}{\sigma_2^2}\Big) \tag{30}$$

In this second simple setting, where the models differ only in terms of the process noise variances, the cross-entropy rate is solely determined by these variances, under the assumption of a common sampling period.

4.3. Comparing First-Order Multi-Variate AR Models

Let us assume that (i) the transition matrix $\Phi_i = \begin{bmatrix} \alpha_i & \gamma_i \\ 0 & \beta_i \end{bmatrix}$ has its eigenvalues α_i and β_i with modulus smaller than 1 (ii) the model-noise covariance matrix $Q_i = \begin{bmatrix} \sigma_{i,1}^2 & 0 \\ 0 & \sigma_{i,2}^2 \end{bmatrix}$ is diagonal and positive definite, with $i = 1, 2$. In that case, $\Delta\Phi = \begin{bmatrix} \alpha_1 - \alpha_2 & \gamma_1 - \gamma_2 \\ 0 & \beta_1 - \beta_2 \end{bmatrix}$. In addition, the solution to (23) exists and it can be shown that its elements denoted as $\{p_{ij}\}_{i=1,2;j=1,2}$ satisfy $p_{11} = \frac{1}{1-\alpha_i^2}\Big(\frac{(1+\alpha_i\beta_i)\gamma_i^2}{(1-\alpha_i\beta_i)(1-\beta_i^2)}\sigma_{i,2}^2 + \sigma_{i,1}^2\Big)$, $p_{22} = \frac{\sigma_{i,2}^2}{1-\beta_i^2}$ and $p_{12} = p_{21} = \frac{\gamma_i\beta_i}{(1-\alpha_i\beta_i)(1-\beta_i^2)}\sigma_{i,2}^2$. Therefore, every element in (28) is defined and the divergence and cross-entropy rates can be deduced. In Figures 2–4, a simple example is considered with $\alpha_1 = \alpha_2 = 0.99$, $\beta_1 = \beta_2 = 0.95$, $\sigma_{1,1}^2 = \sigma_{2,1}^2 = \sigma_{1,2}^2 = \sigma_{2,2}^2 = 10$. In addition, $\gamma_1 = 0.1$, whereas γ_2 takes the values 0.2, 0.5, and 0.8 in order to illustrate how the divergence and cross-entropy, together with their associated rates, change when the multivariate AR process evolves from a weak coupling to a strongly coupled regime, where the first component of the state vector depends heavily

on the second. In Figure 2, examples of trajectories generated from the same realization of the model noise are given for illustration. We can notice that the trajectories differ in the abscissas. At first glance, the reference trajectory (obtained with γ_1) may appear to be stretched along the x-axis when looking at the other trajectories obtained with γ_2. However, this is not entirely accurate. Indeed, let us look at the difference $\delta x(k)$ between the abscissas of the two trajectories, for any value of γ_2. As $\alpha_1 = \alpha_2$, one has

$$\delta x(k) = \alpha_1 \delta x(k) + (\gamma_1 - \gamma_2) y(k-1) \tag{31}$$

As $y(k) = \beta_1 y(k-1) + u_2(k)$ is a 1^{st}-order AR process whose excitation is $u_2(k)$, considering the z-transform of $\delta x(k)$, $y(k)$ and $u_2(k)$ denoted as $\delta X(z)$, $Y(z)$ and $U_2(z)$, one has

$$\frac{\delta X(Z)}{U_2(z)} = \frac{(\gamma_1 - \gamma_2) z^{-1}}{(1 - \alpha_1 z^{-1})(1 - \beta_1 z^{-1})} \tag{32}$$

This transfer function corresponds to a stable delayed 2^{nd}-order AR process $\delta x(k)$ driven by a zero-mean white noise whose variance is equal to $(\gamma_1 - \gamma_2)^2 \sigma_{1,2}^2$. This explains the trajectories that are obtained. As seen in Figures 3 and 4, the KL divergence as well as the cross-entropy rate vary with respect to γ_2. The larger γ_2 is, the larger the rates are and the more dissimilar the models are.

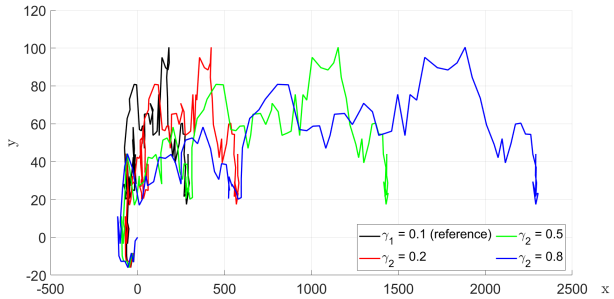


Figure 2. Examples of (x, y) trajectories generated by the four models under study using the same model-noise realization. The models differ solely in the value of γ_i.

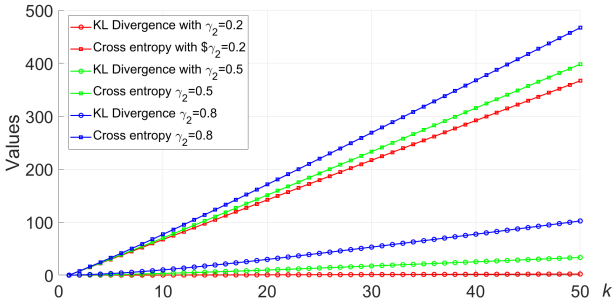


Figure 3. Evolution of the divergence and the cross-entropy with respect to k.

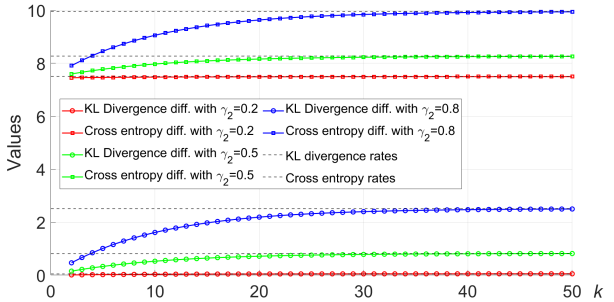


Figure 4. Evolution of the difference in the divergence and the cross-entropy with respect to k compared to corresponding rates.

5. Conclusions

In this paper, we analyzed how entropy, Kullback–Leibler divergence, and the resulting cross-entropy can be derived within a state-space representation framework. We showed that certain conditions must be satisfied for divergence and cross-entropy rates to be well defined. Some model comparison examples, with potential applications to object tracking using interacting multiple models (IMM) in GPS navigation and radar signal processing, were presented for illustrative purposes. If no finite rate exists, model comparison relies on the choice of the number of data points considered. This later will be determined by the constraints of the application. Nevertheless, the proposed methodology can be extended to other application contexts. As a perspective, we plan to study how to deal with nonlinear or non-Gaussian state-space models.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset generated and analyzed during the current study is available from the corresponding author upon reasonable request.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AR	Autoregressive
CA	Constant acceleration
CDKF	Central difference Kalman filter
CV	Constant velocity
EKF	Extended Kalman filter
IMM	Interactive multiple model
KL	Kullback–Leibler
pdf	Probability density function
SPKF	Sigma point Kalman filter
SSR	State-space representation
UKF	Unscented Kalman filter

Appendix A. Singer Case (Optional and Mainly Intended for Completeness)

In (20), if $P_{1,k}$ converges to a limit P_1, the Lyapunov equation that should be considered for the Singer case would be $P_1 = \Phi P_1 + Q_1$. Let us rewrite the transition matrix for the Singer case in a block matrix:

$$\Phi = \begin{bmatrix} 1 & T & \phi_{13} \\ 0 & 1 & \phi_{23} \\ 0 & 0 & \phi_{33} \end{bmatrix} = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \mathbf{0_{2\times 1}} & \phi_{33} \end{bmatrix} \tag{A1}$$

with the matrix blocks $\Phi_{11} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$, $\Phi_{21} = \begin{bmatrix} \phi_{13} \\ \phi_{23} \end{bmatrix}$, Φ_{21} a row null vector and $\Phi_{22} = \phi_{33} = \rho$. Similarly, let us express P_1 and Q_1 in block matrices whose block dimensions are the same as those used to define Φ. The blocks will be denoted as P_{ij} and Q_{ij}, respectively. Note that $Q_{22} = q_{33}$ and $P_{21} = P_{12}^T$. This leads to the following equalities:

$$\begin{cases} P_{22} = \frac{q_{33}}{1-\rho^2} = \frac{(1-\rho^2)}{2\alpha(1-\rho^2)} = \frac{1}{2\alpha} \\ P_{12} = \rho(\phi_{11}P_{12} + Q_{12}P22) + Q_{12} \\ P_{11} = \Phi_{11}P_{11}\Phi_{11}^T + \Phi_{12}P_{12}^T\Phi_{11}^T + \Phi_{11}P_{12}\Phi_{12}^T + P_{22}\Phi_{12}\Phi_{12}^T + Q_{11} = \Phi_{11}P_{11}\Phi_{11}^T + S \end{cases} \tag{A2}$$

with $S = \Phi_{12}P_{12}^T\Phi_{11}^T + \Phi_{11}P_{12}\Phi_{12}^T + P_{22}\Phi_{12}\Phi_{12}^T + Q_{11}$.

Given the above set of equations, we can show that P_{12} can be expressed as follows:

$$P_{12} = \sum_{l=0}^{+\infty} \rho^l \Phi_{11}^l (\rho\Phi_{12}P_{22} + Q_{12}) \tag{A3}$$

The proof consists in substituting the expression (A3) in the second equation given in (A2). By considering that the Froebenius norm of $\Phi_{11}^l = \begin{bmatrix} 1 & lT \\ 0 & 1 \end{bmatrix}$ is equal to $\sqrt{2 + (lT)^2} \leq \sqrt{2} + lT$ and that $||(\rho\Phi_{12}P_{22} + Q_{12})|| = K_1$, the Froebenius norm of P_{12} can be upper-bounded this way:

$$||P_{12}||_F \leq K_1 \sum_{i=0}^{+\infty} \rho^l (\sqrt{2} + lT) < +\infty \tag{A4}$$

If the Froebenius norms of P_{22} and P_{12} are bounded, we are going to see this is not the case for P_{11}. Indeed, if P_{11} exists, it should satisfy as S is symmetric:

$$\begin{aligned} P_{11} &= \sum_{l=0}^{+\infty} \Phi_{11}^l S (\Phi_{11}^T)^l = \sum_{l=0}^{+\infty} \begin{bmatrix} 1 & lT \\ 0 & 1 \end{bmatrix} \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix} \begin{bmatrix} 1 & O \\ lT & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sum_{l=0}^{+\infty} s_{11} + 2s_{12}lT + s_{22}(lT)^2 & \sum_{l=0}^{+\infty} s_{12} + s_{22}lT \\ \sum_{l=0}^{+\infty} s_{12} + s_{22}lT & \sum_{l=0}^{+\infty} s_{22} \end{bmatrix} \end{aligned} \tag{A5}$$

However, the elements of P_{11} correspond to infinite sums that diverge. Therefore, $P_{1,k}$ does not tend to a finite limit P_1.

References

1. Najim, M. *Modeling, Estimation and Optimal Filtering in Signal Processing*; Wiley: Hoboken, NJ, USA, 2010.
2. Bar-Shalom, Y.; Li, X.R.; Kirubarajan, T. *Estimation with Applications to Tracking and Navigation*; Wiley: Hoboken, NJ, USA, 2001.
3. Wan, E.A.; van der Merwe, R. Chapter 7, The Unscented Kalman Filter. In *Adaptive and Learning Systems for Signal Processing, Communications and Control*; Wiley: Hoboken, NJ, USA, 2001.
4. Ito, K.; Xiong, K. Gaussian Filters for Nonlinear Filtering Problems. *IEEE Trans. Autom. Control* **2000**, *5*, 910–927. [CrossRef]
5. Van der Merwe, R.; Wan, E.A.; Julier, S.J. Sigma-point Kalman filters for nonlinear estimation and sensor fusion: Applications to Integrated Navigation. *AIAA-GNC* **2004**, *3*, 1–30.
6. Mazor, E.; Averbuch, A.; Bar-Shalom Y.; Dayan, J. Interacting multiple model methods in target tracking: A survey. *IEEE Trans. Aerosp. Electron. Syst.* **1998**, *34*, 103–123. [CrossRef]
7. Li, X.R.; Bar-Shalom, Y. Multiple-model estimation with variable structure, *IEEE Trans. Autom. Control* **1996**, *41*, 478–493. [CrossRef]
8. Jilkov, V.P.; Rong Li, X. Online Bayesian Estimation of Transition Probabilities for Markovian Jump Systems. *IEEE Trans. Signal Process.* **2004**, *52*, 1620–1630. [CrossRef]
9. Nielsen, F.; Nock, R. A closed-form expression for the Sharma-Mittal entropy of exponential families. *J. Phys. A Math. Theor.* **2012**, *45*, 032003. [CrossRef]
10. Nielsen, F.; Nock, R. On Rényi and Tsallis entropies and divergences for exponential families. *J. Phys. A Math. Theor.* **2012**, *45*, 3.
11. Nguyen, X.; Wainwright, M.J.; Jordan, M.I. Estimating divergence functionals and the likelihood ratio by convex risk minimization. *IEEE Trans. Inf. Theory* **2010**, *56*, 5847–5861. [CrossRef]
12. Contreras-Reyes, J.E.; Arellano-Valle, R.B. Kullback–Leibler divergence measure for multivariate skewnormal distributions. *Entropy* **2012**, *14*, 1606–1626. [CrossRef]
13. Sugiyama, M.; Liu, S.; Christoffel du Plessis, M.; Yamanaka, M.; Suzuki, T.; Kanamori, T. Direct divergence approximation between probability distributions and its applications in machine learning. *J. Comput. Sci. Eng.* **2013**, *7*, 99–111. [CrossRef]
14. Contreras-Reyes, J.E. Analyzing fish condition factor index through skew-gaussian information theory quantifiers. *Fluct. Noise Lett.* **2016**, *15*, 1–16. [CrossRef]

15. Magnant, C.; Grivel, E.; Giremus, A.; Joseph, B.; Ratton, L. Jeffrey's divergence for state-space model comparison. *Elsevier Signal Process.* **2015**, *114*, 61–74. [CrossRef]
16. Regnault, P.; Girardin, V.; Lhote, L. Weighted Closed Form Expressions Based on Escort Distributions for Rényi Entropy Rates of Markov Chains. In *International Conference on Geometric Science of Information*; Springer International Publishing: Cham, Switzerland, 2017.
17. Ihara, S. *Information Theory for Continuous Systems*; World Scientific: Singapore, 1993.
18. Gil, M. On Rényi Divergence Measures for Continuous Alphabet Sources. Ph.D. Thesis, Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada, 2011.
19. Magnant, C.; Giremus, A.; Grivel, E. On Computing Jeffrey's Divergence Between Time-Varying Autoregressive Models. *IEEE Signal Process. Lett.* **2015**, *22*, 915–919. [CrossRef]
20. Saleh, M.; Grivel, E.; Omar, S.-M. Jeffrey's divergence between ARFIMA processes. *Digit. Signal Process.* **2018**, *82*, 175–186 . [CrossRef]
21. Legrand, L.; Grivel, E. Jeffrey's Divergence between Autoregressive Processes Disturbed by Additive White Noises. *Elsevier Signal Process.* **2018**, *149*, 162–178. [CrossRef]
22. Grivel, E.; Diversi, R.; Merchan, F. Kullback- Leibler and Rényi divergence rate for Gaussian stationary ARMA processes comparison. *Elsevier Digit. Signal Process.* **2021**, *116*, 103089. [CrossRef]
23. Grivel, E. Analysis of families of divergences to compare Gaussian processes modeled by sums of complex exponentials disturbed by additive white noises. *Elsevier Digit. Signal Process.* **2022**, *123*, 103436. [CrossRef]
24. Grivel, E.; Legrand, P.; Berthelot, B. Multiscale entropy rates: A study on different stochastic processes. *Elsevier Digit. Signal Process.* **2025**, *165*, 105303. [CrossRef]
25. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102. [CrossRef] [PubMed]
26. Grivel, E.; Legrand, P.; Berthelot, B. Multiscale cross entropy rate as a way to compare signals: Application to Gaussian ARMA processes. In *2025 33rd European Signal Processing Conference (EUSIPCO)*; IEEE: Piscataway, NJ, USA, 2025.
27. Rong Li, X.; Jilkov, V.P. Survey of maneuvering target tracking. Part I. Dynamic models. *IEEE Trans. Aerosp. Electron. Syst.* **2003**, 39, 1333–1364.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.