Pole-based Distance Measure for Change Detection in Linear Dynamic Systems

Christof Chlebek and **Uwe D. Hanebeck**
Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT), Germany
christof.chlebek@kit.edu, uwe.hanebeck@ieee.org

***Abstract*—In this work, we derive a distance measure for the detection of changes in the behavior of linear dynamic single-input-single-output (SISO) systems based on input-output data. The distance is calculated as a function of the system poles, which are directly estimated from the given data. Poles represent a system as a set and have no identities, which is analogous to the nature of association-free multi-target tracking. This motivates the application of set distances known from multi-target tracking, namely the optimal subpattern assignment (OSPA) distance. Thus, the OSPA distance as well as a modification, the MAX-OSPA distance, are formulated as pole-distances between dynamic systems. In this formulation, the OSPA distance finds the optimal assingment by minimizing over the sum of distances between poles. The MAX-OSPA chooses an optimal assignment by minimizing the maximum distance between two poles. The proposed distances are evaluated in several simulations comparing the deterministic OSPA and MAX-OSPA to a state-of-the-art metric for autoregressive-moving-average (ARMA) processes, as well as OSPA and MAX-OSPA using the direct pole estimation and a two step-pole estimation utilizing recursive ARX (AutoRegressive model with eXogenous input) system identification.**

I. Introduction

Change detection considers the problem of finding modifications to the system based on observations. Depending on the origin, specific formulation, and application, there are several active research fields addressing this highly relevant problem. The field of anomaly detection is concerned with finding elements or behavioral patterns deviating from elements or patterns that are defined as normal [1]. The detection of outliers addresses the problem of finding elements of a set that deviate markedly from other set members [2]. Even though the expressions and the interpretation of anomaly and outlier are distinct, both terms are sometimes used interchangeably in literature. Originated in industrial applications, the research on fault detection is concerned with the problem of recognizing system behavior or states that exceed a permitted region of operation [3], [4]. Very slow and gradual changes in processes are studied under the expression of concept drift [5]. Finally, novelty detection emerged from the field of machine learning and has its focus on discerning incoming information with respect to its coverage by already learned models and pattern [6]. In all these research topics, change detection can be seen as a neutral interpretation of the stated problem.

In this work, we are focusing on the detection of changes in the parameters of dynamic systems, where the system output depends on the system state and input. Therefore, the system output signal cannot be compared directly to a reference signal. There are three basic approaches to address this problem. First, a reference system is identified and the system state is tracked over time. Using state and input, the expected system output is used to calculate the innovation or residual of the measured system output, which is then utilized as basis for decision-making. This validation of output data to the current state and input, subject to the identified model, is effective for tests of abrupt changes as well as single outlier detection. In order to not only consider one state and input-output pair at a time, but also to be able to detect slow and subtle changes in the system cumulative sum approaches are applied [7]. The second approach is a two system test, where a test system is continuously identified based on actual system input-output data, additional to the reference system. Then a sliding time frame of constant length consisting of the latest data samples is tested on both systems, which gives the basis for decision-making. This was introduced first using a likelihood ratio test [8]. In this work, we consider the third approach, which is to compare two systems based on their parameters.

A direct comparison of systems instead of using available signals abstracts from individual realizations, i.e., state and input-output pairs, and thus, facilitates the comparison of the system properties itself. One limitation of this method is that this may not be possible for every system representation. Thus, key to this approach is a set of representative and intuitive parameters, in order to define a distance measure as a basis for decision-making. There are two main requirements for the parameters. First, two different sets of parameters have to represent two different input-output systems, and second, the magnitude of the behavior variation has to be evident in the parameter variation. One common distance measure between systems is based on the spectral characteristics, namely a distance between the cepstra, the inverse Fourier transform of the logarithm of the power spectrum [9]. In [10], Martin derived the ARMA distance, a metric for ARMA processes as the distance of two cepstra, which can be calculated by a function of the system poles. Since poles also have an intuitive interpretation and meet the stated requirements, we are interested in this system representation and will derive a new distance measure for linear dynamic single-input-single-output (SISO) systems. In this type of system, the poles are equivalent to the eigenvalues of the system matrix in state-space formulation.

One approach for the estimation of system poles is to estimate the coefficients of the system matrix in a first step, then to determine the eigenvalues of the system matrix, which requires polynomial root calculation. For small variations in

the parameters, an effective way is to track the roots using homotopy continuation [11]. Another approach is to calculate the poles directly from data. The main benefit is that information about identification uncertainty is obtained directly in pole space. The first method for direct pole estimation was introduced by Nehorai et al. utilizing a recursive least squares method (RLS) [12]. Weruaga derived a method for direct pole estimation on spectral samples [13]. Third, a Bayesian approach to direct pole estimation was derived in [14], which will be employed in this work.

There are several proposals in literature using information on the position of poles as an indicator of changes and anomalies. In [15], a vibration test for fault detection of ball bearings is introduced. Amplitudes of characteristic frequencies are examined, which are represented by the closeness of a system pole to the unit circle. Transformed into subspace formulation, several applications of the ARMA distance were proposed [16]. In the field of civil engineering, it is applied as a damage indicator for structural health monitoring (SHM) [17]. Utilizing two sliding windows, a two-model approach for change point detection was derived using this distance in [18]. In [19] a surveillance application with detection of anomalies in crowd motion is presented.

In this work, the main contribution is the derivation of two novel distance measures for linear dynamic SISO systems. The analogy between a set poles, which are free of identities, and association-free multi-target tracking, motivates the key idea. That is the application of a set distance known from the context of multi-target tracking and adopt it as a distance of dynamic systems. Using an association-free distance measure, namely the optimal subpattern assignment (OSPA) distance [20], we define the system distance as a distance between their sets of poles representing the spectral characteristics of the system. The distance is derived as a function of point estimates as well as random variables. This facilitates a direct representation of the parameter estimation quality as part of the novel system distance. The need to consider changes in single frequencies of the spectral characteristics (i.e., movements of single poles) necessitates a distance based on the maximum distance between two single poles, which is derived from the OSPA distance and will be referred to as MAX-OSPA.

The remainder of this paper is structured as follows. First, in Sec. II we give a formal description of the considered problem. Sec. III introduces the applied system identification method, a Bayesian approach to direct pole estimation, where the information about the parameter uncertainty is available in pole-space. We derive a novel metric in pole-space for the comparison of two dynamic systems in Sec. IV, exploiting this uncertainty information. The effectiveness of the presented algorithm is evaluated in Sec. V by means of several simulations. Finally, Sec. VI concludes the work.

The following notation is used: Deterministic quantities a and random variables $\boldsymbol{a}$ are distinguished by normal lettering and bold face letters, respectively. The notation $\boldsymbol{a} \sim f(a)$ describes the characterization of $\boldsymbol{a}$ by its probability density function $f(a)$. Finally, a vector $\underline{a}$ is indicated by an underscore and a matrix $\mathbf{A}$ will be denoted as a bold face capital letter.

II. Problem Formulation

In this work, we consider the problem of finding a distance function characterizing the distance d_k between a reference system $S_k^{(r)}$ and a test system $S_k^{(t)}$ calculated on basis of their parameter vectors $\underline{\theta}_k^{(r)}$ and $\underline{\theta}_k^{(t)}$, which are estimated from input-output data, i.e.,

$$D\left(S_k^{(r)}, S_k^{(t)}\right) := D\left(\underline{\theta}_k^{(r)}, \underline{\theta}_k^{(t)}\right) = d_k \ . \tag{1}$$

The superscripts (r) and (t) indicate the reference and the test system, respectively. Throughout the work, we consider linear time-variant single-input-single-output (SISO) systems of the form $S_k = \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ given in the control canonical state-space form

$$\begin{aligned} \underline{\boldsymbol{x}}_{k+1} &= \mathbf{A}\underline{\boldsymbol{x}}_k + \mathbf{B}(u_k + \boldsymbol{w}_k) \ , \\ \boldsymbol{y}_k &= \mathbf{C}\underline{\boldsymbol{x}}_k + \boldsymbol{v}_k \ , \end{aligned} \tag{2}$$

with the state vector $\underline{\boldsymbol{x}}_k$, the system input u_k and output $\boldsymbol{y}_k$, zero-mean white system noise $\boldsymbol{w}_k$, and zero-mean white measurement noise $\boldsymbol{v}_k$. The matrices $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are characterized by

$$\mathbf{A} = \begin{bmatrix} -a_1 & -a_2 & \dots & -a_{p-1} & -a_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} \ , \tag{3}$$

$\mathbf{B} = [1, 0, 0, \dots, 0]^\top$, and $\mathbf{C} = [b_1, b_2, b_3, \dots, b_p]$, assuming that $\mathbf{C}$ is given. We would like to emphasize that we are considering time-variant systems. For the reason of better readability, we will omit the time index k in the matrices and the single parameters. In general, a parameter vector $\underline{\theta}_k$ of a system $S_k = \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ comprises a function of S_k and will be further specified later in this work.

In the following, we specify two problem formulations. First, a general formulation of the distance between two systems given input-output data is regarded. Second, a more specific formulation for change detection in the system behavior is derived.

A. Distance Between Systems

Given are two input-output time series $W_{1:N}^{(r)} = \left[u_{1:N}^{(r)}, y_{1:N}^{(r)}\right]$ and $W_{1:N}^{(t)} = \left[u_{1:N}^{(t)}, y_{1:N}^{(t)}\right]$ subject to $S_k^{(r)} = \left\{\mathbf{A}^{(r)}, \mathbf{B}^{(r)}, \mathbf{C}^{(r)}\right\}$ and $S_k^{(t)} = \left\{\mathbf{A}^{(t)}, \mathbf{B}^{(t)}, \mathbf{C}^{(t)}\right\}$, respectively. The superscripts (r) and (t) again indicate the reference and the test system, respectively. The problem is to constantly estimate the parameter sets $\underline{\theta}_k^{(r)}$ and $\underline{\theta}_k^{(t)}$ simultaneously with the unknown system states $\underline{\boldsymbol{x}}_k^{(r)}$ and $\underline{\boldsymbol{x}}_k^{(t)}$ from $W_{1:N}^{(r)}$ and $W_{1:N}^{(t)}$ and calculate a distance d_k as defined in (1).

In system identification, the system matrix $\mathbf{A}$ is typically estimated based on the input-output time series $W_{1:N}$ by traditional approaches (e.g., Burg's method [21]) using a sliding window or a recursive formulation. Note that we are omitting the superscripts here, meaning both reference and the test system. A formulation of the parameter vector $\underline{\theta}_k$ directly based on coefficients a_i does not allow for a representative distance function, and thus, has to be transformed appropriately.

B. Change Detection

In change detection, only a single system $S_k = \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ is involved, given by the input-output time series $W_{1:N} = [u_{1:N}, y_{1:N}]$. The time series $W_{1:N}$ can be subdivided into

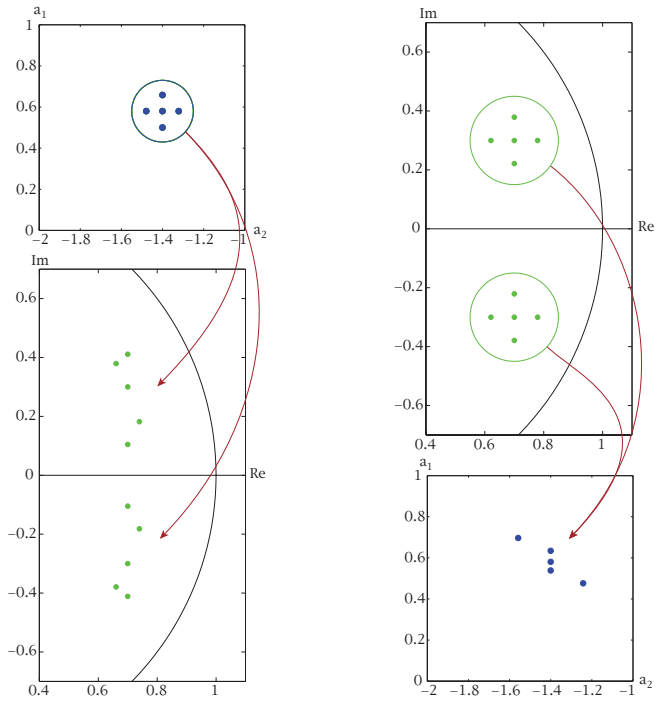


Fig. 1. Visualization of the sample-based transformations between pole and coefficent space. Before transformation of an estimated parameter, the density describing this parameter has to be sampled first, then transformed sample-wise. The coefficient space is marked in blue, the pole space in green. Each parameter sample corresponds to two complex conjugate poles.

$W_{1:n-1}$ and $W_{n:N}$, where time step n is the change point. We define the reference system as $S_{\tilde{n}}^{(r)}$ subject to $W_{1:\tilde{n}}$, where $\tilde{n}$ is sufficiently large for the identification of $\underline{\theta}_{\tilde{n}}^{(r)}$ and the last considered timestep of the reference system $\tilde{n} \ll n$. Furthermore, we define the test system as $S_k^{(t)}$ subject to $W_{\tilde{n}+1:N}$. The problem is to calculate the distance between systems $d_k = D\big(\underline{\theta}_{\tilde{n}}^{(r)}, \underline{\theta}_k^{(t)}\big)$ analogous to Sec. II-A. In other words, we first need to identify a reference system using enough data, using only data prior to the change point. The rest of the data is used to continuously estimate the test system, which is compared to the reference system every time step.

III. Direct Pole Estimation

There are two approaches for the estimation of the poles of a system. First, there is a two-step approach, where initially the coefficients a_i of the matrix $\mathbf{A}$ are estimated. Then, the poles are obtained by calculation of the eigenvalues of $\mathbf{A}$. Second, there is a one-step approach. In this so-called *direct pole estimation* approach, the poles are directly estimated from input-output data.

The disadvantage of the two-step approach is that information about the estimation quality is at first available in the coefficient space. This information is given in the form of a probability density function and has also to be transformed into pole space, which can be performed by sampling from the density and sample-wise transformation. Thereby, every sample corresponds to one set of all p coefficents. Thus, for each sample the zeros of the p-order characteristic polynomial have to be calculated. On the contrary, using the one-step approach, the poles are obtained including available information about the estimation quality directly in pole-space.

In the following, we introduce the Bayesian direct pole estimation in a general form for linear SISO systems [14]. In systems with real valued output, poles occur either real or in complex conjugate pairs. In order to ensure this constraint in an effective way, we later give a formulation for a known proportion of real and complex poles. This assumption is not unrealistic for change detection problems, since an a prior estimate is easy to attain from a small number of data samples and a change in the pole proportion results in a system change.

Let a system S_k be given by (2) and $\underline{\theta}_k$ be the vector of unknown poles characterizing S_k given by

$$\underline{\theta}_k := [\lambda_1, \ldots, \lambda_p]^\top \ , \tag{4}$$

where every $\lambda_i = \sigma_i + j\omega_i$ is one complex pole composed of the real part σ_i and the imaginary part ω_i, with the imaginary unit $j^2 = -1$. Note that the time index is omitted in every λ_i and its components for better readability. Also, we would like to emphasize that all σ_i and ω_i have to be estimated separately, thus, the length of the parameter vector to be estimated is $2p$, where p denotes the system order. Using state augmentation in order to calculate the state and parameter $\underline{x}_k$ and $\underline{\theta}_k$ simultaneously, we restate the system formulation by

$$\begin{bmatrix} \underline{x}_{k+1} \\ \underline{\theta}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \underline{x}_k \\ \underline{\theta}_k \end{bmatrix} + \begin{bmatrix} 1 \\ \underline{0} \end{bmatrix} (u_k + w_k) \ , \tag{5}$$

$$y_k = [\mathbf{C} \quad \underline{0}] \begin{bmatrix} \underline{x}_k \\ \underline{\theta}_k \end{bmatrix} + \begin{bmatrix} v_k \\ \underline{0} \end{bmatrix} ,$$

where the matrix $\mathbf{A}$ is given by (3), with the parameters a_i directly calculated from $\underline{\theta}_k$ by

$$\begin{aligned}
a_0 &= (-1)^0 \ , \\
a_1 &= (-1)^1(\lambda_1 + \lambda_2 + \ldots + \lambda_{p-1} + \lambda_p) \ , \\
a_2 &= (-1)^2(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \ldots + \lambda_1\lambda_p + \lambda_2\lambda_3 + \ldots) \ , \\
a_3 &= (-1)^3(\lambda_1\lambda_2\lambda_3 + \ldots + \lambda_1\lambda_2\lambda_p + \lambda_2\lambda_3\lambda_4 + \ldots) \ , \\
&\vdots \\
a_p &= (-1)^p(\lambda_1\lambda_2\lambda_3 \cdot \ldots \cdot \lambda_{p-1}\lambda_p) \ ,
\end{aligned}$$

which can be rewritten as

$$a_i = (-1)^i \cdot \sum_{\substack{M \subseteq \{1..p\} \\ |M| = i}} \left(\prod_{m \in M} \lambda_m \right) \ , \tag{6}$$

i.e., the coefficients a_i are replaced by the sum of all products of poles consisting of i factors of all subsets M with i elements.

The matrix $\mathbf{A}$ can be easily constructed by the power set of all poles. Each a_i is calculated from subsets with i elements of the power set, which are element-wise multiplied and summed up. Note that a power set has 2^p subsets. Thus, the complexity of one evaluation of (5) is in $\mathcal{O}(2^p)$ for a dynamic system with order p. We recommend using this approach only for systems with $p < 20$. We also want to emphasize that a power set is independent of the sorting of its elements. This results in a function comprising a symmetry over the ordering of all elements in $\underline{\theta}_k$. As a consequence, there are $p!$ potentially correct results of estimating all λ_i, namely all permutations of

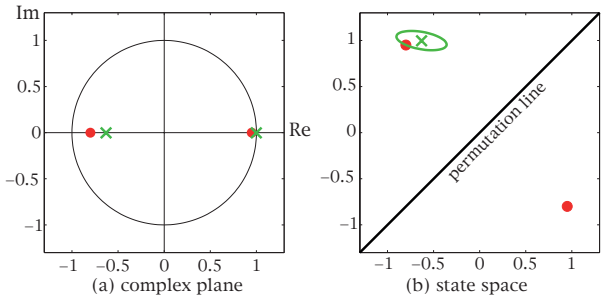


Fig. 2. Illustration of the challenge occuring in the representation of sets. If a set is represented as a vector, all order permutations are equivalent. Geometrically this can be represented by $p!$ permutation hyperplanes in state space, at which the estimate comprising all poles can be reflected. For $p = 2$ there is one permutation line. The green cross shows the mean of an estimated pole set and the ellipsoid its uncertainty. The red dots are the actual system poles. Note that in the state space the red dot is on both sides of the permutation line, whereas the estimation uses an ordered vector as representation and thus, is only on one side.

$\underline{\theta}_k$. This shows that actually we are working on a set of poles with no specific identities. However, this set is arranged in the vector $\underline{\theta}_k$. This problem is well known from association-free multi-target tracking, where one is interested in the positions of a set of targets, disregarding their identities. We have visualized both potential correct results of a system with two real poles in Fig. 2. The two-dimensional state space is projected onto the complex plane with two poles.

Under the assumption of a known proportion of real and complex poles, and exploiting the fact that the output is real-valued, we can give a more efficient formulation. In this case, complex poles are given by complex conjugate pairs, and thus, we can formulate a parameter vector $\tilde{\underline{\theta}}_k$ with reduced length p. We define real poles as $\lambda_i^R := \sigma_i^R$ and complex conjugate pole pairs as $\lambda_i^C := \sigma_i^C + j \cdot \omega_i^C$ and $\lambda_i^{C*} := \sigma_i^C - j \cdot \omega_i^C$, respectively, with $\omega_i^C > 0$. The set of poles is then given by

$$\{\lambda_i^R | i = 1, \ldots, S\} \cup \{\lambda_i^C, \lambda_i^{C*} | i = 1, \ldots, T\} \ , \qquad (7)$$

with $p = S + 2T$. Using this formulation, the parameter vector can be written as

$$\tilde{\underline{\theta}}_k = \left[\sigma_1^R, \ldots, \sigma_S^R, \sigma_1^C, \ldots, \sigma_T^C, \omega_1^C, \ldots, \omega_T^C\right]^\top \ , \qquad (8)$$

which can be mapped back to $\underline{\theta}_k$ in the form (4) in order to calculate (6). This formulation allows to ensure that the system output is real-valued and removes all redundancy in the estimation.

Finally, for a given input-output time series $W_{1:N}$, we can estimate $[\underline{x}_k, \underline{\theta}_k]^\top$ by recursive Bayesian estimation. We would like to point out that the stated problem is nonlinear with the matrix $\mathbf{A}$ being directly dependent on $\underline{\theta}_k$. Hence, the commonly used Kalman Filter can only be used in the measurement update step. For the time update step an effective choice is, for example, a Linear Regression Kalman Filter (LRKF), such as the Unscented Kalman Filter (UKF) [22] or the Smart Sampling Kalman Filter (S^{2}KF) [23]. Using this approach the parameters are obtained as a Gaussian distributed random variable $\underline{\boldsymbol{\theta}}_k \sim \mathcal{N}(\hat{\underline{\theta}}_k, \mathbf{C}_k^\theta)$. In the next section we use this obained random variable for the calculation of a distance between systems. The parameter vector $\underline{\boldsymbol{\theta}}_k$ can be employed using the mean, or has to be transformed into a sample-based form. This can, for example, be done by the sampling technique of the UKF or S^{2}KF.

IV. Pole-based Distance

In the previous section, we have introduced the direct estimation of poles and have shown that a system representation in pole space has no fixed ordering in the estimated parameter vector, i.e., we are actually estimating a set. Comparing two sets requires specific distance measures, where the correspondence of the present elements is implicitly regarded. This fact is illustrated in Fig. 2. This problem is well known from multi-target tracking, which was a major inspiration for the exploitation of association free distance measures for pole-based distances between systems. We would like to emphasize that the correspondence problem of pole sets occurs independent of the applied pole estimation method.

In the following, we derive the base distance of poles, which is the distance between two single poles in the complex plane. Then, we introduce two association free distance measures, namely the optimal subpattern assignment (OSPA) distance [20] and a modification, the MAX-OSPA. The MAX-OSPA uses the OSPA assignment minimizing the maximum base distance, then calculates this maximum base distance.

A. Base Distance of Poles

First, we introduce the so called base distance $b(\cdot, \cdot)$, with $b : \mathbb{C} \times \mathbb{C} \to \mathbb{R}_0^+$ as a function satisfying identity, symmetry, and the triangle inequality. The base distance is the distance between two complex poles poles $\lambda_i^{(r)}$ and $\lambda_i^{(t)}$. The z-plane, in which the poles of a discrete-time system are defined, can be interpreted as a specific Möbius transformation from the more intuitive s-plane of a continuous-time system that transforms the Cartesian coordinate system of the s-plane into a polar coordinate system. The complete left hand side of the s-plane is thereby compressed into the unit circle. Thus, we will define the distance function in the s-plane and transform it into the z-plane. In order to ensure unambiguity, we write s for a pole in the s-plane and z for a pole in the z-plane in the following definitions.

Definition 1 (Continuous-Time Base Distance) Let s and $\tilde{s}$ be two poles of a continuous-time system in the s-plane. The *base distance* $b(\cdot, \cdot)$ between s and $\tilde{s}$ is defined as

$$b(s, \tilde{s}) = \sqrt{(\mathrm{Re}(s) - \mathrm{Re}(\tilde{s}))^2 + (\mathrm{Im}(s) - \mathrm{Im}(\tilde{s}))^2} \ ,$$

i.e., the Euclidean distance between s and $\tilde{s}$ in the complex plane.

We will now transform this distance into the z-plane using the general transformation

$$s = \frac{1}{T} \ln(z) \ , \qquad (9)$$

where s is a point in the s-plane, z is a point in the z-plane, and T is the sampling rate. For s given in complex form $s = \sigma + j\omega$ and z in polar form $z = r \cdot e^{j\rho}$ we can rewrite (9) by

$$\sigma + j\omega = \frac{1}{T} \ln(r \cdot e^{j\rho}) = \frac{1}{T} \ln(r) + \frac{1}{T} j\rho \ .$$

Hence, the transformation yields

$$\sigma = \frac{1}{T} \ln(r), \ \omega = \frac{1}{T} \rho \ ,$$

which motivates the following definition.

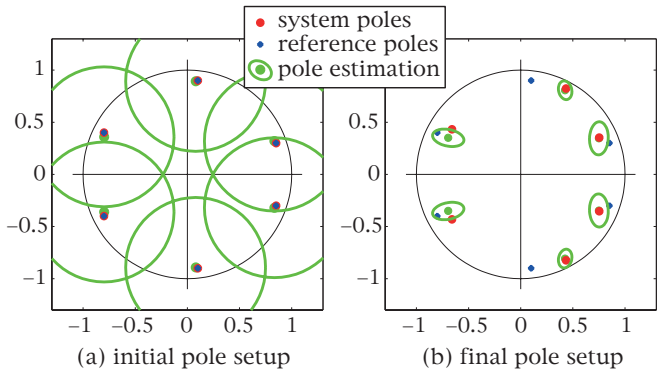


Fig. 3. Initial and resulting pole setups describing the system, as well as the projection onto the complex plane of the direct pole estimation with its initializiation and an exemplary outcome showing the final time step of the run depicted in Fig. 5. The ellipsoids visualize the 3σ bound of the pole estimation uncertainty.

Definition 2 (Discrete-Time Base Distance) Let z and $\tilde{z}$ be two poles of a discrete-time system given in z-plane. The base distance between the poles z and $\tilde{z}$ is defined as

$$b(z,\tilde{z}) = \frac{1}{T}\sqrt{\left(\ln(r)-\ln(\tilde{r})\right)^2+(\rho-\tilde{\rho})^2}\ , \qquad (10)$$

where r and $\tilde{r}$, and ρ and $\tilde{\rho}$ are the polar coordinates of z and $\tilde{z}$, respectively. The constant T is the sampling rate.

Since the distance was derived from the Euclidean distance L_2 by substitution, it can be easily shown that identity, symmetry, and the triangle inequality are satisfied.

From stability analysis, it is well known that poles close to the imaginary axis in the s-plane or close to the unit circle in the z-plane have a strong effect on the system behavior. We propose to weight (10) by

$$\tilde{b}\left(\lambda^{(r)},\lambda^{(t)}\right) = w\left(\lambda^{(r)},\lambda^{(t)}\right)\cdot b\left(\lambda^{(r)},\lambda^{(t)}\right)\ ,$$

with the exponential weighting function

$$w\left(\lambda^{(r)},\lambda^{(t)}\right) = e^{r^{(r)}r^{(t)}}\ ,$$

in order to represent these relevant properties in the distance measure.

B. OSPA Distance

As already mentioned, the challenge of calculating distances between sets is especially known from association-free multi-target tracking. Considering the association of pole tracking and multi-object tracking, we are introducing a label-free distance measure as a distance between systems, namely the optimal subpattern assignment (OSPA) distance [20].

Definition 3 (OSPA [20]) The optimal subpattern assignment (OSPA) distance between two sets of poles $\left\{\lambda_1^{(r)},\dots,\lambda_p^{(r)}\right\}$ and $\left\{\lambda_1^{(t)},\dots,\lambda_p^{(t)}\right\}$, each comprising p elements and given by the ordered vectors $\underline{\theta}_k^{(r)} = \left[\lambda_1^{(r)},\dots,\lambda_p^{(r)}\right]^\top$ and $\underline{\theta}_k^{(t)} = \left[\lambda_1^{(t)},\dots,\lambda_p^{(t)}\right]^\top$ is defined by

$$D(\underline{\theta}_k^{(r)},\underline{\theta}_k^{(t)}) = \left(\frac{1}{p}\min_{\pi\in\Pi_p}\left(\sum_{i=1}^{p} b\left(\lambda_i^{(r)},\lambda_{\pi(i)}^{(t)}\right)^q\right)\right)^{1/q}\ , \qquad (11)$$

with $q\in\mathbb{R}$ and $q\geq 1$. The function $b\left(\lambda_i^{(r)},\lambda_j^{(t)}\right)$ is the base distance between the two elements $\lambda_i^{(r)}$ and $\lambda_j^{(t)}$. The expression Π_p describes all permutations of the set $\{1,\dots,p\}$ and the notation $\lambda_{\pi(i)}^{(t)}$ is the i-th element of the permutation π of $\underline{\theta}_k^{(t)}$ that is generated by reordering the vector.

Hence, the OSPA distance is a function of two vectors, where the association problem is solved implicitly by finding the minimal sum of all pairwise distances and can be implemented effectively by the Hungarian algorithm [24]. In this work, we are using the quadratic measure ($q=2$). Considering the result of Sec. III, the estimated parameter vector leads to a formulation of a distance between random vectors. In the following, we will omit the time index k in favor of readability.

Let $\underline{\boldsymbol{\theta}}^{(r)}\sim f\left(\underline{\theta}^{(r)}\right)$ and $\underline{\boldsymbol{\theta}}^{(t)}\sim f\left(\underline{\theta}^{(t)}\right)$ be two independent random vectors. Thus, the OSPA distance

$$D\left(\underline{\boldsymbol{\theta}}^{(r)},\underline{\boldsymbol{\theta}}^{(t)}\right) = \boldsymbol{d}$$

is itself a random variable $\boldsymbol{d}\sim f(d)$, with the probability density $f(d)$ given by

$$f(d) = \iint f\left(d|\underline{\theta}^{(r)},\underline{\theta}^{(t)}\right)\cdot f\left(\underline{\theta}^{(r)}\right) f\left(\underline{\theta}^{(t)}\right)\ \mathrm{d}\underline{\theta}^{(r)}\mathrm{d}\underline{\theta}^{(t)}.$$

Using $f\left(d|\underline{\theta}^{(r)},\underline{\theta}^{(t)}\right) = \delta\left(d-D\left(\underline{\theta}^{(r)},\underline{\theta}^{(t)}\right)\right)$, we can then write

$$f(d) = \int\int \delta\left(d-D\left(\underline{\theta}^{(r)},\underline{\theta}^{(t)}\right)\right)\cdot f\left(\underline{\theta}^{(r)}\right) f\left(\underline{\theta}^{(t)}\right)\ \mathrm{d}\underline{\theta}^{(r)}\mathrm{d}\underline{\theta}^{(t)}\ . \qquad (12)$$

Unfortunately, the distance function $D(\cdot,\cdot)$ is nonlinear and thus, this integral is not analytically solvable for arbitrary probability densities. Hence, we will specify the calculation of the $f(d)$ for the special case of $f\left(\underline{\theta}^{(r)}\right)$ and $f\left(\underline{\theta}^{(t)}\right)$ given as Dirac mixture densities, which can be obtained by deterministic or random sampling from arbitrary probability density functions.

Let an estimated parameter vector $\underline{\boldsymbol{\theta}}$ be characterized by a Dirac mixture density of the form

$$\tilde{f}(\underline{\theta}) = \sum_{j=1}^{L} w_j\cdot\delta(\underline{\theta}-\underline{\theta}_j)\ , \qquad (13)$$

where $\delta(\cdot)$ denotes the Dirac-delta function and $\underline{\theta}_j$ with $j=1,\dots,L$ is the position of the j-th Dirac component. The component weights are given by $w_j>0$ with $\sum_{j=1}^{L} w_j = 1$. Each Dirac component $\underline{\theta}_j$ can be interpreted as one realization of the random vector $\underline{\boldsymbol{\theta}}$, i.e., represents one set of poles.

Using $\underline{\boldsymbol{\theta}}^{(r)}\sim\tilde{f}\left(\underline{\theta}^{(r)}\right)$ and $\underline{\boldsymbol{\theta}}^{(t)}\sim\tilde{f}\left(\underline{\theta}^{(t)}\right)$ in accordance with the density formulation (13), we can simplify (12) to

$$\tilde{f}(d) = \sum_{j=1}^{L^{(r)}}\sum_{i=1}^{L^{(t)}} w_j^{(r)}w_i^{(t)}\cdot\delta\left(d-D\left(\underline{\theta}_j^{(r)},\underline{\theta}_i^{(t)}\right)\right)\ ,$$

which corresponds to the Cartesian product of both Dirac mixture densities representing the poles propagated through $D(\cdot,\cdot)$. Summing up, the OSPA distance between two random vectors, characterized by Dirac mixture densities, is itself a

random variable, also characterized by a Dirac mixture density. The distance function is evaluated at every combination of Diracs $\left[\underline{\theta}_j^{(r)}, \underline{\theta}_i^{(t)}\right]$. In the common case, where the reference model $S^{(r)}$ is identified precisely enough to be represented by a point estimate, this reduces to $L^{(t)}$ evaluations of (11). The resulting random variable can then be used in the change detection decision making process, e.g., by evaluating the mean and variance of the distance.

C. MAX-OSPA Distance

The OSPA distance calculates the sum of point to point distances, while resolving the assignment problem, by choosing the minimum over all possible sums. Problems with this approach can arise when many points are tested, but only few – in the extreme only one – are strongly moving. This reduces the visibility of the strong movement. In the case of testing system poles, we get a reduced visibility of strong changes in single frequencies, when the rest of the spectral characteristics stays unchanged. Hence, we give an alternative implementation using the maximum measure. In the following, we define the MAX-OSPA distance.

Definition 4 (MAX-OSPA) Given two sets of poles by the two vectors $\underline{\theta}_k^{(r)}$ and $\underline{\theta}_k^{(t)}$, the MAX-OSPA distance is defined by

$$D_{\max}\left(\underline{\theta}_k^{(r)}, \underline{\theta}_k^{(t)}\right) := \max_{i=1,\dots,p} \left(b\left(\lambda_i^{(r)}, \lambda_{\pi^*(i)}^{(t)}\right)\right) \ , \qquad (14)$$

where π^* is the optimal OSPA permutation, given by

$$\pi^* := \text{OSPA}\left(\underline{\theta}_k^{(r)}, \underline{\theta}_k^{(t)}\right) = \arg\min_{\pi\in\Pi_p} \max_{i=1,\dots,p} b\left(\lambda_i^{(r)}, \lambda_{\pi(i)}^{(t)}\right) \ .$$

In other words, we are choosing the maximum base distance between two single elements of a pair $\underline{\theta}_k^{(r)}$ and $\underline{\theta}_k^{(t)}$, while using the optimal subpattern assignment over the maximum norm ($q \to \infty$). Thus, we are minimizing the maximal base distance. Using this formulation, a higher sensitivity to single frequency changes in the spectral characteristics of a system can be attained.

V. Evaluation

In order to demonstrate the value of the presented distance measure, we evaluate it in several simulations. We have used a linear system comprising six complex poles (i.e., system order $p = 6$) and examined both gradual and sudden changes. We have used the systems given by their poles

$$S^{(1)} = \begin{bmatrix} -0.80 \pm 0.40j \\ 0.10 \pm 0.90j \\ 0.85 \pm 0.30j \end{bmatrix}, \quad S^{(2)} = \begin{bmatrix} -0.65 \pm 0.45j \\ 0.45 \pm 0.80j \\ 0.75 \pm 0.35j \end{bmatrix},$$

where we assume that $S^{(r)} = S^{(1)}$ is known and after the change point n the system shifts to $S^{(2)}$. A gradual change is performed by linear progression of the position of system poles from $S^{(1)}$ to $S^{(2)}$. This setup is visualized in Fig. 3 with the initial and final pole setups, as well as a projection of the estimation initialization and an exemplary outcome. The measurement noise and the input noise are given by $\boldsymbol{v_k} \sim \mathcal{N}(0, 0.1)$ and $\boldsymbol{w_k} \sim \mathcal{N}(0, 0.1)$, respectively. Using this setup we have randomly generated two types of data sets for the evaluations. First, we generated an input-output data set with the input $u_k \in [-0.5, 0.5]$ alternating every 100 time steps comprising a gradual change between time steps $k = 2000, \dots, 4000$. Second, we generated an output only data set with a sudden change at $k = 3000$. In this case, system excitation arises from extrinsic process noise only. An exemplary outcome of the data sets is depicted in Fig. 4 (a) and Fig. 4 (b), respectively.

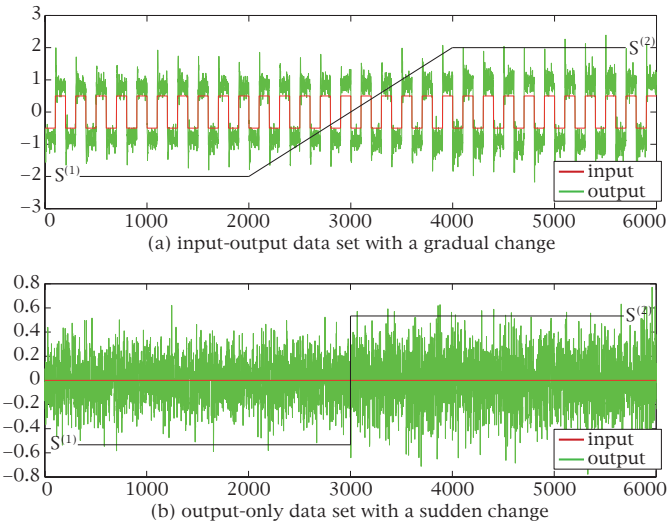


Fig. 4. Two exemplary data sets, with (a) showing an input-output data set with a gradual change from $S^{(1)}$ to $S^{(2)}$ with a linear transition over the time period of 2000 time steps and (b) depicting an output only data set with an abrupt change point at $k = 3000$. The changes are not visually apparent in the noise-corrupted system output.

In the following, we compare the novel measure to the ARMA distance introduced in [10]. Subsequently, we evaluate the presented method using direct pole estimation against the traditional two step pole estimation approach.

A. Comparison of OSPA, MAX-OSPA and ARMA Distance

We are evaluating the presented distance in comparison to the ARMA distance [10] using the two data sets depicted in Fig. 4 and the mean of the estimated parameters. The estimation of the system is performed by the direct pole estimation described in Section III. The initial mean $\hat{\underline{\theta}}_0^{(t)}$ of the pole estimation vector is a random displacement from the reference system $\underline{\theta}^{(r)}$ and is calculated by drawing a sample from the distribution $\mathcal{N}(\underline{\theta}^{(r)}, \mathbf{C}_0^{(t)})$, with $\mathbf{C}_0^{(t)} = 0.05\mathbf{I}$ the initial covariance matrix of $\underline{\boldsymbol{\theta}}_0^{(t)}$. A visualization by projection onto the complex plane of one exemplary initialization is shown in Fig. 3 (a). The system state is initialized with $\underline{\boldsymbol{x}}_0 \sim \mathcal{N}(\underline{0}, 0.5\mathbf{I})$. The additive noise variances for measurement and prediction are given by $\mathbf{C}^v = 0.1$ and $\mathbf{C}^w = 0.1\mathbf{I}$, respectively. The parameter prediction variance is given by $\mathbf{C}^\theta = 10^{-6}\mathbf{I}$, which is a good trade-off between identification quality and parameter tracking ability. We would like to emphasize that when performing simultaneous state and parameter estimation a reasonable choice of the proportion between $\mathbf{C}^w$ and $\mathbf{C}^\theta$ is important. The estimation was performed by the Unscented Kalman Filter (UKF) [22].

The results of this evaluation are depicted in Fig. 5 (a)

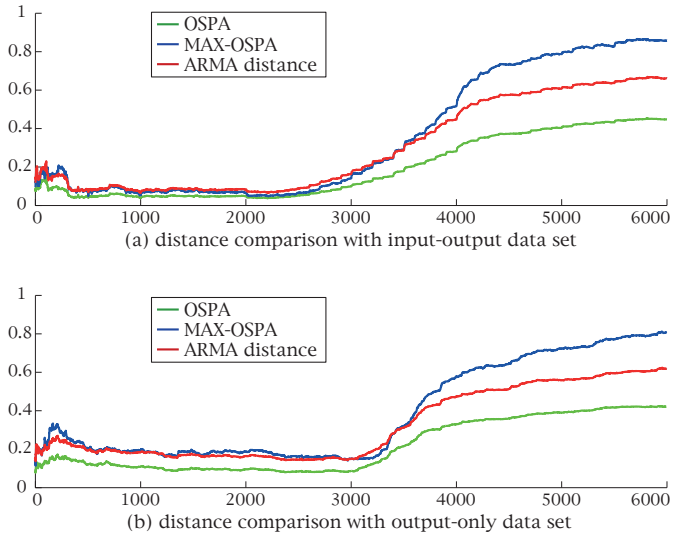


Fig. 5. Exemplary evaluation of the presented OSPA and MAX-OSPA distances in pole space, and the ARMA distance [10], which is scaled by the factor of 1/3. The comparison shows a very similar performance of the measures, but all have different properties and interpretations. In (a) we used the input-output data set depicted in Fig. 4 (a) comprising a gradual change. In (b) the distances were evaluated with the output-only data set with an abrupt change point, which is visualized in Fig. 4 (b).

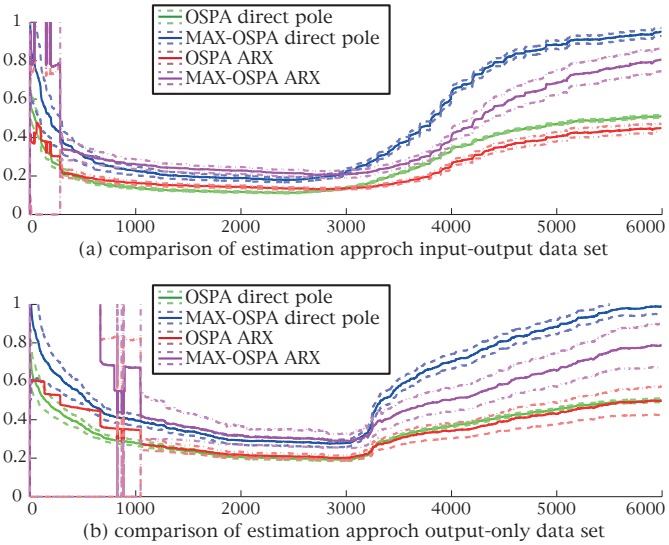


Fig. 6. Evaluation of OSPA and MAX-OSPA distances based on direct pole estimation and two-step pole estimation using a recursive ARX system identification method. (a) depicts the resulting distances for the input-output data set with a gradual change. (b) shows the result for the output-only data set with an abrupt change point. The behavior of the distance variance of the two-step pole estimation comes due to the sample-based transformation of estimated polynomial coefficients. In certain cases the transformed poles sheer off resulting in large distances.

for the input-output data set and in Fig. 5 (b) for the input only data set. In order to ensure a better comparability in the plots, we scaled the ARMA distance by $1/3$. It can be seen that the performances of the distance measures are similar, but there are differences in their properties. In the literature there exist several different transformations of the ARMA distance [10], [16], [25], but it is only valid for stable systems. On the other hand, the OSPA and MAX-OSPA measures have a more natural meaning in the pole-space and can be adjusted and interpreted more easily. Using the maximum single pole distance instead of the average of all poles, the MAX-OSPA is always larger than the OSPA distance, as expected. This helps to detect single frequency changes. The performance in detecting changes itself depends on the identifiability of the parameters, i.e., a sufficent excitation of the system, and on the parameterization of $\mathbf{C}^{\theta}$. If the possible time-variance of the system is not known, it is possible to use a multi model approach with several different parameterizations [26].

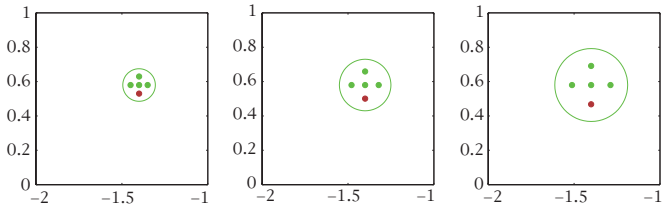
B. Comparison of Distances Calculated Based on Direct Pole Estimation and Two-Step Pole Estimation

In this second simulation, we are evaluating the OSPA and MAX-OSPA distances comprising the parameter uncertainty based on the Dirac mixture formulation derived in Sec. IV-B. We are comparing the direct pole estimation as presented in Sec. III to the commonly used two-step approach, where the system coefficients are estimated first and then transformed into pole-space. The direct pole estimation was initialized equally to Sect. V-A. The estimation of coefficients was performed using the recursive ARX estimator `rarx()` from the MATLAB® system identification toolbox with Kalman Filter implementation. The mean of the initial state was determined by transformation of the mean of the initial poles into coefficients by application of (6). The initial uncertainty was also chosen as $\mathbf{C}_0 = 0.05\mathbf{I}$, whereas the prediction uncertainty was set to $\mathbf{C}_w = 5 \cdot 10^{-6}$ to match the parameter tracking characteristic. The poles, together with their uncertainty, were calculated by UKF-sampling of the uncertain coefficients and sample-wise transformed into pole-space. Thus, $2 \cdot p + 1$ root calculations have to be performed.

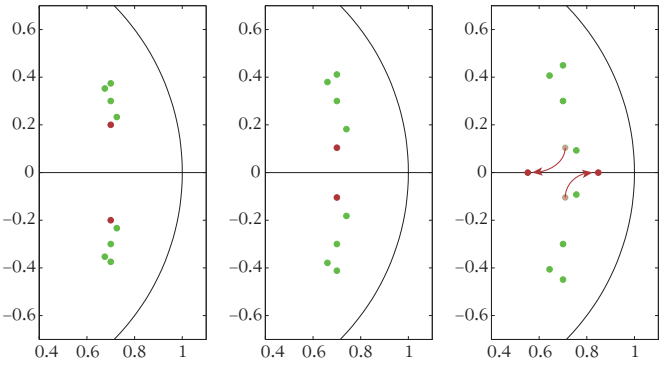
The evaluation is depicted in Fig. 6, where again (a) shows the results based on a realization of the input-output data set with a gradual change as shown in Fig. 4 (a), and (b) shows the results of the output-only data with an abrupt change as depicted in Fig. 4 (b). Prominent in this evaluation is that the variance of the distances calculated by the two step approach has a very extreme behavior in the initialization phase, where the variance is high. This is due to the effect that whenever the uncertainty of two complex poles rises to actually cross the real axis, the two complex pole-samples become two real poles sheering off. This leads to strong changes in the distance measure. We have visulalized this effect in Fig. 7.

VI. Conclusions

In this work, we have derived a novel distance measure for linear SISO systems in the context of change detection by exploiting the strong analogies of pole identification to association-free mulit-target tracking. The novel pole-based distance comprises parameter uncertainty resulting from system identification, which can be used in the decision process of change detection. In the evaluation, we have shown that the performance of the mean distance is similar to the well comparable ARMA distance [10], but the interpretation in pole space is more natural as the poles have an intuitive interpretation. Evaluating direct pole estimation against two-step pole estimation using standard system identification and explicit sample-based root calculation revealed problems of the explicit root calculation in case of large parameter uncertainty



(a) uncertain coefficients of an order 2 system and UKF sampling



(b) sample-wise transformed into pole space by explicit root calculation

Fig. 7. Visualization of the two-step pole estimation by sample-based explicit root calculation of an $p = 2$ order system with the parameters $a_1 = -1.4$ and $a_2 = 0.58$. (a) shows the parameter space with the 3σ-bounds displayed by the covariance ellipses of different size, as well as the UKF-samples. (b) shows the transformed samples in pole space. Each 2-D coefficient sample results in two pole samples. We have marked one sample red, in order to emphasize its correspondence. As soon as the sample reaches the real axis in pole space, it transforms from two complex poles into two real poles sheering off on the real axis.

or complex poles close to the real axis. Probability mass of complex poles growing to cross the real axis, lead to real poles moving apart (see Fig. 7). Hence, using this approach only shows good results for relatively small parameter uncertainty. On the other hand, the computational complexity of direct pole identification grows exponentially in the number of pole, and thus, it is not applicable to very high order systems.

REFERENCES

[1] V. Chandola, A. Banerjee, and Vipin Kumar, "Anomaly Detection: A Survey", *in ACM Computing Surveys (CSUR)*, Vol. 41, No. 3, pp. 559–564, July, 2009.

[2] V. Hodge and J. Austin, "A Survey of Outlier Detection Methodologies", *in Artificial Intelligence Review*, Vol. 22, No. 2, pp. 85–126, October, 2004.

[3] D. Miljkovic, "Fault Detection Methods: A Literature Survey", *in Proceedings of the 34th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2011)*, pp. 750–755, May, 2011.

[4] R. Isermann, "Model-based Fault-Detection and Diagnosis - Status and Applications", *in Annual Reviews in Control*, Vol. 29, No. 1, pp. 71-85, 2005.

[5] A. Tsymbal, "The Problem of Concept Drift: Definitions and Related Work", *TCD-CS-2004-15*, The University of Dublin, Trinity College, Department of Computer Science, Dublin, Ireland, 2004.

[6] M. Markou and S. Singh, "Novelty Detection: A Review – Part 1: Statistical Approaches", *in Signal Processing*, Vol. 83, No. 12, pp. 2481–2497, December, 2003.

[7] C. Inclan and G. C. Tiao, "Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance", *in Journal of the American Statistical Association*, Vol. 89, No. 427, pp. 913–923, September, 1994.

[8] U. Appel and A. V. Brandt, "Adaptive Sequential Segmentation of Piecewise Stationary Time Series", *in Information Sciences*, Vol. 29, No. 1, pp. 27–56, February, 1983.

[9] M. Basseville, "Distance Measures for Signal Processing and Pattern Recognition", *in Signal Processing*, Vol. 18, No. 4, pp. 349–369, December, 1989.

[10] R. J. Martin, "A Metric for ARMA Processes", *in IEEE Transactions on Signal Processing*, Vol. 48, No. 4, pp. 1164-1170, April, 2000.

[11] T. S. Alexander and V. L. Stonick, "Fast Adaptive Polynomial Root Tracking Using a Homotopy Continuation Method", *in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1993)*, Vol. 3, pp. 480-483, April, 1993

[12] A. Nehorai and D. Starer, "Adaptive Pole Estimation", *in IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 38, No. 5, pp. 825-838, May 1990

[13] L. Weruaga, "All-Pole Estimation in Spectral Domain", *in IEEE Transactions on Signal Processing*, Vol. 55, No. 10, pp. 4821-4830, October 2007

[14] C. Chlebek and U. D. Hanebeck, "Bayesian Approach to Direct Pole Estimation", *in Proceedings of the 2014 European Control Conference (ECC 2014)*, Strasbourg, France, June 2014.

[15] S. Thanagarundram, S. Spurgeon, and F. S. Schlindwein, "A Fault Detection Tool Using Analysis from an Autoregressive Model Pole Trajectory", *in Journal of Sound and Vibration*, Vol. 317, No. 3-5, pp. 975–993, November, 2008

[16] K. De Cock and B. De Moor, "Subspace Angles Between ARMA Models", *in Systems & Control Letters*, Vol. 46, No. 4, pp. 265-270 July, 2002.

[17] H. Zheng and A. Mita, "Damage Indicator Defined as the Distance between ARMA Models for Structural Health Monitoring", *in Structural Control and Health Monitoring*, Vol. 15, No. 7, pp. 992-1005, November, 2008.

[18] Y. Kawahara, T. Yairi, and K. Machida, "Change-Point Detection in Time-Series Data Based on Subspace Identification", *in Proceedings of the Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pp. 559-564, October, 2007.

[19] F. Jiang, W. Ying, and A. K. Katsaggelos, "Detecting Contextual Anomalies of Crowd Motion in Surveillance Video", *in Proceedings of the 16th IEEE International Conference on Image Processing (ICIP)*, pp.1117-1120, November, 2009.

[20] M. Guerriero, L. Svensson, D. Svensson, and P. Willett, "Shooting Two Birds with Two Bullets: How to Find Minimum Mean OSPA Estimates", *in Proceedings of the 13th International Conference on Information Fusion (FUSION 2010)*, July, 2010

[21] J. P. Burg, "A New Analysis Technique for Time Series Data", *in NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics*, Enschede, The Netherlands, August, 1968, reprinted *in Modern Spectrum Analysis*, IEEE Press, New York, 1978.

[22] Simon J. Julier and Jeffrey K. Uhlmann, "Unscented Filtering and Nonlinear Estimation", *in Proceedings of the IEEE*, Vol. 92, No. 3, pp. 401-422, March, 2004.

[23] J. Steinbring and U. D. Hanebeck, "S2KF: The Smart Sampling Kalman Filter", *in Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, July, 2013.

[24] H. W. Kuhn, "The Hungarian Method for the Assignment Problem", *Naval Research Logistics Quarterly*, 2:83–97, 1955.

[25] K. De Cock, B. Hanzon, and B. De Moor, "On a Cepstral Norm for an ARMA Model and the Polar Plot of the Logarithm of its Transfer Function", *in Signal Processing*, Vol. 83, No. 2, pp. 439–443, February, 2003.

[26] F. Gustafsson, *Adaptive Filtering and Change Detection*, John Wiley & Sons, ISBN: 978-0-470-85340-5, October, 2000.