Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

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 $P_\star$. 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 the space of stable discrete LTI dynamic systems. Consider a set of stable discrete LTI dynamic systems denoted by $\mathfrak{P}$; A topology on $\mathfrak{P}$ is a collection $\mathfrak{T}$ of subsets of $\mathfrak{P}$ 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 $\mathfrak{P}$ is a function

$$m : \mathfrak{P} \times \mathfrak{P} \to \mathbb{R}^+$$

having the following properties

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

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 refereed 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 dynamic systems is an old problem which has gained more attention and been revisited in recent years, for example see Afsari 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] It is

⋆ This work was supported in part by project "Sea Trials and Model Tests for Validation of Shiphandling Simulation Models" (of MAROFF KPN projects) and the Norwegian Marine Technology Research Institute (MARINTEK); the Norwegian research council is acknowledged as sponsor of MARINTEK.

[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,

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + Gw(t), \\ y(t) &= Cx(t) + v(t), \end{aligned} \tag{1}$$

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, $w(t) \in \mathbb{R}^r$ an input plant disturbance that can not be measured, and $v(t) \in \mathbb{R}^q$ is the measurement noise. The vectors $w(t)$ and $v(t)$ are zero-mean, mutually independent white Gaussian sequences, with covariances $cov[w(t); w(\tau)] = Q\delta_{t\tau}$ and $cov[v(t); v(\tau)] = R\delta_{t\tau}$, respectively. The initial condition $x(0)$ of (1) 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 $s \in \mathfrak{S}$. For example, we could have

$$\mathfrak{P} := \left\{ \begin{array}{l} x_s(t+1) = A_s x_s(t) + B_s u(t) + G_s w(t) \\ y(t) = C_s x_s(t) + v(t) \end{array} : s \in \mathfrak{S} \right\}, \tag{2}$$

with the parameterizing set $\mathfrak{S}$ 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 $(A_s; B_s; G_S; C_S)$ (which is called a realization of P_s). A steady state Kalman filter, see Anderson and Moore [1979] for details, can be designed:

$$\hat{x}_s(t+1) = A_s \hat{x}_s(t) + B_s u(t) + H_s\big(y(t) - C_s \hat{x}_s(t)\big), \tag{3a}$$

$$\hat{y}_s(t) = C_s \hat{x}_s(t), \tag{3b}$$

$$H_s = A_s \Sigma_s C_s^T [C_s \Sigma_s C_s^T + R]^{-1} \tag{3c}$$

where Σ_s is the solution of the discrete Riccati equation

$$\begin{aligned} \Sigma_s = {} & A_s \Sigma_s A_s^T + G_s Q G_s^T \\ & - A_s^T \Sigma_s C_s^T [C_s \Sigma_s C_s^T + R]^{-1} C_s \Sigma_s A_s. \end{aligned} \tag{4}$$

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

$$\begin{aligned} x_{s^\star}(t+1) &= A_{s^\star} x_{s^\star}(t) + B_{s^\star} u(t) + G_{s^\star} w(t), \\ y(t) &= C_{s^\star} x_{s^\star}(t) + v(t), \end{aligned} \tag{5}$$

and the real value of the set of matrices $(A_{s^\star}; B_{s^\star}; G_{s^\star}; C_{s^\star})$ is unknown. Let us further assume that two possible model of the physical system are suggested as $(A_r; B_r; G_r; C_r)$ 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 $P_{s^\star}$, described in (5), respectively. Moreover, let $\tilde{y}_r(t)$ and $\tilde{y}_f(t)$ denote the output innovation sequence (residual) from two KFs designed for the model P_r (with realization $(A_r; B_r; G_r; C_r)$) and P_s (with realization $(A_f; B_f; G_f; C_f)$), respectively, and given by

$$\begin{aligned} \tilde{y}_r(t) &= y(t) - \hat{y}_r(t), \\ \tilde{y}_f(t) &= y(t) - \hat{y}_f(t). \end{aligned}$$

We will assume that the residual sequences in all the KFs are stationary and ergodic (see Hassani et al. [2013] for necessary conditions).
Let $Y_t \equiv \{y(0), y(1), \cdots, y(t), u(1), \cdots, u(t)\}$ condense the history of the measurements from the beginning up to time t. Consider the conditional probability density function $f_s\big(y(t)|Y_{t-1}, P_s\big)$ (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 $f_s\big(Y_t|P_s\big) = \prod_{k=1}^{t} f_s\big(y(k)|Y_{k-1}, P_s\big)$.
For two different KFs based on P_r and P_f, if

$$f_f\big(Y_t|P_f\big) > f_r\big(Y_t|P_r\big), \tag{6}$$

or, equivalently, if

$$\log f_f\big(Y_t|P_f\big) > \log f_r\big(Y_t|P_r\big),$$

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 probable than) the KF designed based on model P_r. Define the likelihood ratio for the sequence of Y_t

$$k_r^f\big(Y_t\big) = \frac{f_f\big(Y_t|P_f\big)}{f_r\big(Y_t|P_r\big)} \tag{7}$$

or, equivalently,

$$\log k_r^f\big(Y_t\big) = \log f_f\big(Y_t|P_f\big) - \log f_r\big(Y_t|P_r\big),$$

where $\log k_r^f\big(Y_t\big)$ 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_f and P_r.[2] Similarly, one can compute the conditional likelihood ratio

$$k_r^f\big(y(t)|Y_{t-1}\big) = \frac{f_f\big(y(t)|Y_{t-1}, P_f\big)}{f_r\big(y(t)|Y_{t-1}, P_r\big)} \tag{8}$$

[2] Positive values of $\log k_r^f\big(Y_t\big)$ 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,

$$\log k_r^f\big(y(t)|Y_{t-1}\big) = \log f_f\big(y(t)|Y_{t-1},P_f\big) - \log f_r\big(y(t)|Y_{t-1},P_r\big)$$

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

$$d_t(P_f,P_r) = E\{\log k_r^f\big(y(t)|Y_{t-1}\big)\}. \tag{9}$$

When $d_t(P_f,P_r)$ [3] 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 $P_{s^\star}$ is always preferred over other KFs).

Proposition 1. Let Y_t be the measured data from a system described by model $P_{s^\star}$, then for all KFs designed based on model P_r ($P_r \neq P_{s^\star}$)

$$d(P_{s^\star},P_r) \geq 0, \tag{10}$$

with equality if & only if $f_{s^\star}\big(y_t|Y_{t-1},P_{s^\star}\big) = f_r\big(y_t|Y_{t-1},P_r\big)$.

Proof. Since Y_t is the measured data from a system described by model $P_{s^\star}$, using the KF designed based on model $P_{s^\star}$, $(f_\star\big(y_t|Y_{t-1},\theta_\star\big))$ 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

$$E\{f_{s^\star}\big(y_t|Y_{t-1},P_{s^\star}\big)\} \geq E\{f_r\big(y_t|Y_{t-1},P_r\big)\},$$

with equality if & only if $f_{s^\star}\big(y_t|Y_{t-1},P_{s^\star}\big) = f_r\big(y_t|Y_{t-1},P_r\big)$. Now it is straightforward to get the result. $\square$

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 $P_{s^\star}$ has the form (see Anderson and Moore [1979])

$$f_{s^\star}\big(y(t)|Y_{t-1},P_{s^\star}\big) = \frac{exp\{-\frac{1}{2}\tilde{y}_{s^\star}(t)^T S_{s^\star}^{-1}\tilde{y}_{s^\star}(t)\}}{\sqrt{(2\pi)^q|S_{s^\star}|}}, \tag{11}$$

where q is the dimension of $\tilde{y}_{s^\star}(t)$ and $S_{s^\star} = C_{s^\star}\Sigma_{s^\star}C_{s^\star}^T + R$ 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 $P_{s^\star}$, $f_{s^\star}\big(y(t)|Y_{t-1},P_{s^\star}\big)$, is a gaussian distribution with mean $\hat{y}_{s^\star}(t)$ and covariance $E\{\tilde{y}_{s^\star}(t)\tilde{y}_{s^\star}^T(t)\}$, which we denote by $S_{s^\star}$. [4]

Now let us consider the case that Y_t is the measured data from a system described by model $P_{s^\star}$, but a KF designed based on the model P_r is used for estimation of $y(t)$. It follows that

$$\begin{aligned} &E\log\{f_r\big(y(t)|Y_{t-1},P_r\big)\} \\ &= -\frac{q}{2}\log(2\pi) - \frac{1}{2}\log(|S_r|) - \frac{1}{2}tr(S_r^{-1}E\{\tilde{y}_r^T(t)\tilde{y}_r(t)\}) \\ &= -\frac{q}{2}\log(2\pi) - \frac{1}{2}\log(|S_r|) - \frac{1}{2}tr(S_r^{-1}S_r^{s^\star}) \end{aligned} \tag{12}$$

where $S_r^{s^\star}$ is the covariance of output estimation sequence when the true plant model is $P_{s^\star}$ but the KF is designed based on model P_r. [5]

Now, let us go back to the very last question of the previous section, where Y_t is measured from a physical system whose dynamics can be modeled as $P_{s^\star}$ ($P_{s^\star}$ is not known), and two possible model for the physical system are suggested as P_r and P_f. It is easy to write $d(P_f,P_r)$ as

$$\begin{aligned} &d(P_f,P_r) = \\ &+E\log\{f_f\big(y(t)|Y_{t-1},P_f\big)\} - E\log\{f_r\big(y(t)|Y_{t-1},P_r\big)\} \\ &+\frac{1}{2}\log(|S_r|) + \frac{1}{2}tr(S_r^{-1}S_r^{s^\star}) \\ &-\frac{1}{2}\log(|S_f|) - \frac{1}{2}tr(S_f^{-1}S_f^{s^\star}). \end{aligned} \tag{13}$$

Let [6]

$$\Gamma_r^{s^\star} \equiv \frac{1}{2}\log(|S_r|) + \frac{1}{2}tr(S_r^{-1}S_r^{s^\star}), \tag{14}$$

from which it follows that

$$d(P_f,P_r) = \Gamma_r^{s^\star} - \Gamma_f^{s^\star}. \tag{15}$$

It is also useful to mention that

$$d(s^\star,r) - d(s^\star,f) = \Gamma_r^\star - \Gamma_f^\star$$

so that

$$d(s^\star,r) \geq d(s^\star,f),$$

if and only if

$$\Gamma_r^{s^\star} \geq \Gamma_f^{s^\star}.$$

Theorem 2. For the KFs designed based on models P_r and P_f, under the assumption of ergodicity and stationarity of the residuals (see Hassani et al. [2013]), we have

$$\lim_{t\to\infty} k_f^r\big(Y_t\big) = 0 \tag{16}$$

if and only if

$$\Gamma_r^{s^\star} \geq \Gamma_f^{s^\star} \tag{17}$$

Proof. Note that

$$\log k_f^r\big(Y_t\big) = \sum_{n=1}^{t}\log k_f^r\big(y(n)|Y_{n-1}\big). \tag{18}$$

[3] When the dynamics of the system $P_{s^\star}$ 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 $d_t(P_f,P_r)$ and use $d(P_f,P_r)$ instead.

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

[5] We should highlight here that the notation of the term $\tilde{y}_r(t)$ 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) $\tilde{y}_r(t)$ has the second meaning.

[6] Hassani et al. [2009, 2011] used $\Gamma_r^{s^\star}$ 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 $\log k_f^r\big(y(n)|Y_{n-1}\big)$ as

$$\lim_{t\to\infty}\frac{1}{t}\sum_{n=1}^{t}\log k_f^r\big(y(n)|Y_{n-1}\big) = E\{\log k_f^r\big(y(n)|Y_{n-1}\big)\}$$
$$= d_n(P_r,P_f) = \Gamma_f^{s^\star} - \Gamma_r^{s^\star}. \tag{19}$$

If

$$\Gamma_f^{s^\star} \le \Gamma_r^{s^\star} \tag{20}$$

then by comparing (18), (19), and (20), it follows that

$$\lim_{t\to\infty}\log k_f^r\big(Y_t\big) = \lim_{t\to\infty}\sum_{n=1}^{t}\log k_f^r\big(y(n)|Y_{n-1}\big) = -\infty \tag{21}$$

which implies that

$$\lim_{t\to\infty} k_f^r\big(Y_t\big) = 0. \tag{22}$$

□

This theorem shows that the KF which has the minimum $\Gamma_v^{s^\star}$ (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 $P_{s^\star}$ in some sense. Based on the results of Theorem 2 we can define a Pseudo-norm on the set of LTI systems given by

$$m(P_r,P_f) := |\Gamma_r^{s^\star} - \Gamma_f^{s^\star}|. \tag{23}$$

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

Proof. It is not difficult to see that

$$m(P_r,P_r) = |\Gamma_r^{s^\star} - \Gamma_r^{s^\star}| = 0.$$

To prove the symmetry property, use the fact that

$$m(P_r,P_f) = |\Gamma_r^{s^\star} - \Gamma_f^{s^\star}| = |\Gamma_f^{s^\star} - \Gamma_r^{s^\star}| = m(P_f,P_r).$$

The triangle inequality follows from

$$\begin{aligned} m(P_r,P_p)+m(P_p,P_f) &= \\ |\Gamma_r^{s^\star} - \Gamma_p^{s^\star}| + |\Gamma_p^{s^\star} - \Gamma_f^{s^\star}| &\ge \\ |\Gamma_r^{s^\star} - \Gamma_p^{s^\star} + \Gamma_p^{s^\star} - \Gamma_f^{s^\star}| &= \\ |\Gamma_i^{\star} - \Gamma_f^{s^\star}| &= m(P_r,P_f). \end{aligned}$$

□

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

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

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:

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

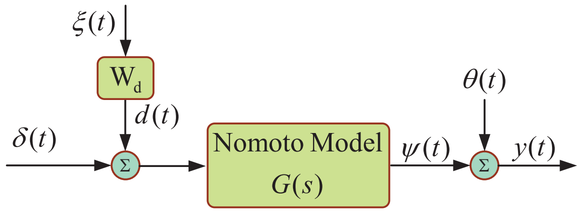

Fig. 1. The Stable Discrete-Time Linear Time-Invariant (LTI) Plant.

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 $-1 \le k \le -0.1$ and $6 \le T \le 14$.

Let us denote by P_k^T the system described in (24). For example $P_{-.15}^8$ denotes the system described in (24) with $k = -0.15$ and $T = 8$. It can be noticed that $m(P_{-.15}^8, P_k^T)_{|_{k=-0.15}^{T=8}} = 0$ 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.

[7] A pseudo-norm or seminorm is a norm that does not satisfy the identity of indiscernibles.

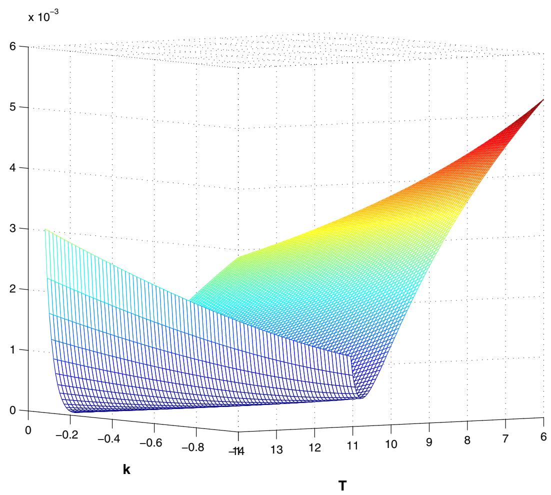

Fig. 2. Distance of the systems with different k and T values from the system with $k = -0.15$ and $T = 8$ as a function of k and T.

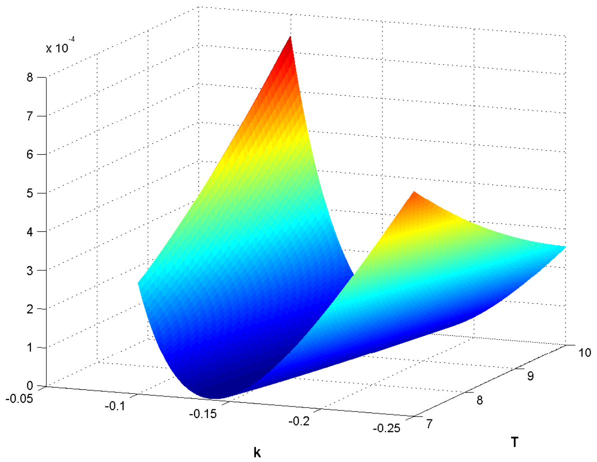

Fig. 3. A zoomed in view of Fig. 2.

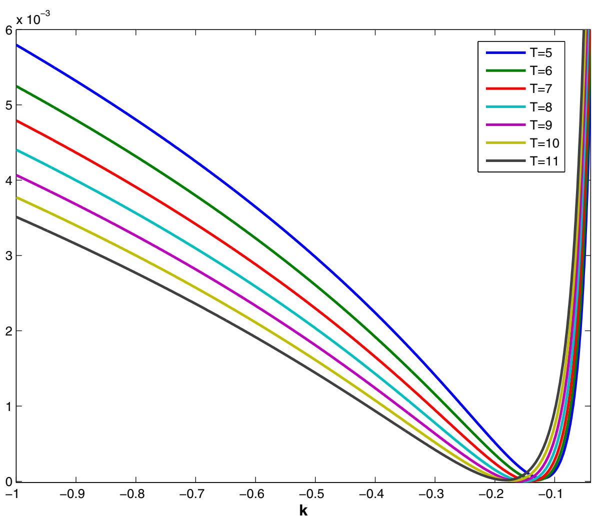

Fig. 4. distance of the systems with different k and T values from the system with $k = -0.15$ and $T = 8$ as a function of k (fixed T).

ACKNOWLEDGEMENTS

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

REFERENCES

Bijan Afsari and René Vidal. The alignment distance on spaces of linear dynamical systems. In *Proc. IEEE Conference on Decision and Control (CDC'13)*, Florence, Italy, December 2013a.

Bijan Afsari and René Vidal. Group action induced distances on spaces of high-dimensional linear stochastic processes. In *Geometric Science of Information*, pages 425–432. Springer, 2013b.

B. D. O. Anderson and J. B. Moore. *Optimal Filtering.* Prentice-Hall, New Jersey, USA, 1979.

Y. Baram and N.R. Sandell. An information theoretic approach to dynamical systems modeling and identification. *IEEE Trans. on Automat. Contr.*, 23:61–66, 1978.

Stephen P Boyd and Craig H Barratt. *Linear controller design: limits of performance.* Prentice Hall Englewood Cliffs, NJ, 1991.

Katrien De Cock and Bart De Moor. Subspace angles between ARMA models. *Systems & Control Letters*, 46 (4):265–270, 2002.

Bernard Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems.* PhD thesis, Erasmus Universiteit Rotterdam, Netherlands, 1986.

Vahid Hassani, A. Pedro Aguiar, Antonio M. Pascoal, and Michael Athans. A performance based model-set design strategy for multiple model adaptive estimation. In *ECC'09 - European Control Conference*, Budapest, Hungary, 2009.

Vahid Hassani, J. Hespanha, Michael Athans, and Antonio M. Pascoal. Stability analysis of robust multiple model adaptive control. In *Proc. of The 18th IFAC World Congress*, Milan, Italy, 2011.

Vahid Hassani, Antonio M. Pascoal, and A. Pedro Aguiar. Multiple model adaptive estimation for open loop unstable plants. In *Proc. ECC'13 - European Control Conference*, Zurich, Switzerland, 2013.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

Richard J Martin. A metric for ARMA processes. *Signal Processing, IEEE Transactions on*, 48(4):1164–1170, 2000.

James R Munkres. *Topology*, volume 2. Prentice Hall Upper Saddle River, 2000.

K. Nomoto, T. Taguchi, K. Honda, and S. Hirano. On the steering qualities of ships. *International Shipbuilding Progress*, 4(4):354–370, 1957.

Mihály Petreczky and René Vidal. Metrics and topology for nonlinear and hybrid systems. In *Hybrid Systems: Computation and Control*, pages 459–472. Springer, 2007.