Preprints of the 18th IFAC World Congress
Milano (Italy) August 28 - September 2, 2011

Quantitative Measure of Observability for Stochastic Systems

Yüksel SUBAŞI*, Mübeccel DEMİREKLER**

**Turkey Scientific and Technological Research Agency-Space Technologies Research Institute, Ankara, TURKEY (Tel: +90-312-210-1310/1200; e-mail: yuksel.subasi@uzay.tubitak.gov.tr)*
*** Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, TURKEY (e-mail: demirek@eee.metu.edu.tr).*

Abstract: The observability measure based on the mutual information between state and/or state sequence and measurements, originally proposed by Mohler and Hwang (1988), is analyzed in detail and improved further for linear time invariant Gaussian systems. It is shown that unobservable states of the deterministic system have no effect on this measure and any observable part with no measurement uncertainty makes it infinite. Utilizing this fact, the observability measure is extended to subspaces of the state space.

Bhattacharyya and Hellinger distances are also analyzed as observability measures. The relationships between the observability measures and Kalman filter covariance matrix are investigated.

Keywords: Observability, observability measure, probabilistic distance measures, mutual information.

1. INTRODUCTION

For deterministic systems observability is handled by determining the rank of the observability Gramian. The output of the process is binary, the system is either completely observable or not and the process does not say anything about the degree of the observability. The aim of this paper is to answer the question of how much observable a stochastic system is and to give an observability measure.

Quantitative measures for observability of deterministic systems are first proposed by Müller and Weber (1972). They propose three observability measures as determinant, trace, and maximal eigenvalue of the inverse of observability Gramian. Several observability measures, which are related to time and frequency domain characteristics of a linear multivariable deterministic system, are proposed by Tarokh (1992).

Mode observability is examined for a deterministic linear multivariable system by Porter and Crossley (1970). Measures of mode observability are proposed for a deterministic system by Lindner et al. (1989) and Hamdan and Nayfeh (1989).

Two stochastic observability definitions, strict sense and wide sense observability, are proposed by Aoki (1967). Similar definitions are given by Han-Fu (1980) and Bageshwar et al. (2009). Bageshwar et al. (2009) state that a necessary condition for the stochastic observability is the observability of the deterministic system, and give upper and lower bounds for the covariance matrix of the conditional probability density function.

Kam et al. (1987) define an observability measure by using the mutual information and entropy concepts. The system is declared as a linear time-invariant stochastic system. However, the only uncertainty is in the initial state and the representation has a deterministic nature. Kam et al. (1987) propose the entropy correlation coefficient as an observability measure for a discrete state system. Chen et al. (2007) extend the results of Kam et al. (1987) to continuous state systems by using the quantized versions of continuous variables.

By Mohler and Hwang (1988) the mutual information between the state and the observation is used as the measure of the observability of a continuous-time stochastic system. They propose an estimated second moment approximation for the observability measure due to the difficulty in solving the density equation. The relationship between the mutual information and Kalman filter covariance is given. The same measure is also proposed for individual states. By Logothetis et al. (1997) the results of Mohler and Hwang (1988) are applied to observer path design for bearings-only tracking. Optimal paths are derived by maximizing the mutual information between the measurement sequence and the final target state or the entire target trajectory.

The goal of this paper is to improve the results of Mohler and Hwang (1988). We have used mutual information as a measure of observability for any stochastic system and then applied this measure to LTI (linear time invariant) discrete-time Gaussian systems to get more concrete results. The mutual information is a special case of the Kullback-Leibler (KL) distance. So other distance measures i.e., Bhattacharyya and Hellinger distances are analyzed as other observability measures. The relationships between the observability measures and Kalman filter covariance matrix are derived. Observability measure of individual state of the system is obtained by using the mutual information. This concept can be extended to the observability of the subspaces of the state space and so to the modes of the system, and also by Bhattacharyya and Hellinger distances.

The paper is organized as follows. In section 2 the system equations are given for a LTI discrete-time Gaussian system, the definitions of mutual information, Kullback-Leibler, Bhattacharyya and Hellinger distances are given, including some basic properties. In section 3, the explicit formulas of the measure of observability are given for all mentioned measures. Furthermore, the results are discussed in detail. In section 4, the relationships between the observability measures and Kalman filter covariance matrix are given. In section 5 observability measure of individual state of the system is given by using the mutual information. Section 6 contains the conclusion remarks.

2. MUTUAL INFORMATION, KULLBACK-LEIBLER, BHATTACHARYYA AND HELLINGER DISTANCES

The system that we have analyzed is represented by the following equations:

$$x_{k+1} = Ax_k + Gw_k \tag{1}$$

$$y_k = Cx_k + Hv_k \tag{2}$$

where, $x_k \in R^n$ is the state of the system at time k, $y_k \in R^m$ is the measurement of the system at time k. A, G, C, H are constant matrices. It is assumed that, the basic random variables are independent identically distributed.

$$x_0 \sim N(\overline{x_0}, \Sigma_0), \quad w_k \sim N(0, Q), \quad v_k \sim N(0, R)$$

Note that $$x_k = A^k x_0 + \sum_{i=0}^{k-1} A^{k-1-i} G w_i \tag{3}$$

and $$y_k = CA^k x_0 + C\sum_{i=0}^{k-1} A^{k-1-i} G w_i + Hv_k \tag{4}$$

Let us define X^k, Y^k, A_k, C_k, H_k, G_k, Q_k and R_k as,

$$X^k = \begin{bmatrix} x_0 \\ \vdots \\ x_k \end{bmatrix} \quad Y^k = \begin{bmatrix} y_0 \\ \vdots \\ y_k \end{bmatrix} \quad A_k = \begin{bmatrix} I \\ \vdots \\ A^k \end{bmatrix} \quad C_k = \begin{bmatrix} C & 0 & \cdots & 0 \\ 0 & C & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C \end{bmatrix}$$

$$H_k = \begin{bmatrix} H & 0 & \cdots & 0 \\ 0 & H & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H \end{bmatrix} \quad G_k = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ G & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A^{k-1}G & A^{k-2}G & \cdots & G \end{bmatrix}$$

$$Q_k = \begin{bmatrix} Q & 0 & \cdots & 0 \\ 0 & Q & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Q \end{bmatrix} \quad R_k = \begin{bmatrix} R & 0 & \cdots & 0 \\ 0 & R & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R \end{bmatrix}$$

so that, X^k and Y^k have also normal densities:

$$X^k \sim N\left(A_k\overline{x_0},\ A_k\Sigma_0{A_k}^T + G_kQ_k{G_k}^T\right) \tag{5}$$

$$Y^k \sim N\left(C_kA_k\overline{x_0},\ C_k\Sigma_{X^k}{C_k}^T + H_kR_k{H_k}^T\right) \tag{6}$$

In addition the covariance matrix for joint probability density function of X^k and Y^k is,

$$\Sigma_{[X^k,Y^k]} = \begin{bmatrix} \Sigma_{X^k} & \Sigma_{X^kY^k} \\ \Sigma_{Y^kX^k} & \Sigma_{Y^k} \end{bmatrix} \tag{7}$$

where $$\Sigma_{X^kY^k} = {\Sigma_{Y^kX^k}}^T = \Sigma_{X^k}{C_k}^T \tag{8}$$

The mutual information $I(X,Y)$ between two continuous random variables with joint density $f(x,y)$ is defined as (Cover and Thomas (2006))

$$I(X,Y) = \iint f(x,y)\log\frac{f(x,y)}{f(x)f(y)}\,dxdy \tag{9}$$

Mutual information is a special case of Kullback-Leibler distance which is defined by

$$D\big(f(x), g(x)\big) = \int f(x)\log\frac{f(x)}{g(x)}\,dx \tag{10}$$

between two densities $f(x)$ and $g(x)$. Kullback-Leibler distance is always nonnegative. However, it is not a true distance since it is not symmetric and it does not satisfy the triangle inequality.

We will define two more popular distances here, the Bhattacharyya and Hellinger distances. Bhattacharyya distance between two probability density functions has the general expression (Basseville (1989)):

$$BD\big(f(x), g(x)\big) = -\log\Big(BC\big(f(x), g(x)\big)\Big); \qquad 0 \le BD \le \infty \tag{11}$$

where, Bhattacharyya Coefficient is defined as follows.

$$BC\big(f(x), g(x)\big) = \int \sqrt{f(x)g(x)}\,dx;\ 0 \le BC \le 1 \tag{12}$$

Bhattacharyya distance is symmetric, however, does not satisfy triangle inequality (Upadhyaya and Sorenson (1977)).

The square of the Hellinger distance between two probability density functions has the general expression (Basseville (1989)):

$$HD^2\big(f(x), g(x)\big) = \frac{1}{2}\int\left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx \tag{13}$$

Note that the Hellinger distance can be written in terms of the Bhattacharyya distance as:

$$HD\big(f(x), g(x)\big) = \sqrt{1 - BC\big(f(x), g(x)\big)};\ 0 \le HD \le 1 \tag{14}$$

$$= \sqrt{1 - e^{-BD(f(x),g(x))}} \tag{15}$$

Hellinger distance does obey the triangle inequality (Upadhyaya and Sorenson (1977), Basseville (1989)).

3. OBSERVABILITY MEASURES

3.1 Observability Measure Based On the Mutual Information

By substituting the density functions (5-8) into (9), taking the logarithm and by using the derivation given in Appendix A, the mutual information between X^k and Y^k can be calculated as,

$$I\big(X^k, Y^k\big) = \frac{1}{2}\log\frac{\left|C_kA_k\Sigma_0{A_k}^T{C_k}^T + C_kG_kQ_k{G_k}^T{C_k}^T + H_kR_k{H_k}^T\right|}{\left|H_kR_k{H_k}^T\right|} \tag{16}$$

Note that, $$\left|H_kR_k{H_k}^T\right| = \left|HRH^T\right|^{k+1} \tag{17}$$

When HRH^T is singular, the observability measure goes to infinity. In the calculation of the mutual information, the term $\left|\Sigma_{X^k}\right|$ that appears both at the numerator and at the denominator is cancelled. When $|\Sigma_0|$ and/or $|GQG^T|$ are singular, the mutual information becomes indefinite as $\frac{0}{0}$. For this case, the limit of $I\big(X^k, Y^k\big)$ the mutual information as $|\Sigma_0|$ and/or $|GQG^T|$ approaches to zero can be considered as the actual measure.

The mutual information can be written in another form as

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_0||GQG^T|^k}{|\Sigma_{X^k} - \Sigma_{X^kY^k}\Sigma_{Y^k}^{-1}\Sigma_{Y^kX^k}|} \tag{18}$$

Although (16) shows the mutual information with basic matrices, (18) is useful when considering the relationship between the mutual information and Kalman filter covariance matrix.

The mutual information, between the states and the measurements for the single time k, can be found similarly as,

$$I(x_k, y_k) = \frac{1}{2}\log\left|C\Sigma_{x_k}C^T + HRH^T\right| - \frac{1}{2}\log|HRH^T| \tag{19}$$

3.1.1 Discussion on the Proposed Observability Measure

We can make several observations to criticize the proposed observability measure.

1) Unobservable states of the pair (C, A) have no effect on the observability measure.

   Proof: Let (C, A) be written in the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \tag{20}$$

$$C = [0 \quad C_{12}] \tag{21}$$

   where A_{11} contains the unobservable modes of (C, A), and let us assume that,

$$\Sigma_0 = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{22}$$

$$GQG^T = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \tag{23}$$

   In (16), the terms, related with initial state covariance and process noise covariance are functions of the matrices $CA^r\Sigma_0 A^{p^T}C^T$ and $CA^rGQG^TA^{p^T}C^T$ ($p = 0, \dots, k;\ r = 0, \dots, k$).

   Replacing the matrices given in (20-23) into these expressions the following result can be found.

$$CA^r\Sigma_0A^{p^T}C^T = C_{12}A_{22}^r\Sigma_{22}A_{22}^{p\,T}C_{12}^{\,T};$$

$$p = 0, \dots, k; \qquad r = 0, \dots, k;$$

   So, the term $C_kA_k\Sigma_0A_k^{\,T}C_k^{\,T}$ is independent of unobservable part. Similarly, $C_kG_kQ_kG_k^{\,T}C_k^{\,T}$ contains only the terms of $C_{12}A_{22}^rGQG^TA_{22}^{p\,T}C_{12}^{\,T}$. So, again the unobservable part is not involved.

2) When the observation noise is zero, the mutual information is equal to infinity. Even if there is only one unnoisy measurement, i.e. HRH^T is singular, this fact holds. This property can suggest the following treatments:

   a) The mutual information gives infinity because of the existence of outputs with no noise that provide a perfect estimation of a part of the state vector. For this case the aim of obtaining a measure of observability is modified so a measure is computed for the remaining states by eliminating the perfectly estimated part. The removing of the unnoisy observable modes is done as follows.

      i) Define $\overline{y_k} \triangleq My_k$ so that $MHRH^TM^T = \begin{bmatrix} R_1 & 0 \\ 0 & 0 \end{bmatrix}$, where M is nonsingular. The new measurement equation is $\overline{y_k} = MCx_k + MHv_k$, where $MHv_k = \begin{bmatrix} H_1 \\ 0 \end{bmatrix} v_k = \begin{bmatrix} v_1 \\ 0 \end{bmatrix}$.

      ii) Define $\overline{C_1}$ and $\overline{C_2}$ such that $\bar{C} = MC = \begin{bmatrix} \overline{C_1} \\ \overline{C_2} \end{bmatrix}$.

      iii) Define a transformation on the state of the system as $\widetilde{x_k} = Tx_k$ so that $\tilde{C} = \bar{C}T^{-1} = \begin{bmatrix} \overline{C_1} \\ \overline{C_2} \end{bmatrix} T^{-1} = \begin{bmatrix} \widetilde{C_{11}} & \widetilde{C_{12}} \\ I & 0 \end{bmatrix}$, where I is the identity matrix. The new system equations can be written as

$$\begin{bmatrix} \widetilde{x_{k+1}}^1 \\ \widetilde{x_{k+1}}^2 \end{bmatrix} = \begin{bmatrix} \widetilde{A_{11}} & \widetilde{A_{12}} \\ \widetilde{A_{21}} & \widetilde{A_{22}} \end{bmatrix} \begin{bmatrix} \widetilde{x_k}^1 \\ \widetilde{x_k}^2 \end{bmatrix} + \begin{bmatrix} \widetilde{w_k}^1 \\ \widetilde{w_k}^2 \end{bmatrix}$$

$$\begin{bmatrix} \widetilde{y_k}^1 \\ \widetilde{y_k}^2 \end{bmatrix} = \begin{bmatrix} \widetilde{C_{11}} & \widetilde{C_{12}} \\ I & 0 \end{bmatrix} \begin{bmatrix} \widetilde{x_k}^1 \\ \widetilde{x_k}^2 \end{bmatrix} + \begin{bmatrix} v_1 \\ 0 \end{bmatrix}$$

      where, $\begin{bmatrix} \widetilde{w_k}^1 \\ \widetilde{w_k}^2 \end{bmatrix} = TGw_k$.

      iv) By applying elementary row operations to the measurement equation we have

$$\begin{bmatrix} \widetilde{y_k}^1 - \widetilde{C_{11}}\widetilde{y_k}^2 \\ \widetilde{y_k}^2 \end{bmatrix} = \begin{bmatrix} 0 & \widetilde{C_{12}} \\ I & 0 \end{bmatrix} \begin{bmatrix} \widetilde{x_k}^1 \\ \widetilde{x_k}^2 \end{bmatrix} + \begin{bmatrix} v_1 \\ 0 \end{bmatrix}$$

      $\widetilde{x_k}^1$ is measured directly with $\widetilde{y_k}^2$. The state equation of $\widetilde{x_k}^2$ is

$$\widetilde{x_{k+1}}^2 = \widetilde{A_{21}}\widetilde{x_k}^1 + \widetilde{A_{22}}\widetilde{x_k}^2 + \widetilde{w_k}^2$$

      Note that $\widetilde{A_{21}}\widetilde{x_k}^1$ is known and it can be treated as a known input of the system.

      As a summary the system is reduced to

$$\widetilde{x_{k+1}}^2 = \widetilde{A_{22}}\widetilde{x_k}^2 + u_k + \widetilde{w_k}^2$$

$$\widehat{y_k} = \widetilde{C_{12}}\widetilde{x_k}^2 + v_1$$

      where, $\widehat{y_k} = \widetilde{y_k}^1 - \widetilde{C_{11}}\widetilde{y_k}^2$ and $u_k = \widetilde{A_{21}}\widetilde{x_k}^1$. As a result, the proposed observability measure can be applied to the reduced system.

   b) Another approach to overcome perfect measurement problem is to add a fixed small variance noise to all measurements. This approach helps us to differentiate the observability of two perfectly measured modes. After this treatment they may have different observability indices.

   c) Observability measure of the subspace of the state space explained in Section 5 can be used for the states which are not observed deterministically.

3) Eigenvalues of the matrix A affect the mutual information. If the eigenvalues of A are increasing then

the information gained increases and eventually approaches to infinity. This is intuitively meaningful, but can also be seen from (16). And also when the entries of C and G are increasing the mutual information increases. Note that when the entries of H are increasing the mutual information decreases.

4) When the covariance matrix of X^k increases the mutual information increases. That means, in that case, the measurements are more valuable.

3.2 Observability Measure Based On the Bhattacharyya Distance

Defining observability measure as the mutual information between the measurements and states seems to be natural. Since the mutual information is a special case of KL distance, the observability measure reduces to the KL distance between two Gaussians. This fact leads to the idea of using some other distance measures as observability measure for stochastic systems. In this section, we will derive the equations for Bhattacharyya and Hellinger distances.

Notice that, the only difference between two Gaussian density functions $f(X^k,Y^k)$ and $f(X^k)f(Y^k)$ is that the covariance matrix of $f(X^k)f(Y^k)$ $(\Sigma_{(X^k,Y^k)})$ does not involve cross covariance matrices $\Sigma_{Y^kX^k}$ and $\Sigma_{X^kY^k}$.

By using the derivation given in Appendix C, Bhattacharyya distance between $f(X^k,Y^k)$ and $f(X^k)f(Y^k)$ can be calculated as,

$$BD\left(f(X^k,Y^k),f(X^k)f(Y^k)\right)=\frac{1}{2}\log\frac{\left|\frac{3}{4}C_k\Sigma_{X^k}C_k{}^T+H_kR_kH_k{}^T\right|}{\left|HRH^T\right|^{\frac{k+1}{2}}\left|\Sigma_{Y^k}\right|^{\frac{1}{2}}} \quad (24)$$

By examining (24), it can be seen that the discussions 1 and 2 given in 3.1.1 are also valid for this measure.

For the specific time k the two Gaussian densities $f(x_k,y_k)$ and $f(x_k)\,f(y_k)$, give the Bhattacharyya distance as,

$$BD\big(f(x_k,y_k)\ ,f(x_k)f(y_k)\big)=\frac{1}{2}\log\frac{\left|\frac{3}{4}C\Sigma_{x_k}C^T+HRH^T\right|}{\left|HRH^T\right|^{\frac{1}{2}}\left|\Sigma_{y_k}\right|^{\frac{1}{2}}} \quad (25)$$

3.3 Observability Measure Based On the Hellinger Distance

Hellinger distance between $f(X^k,Y^k)$ and $f(X^k)f(Y^k)$ can be found from (24) and (15) as

$$HD\left(f(X^k,Y^k),f(X^k)f(Y^k)\right)=\sqrt{1-\frac{\left|HRH^T\right|^{\frac{k+1}{4}}\left|\Sigma_{Y^k}\right|^{\frac{1}{4}}}{\left|\frac{3}{4}C_k\Sigma_{X^k}C_k{}^T+H_kR_kH_k{}^T\right|^{\frac{1}{2}}}} \quad (26)$$

(26) implies again that the discussions 1 and 2 given in 3.1.1 are valid.

And also, Hellinger distance between $f(x_k,y_k)$ and $f(x_k)f(y_k)$ can be found by using (25) and (15) as:

$$HD\big(f(x_k,y_k)\ ,f(x_k)f(y_k)\big)=\sqrt{1-\frac{\left|HRH^T\right|^{\frac{1}{4}}\left|\Sigma_{y_k}\right|^{\frac{1}{4}}}{\left|\frac{3}{4}C\Sigma_{x_k}C^T+HRH^T\right|^{\frac{1}{2}}}} \quad (27)$$

4. RELATIONSHIPS BETWEEN THE OBSERVABILITY MEASURES AND KALMAN FILTER COVARIANCE MATRIX

State covariance matrix of Kalman filter certainly represents the uncertainty of the state at any time k so it may be argued that this can also be used as an observability measure. In this section we will give the relationship between the determinant of the covariance matrix of the state obtained from the Kalman filter and our definition.

Fact: Mutual information between x_k and Y^k is related with the ratio between the determinant of marginal density covariance matrix $\Sigma_{x_kx_k}$ of x_k, and the determinant of Kalman filter state covariance matrix $\Sigma_{x_k|Y^k}$ at time k (Logothetis et al. (1997)). That is,

$$I(x_k,Y^k)=\frac{1}{2}\log\frac{\left|\Sigma_{x_k}\right|}{\left|\Sigma_{x_k|Y^k}\right|} \quad (28)$$

As seen from (28), that mutual information compares determinants of state covariances when the measurements are used and not used.

Similarly the mutual information between X^k and Y^k , i.e., the observation measure of the whole state sequence can be found as,

$$I(X^k,Y^k)=\frac{1}{2}\log\frac{\left|\Sigma_{X^k}\right|}{\left|\Sigma_{X^k|Y^k}\right|} \quad (29)$$

where, $\Sigma_{X^k|Y^k}$ is the covariance matrix of the probability density function of $f(X^k|Y^k)$.

And also examining the densities $f(x_k,Y^k)$ and $f(x_k)f(Y^k)$ gives the relationship between the Kalman Filter state covariance matrix at time k and the Bhattacharyya distance as,

$$BD\left(f(x_k,Y^k)\ ,f(x_k)f(Y^k)\right)$$
$$=\frac{1}{2}\log\frac{\left|\Sigma_{Y^k}\right|\left|\Sigma_{x_k}-\frac{1}{4}\Sigma_{x_kY^k}\left(\Sigma_{Y^k}\right)^{-1}\Sigma_{Y^kx_k}\right|}{\sqrt{\left|\Sigma_{x_k}\right|\left|\Sigma_{Y^k}\right|\left|\Sigma_{Y^k}\right|\left|\Sigma_{x_k}-\Sigma_{x_kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kx_k}\right|}}$$

$$BD\left(f(x_k,Y^k)\ ,f(x_k)f(Y^k)\right)=\frac{1}{2}\log\frac{\left|\frac{3}{4}\Sigma_{x_k}+\frac{1}{4}\Sigma_{x_k|Y^k}\right|}{\sqrt{\left|\Sigma_{x_k}\right|\left|\Sigma_{x_k|Y^k}\right|}} \quad (30)$$

Similarly, the relationship between conditional covariance matrix $\Sigma_{X^k|Y^k}$ and the Bhattacharyya distance for the densities $f(X^k,Y^k)$ and $f(X^k)f(Y^k)$ can be written as,

$$BD\left(f(X^k,Y^k)\ ,f(X^k)f(Y^k)\right)=\frac{1}{2}\log\frac{\left|\frac{3}{4}\Sigma_{X^k}+\frac{1}{4}\Sigma_{X^k|Y^k}\right|}{\sqrt{\left|\Sigma_{X^k}\right|\left|\Sigma_{X^k|Y^k}\right|}} \quad (31)$$

As in the Bhattacharyya distance, the relationship between Kalman filter state covariance matrix at time k and Hellinger distance between the densities $f(x_k, Y^k)$ and $f(x_k)f(Y^k)$ can be found by using (15) and (31) as:

$$HD\left(f(x_k, Y^k)\ , f(x_k)f(Y^k)\right) = \sqrt{1 - \frac{\left(|\Sigma_{x_k}||\Sigma_{x_k|Y^k}|\right)^{\frac{1}{4}}}{\left|\frac{3}{4}\Sigma_{x_k} + \frac{1}{4}\Sigma_{x_k|Y^k}\right|^{\frac{1}{2}}}} \tag{32}$$

A similar relationship can also be written for the densities $f(X^k, Y^k)$ and $f(X^k)f(Y^k)$:

$$HD\left(f(X^k, Y^k)\ , f(X^k)f(Y^k)\right) = \sqrt{1 - \frac{\left(|\Sigma_{X^k}||\Sigma_{X^k|Y^k}|\right)^{\frac{1}{4}}}{\left|\frac{3}{4}\Sigma_{X^k} + \frac{1}{4}\Sigma_{X^k|Y^k}\right|^{\frac{1}{2}}}} \tag{33}$$

5. OBSERVABILITY MEASURE FOR INDIVIDUAL STATES, MODES AND SUBSPACES OF STATE SPACE

For individual states of a LTI discrete-time Gaussian system, the mutual information between an individual state and the observations can be used to define the observability measure. By using (28), for the ith state at time k, the mutual information between the ith state and the observations can be found as

$$I(x_{i,k}, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{x_{i,k}}|}{|\Sigma_{x_{i,k}|Y^k}|} \tag{34}$$

And also, by using (29), the mutual information between the ith state for time 0 to k and the observations can be found as

$$I(X_i^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{X_i^k}|}{|\Sigma_{X_i^k|Y^k}|} \tag{35}$$

(34) and (35) can also be used to define the observability measure for a subspace of the state space. In addition, when the state equation is given in Jordan form, these equations can be used to define observability measure of the system modes. The covariance matrices of $\Sigma_{x_{i,k}}$, $\Sigma_{x_{i,k}|Y^k}$, $\Sigma_{X_i^k}$ and $\Sigma_{X_i^k|Y^k}$ can be found from the covariance matrices of Σ_{x_k}, $\Sigma_{x_k|Y^k}$, Σ_{X^k}, $\Sigma_{X^k|Y^k}$. And also, similar derivations can be done by using the equations of Bhattacharyya and Hellinger distances.

Example: We will apply the above ideas to a simple example. Let the system equations be defined as:

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} -0.5 & 0 \\ 0 & -0.7 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} w_k$$

$$y_k = \begin{bmatrix} 0.75 & 0.075 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + v_k$$

where,

$$x_0 \sim N\left(\overline{x_0}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right),\ w_k \sim N\left(0, \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\right),\ v_k \sim N(0, 0.5)$$

Note that the system representation is selected in the diagonal form so that the observability measures of the individual states become the observability measures of the modes. The process noise is selected so that both states are under the same process noise. Since the coefficient of the first state in the matrix C is higher than the coefficient of the second state, the expectation is that the first state is more observable compared to the second one. The computed observability measures of the states $x_{1,k}$ and $x_{2,k}$ according to (34) are: $I(x_{1,k}, Y^k) = 0.3127$ and $I(x_{2,k}, Y^k) = 0.0044$. These values are for the steady state, i.e. a large k and satisfy the expectation. Similarly for the whole state sequence the computed observability measures of X_1^k and X_2^k according to (35) are $I(X_1^k, Y^k) = 25.4810$ and $I(X_2^k, Y^k) = 0.2383$, as expected, for k =100. And also the observability measures of x_k and X^k according to (28) and (16) are $I(x_k, Y^k) = 0.3206$ and $I(X^k, Y^k) = 26.0296$.

6. CONCLUSIONS

The analysis of several observability measures defined in this work shows that they do not contribute much to the observability measure concept over the basic mutual information definition. However Hellinger distance may be preferable because of its boundedness.

The definition is clearly expandable to non Gaussian and nonlinear systems. The use of the definition for a general set up may be done by using particle filters.

REFERENCES

Aoki, M. (1967). *Optimization of Stochastic Systems*. Academic Press Inc. New York, USA.

Bageshwar, V.L., Egziabler, D.G., Garrard, W.L., Georgiou, T.T. (2009). Stochastic Observability Test for Discrete-Time Kalman Filters. *Journal of Guidance, Control, and Dynamics*, volume (32-4), 1357-1370.

Basseville, M. (1989). Distance Measures For Signal Processing and Pattern Recognition. *Signal Processing*, volume (18), 349-369.

Chen, B., Hu, J., Li, H., Sun, Z. (2007). Measure Observability by the Generalized Informational Correlation. *Proceedings of the 46th IEEE Conference on Decision and Control*, 5570-5574.

Cover, T.M., Thomas, J.A. (2006). *Elements of Information Theory, 2nd edition*, A JOHN WILEY & SONS, INC. New Jersey, USA.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition.* Academic Press. San Diego, USA.

Hamdan, A.M.A., Nayfeh, A.H. (1989). Measures of Modal Controllability and Observability for First- and Second-Order Linear Systems. *Journal of Guidance, Control, and Dynamics*, volume (12-3), 421-428.

Han-Fu, C. (1980). On Stochastic Observability and Controllability. *Automatica*, volume (16), 179-190.

Huang, T., Chen, B. (2008). Information Evolutions in the Linear Stochastic Control Systems. *Proceedings of the 7th World Congress on Intelligent Control and Automation, Chongqing, China*, 3384-3387.

Kam, M., Cheng, R., Kalata, P. (1987). An Information Theoretic Interpretation of Stability and Observability. *Proceedings of the American Control Conference, 3*, 1957-1962.

Lindner, D.K., Babendreier, J., Hamdan, A.M.A. (1989). Measures of the Controllability and Observability and

Residues. *IEEE Transactions on Automatic Control*, volume (34-6), 648-650.

Logothetis, A., Isaksson, A., Evans, R.J. (1997). An Information Theoretic Approach to Observer Path Design for Bearings-Only Tracking. *Proceedings of the 36th Conference on Decision & Control*.

Mohler, R.R., Hwang, C.S. (1988). Nonlinear Data Observability and Information. *Journal of the Franklin Institute*, vol. 325, no. 4, pp. 443-464, 1988.

Müller, P.C., Weber, H.I. (1972). Analysis and Optimization of Certain Qualities of Controllability and Observability for Linear Dynamical Systems. *Automatica*, volume (8), 237-246.

Porter, B., Crossley, T.R. (1970). Mode-Observability Structure of a Multivariable Linear System. *Electronics Letters,* volume (6-21), 664-665.

Tarokh, M. (1992). Measures for Controllability, Observability, and Fixed Modes. *IEEE Transactions on Automatic Control*, volume (37-8), 648-650.

Upadhyaya, B.R., Sorenson, H.W. (1977). Bayesian Discriminant Approach to Input Signal Selection in Parameter Estimation Problems. *Information Sciences*, volume (12), 61-97.

Appendix A. MUTUAL INFORMATION BETWEEN X^k AND Y^k

The mutual information between X^k and Y^k can be calculated as (Huang and Chen (2008)),

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{X^k}||\Sigma_{Y^k}|}{|\Sigma_{[X^k,Y^k]}|} \tag{A.1}$$

Determinant of Σ_{X^k,Y^k} can be written as:

$$\left|\Sigma_{[X^k,Y^k]}\right| = |\Sigma_{X^k}|\left|\Sigma_{Y^k} - \Sigma_{Y^kX^k}\Sigma_{X^k}{}^{-1}\Sigma_{X^kY^k}\right| \tag{A.2}$$

By using covariance matrices given in (5-8) and after some manipulation, the following equation can be found:

$$\left|\Sigma_{[X^k,Y^k]}\right| = |\Sigma_{X^k}|\left|H_kR_kH_k{}^T\right| \tag{A.3}$$

And, by substituting (A.3) in (A.1), the mutual information between X^k and Y^k can be found as,

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{\left|C_kA_k\Sigma_0A_k{}^TC_k{}^T + C_kG_kQ_kG_k{}^TC_k{}^T + H_kR_kH_k{}^T\right|}{\left|H_kR_kH_k{}^T\right|} \tag{A.4}$$

By using the derivation ($|\Sigma_{X^k}|$ and $\left|\Sigma_{[X^k,Y^k]}\right|$) given in Appendix B, (A.4) can be written in another form as

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_{X^k}||\Sigma_{Y^k}|}{|\Sigma_{Y^k}|\left|\Sigma_{X^k} - \Sigma_{X^kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kX^k}\right|}$$

$$I(X^k, Y^k) = \frac{1}{2}\log\frac{|\Sigma_0|\left|GQG^T\right|^k}{\left|\Sigma_{X^k} - \Sigma_{X^kY^k}\Sigma_{Y^k}{}^{-1}\Sigma_{Y^kX^k}\right|} \tag{A.5}$$

Appendix B. DETERMINANTS OF Σ_{X^k} AND $\Sigma_{[X^k,Y^k]}$

Σ_{X^k} is $\quad \Sigma_{X^k} = A_k\Sigma_0A_k{}^T + G_kQ_kG_k{}^T$

$$= \begin{bmatrix} \Sigma_0 & \cdots & \Sigma_0 A^{k^T} \\ \vdots & \ddots & \vdots \\ A^k\Sigma_0 & \cdots & A^k\Sigma_0 A^{k^T} + \sum_{i=0}^{k-1} A^{k-1-i}GQG^T A^{k-1-i^T} \end{bmatrix}$$

By using elementary row operations, the following matrix can be found (note that the determinant does not change):

$$= \begin{bmatrix} \Sigma_0 & \Sigma_0 A^T & \cdots & \Sigma_0 A^{k^T} \\ 0 & GQG^T & \cdots & GQG^T A^{k-1^T} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & GQG^T \end{bmatrix}$$

The determinant of this matrix is

$$|\Sigma_{X^k}| = |\Sigma_0|\left|GQG^T\right|^k \tag{B.1}$$

By using (A.3) and (B.1), the determinant of Σ_{X^k,Y^k} can be found as,

$$\left|\Sigma_{[X^k,Y^k]}\right| = |\Sigma_0|\left|GQG^T\right|^k\left|HRH^T\right|^{k+1} \tag{B.2}$$

Appendix C. BHATTACHARYYA DISTANCE BETWEEN THE DENSITIES $f(X^k, Y^k)$ AND $f(X^k)f(Y^k)$

Bhattacharyya distance between the two Gaussian densities $f(X^k, Y^k)$ and $f(X^k)f(Y^k)$ can be found from the following equation (Fukunaga (1990)):

$$BD\left(f(X^k, Y^k), f(X^k)f(Y^k)\right) = \frac{1}{8}\left(\begin{pmatrix} E(X^k) \\ E(Y^k) \end{pmatrix} - \begin{pmatrix} E(X^k) \\ E(Y^k) \end{pmatrix}\right)^T \left(\frac{\Sigma_{[X^k,Y^k]} + \Sigma_{(X^k,Y^k)}}{2}\right)^{-1}\left(\begin{pmatrix} E(X^k) \\ E(Y^k) \end{pmatrix} - \begin{pmatrix} E(X^k) \\ E(Y^k) \end{pmatrix}\right) + \frac{1}{2}\log\frac{\left|\frac{\Sigma_{[X^k,Y^k]} + \Sigma_{(X^k,Y^k)}}{2}\right|}{\sqrt{\left|\Sigma_{[X^k,Y^k]}\right|\left|\Sigma_{(X^k,Y^k)}\right|}} \tag{C.1}$$

Note that the first term in the right side of the equation is zero, since the two densities have the same mean value. Then,

$$BD\left(f(X^k, Y^k), f(X^k)f(Y^k)\right) = \frac{1}{2}\log\frac{\left|\frac{\Sigma_{[X^k,Y^k]} + \Sigma_{(X^k,Y^k)}}{2}\right|}{\sqrt{\left|\Sigma_{[X^k,Y^k]}\right|\left|\Sigma_{(X^k,Y^k)}\right|}} \tag{C.2}$$

where,

$$\left|\frac{\Sigma_{[X^k,Y^k]} + \Sigma_{(X^k,Y^k)}}{2}\right| = |\Sigma_{X^k}|\left|\frac{3}{4}C_k\Sigma_{X^k}C_k{}^T + H_kR_kH_k{}^T\right| \tag{C.3}$$

and, $\left|\Sigma_{[X^k,Y^k]}\right|\left|\Sigma_{(X^k,Y^k)}\right| = |\Sigma_{X^k}|^2|\Sigma_{Y^k}|\left|HRH^T\right|^{k+1}$ (C.4)

By substituting (C.4) and (C.3) into (C.2), the Bhattacharyya distance can be found as,

$$BD\left(f(X^k, Y^k), f(X^k)f(Y^k)\right) = \frac{1}{2}\log\frac{\left|\frac{3}{4}C_k\Sigma_{X^k}C_k{}^T + H_kR_kH_k{}^T\right|}{\left|HRH^T\right|^{\frac{k+1}{2}}|\Sigma_{Y^k}|^{\frac{1}{2}}} \tag{C.5}$$