Defining a Pseudo-Metric Topology on Linear Dynamic Systems*

Vahid Hassani * Andrew Ross *

* Norwegian Marine Technology Research Institute (MARINTEK), Trondheim, Norway
(e-mail: {Vahid.Hassani, Andrew.Ross}@marintek.sintef.no)

Abstract: This paper defines a (pseudo) metric topology on the space of stable discrete linear time-invariant (LTI) dynamic systems. The article seeks to present a solution to the problem of comparing two LTI systems or, more formally, to find a metric for the space of stable discrete LTI dynamic systems. To this effect, by comparing the performance of two Kalman filters designed for two dynamic systems, a distance-like pseudo-norm between two systems is developed. The defined metric topology can be exploited to select the closest model, among several possible models $P_s$, all of which are known, to an observed data sequence modeled as $P_r$. Numerical simulations are provided illustrating the efficacy of the metric derived.

1. INTRODUCTION

The question that this article seeks to answer is how to compare two stable discrete linear time-invariant (LTI) dynamic systems or, more formally, to find a metric topology on the space of stable discrete LTI dynamic systems. Consider a set of stable discrete LTI dynamic systems denoted by $\mathcal{P}$: A topology on $\mathcal{P}$ is a collection $\mathcal{T}$ of subsets of $\mathcal{P}$ that satisfy a set of axioms relating points and neighborhoods (see Munkres [2000]). One of the most important and common ways of imposing a topology on a set is to introduce the topology in terms of a metric on the set. A metric on a set $\mathcal{P}$ is a function $m : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}^+$ having the following properties

1. $m(P_r, P_f) \geq 0$, $\forall P_r, P_f \in \mathcal{P}$; where equality holds if and only if $P_r = P_f$,
2. $m(P_r, P_f) = m(P_f, P_r)$, $\forall P_r, P_f \in \mathcal{P}$,
3. $m(P_r, P_s) + m(P_s, P_f) \geq m(P_r, P_f)$, $\forall P_r, P_s, P_f \in \mathcal{P}$.

In the study of linear algebra and vector spaces, norm induces a metric, and hence a topology, on the vector space. However, this approach does not always make sense, for the stable discrete LTI dynamic systems. For a review on the definition of different norms for signals and systems, the reader is referred to Boyd and Barratt [1991]. Equipping the space of stable discrete LTI dynamic systems with a topology and a metric not only is an interesting theoretical exercise, but also has many important applications. Many applications in system identification can benefit from defining a metric topology on dynamic systems. Another important application comes in model reduction, when one can define a distance between the original system and the reduced order one, and monitor the distance as an index of similarity between two systems. For early studies of topological and metric properties of dynamical systems the reader is referred to Petreczky and Vidal [2007], De Cock and De Moor [2002], Martin [2000], Hanzon [1986], in which for example, Martin [2000] introduced a cepstral distance measure for single-input single-output (SISO) autoregressive moving average (ARMA) models, De Cock and De Moor [2002] defined subspace angles between two ARMA models, and Petreczky and Vidal [2007] defined a distance between two dynamical systems as the distance between the formal power series that encode the input-output behavior of the systems. The problem of defining metrics on the space of dynamical systems is an old problem which has gained more attention and been revisited in recent years, for example see Af-sari and Vidal [2013a,b].

The main contribution of this paper is defining a metric topology on stable discrete LTI dynamic systems focusing on similarity of the input-output behavior of dynamical systems. The proposed metric is developed around the idea of finding the distance of members of a set of discrete-time LTI systems from a reference system. To this effect and to find a distance between two stable discrete LTI dynamic systems, a Kalman filter (KF) associated with each dynamic system is designed (based on the model of the aforementioned dynamic system) and the performance of these KFs are compared with the one associated with the reference system in order to find the distance of their corresponding systems from the reference system. 1

---

1 The methodology proposed here has its root in Baram and Sandell [1978] in which the modeling and identification of dynamic systems where the model set does not necessarily include the observed system, are treated.
worth stating that in this framework, if the distance between two systems is zero, it means that they are at equal distance from the reference system. This does not mean that they are equal from an input-output point of view. In fact, it means that in this framework and based on the proposed metric there is no possible way of distinguishing two systems on the basis of the measurement data.

The structure of the paper is as follows. In section 2 we review the problem formulation. Section 3 summarizes our main results in which a pseudo-metric on stable discrete LTI dynamic systems is defined and the properties of the metric are shown. Section 4 illustrates the defined metric through numerical simulations. The conclusions are summarized in section 5.

2. PROBLEM FORMULATION

This section revisits some basic properties of stable discrete LTI dynamic systems in a stochastic setting. The system is described by a discrete-time difference equation,

\[
x(t+1) = Ax(t) + Bu(t) + Gw(t),
\]

\[
y(t) = Cx(t) + v(t),
\]

where \( x(t) \in \mathbb{R}^n \) denotes the state of the system, \( u(t) \in \mathbb{R}^m \) its control input, \( y(t) \in \mathbb{R}^q \) its measured noisy output, \( w(t) \in \mathbb{R}^p \) an input plant disturbance that can not be measured, and \( v(t) \in \mathbb{R}^r \) is the measurement noise. The vectors \( w(t) \) and \( v(t) \) are zero-mean, mutually independent white Gaussian sequences, with covariances \( \text{cov}[w(t); w(\tau)] = Q_d \delta_{t \tau} \) and \( \text{cov}[v(t); v(\tau)] = R_d \delta_{t \tau} \), respectively. The initial condition \( x(0) \) is a Gaussian random vector with mean and covariance given by \( E[x(0)] = 0 \) and \( E[x(0)x^T(0)] = P(0) \). It is further assumed that \([A, G]\) and \([A, C]\) are controllable and observable, respectively.

Consider a family of stable discrete LTI dynamic systems parameterized by some variable \( s \in \mathcal{S} \). For example, we could have

\[
\mathfrak{P} := \{ x_s(t+1) = A_s x_s(t) + B_s u(t) + G_s w(t) : s \in \mathcal{S} \},
\]

with the parameterizing set \( \mathcal{S} \) finite, infinite but countable, or not even countable; and all the \( x_s, w, v, u, \) and \( y \) with the same dimension, accordingly.

Any member of \( \mathfrak{P} \), say \( P_s \), is represented by the set of matrices \((A_s; B_s; C_s; G_s)\) (which is called a realization of \( P_s \)). A steady state Kalman filter, see Anderson and Moore [1979] for details, can be designed:

\[
\hat{x}_s(t+1) = A_s \hat{x}_s(t) + B_s u(t) + H_s (y(t) - C_s \hat{x}_s(t)),
\]

\[
\hat{y}_s(t) = C_s \hat{x}_s(t),
\]

\[
H_s = A_s \Sigma_s C_s^T [C_s \Sigma_s C_s^T + R_s]^{-1}
\]

where \( \Sigma_s \) is the solution of the discrete Riccati equation

\[
\Sigma_s = A_s \Sigma_s A_s^T + G_s Q_s G_s^T - A_s \Sigma_s C_s^T [C_s \Sigma_s C_s^T + R_s]^{-1} C_s \Sigma_s A_s.
\]

The designed KF can be used to estimate the states of \( P_s \) using \( u(t) \) and \( y(t) \).

Let us consider a special case where the dynamics of a physical system are governed by

\[
x_{s*}(t+1) = A_{s*} x_{s*}(t) + B_{s*} u(t) + G_{s*} w(t),
\]

\[
y(t) = C_{s*} x_{s*}(t) + v(t),
\]

and the real value of the set of matrices \((A_{s*}; B_{s*}; G_{s*}; C_{s*})\) is unknown. Let us further assume that two possible model of the physical system are suggested as \((A_s; B_s; G_s; C_s)\) and \((A_f; B_f; G_f; C_f)\). How can one select between the two possible model (having access to the input-output measurements of the real system)?

In the following section we will develop a pseudo-norm for comparing two stable LTI systems.

3. METRIC DEFINITION

Let \( u(t) \) and \( y(t) \) be the measured input and output of system \( P_{s*} \), described in (5), respectively. Moreover, let \( \hat{y}_r(t) \) and \( \hat{y}_f(t) \) denote the output innovation sequence (residual) from two KFs designed for the model \( P_s \) (with realization \((A_s; B_s; G_s; C_s)\)) and \( P_f \) (with realization \((A_f; B_f; G_f; C_f)\)), respectively, and given by

\[
\hat{y}_r(t) = y(t) - \hat{y}_r(t),
\]

\[
\hat{y}_f(t) = y(t) - \hat{y}_f(t).
\]

We will assume that the residual sequences in all the KFs are stationary and ergodic (see Hassani et al. [2013] for necessary conditions).

Let \( Y_t \equiv \{ y(0), y(1), \ldots, y(t), u(1), \ldots, u(t) \} \) condense the history of the measurements from the beginning up to time \( t \). Consider the conditional probability density function \( f_x(y(t)|Y_{t-1}, P_s) \) (the probability distribution of \( y(t) \) when \( Y_{t-1} \) is known to be a particular value and assuming that the model of the system is \( P_s \)). Furthermore, for each KF we have \( f_s(Y_t|P_s) = \prod_{k=1}^t f_s(y(k)|Y_{k-1}, P_s) \).

For two different KFs based on \( P_s \) and \( P_f \), if

\[
f_f(Y_t|P_f) > f_r(Y_t|P_r),
\]

or, equivalently, if

\[
\log f_f(Y_t|P_f) > \log f_r(Y_t|P_r),
\]

we will say that based on the observation vector \( Y_t \), the KF designed based on model \( P_f \) is preferred over (more likely or probably) than the KF designed based on model \( P_r \). Define the likelihood ratio for the sequence of \( Y_t \)

\[
k_f^f/Y_t = f_f(Y_t|P_f) / f_r(Y_t|P_f),
\]

or, equivalently,

\[
\log k_f^f/Y_t = \log f_f(Y_t|P_f) - \log f_r(Y_t|P_r),
\]

where \( \log k_f^f/Y_t \) can be regarded as a measure of the information contained in \( Y_t \) that can be used to select between the KFs designed based on the models \( P_s \) and \( P_f \). Similarly, one can compute the conditional likelihood ratio

\[
k_r^f/Y_t = f_r(y(t)|Y_{t-1}, P_f) / f_r(y(t)|Y_{t-1}, P_r).
\]

Positive values of \( \log k_f^f/Y_t \) mean that based on the observation vector \( Y_t \), the KF based on \( P_f \) is more likely to be the optimal observer than the KF based on \( P_r \), while negative values show that the KF based on \( P_r \) is preferred over the KF based on \( P_f \).
or, equivalently,
$$
\log k_f^r(y(t)|Y_{t-1}) = \log f_f(y(t)|Y_{t-1}, P_f) - \log f_r(y(t)|Y_{t-1})
$$
which can be interpreted as a measure of the information contained in $y(t)$ that can be used to select between the KFs designed based on models $P_r$ and $P_f$. We can define the mean information in $y(t)$ for preferring the KF designed based on model $P_f$ (or $(A_f; B_f; G_f; C_f)$) over the KF designed based on model $P_r$ (or $(A_r; B_r; G_r; C_r)$) as
$$
d_f(P_f, P_r) = E(\log k_f^r(y(t)|Y_{t-1})).
$$
(9)

When $d_f(P_f, P_r)$ is positive we can conclude that the KF based on $P_f$ is more probable to be the true KF than the KF based on $P_r$. The above variable can be regarded as a yardstick against which to select the “best” KF modeling the behavior of the real system. It is easy to see that the true KF is always preferred over other KFs (i.e. the KF designed based on model $P_{\star}$, is always preferred over other KFs).

**Proposition 1.** Let $Y_t$ be the measured data from a system described by model $P_{\star}$, then for all KFs designed based on model $P_r$ ($P_r \neq P_{\star}$)
$$
d(P_r, P_{\star}) \geq 0,
$$
with equality if & only if $f_{\star}(y_t|Y_{t-1}, P_{\star}) = f_r(y_t|Y_{t-1})$. 

**Proof.** Since $Y_t$ is the measured data from a system described by model $P_{\star}$, using the KF designed based on model $P_{\star}$, $(f_{\star}(y_t|Y_{t-1}, P_{\star}))$ is the true sequence of conditional probability densities of $(y_t)$; it follows immediately from the definition that for each $t \geq 0$ we have
$$
E(f_{\star}(y_t|Y_{t-1}, P_{\star})) \geq E(f_r(y_t|Y_{t-1}, P_r)),
$$
with equality if & only if $f_{\star}(y_t|Y_{t-1}, P_{\star}) = f_r(y_t|Y_{t-1}, P_r)$.

Now it is straightforward to get the result. \qed

The conditional probability density of $y(t)$ given the past observation $Y_{t-1}$ when the true model of the system (from which $Y_{t-1}$ is sampled) is $P_{\star}$, has the form (see Anderson and Moore [1979])
$$
f_{\star}(y(t)|Y_{t-1}, P_{\star}) = \frac{\exp\left(-\frac{1}{2}S_{\star}(t)^{-1}y_{\star}(t)\right)}{\sqrt{(2\pi)^T|S_{\star}|}},
$$
(11)
where $q$ is the dimension of $y_{\star}(t)$ and $S_{\star} = C_{\star}T_{\star}C_{\star}^T + R$ is the covariance of the innovation sequence. In fact, in this case the conditional probability density of $y(t)$ given the past observation $Y_{t-1}$ when the true model of the system is $P_{\star}$, $(f_{\star}(y(t)|Y_{t-1}, P_{\star}))$, is a gaussian distribution with mean $y_{\star}(t)$ and covariance $E\{\tilde{y}_{\star}(t)\tilde{y}_{\star}^T(t)\}$, which we denote by $S_{\star}$. 

Now let us consider the case that $Y_t$ is the measured data from a system described by model $P_{\star}$, but a KF designed

---

3 When the dynamics of the system $P_{\star}$ is constant, it is reasonable to assume that in steady state $y(t)$ and $y(\tau)$, $t \neq \tau$ have the same “amount” of information for selecting between the KFs. So we drop the $t$ in $d_f(P_f, P_r)$ and use $d(P_f, P_r)$ instead.

4 According to the assumption of stationarity, $S_{\star}$ is independent of $t$.

5 We should highlight here that the notation of the term $\tilde{y}_{\star}(t)$ in (12) is ambiguous, since it may denote either the residual of the KF designed based on the assumption that the true plant model is $P_r$, or the residual of the KF designed based on the model $P_r$, irrespective of the true plant model. Clearly, in (12) $\tilde{y}_{\star}(t)$ has the second meaning.

6 Hassani et al. [2009, 2011] used $\Gamma_{\star}$ as a performance index in multiple model adaptive estimator and also as a tool for proper designing of observers in robust multiple model adaptive control methodology. It is also closely related to Kullback and Leibler’s distance (see Kullback and Leibler [1951]) and in fact, can be viewed as a modified version of Kullback and Leibler’s information index, see Baram and Sandell [1978].
Under assumption on ergodicity and stationarity of the residuals, we can compute the expected value of \( \log k_f^T(y(n)|Y_{n-1}) \) as

\[
\lim_{t \to \infty} \frac{1}{t} \sum_{n=1}^{t} \log k_f^T(y(n)|Y_{n-1}) = E\{\log k_f^T(y(n)|Y_{n-1})\} = d_n(P_r, P_f) = \Gamma_r^* - \Gamma^*_f. \tag{19}
\]

If \( \Gamma^*_f \leq \Gamma^*_r \) then by comparing (18), (19), and (20), it follows that

\[
\lim_{t \to \infty} \log k_f^T(Y_t) = \lim_{t \to \infty} \frac{1}{t} \sum_{n=1}^{t} \log k_f^T(y(n)|Y_{n-1}) = -\infty \tag{21}
\]

which implies that

\[
\lim_{t \to \infty} k_f^T(Y_t) = 0. \tag{22}
\]

This theorem shows that the KF which has the minimum \( \Gamma^*_v \) (i.e. the KF designed based on model \( P_v \)) is more preferable over other KFs. In other word, the model \( P_v \) is closer to the true plant model \( P \) in some sense. Based on the results of Theorem 2 we can define a Pseudo-norm on the set of LTI systems given by

\[
m(P_r, P_f) := |\Gamma^*_r - \Gamma^*_f|. \tag{23}
\]

Lemma 3. The defined norm in (23) is a Pseudo-Norm\(^7\).

**Proof.** It is not difficult to see that

\[
m(P_r, P_f) = |\Gamma^*_r - \Gamma^*_f| = 0.
\]

To prove the symmetry property, use the fact that

\[
m(P_r, P_f) = |\Gamma^*_r - \Gamma^*_f| = |\Gamma^*_f - \Gamma^*_r| = m(P_f, P_r).
\]

The triangle inequality follows from

\[
m(P_r, P_p) + m(P_p, P_f) = |\Gamma^*_r - \Gamma^*_p| + |\Gamma^*_p - \Gamma^*_f| \geq |\Gamma^*_r - \Gamma^*_r| = m(P_r, P_f).
\]

\[
4. \text{SIMULATION}
\]

This section illustrates the design methodology described in the previous section. Motivated by Nomoto et al. [1957] we consider the steering equations known as Nomoto model. The model is developed such that the steering dynamics of the yaw mode of marine vessel could be analyzed in isolation, through either a first or second order transfer function. For a large class of marine vessels, Nomoto Model gives a reasonably accurate description of the course-keeping behavior and even today, this simple and thoroughly effective model is used within a multitude

\[
\xi(t) \quad W_d \quad \delta(t) \quad \theta(t)
\]

Fig. 1. The Stable Discrete-Time Linear Time-Invariant (LTI) Plant.

of guidance and control system design papers. The differential equation corresponding to the first order Nomoto model can be written as

\[
\dot{\psi}(t) + \frac{1}{T} \psi(t) = \frac{k}{T} \delta(t), \tag{24}
\]

where \( \psi(t) \) and \( \delta(t) \) denote the yaw angle and rudder angle of the ship, respectively, and \( T \) and \( k \) are the effective time constant and gain constant, respectively. Let us further assume that the steering model is subject to a low-frequency stochastic disturbance input \( d(t) \) obtained by filtering white noise \( \xi(t) \) with zero mean and unit intensity, as follows:

\[
W_d(s) = \frac{d(s)}{\xi(s)} = \frac{0.1}{s + 0.1}. \tag{25}
\]

Fig. 1 shows the block diagram of the example adopted where \( y(t) \) is the observed output (measured heading of the vessel), \( \delta(t) \) is the control input (rudder angle), \( d(t) \) is the plant disturbance, and \( \theta(t) \) is the sensor noise assumed to be white noise with zero mean and intensity \( 10^{-3} \).

All simulations for this example were implemented in discrete-time using a zero-order hold with a sampling time of \( T_s = 0.01 \) secs.

Figs. 2, 3, and 4 illustrates the distance between the system with \( k = -0.15 \) and \( T = 8 \) and a system with \(-1 \leq k \leq -0.1 \) and \( 6 \leq T \leq 14 \).

Let us denote by \( P^T_k \) the system described in (24). For example \( P^{15}_{-15} \) denotes the system described in (24) with \( k = -0.15 \) and \( T = 8 \). It can be noticed that \( m(P^{15}_{-15}, P^T_k)_{|T=8,k=-0.15} = 0 \) i.e. the distance of the system described in (24) with \( k = -0.15 \) and \( T = 8 \) from itself is zero.

5. CONCLUSIONS

This paper introduced a pseudo-metric topology on the space of stable discrete LTI dynamic systems. Based on the defined norm one can compare two LTI systems from input-output behavior point of view. The defined metric space have important application in system identification. Future work will aim at extending the current metric to accommodate the space of time varying and nonlinear systems.
Fig. 2. Distance of the systems with different $k$ and $T$ values from the system with $k = -0.15$ and $T = 8$ as a function of $k$ and $T$.

Fig. 3. A zoomed in view of Fig. 2.

Fig. 4. Distance of the systems with different $k$ and $T$ values from the system with $k = -0.15$ and $T = 8$ as a function of $k$ (fixed $T$).

ACKNOWLEDGEMENTS

We wish to thank my colleagues A. Pedro Aguiar, Antonio Pascoal and Michael Athans for many discussions on system theory and control.

REFERENCES


