Bayesian Approach to Direct Pole Estimation

Christof Chlebek[1] and Uwe D. Hanebeck[1]

***Abstract*— In this work, the problem of pole identification of discrete-time single-input single-output (SISO) linear time-invariant (LTI) systems directly from input-output data is considered. The solution to this nonlinear estimation problem is derived in form of the general Bayesian estimation framework, as well as a practical approximate solution by application of statistical linearization. The derived direct pole estimation algorithm by statistical linearization is given in closed-form and regression point based, by the so-called Linear Regression Kalman Filter (LRKF). We consider both, an input-output and a state-space formulation. Two realizations of the LRKF algorithm are tested, namely the Unscented Kalman Filter (UKF) for low computational complexity and thus, for high update rates, and the Smart Sampling Kalman Filter (S2KF) for high precision with faster convergence. Both, the UKF and S2KF are compared to the Adaptive Pole Estimation (APE), a solution by recursive nonlinear least squares minimizing the prediction error gradient.**

I. Introduction

The identification of poles of the transfer function of a system, plays a key role in many applications, such as linear systems identification and analysis [1], speech analysis [2], and signal processing [3]. For example the comparison of two systems can be calculated on basis of the poles of the transfer functions [4]. Several different applications of this pole based distance measure can be found in literature, such as change point detection [5], anomaly detection [6], or structural damage analysis [7]. Another structural damage analysis method based on damage indication calculated from poles of the transfer function can be found in [8]. The damage condition of a bearing is determined by vibration-based monitoring and calculated from the position of the poles in [9]. Biomedical applications can be found in [10] and [11]. In [10], the poles of voice signals show a significantly different behavior in terms of pole deviations between healthy and unhealthy glottal tracts of test subjects. In [11], it is shown that a significant change of poles of EEG signals indicates a specific change in the mental state of rats.

The typical approach in pole estimation, is to identify the polynomial coefficients or the system matrix by a well-established method like the Recursive Least Squares (RLS) algorithm [1] or the Yule-Walker equations [12], [13] (for a modern formulation see also, e.g., [1]). Then, in a second procedure, the roots of the (characteristic) polynomial are determined by standard factorization methods. With this two-step approach, information about the identification quality, i.e., uncertainty of the identified parameters, is present in the polynomial coefficient space. Usually only a point estimate is then transformed into pole-space and the information about the parameter uncertainty is disregarded. One possible way to determine the uncertainty in pole-space is by application of a sampling approach, where not only the point estimate is transformed, but a set of samples. This approach has a heavy computational burden if poles are required to be updated each time step with high precision.

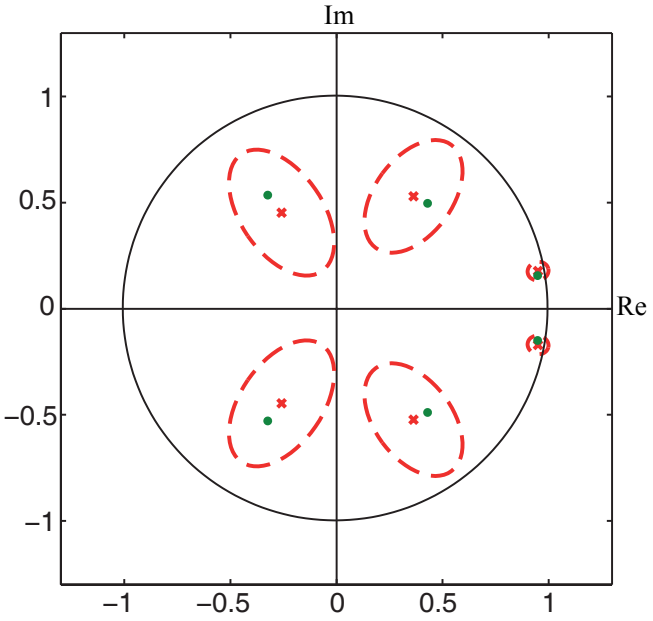

Fig. 1. Direct estimation of poles of the transfer function of an AR process. Green dots show the actual position of the poles of the system, red crosses and dashed ellipses show the mean and Gaussian uncertainty region displayed as the 3σ bound of the estimated poles. As can be seen, poles close to the unit circle are identified faster, which can be explained by their stronger impact on the input-output bahavior.

Hence, an effective on-line pole estimation method has to satisfy the following requirements. (i) The pole estimates have to be updated at every time step directly from the data, providing information about the identification uncertainty. Furthermore, (ii) the estimates have to improve with every information gain. Finally, (iii) the method has to work with fixed amount of memory and computation time, independent of the amount of available data samples, i.e., saving all captured data points and processing them explicitly at every time step has to be avoided. In Fig. 1 an intermediate result of the direct pole estimation is illustrated, showing the pole constellation of a system depicted by green dots and the estimated poles depicted by red crosses together with their uncertainty region shown as red ellipsoids.

[1]The authors are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics and Robotics (IAR), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, Email: christof.chlebek@kit.edu, uwe.hanebeck@ieee.org .

There are few proposals in literature solving this problem, while satisfying the stated requirements (i) to (iii). In [14] the poles of an AR process are directly estimated from data by solving a Nonlinear Recursive Least Square (NRLS) problem formulation. By application of a Gauss-Newton Recursive Prediction Error (RPE) method a quadratic error is minimized recursively over time, where the error gradient is calculated with respect to the poles. Though, the derivation the adaptive pole estimation (APE) algorithm in [14] comes by a completely different course, careful analysis reveals that APE is an approximate solution to the general Bayesian direct pole estimation presented in this work, namely calculation by the Extended Kalman Filter (EKF). Only an optional forgetting factor introduced to the APE shows some differences to the EKF solution. Given that data samples are in spectral domain, pole estimation can be performed directly using spectral all-pole estimation (SAPE) presented in [15]. Though designed for the special case of process output realizations in spectral domain, SAPE can also be applied for samples in time domain by using windowing over a set of samples and transforming into spectral domain utilizing the Discrete Fourier Transform (DFT). This approach is useful, for example, with non-uniform sampling intervals, but with the disadvantage that the last several hundred samples have to be saved and processed every time step.

A. Contribution

The main contribution of this work is the derivation of a general Bayesian solution to the problem of the estimation of poles of the transfer function that characterize the system behavior. We assume that the system order is known. The key idea is to obtain the poles by directly processing the input-output data without prior identification of polynomial coefficients or the system matrix. The presented Bayesian solution can not be computed in closed form, but it enables the application of many approximate estimation methods. Methods can be selected depending on given requirements, such as available computational resources or estimation quality. This is demonstrated by application of statistical linearization, which can be calculated in a closed-form for the present case [18], or by means of the LRKF [16], [17]. We will evaluate two approaches utilizing linear regression, namely the computationally inexpensive Unscented Kalman Filter (UKF) [19] and the Smart Sampling Kalman Filter (S2KF) [20] with adjustable precision.

B. Outline and General Notation

The remainder of this paper is structured as follows: In the following section, the formal problem formulation is given, introducing the considered system model. In Sec. III, the general solution is derived using the recursive Bayesian estimation by directly mapping the data to system poles and hence, identifying poles from data. A practical algorithm by application of statistical linearization is presented in Sec. IV. In Sec. V the solution to simultaneous state and pole estimation in the state-space form is given based on state augmentation. In Sec. VI, the LRKF-based solution is evaluated comparing the UKF and the S2KF based LRKF-realizations to the Adaptive Pole Estimation (APE) [14]. The last section concludes the paper and gives an outlook on future work.

Throughout this work the following notation will be used: We distinguish deterministic quantities a and random variables $\boldsymbol{a}$ by normal lettering and bold face letters, respectively. The notation $\boldsymbol{a} \sim f(a)$ describes the characterization of the random variable $\boldsymbol{a}$ by it's 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 paper, we consider a single-input single-output (SISO) linear time-invariant (LTI) system $S = \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ given in the control canonical form

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

with the state vector $\underline{\boldsymbol{x}}_k = [\boldsymbol{x}_k, \boldsymbol{x}_{k-1}, \dots \boldsymbol{x}_{k-p+1}]^\top$, the system input u_k and output $\boldsymbol{y}_k$, and the zero-mean white system noise $\boldsymbol{w}_k$ and zero-mean white measurement noise $\boldsymbol{v}_k$. The system matrix $\mathbf{A}$ is characterized by

$$\mathbf{A} = \begin{bmatrix} -a_1 & -a_2 & \dots & -a_{p-1} & -a_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} , \tag{2}$$

with the system order $p \in \mathbb{N}_+$. Since we are only interested in the poles, which are characterized by eigenvalues of the system matrix $\mathbf{A}$, for the sake of simplicity the matrices $\mathbf{B}$ and $\mathbf{C}$ are given by $\mathbf{B} = [1, 0, 0, \dots, 0]$ and $\mathbf{C} = [1, 0, 0, \dots, 0]^\top$, which can easily be extended to an arbitrary, but known case. We consider the system to be discrete-time with the time index k.

Given that the input u_k is known and the state is directly accessible (i.e., $\boldsymbol{v}_k = 0$), we can restate the system in an input-output form, by an AR process of the form

$$\boldsymbol{y}_k = -\sum_{i=1}^{p} a_i \boldsymbol{y}_{k-i} + \boldsymbol{w}_k \ , \tag{3}$$

where the parameters a_i with $i = 1, \dots, p$ are AR coefficients and the known input u_k is subtracted. In z-domain the AR process is described by the transfer function

$$H(z) = \frac{1}{\sum_{i=0}^{p} a_i z^{-i}} \nu_w = \frac{1}{\prod_{i=1}^{p} (1 - \alpha_i z^{-1})} \nu_w \ , \tag{4}$$

where $a_0 := 1$. The transfer function $H(z)$ can be represented as a rational function of a sum or a product, where the parameters α_i are known as the poles of the model. Considering a real valued AR process implies that all α_i occur in complex conjugate pairs.

The main problem addressed in this work is to identify the poles α_i directly from system input-output data without prior identification of the coefficients a_i. This is performed in a

recursive way by application of Bayesian inference, which is introduced in the following section.

III. Bayesian Pole Identification

Parameters of a stationary process or an LTI system can be identified by application of the recursive Bayesian estimation, whereby one Bayesian inference step is performed every time step.

Remark I: In the case of a non-stationary process or a time-variant system, the prediction step can be exploited as a data forgetting step. The forgetting factor can be realized in a random walk fashion, by using the identity as prediction function adding a small noise in each time step. Thus, the presented pole identification algorithm is able to track moving poles over time at the cost of a slower convergence rate. Since there are several valid ways of adding a forgetting factor to allow for the tracking of non-stationary parameters, we will not further elaborate on this case in the presented work.

Remark II: In the given algorithm, there is no restriction on stable systems. Poles outside the unit circle can be identified in the same way as poles inside the unit circle. The only restriction is the numerical stability. Unstable systems have increasing output values over time, which may result in surpassing the numerical range.

A. General Parameter Identification

Let $\underline{\theta}$ be a constant vector with dimension n_θ and let $\boldsymbol{Y} = \boldsymbol{y}_k$ with $k \in \mathbb{N}_+$ be a discrete-time stochastic process. The vector $\underline{\theta}$ represents the set of parameters to be identified and $\boldsymbol{Y}$ is typically called observation process. The single observations are conditionally independent of each other and have the marginal distribution

$$\mathrm{P}(\boldsymbol{y}_k \in A|\underline{\theta}) = \int_A f^L(y_k|\underline{\theta})\ \mathrm{d}y_k \ ,$$

where A is a subset of $\mathbb{R}$. Commonly, $f^L(y_k|\underline{\theta})$ is called the likelihood of the measurement y_k. In the following, we define the probability density function

$$f_k(\underline{\theta}) = f(\underline{\theta}|y_k, y_{k-1}, \ldots, y_1) \ ,$$

being the estimate of $\underline{\theta}$ at time step k comprising all available observations $\underline{y}_k, \underline{y}_{k-1}, \ldots, \underline{y}_1$. Given a prior estimate $f_0(\underline{\theta})$ the Bayes theorem allows us to update $f_k(\underline{\theta})$ with new observations using the recursion

$$f_k(\underline{\theta}) = c_k \cdot f^L(y_k|\underline{\theta}) f_{k-1}(\underline{\theta}) \ , \tag{5}$$

where $c_k = 1/\int_{\mathbb{R}^{n_\theta}} (f^L(y_k|\underline{\theta}) f_{k-1}(\underline{\theta}))\ \mathrm{d}\underline{\theta}$ is a normalization factor.

B. AR Model Identification

Let us consider $\underline{\theta}$ to be a set of parameters to be estimated and let the observation process be given by the scalar dynamic model

$$\boldsymbol{y}_k = h(\underline{\theta}, \boldsymbol{y}_{k-1:k-p}) + \boldsymbol{w}_k \ , \tag{6}$$

where $\boldsymbol{y}_{k-1:k-p} = [\boldsymbol{y}_{k-1}, \boldsymbol{y}_{k-2}, \ldots, \boldsymbol{y}_{k-p}]^T$ is a vector of the last p consecutive observations. The observation process will be called measurement function in the following. The process noise $\boldsymbol{w}_k \sim f^w(w)$ is an independent identically distributed (i.d.d.) random variable. By factorization of the joint density $f(y_k, w)$ and exploiting the fundamental property of the Dirac delta function, we can further specify the likelihood of (5) by

$$\begin{aligned} f^L(y_k|\underline{\theta}) &= \int_{\mathbb{R}} \delta(y_k - h(\underline{\theta}, y_{k-1:k-p}) - w) f^w(w)\ \mathrm{d}w \\ &= f^w(y_k - h(\underline{\theta}, y_{k-1:k-p})) \ . \end{aligned} \tag{7}$$

C. Direct Pole Observation

The formulation given above can be applied to general parameter identification problems of an AR process. We now derive the measurement function for the direct pole identification of an AR process.

We define the random vector $\underline{\boldsymbol{\theta}}_k \sim f_k(\underline{\theta})$ as estimation of the set of unknown parameters

$$\underline{\theta} := [\alpha_1, \ldots, \alpha_p]^T \ , \tag{8}$$

where every $\alpha_i = \sigma_i + j\omega_i$ is one complex pole composed of the real part σ_i and the imaginary part ω_i, with the imaginary unit $j^2 = -1$. Thus, the length of the parameter vector $\underline{\theta}$ is $2p$.

Comparing (3) and (6), exploiting the relation of a_i and α_i in (4), and substitution of a_i leads to the measurement function

$$h(\underline{\theta}_k, \boldsymbol{y}_{k-1:k-p}) = -\sum_{i=1}^{p} \left((-1)^i \boldsymbol{y}_{k-i} \cdot \sum_{\substack{M \subseteq \{1..p\} \\ |M|=i}} \left(\prod_{m \in M} \alpha_m \right) \right), \tag{9}$$

mapping the poles directly to the process data. 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 complete derivation of (9) is given in the Appendix. The function can be generated by the power set of all poles. The subsets with i elements are summed up and multiplied by $(-1)^i \boldsymbol{y}_{k-i}$. Note 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 $\underline{\theta}$. As a consequence, there are $n!$ potential correct results of estimating all α_i, namely all permutations of $\underline{\theta}$. We illustrate this with a simple example system.

Example System 1 (Two Poles) Let us consider a simple system with two real valued poles given by

$$\boldsymbol{y}_k = -0.15\boldsymbol{y}_{k-1} - 0.76\boldsymbol{y}_{k-2} + \boldsymbol{w}_k \ ,$$

with the poles located at the positions $[-0.8, 0.95]$ and the process noise $\boldsymbol{w}_k \sim \mathcal{N}(0, 1)$. The measurement function in this case is given by

$$h([\alpha_1, \alpha_2], \boldsymbol{y}_{k-1:k-2}) = (\alpha_1 + \alpha_2) \cdot \boldsymbol{y}_{k-1} - (\alpha_1 \cdot \alpha_2) \cdot \boldsymbol{y}_{k-2} \ .$$

It is easy to see that α_1 and α_2 can be swapped without impact on the result of $h(.,.)$. Fig. 2 illustrates how the symmetry in

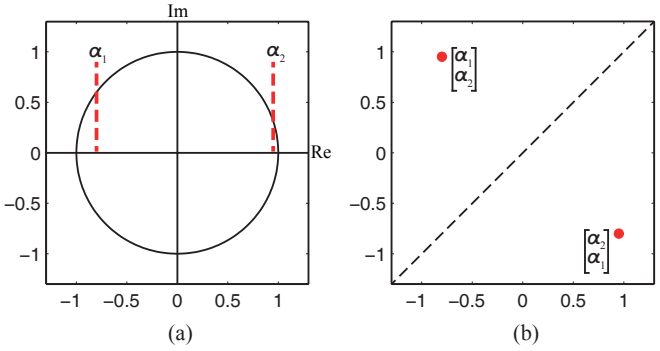

Fig. 2. This figure illustrates Example System 1. Subfig. (a) shows the pole constellation on the complex unit circle with the poles on the real axis visualized by red dashed lines. Subfig. (b) shows the actual estimation space, with the symmetry axis illustrated by the black dashed line, which is located on $\alpha_1 = \alpha_2$. Each of the red dots shows the position of both poles α_1 and α_2, but in reverse order.

the measurement function results in multiple correct solutions, occurring as a consequence of possible permutations of $\underline{\theta}$.

For a given realization series of outputs $y_k, y_{k-1}, \ldots, y_{k-p}$ we can perform the recursive Bayes inference step by evaluation of (5). Although, (9) is nonlinear and thus, it is not possible to evaluate the Bayesian inference step exactly. Therefore, we introduce a practical approximate solution in Sec. IV by application of statistical linearization.

D. Known Proportion of Real and Complex Poles

Exploiting the fact that the output is real-valued, we can give a more efficient formulation, if the proportions of complex and real-valued poles are known. Since in this case complex poles are given by complex conjugate pairs, the length of the parameter vector $\underline{\theta}$ can be reduced by half.

We define real and complex valued poles as

$$\alpha_i = \begin{cases} \alpha_s^R = \sigma_s^R + j \cdot 0 \ , & \text{if } \mathrm{Im}(\alpha_i) = 0 \\ \alpha_t^C = \sigma_t^C + j \cdot \omega_t^C \ , & \text{if } \mathrm{Im}(\alpha_i) > 0 \end{cases} \ , \tag{10}$$

where for every α_t^C there is a complex conjugate α_t^{C*}, i.e. $\alpha_t^{C*} = \sigma_t^C - j\omega_t^C$. Both $s = 1, \ldots, n$ and $t = 1, \ldots, m$ are new consecutive indices for the real and complex poles, respectively, with

$$p = n + 2m \ .$$

Using definition (10), we can rewrite the denominator of (4) by

$$\prod_{i=1}^{p}(1 - \alpha_i z^{-1}) = \prod_{s=1}^{n}(1 - \alpha_s^R z^{-1}) \cdot \prod_{t=1}^{m}(1 - \alpha_t^C z^{-1})(1 - \alpha_t^{C^*} z^{-1}) \ .$$

The new parameter vector is then given by

$$\underline{\theta} = [\sigma_1^R, \ldots, \sigma_n^R, \sigma_1^C, \ldots, \sigma_m^C, \omega_1^C, \ldots, \omega_m^C]^T \ . \tag{11}$$

Using (10) we can map (11) back to the form (13) and thus, we can proceed as in III-C. This formulation allows to ensure the constraint that the system output is real-valued and at the same time remove all redundancy in the estimation.

Data: $\hat{\underline{\theta}}_0$, $\mathbf{C}_0$, $y_{1,2,\ldots}$, ν_w
Result: $\hat{\underline{\theta}}_k$, $\mathbf{C}_k$
$[\hat{\underline{\theta}}_p, \mathbf{C}_p] \leftarrow [\hat{\underline{\theta}}_0, \mathbf{C}_0]$
for $k \leftarrow p+1$ **to** *end* **do**
  // calculate regression point (e.g., [19], [20]) and
  // apply observation function (9)
  $\theta_{k-1} \leftarrow getRegressionPoints(\hat{\theta}_{k-1}, \mathbf{C}_{k-1})$;
  $\mathcal{Y}_k \leftarrow h(\theta_{k-1}, \underline{y}_{k-1:k-p})$;
  // calculate expected observation mean and varinace,
  // joint state-observation cross-covariance, and gain
  $\hat{y}_k = \sum_{i=1}^{r} \omega_{k-1}^{(i)} \tilde{y}_k^{(i)}$
  $\nu_k^{(y)} \leftarrow \sum_{i=1}^{r} \omega_{k-1}^{(i)} (\tilde{y}_k^{(i)} - \hat{y}_k)^2$;
  $\underline{c}_k^{(\theta y)} \leftarrow \sum_{i=1}^{r} \omega_{k-1}^{(i)} (\tilde{\underline{\theta}}_{k-1}^{(i)} - \hat{\underline{\theta}}_{k-1})(\tilde{y}_k^{(i)} - \hat{y}_k)^T$;
  $\mathcal{K}_k \leftarrow (\underline{c}_k^{(\theta y)})^T (\nu_k^{(y)} + \nu_w)^{-1}$;
  // update estimate and estimation uncertainty
  $\hat{\underline{\theta}}_k \leftarrow \hat{\underline{\theta}}_{k-1} + \mathcal{K}_k (y_k - \hat{y}_k)$;
  $\mathbf{C}_k \leftarrow \mathbf{C}_{k-1} - \mathcal{K}_k \nu_k^{(y)} \mathcal{K}_k^T$;
end

Fig. 3. Direct pole identification from data using an LRKF update.

IV. Pole Estimation based on Statistical Linearization

In the case of a nonlinear observation function, it is common to utilize a modification of the Kalman filter. Nonlinear functions are handled by linearization, either by function linearization around a single point (e.g., Taylor-series based linearization) or by statistical linearization, which can be performed analytically for polynomials [18] or based on regression points [16], [17]. This results in the general assumption of a Gaussian distributed joint density $f(\underline{\theta}, y, w)$. Consequently, the estimate $\underline{\boldsymbol{\theta}}_k \sim f_k(\underline{\theta})$ is approximated by a Gaussian distribution $\tilde{\underline{\boldsymbol{\theta}}}_k$ given by the mean $\hat{\underline{\theta}}_k$ and covariance matrix $\mathbf{C}_k$. First, we will introduce the LRKF [16], [17], which uses regression point based statistical linearization, also called weighted statistical linear regression. Then, we introduce the analytic statistical linearization [18]. As already mentioned in the introduction, the adaptive pole estimation (APE) [14] equals the application of the EKF in method, but for the optional forgetting factor, hence, the introduction to this approach will be omitted.

Using the regression point based statistical linearization, $\tilde{\underline{\boldsymbol{\theta}}}_{k-1}$ is deterministically sampled by a Dirac mixture density

$$\mathcal{X}_{k-1} = \sum_{i=1}^{r} \omega_{k-1}^{(i)} \delta\left(\underline{\theta} - \tilde{\underline{\theta}}_{k-1}^{(i)}\right) \ ,$$

comprising r regression points at positions $\tilde{\underline{\theta}}_{k-1}^{(i)}$ and weights $\omega_{k-1}^{(i)}$, with $\sum_{i=1}^{r} \omega_{k-1}^{(i)} = 1$, such that the mean and the covariance of $\mathcal{X}_{k-1}$ equal $\hat{\underline{\theta}}_{k-1}$ and $\mathbf{C}_{k-1}$, respectively. The sampling method determines quality and complexity of

each estimation step. We recommend the so-called Unscented Transformation used by the Unscented Kalman Filter (UKF) [19] for few samples and thus, low complexity and high sampling rates, or the Smart Sampling Kalman Filter (S2KF) [20] for high quality estimation with faster convergence. The observation function $h(\cdot)$ is then evaluated at every regression point of $\mathcal{X}_{k-1}$ and linearized by

$$h(\tilde{\underline{\theta}}_{k-1}^{(i)}, y_{k-1:k-p}) = \mathbf{H}_k \tilde{\underline{\theta}}_{k-1}^{(i)} + b_k + e_k^{(i)} = \tilde{y}_k^{(i)} \ ,$$

where $e_k^{(i)}$ is an additional linearization error for every regression point. Note, that $y_{k-1:k-p}$ is a realization of $\boldsymbol{y}_{k-1:k-p}$, namely, the actual measurements. We will denote the Dirac mixture density of all transformed regression points by $\mathcal{Y}_k$, i.e.,

$$\mathcal{Y}_k = h(\mathcal{X}_{k-1}, \underline{y}_{k-1:k-p}) = \sum_{i=1}^{r} \omega_{k-1}^{(i)} \delta\left(y - \tilde{y}_k^{(i)}\right) \ .$$

Minimization of the sum of the least squares of the linearization error yields

$$[\mathbf{H}_k, b_k] = \underset{H,b}{\arg\min} \sum_{i=1}^{r} \omega_k^{(i)} \left(e_k^{(i)}\right)^2 \ ,$$

which can be solved by

$$\begin{aligned} \mathbf{H}_k &= (\underline{c}_k^{(\theta y)})^T \left(\mathbf{C}_{k-1}\right)^{-1} \\ b_k &= \hat{y}_k - \mathbf{H}_k \hat{\underline{\theta}}_{k-1} \ , \end{aligned}$$

with

$$\underline{c}_k^{(\theta y)} = \sum_{i=1}^{r} \omega_{k-1}^{(i)} \left(\tilde{\underline{\theta}}_{k-1}^{(i)} - \hat{\underline{\theta}}_{k-1}\right) \left(\tilde{y}_k^{(i)} - \hat{y}_k\right)^T$$

the regression point based joint state-measurement covariance. The linearization error can be characterized by the error variance

$$\nu_k^{(er)} = \nu_k^{(y)} - \mathbf{H}_k \mathbf{C}_{k-1} \mathbf{H}_k^T \ ,$$

with $\nu_k^{(y)} = \sum_{i=1}^{r} \omega_{k-1}^{(i)} (\tilde{y}_k^{(i)} - \hat{y}_k)^2$, which is used as an additional noise source in order to ensure a consistent estimate. With the linearized formulation, the insertion of $\hat{\underline{\theta}}_{k-1}$, $\mathbf{C}_{k-1}$, $\mathbf{H}_k$ and $\nu_k^{(er)}$ into the Kalman filter measurement update [21] yields

$$\begin{aligned} \hat{\underline{\theta}}_k &= \hat{\underline{\theta}}_{k-1} + \mathbf{K}_k \cdot (\hat{y}_k - \mathbf{H}_k \hat{\underline{\theta}}_{k-1} + b_k) \\ \mathbf{C}_k &= \mathbf{C}_{k-1} - \mathbf{K}_k \mathbf{H}_k \mathbf{C}_{k-1} \ , \end{aligned}$$

with the Kalman gain

$$\mathbf{K}_k = \mathbf{C}_{k-1} \mathbf{H}_k^T \cdot \left(\mathbf{H}_k \mathbf{C}_{k-1} \mathbf{H}_k^T + \nu_w + \nu_k^{(er)}\right)^{-1} \ .$$

Together with the regression point based expected measurement mean

$$\hat{y}_k = \sum_{i=1}^{r} \omega_{k-1}^{(i)} \tilde{y}_k^{(i)}$$

this can be transformed to

$$\begin{aligned} \hat{\underline{\theta}}_k &= \hat{\underline{\theta}}_{k-1} + \mathcal{K}_k \left(y_k - \hat{y}_k\right) \\ \mathbf{C}_k &= \mathbf{C}_{k-1} - \mathcal{K}_k \nu_k^{(y)} \mathcal{K}_k^T \ , \end{aligned}$$

with $\mathcal{K}_k = (\underline{c}_k^{(\theta y)})^T (\nu_k^{(y)} + \nu_w)^{-1}$ the statistically linearized Kalman gain.

The complete algorithm for direct pole estimation using an LRKF is summarized in Fig. 3. We assume that the system order is known a priori. As initial value, it is possible to use any prior knowledge to initialize $\hat{\underline{\theta}}_0$, which includes, for example, a prior analysis of a set of the first data samples by other well established methods like the Yule-Walker equations [12], [13] without the necessity of a priori knowledge. Without prior knowledge it is also possible to use some distributed set of poles over the unit disc as initialization. For stable systems we recommend the concentric uniform distribution on a circle with radius r, with $0.5 \leq r \leq 0.9$. The covariance matrix $\mathbf{C}_0$ represents the uncertainty of the chosen initial value of $\hat{\underline{\theta}}_0$. If no detailed prior knowledge is at hand, we recommend an initialization by $\mathbf{C}_0 = c \cdot \mathbf{I}$, with a constant $0.2 \leq c \leq 0.5$.

For analytic calculation of the statistical linear regression, all the regression point based calculations need to be replaced by analytic calculations [18]. These are the expected measurement mean $\hat{y}_k$, the observation variance $v_k^{(y)}$, and the joint state-observation cross-covariance $c_k^{(\theta y)}$. For the present case this is performed by

$$\begin{aligned} \hat{y}_k &= \int h(\underline{\theta}_k, y_{k-1:k-p}) \cdot f(\underline{\theta}_k) \, \mathrm{d}\underline{\theta}_k \\ v_k^{(y)} &= \int h(\underline{\theta}_k, y_{k-1:k-p})^2 \cdot f(\underline{\theta}_k) \, \mathrm{d}\underline{\theta}_k - \hat{y}_k^2 + v^{(w)} \\ c_k^{(\theta y)} &= \int \underline{\theta}_k \cdot h(\underline{\theta}_k, y_{k-1:k-p}) \cdot f(\underline{\theta}_k) \, \mathrm{d}\underline{\theta}_k - \hat{y}_k^2 + v^{(w)} \ . \end{aligned} \tag{12}$$

The dependence of (12) on the function $h(\underline{\theta}_k, y_{k-1:k-p})$, causes the necessity of a specific set of formulas (12) for different system orders. Unfortunately, the application of the analytic solution is only reasonable for systems with order $p < 4$, due to high complexity.

V. Direct Eigenvalue Estimation

Let a system S be given by (1) and $\underline{\theta}$ be the vector of unknown eigenvalues characterizing $\mathbf{A}$ given by

$$\underline{\theta} := [\alpha_1, \ldots, \alpha_p]^\top \ , \tag{13}$$

where every $\alpha_i = \sigma_i + j\omega_i$ is one complex eigenvalue composed of the real part $\sigma_{k,i}$ and the imaginary part ω_i. Thus, written in the form $\alpha_i = [\sigma_i, \omega_i]^\top$ the length of the parameter vector $\underline{\theta}$ is $2p$, where p denotes the system order. We restate the system formulation using state augmentation, in order to calculate the state and parameter estimates $\underline{\boldsymbol{x}}_k$ and $\underline{\boldsymbol{\theta}}_k$ simultaneously at every time step k by

$$\begin{bmatrix} \underline{\boldsymbol{x}}_{k+1} \\ \underline{\boldsymbol{\theta}}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{x}}_k \\ \underline{\boldsymbol{\theta}}_k \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \underline{0} \end{bmatrix} (u_k + \boldsymbol{w}_k)$$

$$\boldsymbol{y}_k = \begin{bmatrix} \mathbf{C} & \underline{0} \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{x}}_k \\ \underline{\boldsymbol{\theta}}_k \end{bmatrix} + \boldsymbol{v}_k \ ,$$

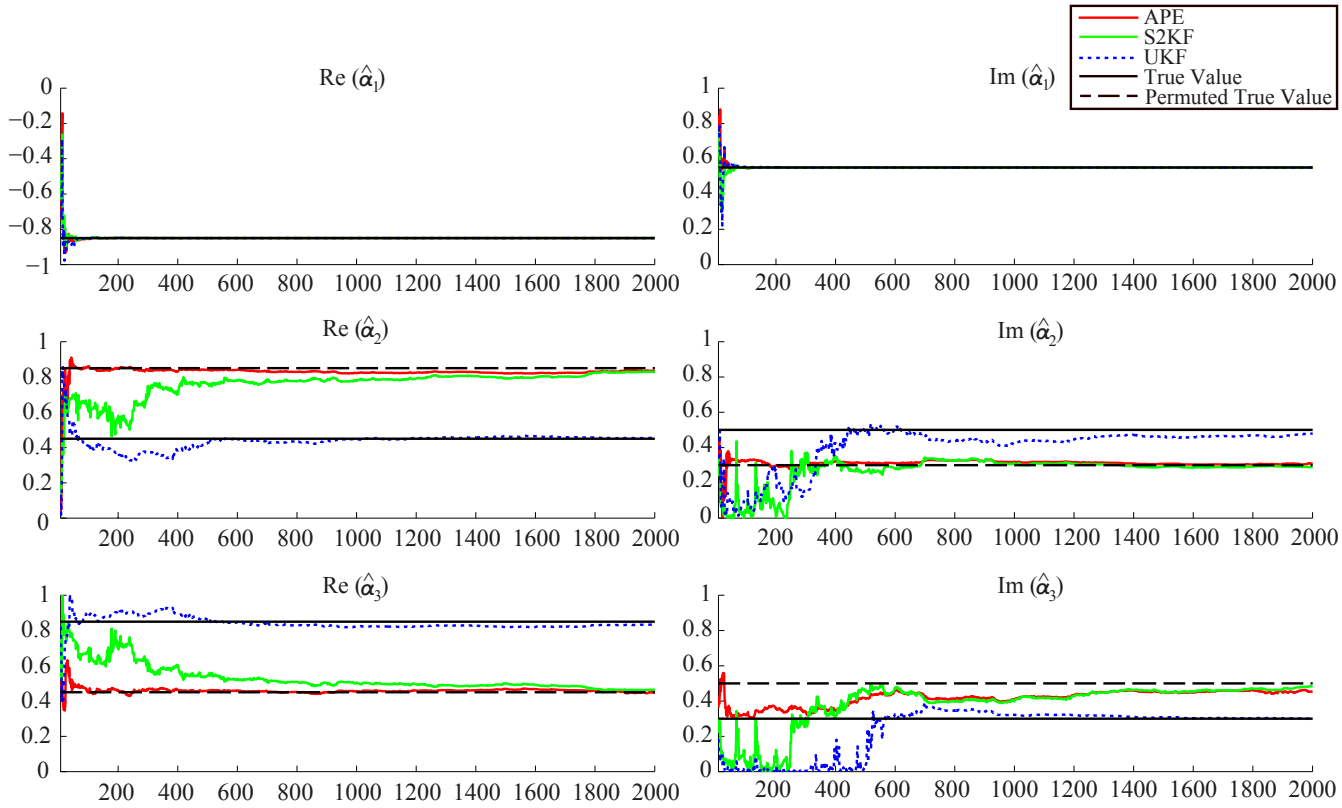

Fig. 4. This figure shows the estimated mean of the simulated system (15). The horizontal plot pairs show the real and imaginary part of one pole. Because of the occurrence of complex conjugate pairs, only three of the six poles are plotted.

where the matrix $\mathbf{A}_k$ is given by (2), with the parameters a_i directly calculated from $\underline{\boldsymbol{\theta}}_k$ by

$$a_i = (-1)^i \cdot \sum_{\substack{M \subseteq \{1..p\} \\ |M|=i}} \left(\prod_{m \in M} \boldsymbol{\alpha}_m \right) \quad , \tag{14}$$

i.e., the same transformation as derived from the AR-coefficients a_i by replacing the sum of all products of poles consisting of i factors of all subsets M with i elements. With this formulation we can estimate $[\underline{\boldsymbol{x}}_k, \underline{\boldsymbol{\theta}}_k]^\top$ for a given input-output time series $[y_{1\ldots k}, u_{1\ldots k}]$ analogously by application of recursive Bayesian estimation.

VI. Simulation

In order to demonstrate the performance of the presented approach to direct pole identification by Bayesian estimation, we evaluate the presented LRKF solution with UKF sampling [19] and S2KF sampling [20] using 50 regression points per dimension, and compare it to the adaptive pole estimation (APE) [14].

For the evaluation we generated a dataset of 2000 data samples by simulation of the 6th order stable system with

$$\begin{aligned} \boldsymbol{y}_k = {} & 0.9\boldsymbol{y}_{k-1} + 0.6\boldsymbol{y}_{k-2} - 0.586\boldsymbol{y}_{k-3} - 0.6817\boldsymbol{y}_{k-4} \\ & + 0.913\boldsymbol{y}_{k-5} - 0.3768\boldsymbol{y}_{k-6} + \boldsymbol{w}_k \quad , \end{aligned} \tag{15}$$

with the process noise $\boldsymbol{w}_k \sim \mathcal{N}(0, 1)$. The poles are located at positions $[-0.85 \pm 0.55j, 0.45 \pm 0.5j, 0.85 \pm 0.3j]$. We have visualized the pole constellation together with an exemplary intermediate result of the identification process in Fig. 1. Two of the three pole pairs have a position far from the unit circle, which results in slower impact of data on their identification process, and thus a slower convergence. The system is initialized with a zero state. The pole identification process is started at time step $k = 7$, using output data of time steps $k = 1 \ldots 6$ as initial input $y_{6:1}$. The prior $x_0 = [0 \pm 0.8j, 0.6928 \pm 0.4j, -0.6928 \pm 0.4j]$ was chosen by an uniformly distributed set of poles over the complex circle with radius 0.8. As initial uncertainty area C_0, we chose $C_0 = 0.3 \cdot \mathbf{I}$, which covers a sufficient area of the unit circle for each pole considering the 3σ-bound of a Gaussian density.

Fig. 4 shows an exemplary outcome of the point estimate of the poles with their real (left column) and imaginary (right column) part. On the first impression the first pole, which is close to the unit circle shows fast convergence by all of the tested methods, as it was expected. Closer analysis shows that APE is slightly faster in the initial phase of the pole estimation process. It shows that permutations of the poles in the estimation vector actually do occur in the present case on the poles 2 and 3. Thus, instead of the more commonly used Root Mean Square Error (RMSE), we utilize the so-

called Mean Optimal Subpattern Assignment (MOSPA) metric with the Euclidean distance as base distance for further quantitative evaluation of the proposed method. The MOSPA metric uses the smallest distance over all possible assignment permutations between estimated and actual poles and thus, calculates the estimation error, while simultaneously solving the assignment problem [22].

The evaluated MOSPA metric over 1000 identification runs is displayed in Fig. 5. As in the single run analysis APE shows a fast convergence rate in the initial identification phase. The proposed LRKF identification method outperforms APE in terms of convergence rate on the long run, using the UKF as well as the S2KF regression point calculation.

In order to explain this behavior, we take a look at the Example System 1. Fig. 6 shows the estimation result after 15 time steps. We used a particle based approach, where we distributed 10000 particles and updated the particle weights by the likelihood (7). The light blue particle cloud shows 97,73% of the probability mass. It can be seen, that the particle mass is starting to split into two masses being drawn to the actual pole positions. This results in a bimodal probability density, where each mode lies over one of the two permutations. Now considering the Gaussian probability density resulting from a Kalman filter based approach, the worst case initialization having the mean on the symmetry axis (i.e. $\alpha_1 = \alpha_2$), results in a good estimate of the whole bimodal density, instead of an estimate of one mode. Note that this happens also to the APE, if the initial value is chosen exactly on the symmetry axis. The statistical linearization considering the whole uncertainty region for the linearization, may cause an attraction towards the symmetry axis, if the uncertainty is having large probability masses on both sides of the symmetry axis. Thereby, the estimation convergence is slowed in the beginning.

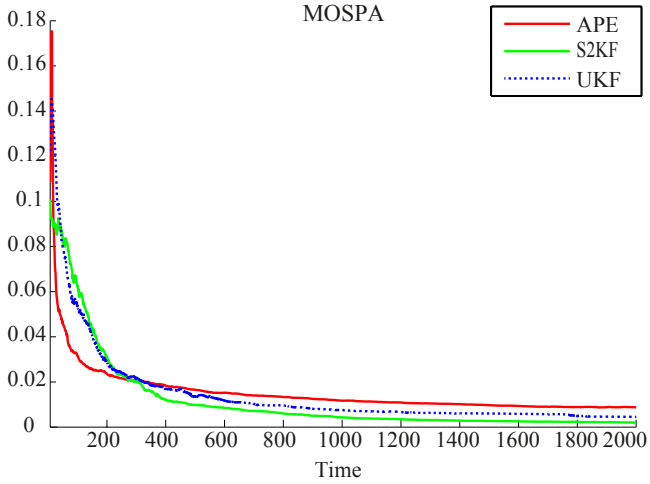

Fig. 5. Evaluated Mean Optimal Subpattern Assignment (MOSPA) metric [22] over 1000 identification runs for a simulated data set of 2000 data points. The identification is performed by the proposed LRKF method with UKF (blue-dotted) and S2KF (green) regression point sampling, as well as the APE [14] (red) for comparison.

VII. Conclusions and Future Work

A general recursive Bayesian solution to direct identification of poles of SISO-LTI systems has been derived. With the general Bayesian approach at hand, it is possible to apply many different Bayesian estimation methods for fast and effective recursive pole identification. An easily applicable solution by utilizing statistical linearization was introduced, which can be calculated analytically or based on regression points. We have derived the likelihood function and the direct mapping from the poles to incoming data and thus, other solutions as the commonly used sequential Monte Carlo filter, are also easily applicable.

We pointed out a systematic problem, caused by the symmetry of the measurement function mapping the poles directly to the process output. In future work, we will address this problem by the development of constrained filtering restricting the estimation space to one possible mode, and multiple-shooting-based filtering splitting the initial estimate and launching multiple estimation processes, fusing them as soon as convergence is reached. Though we presented both a solution for a known and unknown proportion of real and complex poles, if the proportion is unknown the identification result does not necessarily guarantee complex conjugate pole pairs and hence, does not guarantee a real valued system. In future work we will also focus on a Bayesian solution where the identification problem is solved under the constraint of real valued system output and unknown proportion of real and complex poles.

Appendix

In this part we derive the measurement function (6). Comparing (3) and (6) yields

$$h(\underline{x}, \boldsymbol{y}_{k-1:k-p}) = -\sum_{i=1}^{p} a_i \boldsymbol{y}_{k-i} \ . \tag{16}$$

By canceling out ν_w and z^{-p} equation (4) reveals the relation between a_i and α_i, which is given by

$$z^p + \sum_{i=1}^{p} (a_i z^{p-i}) = \prod_{i=1}^{p} (z - \alpha_i) \ .$$

Expansion of the right-hand side yields

$$\begin{aligned} z^p + \textstyle\sum_{i=1}^{p} (a_i z^{p-i}) = z^p + \\ -(\alpha_1 + \alpha_2 + \ldots + \alpha_{p-1} + \alpha_p) z^{p-1} \\ +(\alpha_1\alpha_2 + \alpha_1\alpha_3 + \ldots + \alpha_1\alpha_p + \alpha_2\alpha_3 + \ldots) z^{p-2} \\ -(\alpha_1\alpha_2\alpha_3 + \ldots + \alpha_1\alpha_2\alpha_p + \alpha_2\alpha_3\alpha_4 + \ldots) z^{p-3} \\ \vdots \\ (-1)^p (\alpha_1\alpha_2\alpha_3 \cdot \ldots \cdot \alpha_{p-1}\alpha_p) z^0 \end{aligned}$$

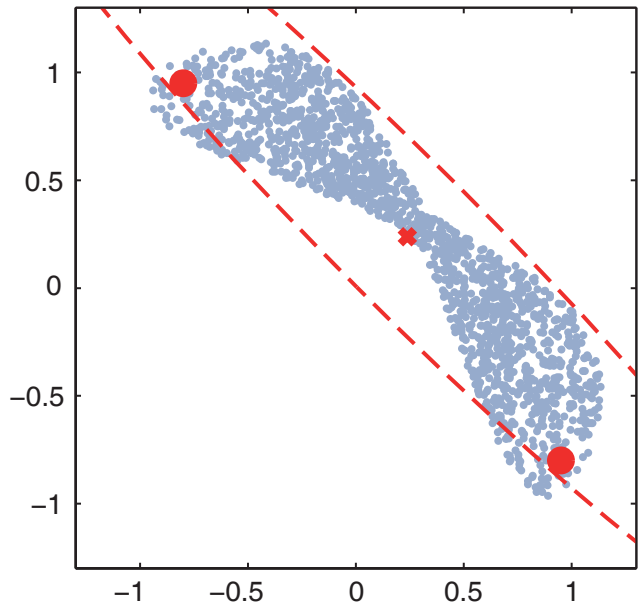

Fig. 6. Result of the pole estimation of Example System 1 after 15 time steps. Each of the red dots illustrates the pole positions of both poles, but in reverse orderring, as in Fig. 2 (b). The light blue points show $99,73\%$ of the probability mass of the result of a sample-based estimation apporach, adjusting the sample weight by multiplication with the likelihood function (7). Over time the cloud splits and forms two modes, both giving a correct result. The red cross and dashed ellipsoid shows exemplary mean and covariance of one of the described Kalman filter based approaches, estimating both modes. Thus, it is not converging to one of the two correct results.

and it follows

$$
\begin{aligned}
a_0 &= (-1)^0 \\
a_1 &= (-1)^1(\alpha_1 + \alpha_2 + \ldots + \alpha_{p-1} + \alpha_p) \\
a_2 &= (-1)^2(\alpha_1\alpha_2 + \alpha_1\alpha_3 + \ldots + \alpha_1\alpha_p + \alpha_2\alpha_3 + \ldots) \\
a_3 &= (-1)^3(\alpha_1\alpha_2\alpha_3 + \ldots + \alpha_1\alpha_2\alpha_p + \alpha_2\alpha_3\alpha_4 + \ldots) \\
&\vdots \\
a_p &= (-1)^p(\alpha_1\alpha_2\alpha_3 \cdot \ldots \cdot \alpha_{p-1}\alpha_p) \ .
\end{aligned}
$$

We will write this in short form as

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

Substitution of a_i in (16) finally yields the measurement function

$$
h(\underline{x}_k, \underline{y}_{k-1:k-p}) = -\sum_{i=1}^{p} \left((-1)^i y_{k-i} \cdot \sum_{\substack{M \subseteq \{1..p\} \\ |M|=i}} \left(\prod_{m \in M} \alpha_m \right) \right) \ .
$$

References

[1] Lennart Ljung, "System Identification: Theory for the User", *in Information and System Sciences Series*, Prentice Hall PTR, December, 1998.

[2] Ron Cole, "Survey of the State of the Art in Human Language Technology", Cambridge University Press, 1997

[3] Steven W. Smith, "Scientist and Engineer's Guide to Digital Signal Processing", Bertrams University Press, January 1997

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

[5] Yoshinobu Kawahara, Takehisa Yairi, and Kazuo 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.

[6] Fan Jiang, Wu Ying, Aggelos 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.

[7] Haitao Zheng and Akira 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.

[8] Ruigen Yao and Shamim N. Pakzad, "Autoregressive Statistical Pattern Recognition Algorithms for Damage Detection in Civil Structures", *in Mechanical Systems and Signal Processing*, Vol. 31, pp. 355-368, August 2012.

[9] Suguna Thanagarundram, Sarah Spurgeon, and Fernando 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.

[10] Paulo R. Scalassara, Maria E. Dajer, and Carlos D. Maciel, "Application of Autoregressive Decomposition and Pole Tracking to Pathological Voice Signals", *in Seventh IEEE International Symposium on Multimedia*, December 2005

[11] Lauri Patomäki, Pasi J. Kaipio, and Pasi A. Karjalainen, "Tracking of nonstationary EEG with the roots of ARMA models", *in 17th IEEE Annual Conference on Engineering in Medicine and Biology Society*, September 1995.

[12] George U. Yule, "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers", *in Philosophical Transactions of the Royal Society of London* Vol. 226, pp. 267-298, 1927.

[13] Gilbert Walker, "On Periodicity in Series of Related Terms", *in Proceedings of the Royal Society of London* Vol. 131, pp. 518-532, 1931.

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

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

[16] Tine Lefebvre, Herman Bruyninckx and Joris De Schutter, "Kalman Filters for Nonlinear Systems: A Comparison of Performance", *in International Journal of Control*, Vol. 77, No. 7, pp. 639-653, 2004

[17] Tine Lefebvre, Herman Bruyninckx and Joris De Schutter, "Nonlinear Kalman Filtering for Force-Controlled Robot Tasks", *ser. Springer Tracts in Advanced Robotics*, Vol. 19, Springer, 2005.

[18] Marco F. Huber, Frederik Beutler, and Uwe D. Hanebeck, "Semi-Analytic Gaussian Assumed Density Filter", *in Proceedings of the 2011 American Control Conference (ACC 2011)*, June 2011

[19] 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

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

[21] Rudolf E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *in Transaction of the ASME - Journal of Basic Engineering*, pp. 35-45, 1960

[22] Marco Guerriero, Lennart Svensson, Daniel Svensson, and Peter 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