A mutual information based distance for multivariate Gaussian processes

Jeroen Boets, Katrien De Cock and Bart De Moor
{jeroen.boets,katrien.decock,bart.demoor}@esat.kuleuven.be

***Abstract*— In this paper a distance on the set of multivariate Gaussian linear stochastic processes is proposed based on the concept of mutual information. The definition of the distance is inspired by various properties of the mutual information of past and future of a stochastic process. For two special classes of models a link exists between this mutual information distance and a previously defined scalar cepstral distance. Finally, it is demonstrated that the distance shows similar behavior to an *ad hoc* defined multivariate cepstral distance.**

I. INTRODUCTION

In this paper a distance based on mutual information is defined on the set of multivariate Gaussian linear stochastic processes. By considering such a process as an infinite-dimensional random variable it is possible to define distances based on information-theoretic measures, e.g. as the (asymptotic) Kullback-Leibler (K-L) divergence, Chernoff divergence or Bhattacharyya divergence of the two processes [8], [9]. Mutual information is an information-theoretic measure too. However, it is not applicable in the same sense as the above measures since the mutual information of two random variables does not measure the similarity (or dissimilarity) of their probability densities. Instead it is a measure for the *dependence* of two random variables. Inspired by previous work in [2], [4] we explain further on in the text how we achieve from this a distance on the set of stochastic processes (without assuming information on their mutual dependencies).

The paper is organized as follows. In Section II we describe the model class we work with: Gaussian linear stochastic models. Section III recalls the notion of mutual information of two random variables, and applies this in the context of stochastic processes. In Section IV a distance between multivariate Gaussian processes based on the concept of mutual information is defined and its properties are investigated. Section V applies this newly defined distance in two simulation experiments with the aim of better characterizing it. Section VI finally states the conclusions of this paper.

Jeroen Boets, Dr. Katrien De Cock and Prof. Dr. Bart De Moor are with the K.U.Leuven, Dept. of Electrical Engineering (ESAT-SCD), Kasteelpark Arenberg 10, B-3001 Leuven, Belgium. Jeroen Boets is a research assistant with the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

II. MODEL CLASS

In this paper we consider stationary stochastic processes $y = \{y(k)\}_{k\in\mathbb{Z}}$ whose first and second order statistics can be described by the following state space equations:

$$\left\{ \begin{array}{rcl} x(k+1) &=& Ax(k) + Bu(k)\ , \\ y(k) &=& Cx(k) + Du(k)\ , \end{array} \right. \quad (1)$$

$$E\{u(k)\} = 0\ , \quad E\{u(k)u^\top(l)\} = I_p\delta_{kl}\ . \quad (2)$$

where $y(k) \in \mathbb{R}^p$ is the value of the process at time k and is called the output of the model (1)-(2). The state process $\{x(k)\}_{k\in\mathbb{Z}} \in \mathbb{R}^n$ and the stationary and ergodic (normalized) white noise process $\{u(k)\}_{k\in\mathbb{Z}} \in \mathbb{R}^p$ are auxiliary processes used to describe the process y in this representation. The matrix $D \in \mathbb{R}^{p\times p}$ is assumed to be of full rank. Unless otherwise stated, we assume throughout this paper that u and consequently also y is a *Gaussian* process. This means that the process y is fully described by (1)-(2).

The model (1) is assumed to be strictly stable and minimum-phase, meaning that its poles (eigenvalues of A) and zeros (eigenvalues of $A - BD^{-1}C$) lie strictly inside the unit circle of the complex plane. The inverse model (from y to u) can then be derived from (1) and is denoted with a subscript $(\cdot)_z$: $(A_z, B_z, C_z, D_z) = (A - BD^{-1}C, BD^{-1}, -D^{-1}C, D^{-1})$.

The infinite controllability and observability matrix of the model (1) are defined as:

$$\begin{array}{rcl} \mathcal{C} &=& \left(B \quad AB \quad A^2B \quad \cdots \right), \\ \Gamma &=& \left(C^\top \quad (CA)^\top \quad (CA^2)^\top \quad \cdots \right)^\top , \end{array}$$

respectively. The model (1) is assumed to be minimal, meaning that $\mathcal{C}$ and Γ are of full rank n. The Gramians corresponding to $\mathcal{C}$ and Γ are the unique and positive definite solution of the controllability and observability Lyapunov equation, respectively:

$$\begin{array}{rcccl} \mathcal{C}\mathcal{C}^\top & = & P & = & APA^\top + BB^\top , \\ \Gamma^\top\Gamma & = & Q & = & A^\top QA + C^\top C . \end{array} \qquad (3)$$

The corresponding quantities for the inverse model are denoted by $\mathcal{C}_z, \Gamma_z, P_z$ and Q_z.

Along with the description (1) and its inverse (A_z, B_z, C_z, D_z), a transfer function can be defined from u to y and from y to u, respectively:

$$h(z) = C(zI - A)^{-1}B + D , \qquad (4)$$

and analogously for $h^{-1}(z)$. Modulo a similarity transformation of the state space model (A, B, C, D) into $(T^{-1}AT, T^{-1}B, CT, D)$ with nonsingular T, there is a one-to-one correspondence between the descriptions (1) and (4). From each of both descriptions, augmented with (2), the second order statistics of the process y can be derived, i.e. its autocovariance sequence

$$\Lambda(s) \;=\; E\left\{y(k)y^\top(k-s)\right\} \;,\quad s \in \mathbb{Z} \;, \qquad (5)$$

or equivalently its spectral density function

$$\Phi(z) = \sum_{s=-\infty}^{+\infty} \Lambda(s)z^{-s} = h(z)h^\top(z^{-1}) \;. \qquad (6)$$

Gaussian processes (which we assume) are fully described by their first and second order statistical properties. Therefore a zero-mean process y is fully described by (5) or (6). From (6) it can thus be seen that $h(z)$ is not uniquely defined for the process y since the transfer functions $h(z)$ and $h(z)V$ with unitary $V \in \mathbb{R}^{p\times p}$ correspond to the same spectral density function $\Phi(z)$. This is the only non-uniqueness in $h(z)$ under the given assumptions and must be kept in mind while we denote a process in this paper by *one of its* foursomes (A, B, C, D) or *one of its* transfer functions $h(z)$.

Next to the process $y = \{y(k)\}_{k\in\mathbb{Z}}$ we also define the processes $y_p = \{y(-k)\}_{k\in\mathbb{N}_0}$ and $y_f = \{y(k)\}_{k\in\mathbb{N}}$, where the subscript p stands for 'past' and f for 'future', and analogously the past and future processes u_p and u_f of the process u.

III. Mutual Information of Past and Future of a Process

A. Definition of mutual information

Definition 3.1: Let V and W be random variables with joint probability density function $f(v, w)$ and marginal densities $f_V(v)$ and $f_W(w)$, respectively. Then, the *mutual information* of V and W is defined as

$$I(V;W) = \iint f(v,w) \log \frac{f(v,w)}{f_V(v)f_W(w)}\, dv\, dw \;,$$

if the integral exists.

The mutual information of two variables is a measure for the amount of information one variable contains about the other and is always nonnegative. In the case of two zero-mean jointly Gaussian random variables V and W, $I(V;W)$ is related to the *canonical correlations* [5] of V and W, here denoted by σ_k $(k = 1, \ldots, \min(p, q))$ [6]:

$$I(V;W) = -\frac{1}{2} \log \prod_{k=1}^{\min(p,q)} (1 - \sigma_k^2) \;. \qquad (7)$$

B. Mutual information of past and future of a process

In this section we apply the notion of mutual information to the stochastic processes y_p, y_f, u_p and u_f. A stochastic process, e.g. $\{y(k)\}_{k\in\mathbb{Z}}$, can be seen as an infinite-dimensional random variable consisting of the (ordered) concatenation of the random variables $\ldots$, $y(-2)$, $y(-1)$, $y(0)$, $y(1)$, $\ldots$ This way we can compute the mutual information for any pair of these processes.

A pair of processes that have at least one canonical correlation equal to 1 can be seen from (7) not to have a finite amount of mutual information. Conversely, processes that are orthogonal to each other (all canonical correlations equal to 0) have mutual information equal to zero. This is for instance the case for u_p and u_f. However, processing this white noise u through the filter $h(z)$ (in general) introduces a time dependence in the resulting process y, which appears as a certain amount of mutual information between its past y_p and future y_f, denoted interchangeably by I_{pf}, $I_{\mathrm{pf}}\{y\}$ or $I_{\mathrm{pf}}\{h(z)\}$, equal to [3, Chap. 4]:

$$I_{\mathrm{pf}} = I(y_p; y_f) = \frac{1}{2} \log\det\left(I_n + Q_z P\right) \;. \qquad (8)$$

This and the values of the mutual information of the other pairs of processes can be found in Table I. Note that I_{pf} is unique for a given stochastic process. So if we write $I_{\mathrm{pf}}\{h(z)\}$ this must be understood as a characteristic of the *process* with spectral density $\Phi(z) = h(z)h^\top(z^{-1})$.

C. Properties of I_{pf}

We now state some of the properties of I_{pf}. Most of these can be proved using (7) and (8).

(a) $I_{\mathrm{pf}} = 0 \Leftrightarrow h(z) = D$ (see (1)).

TABLE I

THE MUTUAL INFORMATION OF EACH PAIR OF PROCESSES.

	u_p	u_f	y_p	y_f
u_p	$+\infty$	0	$+\infty$	I_{pf}
u_f	0	$+\infty$	0	$+\infty$
y_p	$+\infty$	0	$+\infty$	I_{pf}
y_f	I_{pf}	$+\infty$	I_{pf}	$+\infty$

(b) I_{pf} increases with each increase of a canonical correlation between y_p and y_f.

(c) $I_{pf}\{h(z)\} = I_{pf}\{Th(z)\}$ for a nonsingular constant matrix $T \in \mathbb{R}^{p\times p}$.

(d) $I_{pf}\{h(z)\} = I_{pf}\{h^{-\top}(z)\}$.

(e) For $\Phi(z) = \begin{pmatrix} \Phi_1(z) & 0_{p_1\times p_2} \\ 0_{p_2\times p_1} & \Phi_2(z) \end{pmatrix}$, it holds that $I_{pf}\{y\} = I_{pf}\{y_1\} + I_{pf}\{y_2\}$.

Properties (a)-(b) and relation (7) indicate that I_{pf} measures the amount of correlation that exists between y_p and y_f, being zero for a white noise process and increasing with each increase of a correlation between y_p and y_f. This suggests that I_{pf} can be used as a measure for the amount of *dynamics* in the process y.

IV. A DISTANCE BETWEEN MULTIVARIATE GAUSSIAN PROCESSES

In Section IV-A we define a distance between multivariate Gaussian processes based on the concept of mutual information. In Section IV-B a way to compute the distance is shown, while Sections IV-C and IV-D treat two special classes of stochastic processes, for which the distance can be shown to possess some additional properties.

A. Definition and metric properties

Definition 4.1: The *mutual information distance* between two p-variate Gaussian linear stochastic processes y_1 and y_2 with transfer function descriptions $h_1(z)$ and $h_2(z)$ is denoted by $d_{mi}(y_1, y_2)$ and is defined as the non-negative square root of:

$$d^2_{mi}(y_1, y_2) = I_{pf}\{h_{12}(z)\}, \text{ with}$$

$$h_{12}(z) = \begin{pmatrix} h_1^{-1}(z)h_2(z) & 0_p \\ 0_p & h_2^{-1}(z)h_1(z) \end{pmatrix}.$$

The first thing to note is that the mutual information distance $d_{mi}(y_1, y_2)$ is a property of the processes y_1 and y_2, and not of the particular transfer functions $h_1(z)$ and $h_2(z)$. Indeed, substituting $\{h_1(z), h_2(z)\}$ by the equivalent $\{h_1(z)V_1, h_2(z)V_2\}$ with V_1, V_2 constant unitary matrices (see (6)), corresponds to left- and right-multiplying $h_{12}(z)$ by a constant unitary matrix. This has no influence on $I_{pf}\{h_{12}(z)\}$.

Following the discussion in Section III-C, $d_{mi}(y_1, y_2)$ is a measure for the amount of dynamics in the process y_{12} associated with the transfer function $h_{12}(z)$. Clearly $d_{mi}\{y_1, y_1\} = 0$ since $h_{12}(z)$ is in that case a constant matrix and y_{12} is consequently white noise. Definition 4.1 also implies that $d_{mi}(y_1, y_2) = d_{mi}(g(z)y_1, g(z)y_2)$ for arbitrary transfer functions $g(z)$ satisfying the conditions stated in Section II.

The following properties hold for d_{mi}:

1) $d_{mi}(y_1, y_2) \geq 0$ (non-negativity)

2) $d_{mi}(y_1, y_2) = 0 \Leftrightarrow h_2(z) = h_1(z)T$ with T a constant square nonsingular matrix.

3) $d_{mi}(y_1, y_2) = d_{mi}(y_2, y_1)$ (symmetry)

As a consequence of property 2), the distance d_{mi} does not in general satisfy the triangle inequality[1]. Consider three processes with transfer functions $h_1(z)$, $h_2(z)$ and $h_3(z)$ where $h_2(z) = h_1(z)T$ with T a constant square nonsingular matrix. Clearly, $d_{mi}(y_1, y_2) = 0$. Therefore, in order for the triangle inequality to hold, it should generally be satisfied that $d_{mi}(y_1, y_3) = d_{mi}(y_2, y_3)$. This is however not the case. The distance thus satisfies only two of the four properties of a true metric (non-negativity and symmetry).

B. Computation

From property (e) in Section III-C it follows that

$$d^2_{mi}(y_1, y_2) = I_{pf}\{h_1^{-1}h_2\} + I_{pf}\{h_2^{-1}h_1\}. \quad (9)$$

Using this property we now show a way to compute $d_{mi}(y_1, y_2)$ making use of the state space descriptions of $h_1(z)$ and $h_2(z)$. Equations (8) and (9) show that we need to compute the controllability and observability Gramians of both $h_1^{-1}(z)h_2(z)$ and $h_2^{-1}(z)h_1(z)$. This can be easily done by solving the Lyapunov equations (3) from the state space descriptions of both transfer functions. As an example we give a possible state space description of $h_1^{-1}(z)h_2(z)$ denoted by $(A_{12}, B_{12}, C_{12}, D_{12})$:

$$A_{12} = \begin{pmatrix} A_2 & 0_{n_2\times n_1} \\ B_{z_1}C_2 & A_{z_1} \end{pmatrix}, \quad B_{12} = \begin{pmatrix} B_2 \\ B_{z_1}D_2 \end{pmatrix},$$

$$C_{12} = \begin{pmatrix} D_{z_1}C_2 & C_{z_1} \end{pmatrix}, \qquad D_{12} = D_{z_1}D_2,$$

[1]In the case of scalar processes or processes with diagonal spectral density function $\Phi(z)$, however, it can be shown that the triangle inequality *is* satisfied (see Sections IV-C and IV-D).

with $(A_{z_1}, B_{z_1}, C_{z_1}, D_{z_1}) = (A_1 - B_1 D_1^{-1} C_1, B_1 D_1^{-1}, -D_1^{-1} C_1, D_1^{-1})$. The procedure concerning $h_2^{-1}(z) h_1(z)$ is analogous. Afterwards it remains to compute (9) using (8) and (3).

C. Class of scalar processes

In the case of scalar processes y_1 and y_2 it was proven in [3, Chap. 6] as well as in [7] that:

$$\mathrm{d}_{\mathrm{mi}}^2(y_1, y_2) = \sum_{k=0}^{+\infty} k(c_1(k) - c_2(k))^2 = \mathrm{d}_{\mathrm{cep}}^2(y_1, y_2) , \quad (10)$$

with c_1 and c_2 the *cepstra* of the scalar processes y_1 and y_2 (see Section V-A for the definition of cepstrum). This constitutes a link between d_{mi} and the cepstral distance defined in [10], and obviously proves that d_{mi} for scalar processes satisfies the triangle inequality. If we further define a set of equivalence classes of scalar stochastic processes, where two processes y_1 and y_2 with transfer functions $h_1(z)$ and $h_2(z)$ are equivalent if and only if there exists a constant real non-zero number a such that $h_2(z) = a h_1(z)$, then the mutual information distance d_{mi} defined on this set of equivalence classes is a true metric. Various expressions for $\mathrm{d}_{\mathrm{mi}}(y_1, y_2)$ can be derived for the scalar case [3].

D. Class of uncorrelated multivariate processes

In the case of two p-variate processes y_1 and y_2 with diagonal spectral density matrices $\Phi_1(z)$ and $\Phi_2(z)$, it is easily shown that:

$$\mathrm{d}_{\mathrm{mi}}^2(y_1, y_2) = \sum_{i=1}^{p} \mathrm{d}_{\mathrm{cep}}^2(y_{1,i}, y_{2,i}) , \quad (11)$$

where $y_{1,i}$ $(i = 1, \ldots, p)$ are the uncorrelated scalar processes constituting y_1, and analogously for $y_{2,i}$ $(i = 1, \ldots, p)$. This equality follows from relation (9), property (e) in Section III-C and relation (10) for scalar processes.

V. Simulation experiments

For scalar processes, several simulation experiments were performed in [1] in order to compare the behavior of the cepstral distance $\mathrm{d}_{\mathrm{cep}}$, which was equal to d_{mi} because of (10), with the behavior of the H_2 distance, denoted by $\mathrm{d}_{\mathrm{h}_2}$:

$$\mathrm{d}_{\mathrm{h}_2}^2(h_1, h_2) = \frac{1}{2\pi} \int_0^{2\pi} \|h_1(e^{j\theta}) - h_2(e^{j\theta})\|_{\mathrm{F}}^2 d\theta ,$$

with $\|\cdot\|_{\mathrm{F}}$ the Frobenius norm of a matrix. In order to make $\mathrm{d}_{\mathrm{h}_2}$ a distance between processes instead of between transfer functions, we agree to fix the transfer function description of a stochastic process by choosing the D-matrix of a model (1) or (4) to be D_{chol}, the unique Cholesky factor of $DD^\top$.

In this section we demonstrate experimentally that also for multivariate processes d_{mi} has a *cepstral character*, although no relation as in (10) or (11) holds for general multivariate processes. To this aim we focus on two aspects that showed in the scalar case a difference in behavior between the cepstral distance and the H_2 distance:

- The inflation of distance values when poles of the models are approaching the unit circle (Section V-B).
- The difference in influence of pole versus zero locations on the distance between two models (Section V-C).

What is however first needed is the definition of a *multivariate* weighted cepstral distance in Section V-A, which we shall denote by $\mathrm{d}_{\mathrm{cep}}$.

A. Multivariate cepstral distance

We define the cepstrum of a multivariate process y as the inverse Fourier transform of the *matrix* logarithm of the spectrum of y:

$$\log \Phi(e^{j\theta}) = \sum_{k=-\infty}^{+\infty} c(k) e^{-jk\theta} ,$$

where $c(k) \in \mathbb{R}^{p \times p}$ is the kth cepstral coefficient of y. For scalar processes this corresponds to the usual definition of the power cepstrum of a process [11]. The sequence $\{c(k)\}_{k \in \mathbb{Z}}$ is real and even. It obviously contains the same information as $\Phi(z)$ and thus also fully characterizes the zero-mean Gaussian process y. Only for the case of scalar processes, analytical expressions for these cepstral coefficients are known to us.

We now define in analogy with (10) a multivariate weighted cepstral distance as

$$\mathrm{d}_{\mathrm{cep}}^2(y_1, y_2) = \sum_{k=0}^{+\infty} k \|c_1(k) - c_2(k)\|_{\mathrm{F}}^2 , \quad (12)$$

with c_1 and c_2 the cepstra of the multivariate processes y_1 and y_2. For scalar processes this distance coincides with the previously defined distance (10). No relation with the mutual information distance as in (10) for scalar processes holds for multivariate processes, except for diagonal $\Phi_1(z), \Phi_2(z)$ (see Section IV-D) where we can rewrite (11) as:

$$\mathrm{d}_{\mathrm{mi}}^2(y_1, y_2) = \mathrm{d}_{\mathrm{cep}}^2(y_1, y_2) .$$

B. An experiment on the influence of poles

For scalar processes it was demonstrated experimentally in [1] that $\mathrm{d}_{\mathrm{h}_2}$ is more sensitive to high peaks in the spectrum of $h_i(z)$ than cepstral distances. In this section we investigate similar behavior for the case of multivariate processes.

Consider two 3-by-3 stochastic models $h_1(z)$ and $h_2(z)$ of order 2 ($p_1 = p_2 = 3$, $n_1 = n_2 = 2$), both of which have a pair of zeros at radius 0.05 and angles $\pm\pi/2$ in the complex plane. The pole pair of $h_1(z)$ is situated at angles $\pm 0.22\pi$, the pole pair of $h_2(z)$ at angles $\pm 0.78\pi$ in the complex plane. The radii of the pole pairs of $h_1(z)$ and $h_2(z)$ are always taken equal and are varied from $r = 0.10$ to $r = 0.99$. This is schematically presented in Table II. The idea of this setting is to check the influence of the *radii of the poles* on the distance between $h_1(z)$ and $h_2(z)$.

TABLE II

POLES AND ZEROS OF $h_1(z)$ AND $h_2(z)$ FOR THE POLES EXPERIMENT (LOCATED AT $r_i e^{\pm j\phi_i}$).

pole$\{h_1(z)\}$	$r_1 = 0.10, \ldots, 0.99$	$\phi_1 = 0.22\pi$
pole$\{h_2(z)\}$	$r_2 = r_1$	$\phi_2 = 0.78\pi$
zero$\{h_1(z)\}$	$r_1 = 0.05$	$\phi_1 = \frac{\pi}{2}$
zero$\{h_2(z)\}$	$r_2 = 0.05$	$\phi_2 = \frac{\pi}{2}$

Since a multivariate model is characterized by more than only its poles and zeros, we generate for each pole radius between $r = 0.10$ and $r = 0.99$, 100 different pairs of models $\{h_1(z), h_2(z)\}$, both always with fixed pole and zero pair as specified in Table II. For each pair of models the distances d_{mi}, d_{cep} and d_{h_2} between $h_1(z)$ and $h_2(z)$ are computed. The median distance value of the 100 repeats is shown in Fig. 1 for the three distances and for varying pole radius. All values have been rescaled to start at a distance 1 for $r = 0.10$.

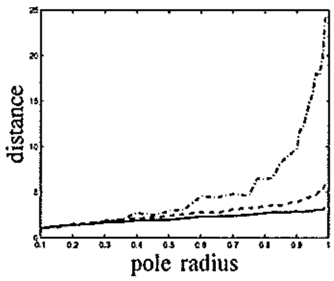

Fig. 1. Evolution of the distance between $h_1(z)$ and $h_2(z)$ as the radius of their pole pairs increases (see Table II), for d_{mi} (full line), d_{cep} (dashed line) and d_{h_2} (dash-dotted line).

The figure shows that d_{h_2} grows much more strongly than the other distances as the pole pairs of $h_1(z)$ and $h_2(z)$ approach the unit circle.

C. *An experiment on the influence of poles & zeros*

In this section we compare poles approaching the unit circle with zeros approaching the unit circle, concerning the influence on the distance between two models. For scalar processes it can be shown that cepstral distances are equally dependent on the poles and zeros of the models.

The experimental setting now consists of two parts. Consider again two 3-by-3 stochastic models $h_1(z)$ and $h_2(z)$ of order 2. For the first part, the location of the poles and zeros of $h_1(z)$ and $h_2(z)$ can be viewed schematically in Table III. It is seen that ϕ_1 is quite close to ϕ_2 for both the pole pair and the zero pair and that the radius of the pole pair of $h_2(z)$ is varied between 0.50 and 0.99.

TABLE III

POLES AND ZEROS OF $h_1(z)$ AND $h_2(z)$ FOR THE FIRST PART OF THE 'POLE VERSUS ZERO' EXPERIMENT.

pole$\{h_1(z)\}$	$r_1 = 0.50$	$\phi_1 = 0.22\pi$
pole$\{h_2(z)\}$	$r_2 = 0.50, \ldots, 0.99$	$\phi_2 = 0.17\pi$
zero$\{h_1(z)\}$	$r_1 = 0.50$	$\phi_1 = 0.78\pi$
zero$\{h_2(z)\}$	$r_2 = 0.50$	$\phi_2 = 0.83\pi$

The location of poles and zeros for the second part of the experiment is shown in Table IV. The only difference with Table III is that now the radius of the *zero* pair (instead of the pole pair) of $h_2(z)$ is varied between 0.50 and 0.99.

TABLE IV

POLES AND ZEROS OF $h_1(z)$ AND $h_2(z)$ FOR THE SECOND PART OF THE 'POLE VERSUS ZERO' EXPERIMENT.

pole$\{h_1(z)\}$	$r_1 = 0.50$	$\phi_1 = 0.22\pi$
pole$\{h_2(z)\}$	$r_2 = 0.50$	$\phi_2 = 0.17\pi$
zero$\{h_1(z)\}$	$r_1 = 0.50$	$\phi_1 = 0.78\pi$
zero$\{h_2(z)\}$	$r_2 = 0.50, \ldots, 0.99$	$\phi_2 = 0.83\pi$

The idea of this experimental setting with two parts is to check the difference between the first and the second part, i.e. the difference in how the distance between $h_1(z)$ and $h_2(z)$ evolves as the radius of pole$\{h_2(z)\}$ versus the radius of zero$\{h_2(z)\}$ is varied between $r_2 = 0.50$ and $r_2 = 0.99$. Just as in the previous section, for each of these radii 100 different pairs of models $\{h_1(z), h_2(z)\}$ are generated, both always with fixed pole and zero pair as specified in Table III or IV. For each pair of models the distances d_{mi}, d_{cep} and d_{h_2} between $h_1(z)$ and $h_2(z)$ are computed and the median distance value of the 100 repeats is selected for each of the three distances.

Fig. 2 shows the results for the distance d_{h_2}. There are two curves in the figure, one for each part of the experiment. The values of both curves have been rescaled by a common factor to have a distance 1 when the radius of zero $\{h_2(z)\}$ in the second part of the experiment is 0.50. The influence on d_{h_2} of pole $\{h_2(z)\}$ approaching the unit circle is much greater than the influence of zero $\{h_2(z)\}$ approaching the unit circle.

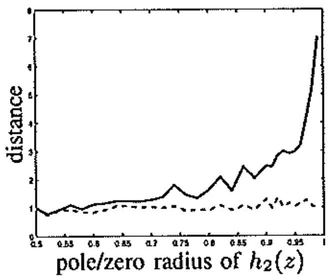

Fig. 2. Evolution of $d_{h_2}(h_1(z), h_2(z))$ as the radius of pole $\{h_2(z)\}$ increases (full line), and as the radius of zero $\{h_2(z)\}$ increases (dashed line) (see Tables III and IV).

Fig. 3 shows in an analogous way the result for the distances d_{mi} and d_{cep}. It demonstrates a very similar influence of pole $\{h_2(z)\}$ approaching the unit circle and zero $\{h_2(z)\}$ approaching the unit circle, on both d_{mi} and d_{cep}.

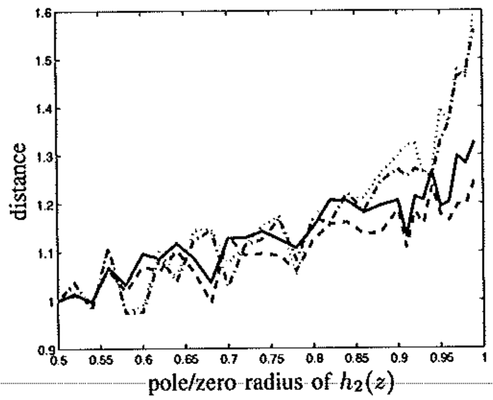

Fig. 3. Evolution of the distance between $h_1(z)$ and $h_2(z)$ as the radius of pole $\{h_2(z)\}$ increases, for d_{mi} (full line) and d_{cep} (dash-dotted line), and as the radius of zero $\{h_2(z)\}$ increases, for d_{mi} (dashed line) and d_{cep} (dotted line) (see Tables III and IV).

With the conclusions of this and the previous experiment we do not claim that one of the distances is *better* than the others. We only wish to point out some differences between them. On the basis of these differences one can choose which distance to use in a specific application.

VI. Conclusions

In this paper we defined the mutual information distance d_{mi} on the set of Gaussian linear stochastic processes, based on the concept of mutual information of past and future of a stochastic process and inspired by the various properties of this notion. We demonstrated how it can be computed from the state space description of the models and investigated its metric properties. For two special classes of stochastic models a link exists between the mutual information distance and a previously defined scalar cepstral distance.

In two simulation experiments d_{mi} was shown to have a behavior similar to an *ad hoc* defined multivariate cepstral distance d_{cep} and dissimilar from the H_2 distance d_{h_2}.

Acknowledgments

Research is supported by: Research Council KUL: GOA AMBioRICS, CoE EF/05/006 Optimization in Engineering(OPTEC), IOF-SCORES4CHEM, several PhD/postdoc & fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects G.0452.04 (new quantum algorithms), G.0499.04 (Statistics), G.0211.05 (Nonlinear), G.0226.06 (cooperative systems and optimization), G.0321.06 (Tensors), G.0302.07 (SVM/Kernel, research communities (ICCoS, ANMMM, MLDM); IWT: PhD Grants, McKnow-E, Eureka-Flite+; Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011); EU: ERNSI.

References

[1] J. Boets, K. De Cock, and B. De Moor. Distances between dynamical models for clustering time series. In *Proc. of the 14th IFAC Symp. on System Identification (SYSID 2006)*, pages 392–397, Newcastle, Australia, 2006.

[2] J. Boets, K. De Cock, M. Espinoza, and B. De Moor. Clustering time series, subspace identification and cepstral distances. *Communications in Information and Systems*, 5(1):69–96, 2005.

[3] K. De Cock. *Principal Angles in System Theory, Information theory and Signal Processing*. PhD thesis, K.U.Leuven, Leuven, Belgium, May 2002.

[4] K. De Cock and B. De Moor. Subspace angles between ARMA models. *Systems & Control Letters*, 46(4):265–270, July 2002.

[5] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–372, 1936.

[6] I. A. Ibragimov and Y. A. Rozanov. *Gaussian Random Processes*. Springer, New York, 1978.

[7] N. P. Jewell, P. Bloomfield, and F. C. Bartmann. Canonical correlations of past and future for time series: bounds and computation. *The Annals of Statistics*, 11(3):848–855, 1983.

[8] Y. Kakizawa, R. H. Shumway, and M. Taniguchi. Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association*, 93:328–340, 1998.

[9] D. Kazakos and P. Papantoni-Kazakos. Spectral distance measures between Gaussian processes. *IEEE Trans. on Automatic Control*, 25(5):950–959, 1980.

[10] R. J. Martin. A metric for ARMA processes. *IEEE Trans. on Signal Processing*, 48(4):1164–1170, 2000.

[11] A. V. Oppenheim and R. W. Schafer. *Digital Signal Processing*. Prentice Hall International, London, 1975.