

Identification of Positive Real Models in Subspace Identification by Using Regularization

Ivan Goethals, Tony Van Gestel, Johan Suykens, Paul Van Dooren, and Bart De Moor

Abstract—In time-domain subspace methods for identifying linear-time invariant dynamical systems, the model matrices are typically estimated from least squares, based on estimated Kalman filter state sequences and the observed outputs and/or inputs. It is well known that for an infinite amount of data, this least squares estimate of the system matrices is unbiased, when the system order is correctly estimated. However, for a finite amount of data, the obtained model may not be positive real, in which case the algorithm is not able to identify a valid stochastic model. In this note, positive realness is imposed by adding a regularization term to a least squares cost function in the subspace identification algorithm. The regularization term is the trace of a matrix which involves the dynamic system matrix and the output matrix.

Index Terms—Positive realness, regularization, ridge regression, stochastic systems, subspace identification.

I. INTRODUCTION

In this note, we will consider stochastic systems and models of the form

$$\begin{aligned} x_{k+1} &= Ax_k + w_k \\ y_k &= Cx_k + v_k \end{aligned} \quad (1)$$

with

$$\mathcal{E} \left\{ \begin{bmatrix} w_p \\ v_p \end{bmatrix} \begin{bmatrix} w_q^T & v_q^T \end{bmatrix} \right\} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \delta_{pq} \geq 0 \quad (2)$$

where $\mathcal{E}\{\cdot\}$ denotes the expected value operator and δ_{pq} the Kronecker delta. The elements of the vector $y_k \in \mathbb{R}^l$ are given observations at the discrete-time index k of the l outputs of the system. The vector

Manuscript received August 22, 2002; revised June 4, 2003. Recommended by Associate Editor E. Bai. This work was supported by Research Council KUL: GOA-Mefisto 666, several Ph.D./Postdoctoral Fellow Grants, the Flemish Government: FWO: Ph.D./Postdoctoral Grants, Projects G.0240.99, G.0407.02, G.0197.02, G.0141.03, G.0491.03, G.0120.03, and G.0800.01, Research Communities (ICCoS, ANMMM), AWI: Bil. Int. Collaboration Hungary/Poland, IWT: Ph.D. Grants, Soft4s, the Belgian Federal Government: DWTC [IUAP IV-02 (1996–2001) and IUAP V-22 (2002–2006), PODO-II (CP/40: TMS and Sustainability)], EU: CAGE, ERNSI, Eureka 2063-IMPACT, Eureka 2419-FlITE, Contract Research/agreements: Data4s, Electrabel, Elia, LMS, IPCOS, and VIB. The scientific responsibility is assumed by the authors.

I. Goethals is with the Department of Electrical Engineering ESAT-SCD, the Katholieke Universiteit Leuven (KULeuven), B-3001 Leuven, Belgium, and also with the Fund for Scientific Research-Flanders (FWO-Vlaanderen) (e-mail: ivan.goethals@esat.kuleuven.ac.be).

T. Van Gestel is with the Department of Electrical Engineering ESAT-SCD, the Katholieke Universiteit Leuven (KULeuven), B-3001 Leuven, Belgium, and also with the FWO-Vlaanderen Dexia Group, Risk Management (e-mail: tony.vangestel@esat.kuleuven.ac.be).

J. Suykens is with the Department of Electrical Engineering ESAT-SCD, the Katholieke Universiteit Leuven (KULeuven), B-3001 Leuven, Belgium, and also with the FWO-Vlaanderen (e-mail: johan.suykens@esat.kuleuven.ac.be).

P. Van Dooren is with the Department of Mathematical Engineering, the Catholic University of Louvain, B-1348 Louvain-la-Neuve, Belgium (e-mail: vdooren@csam.ucl.ac.be).

B. De Moor is with the Department of Electrical Engineering ESAT-SCD, the Katholieke Universiteit Leuven (KULeuven), B-3001 Leuven, Belgium (e-mail: bart.demoor@esat.kuleuven.ac.be).

Digital Object Identifier 10.1109/TAC.2003.817940

$x_k \in \mathbb{R}^n$ is the unknown state vector at time k . The unobserved process and measurement noise $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^l$ are assumed to be white, zero mean, Gaussian with covariance matrices as given in (2). The system matrices A , C and the covariance matrices Q , S , and R have appropriate dimensions.

Denoting the output covariance matrices as $\Lambda_m = \mathcal{E}\{y_{k+m}y_k^T\}$, and the cross-covariance matrix between the states and the observations as $G = \mathcal{E}\{x_{k+1}y_k^T\}$, one can derive that $\Lambda_m = CA^{m-1}G$, $\Lambda_{-m} = \Lambda_m^T$, $m \geq 1$. Hence, the output covariances can be considered as Markov parameters of a deterministic linear time invariant system with system matrices (A, G, C, D) where $D + D^T = \Lambda_0$. Throughout this note, we will refer to (A, G, C, D) as the ‘‘covariance model.’’ The spectral density $\Phi(z)$ of the system (1) can be expressed in terms of the covariance model as $\Phi(z) = S(z) + S^T(z^{-1})$ with $S(z) = D + C(zI_n - A)^{-1}G$ and is assumed to be positive for all z on the unit circle, in which case the model (A, G, C, D) is called positive real.

Stochastic subspace identification methods [1] make extensive use of the covariance model. Typically they start by making an estimate $(\hat{A}, \hat{G}, \hat{C}, \hat{D})$ based on available measurements. In a second step the covariance model is then transformed into a so-called forward innovation model which is statistically equivalent to a model of the form (1). However, it is known that the second step may fail if the estimated model $(\hat{A}, \hat{G}, \hat{C}, \hat{D})$ is not positive real due to modeling errors (see, for instance, [2]). In such cases, no physically meaningful model will be returned by the subspace identification algorithm.

In recent years, several modifications to the standard stochastic subspace identification algorithms have been suggested to solve the positive realness problem. This, however, at the cost of introducing a small bias in the obtained solution. In this note, we impose positive realness by adding a regularization term to a least squares cost function appearing in most stochastic subspace identification algorithms. Although a bias is still introduced, the regularization approach will be seen to outperform those reported in the literature.

The outline of this note is as follows. In Section II, the stochastic subspace identification algorithm will be outlined, and its problems with positive realness will be described. A proposal for a technique to impose positive realness on an identified covariance model will be given in Section III, the performance of this technique will be compared to that of various existing ones in Section IV, and a real-life example will be introduced in Section V. Finally, the conclusions will be drawn in Section VI.

II. SUBSPACE IDENTIFICATION

Stochastic subspace identification algorithms typically start by calculating Kalman filter state sequences $\hat{X}_i \in \mathbb{R}^{n \times j}$ and $\hat{X}_{i+1} \in \mathbb{R}^{n \times j}$ of the system and an estimate of the system order \hat{n} directly from output data. This is done using geometric operations of subspaces spanned by the column or row vectors of block Hankel matrices. For instance, the output data block Hankel matrix $Y_p = Y_{0|i-1}$ of past outputs is constructed from the observations $y_0, y_1, \dots, y_{i+j-2}$ as follows:

$$Y_{0|i-1} = \begin{bmatrix} y_0 & y_1 & \dots & y_{j-1} \\ y_1 & y_2 & \dots & y_j \\ \vdots & \vdots & & \vdots \\ y_{i-1} & y_i & \dots & y_{i+j-2} \end{bmatrix} \quad (3)$$

where i , the number of block rows in the block Hankel matrix, and j , the number of columns, are user defined dimensions with typically $i < j$. In many cases, i will be chosen first whereafter j is adapted so as to use all available observations in the block Hankel matrices.

After the estimation of the Kalman filter state sequences \hat{X}_i and \hat{X}_{i+1} in the first step (see [1] and [3]), an estimate for the covariance model $(\hat{A}, \hat{G}, \hat{C}, \hat{D})$ can be obtained. The system matrices (\hat{A}, \hat{C}) can be identified as the solution to the following least-squares problem:

$$(\hat{A}, \hat{C}) = \arg \min_{A, C} J_1(A, C) \quad (4)$$

with

$$J_1(A, C) = \left\| \begin{bmatrix} \hat{X}_{i+1} \\ Y_{i|i} \end{bmatrix} - \begin{bmatrix} A \\ C \end{bmatrix} \cdot \hat{X}_i \right\|^2. \quad (5)$$

One possible way to obtain an estimate \hat{G} for the matrix $G = \mathcal{E}\{x_{k+1}y_k^T\}$ is by taking the last l columns of the reversed controllability matrix $\hat{\Delta}_i = [\hat{A}^{i-1}\hat{G} \ \hat{A}^{i-2}\hat{G} \ \dots \ \hat{A}\hat{G} \ \hat{G}]$, where $\hat{\Delta}_i$ is calculated as $\hat{\Gamma}_i^\dagger Y_{i|2i-1} Y_{0|i-1}^T$, with $\hat{\Gamma}_i = [\hat{C}^T \ \hat{A}_T \hat{C}^T \ \dots \ \hat{A}^{T(i-1)} \hat{C}^T]^T$, and $\hat{\Lambda}_0$ can immediately be derived as $(1/j)Y_{i|j}Y_{i|i}^T$. This is essentially a square root version of the deterministic realization [4], [5] applied to the observed output covariance matrices $\{\bar{\Lambda}_m\}_{m=1}^{2i-1}$, with

$$\bar{\Lambda}_m = \frac{1}{N} \sum_{k=0}^{N-m-1} y_k y_{k+m}^T. \quad (6)$$

In a last step, the covariance model is used to conceive a model in forward innovation form

$$\begin{aligned} \hat{x}_{k+1} &= \hat{A}\hat{x}_k + \hat{K}e_k \\ y_k &= \hat{C}\hat{x}_k + e_k \end{aligned} \quad (7)$$

obtained by first calculating an estimate $\hat{P} = \mathcal{E}\{\hat{x}_k \hat{x}_k^T\}$ for the forward state covariance matrix of (7) through the solution of the forward algebraic Riccati equation:

$$\hat{P} = \hat{A}\hat{P}\hat{A}^T + (\hat{G} - \hat{A}\hat{P}\hat{C}^T)(\hat{\Lambda}_0 - \hat{C}\hat{P}\hat{C}^T)^{-1}(\hat{G} - \hat{A}\hat{P}\hat{C}^T)^T \quad (8)$$

with the forward Kalman filter gain $\hat{K} = (\hat{G} - \hat{A}\hat{P}\hat{C}^T)(\hat{\Lambda}_0 - \hat{C}\hat{P}\hat{C}^T)^{-1}$. The resulting model matrices of the stochastic system are $(\hat{A}, \hat{K}, \hat{C}, I)$ and the covariance matrix $\mathcal{E}\{e_k e_k^T\}$ is given by $\hat{R} = \hat{\Lambda}_0 - \hat{C}\hat{P}\hat{C}^T$. A transformation from (7) to a system of the form (1) is now straightforward [1].

It is important to note here that a valid forward innovation model (7) can only be found if the estimated covariance model $(\hat{A}, \hat{G}, \hat{C}, \hat{D})$ is positive real. This follows immediately from the positive real lemma [6], that states that a covariance model (A, G, C, D) is positive real if and only if the following matrix inequality is satisfied for at least one positive-definite matrix $P = P^T > 0$:

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} = \begin{bmatrix} P & G \\ G^T & D + D^T \end{bmatrix} - \begin{bmatrix} APA^T & APC^T \\ CPA^T & CPC^T \end{bmatrix} \geq 0. \quad (9)$$

By applying the Schur decomposition to (9) it is clear that no solution to (8) will be found unless the covariance model $(\hat{A}, \hat{G}, \hat{C}, \hat{D})$ is positive real. Also, note from the Lyapunov equation in the upper left block of (9) that a positive real model is necessarily stable. In the following section, we will introduce a regularization term in the least-squares cost function (4) to impose positive realness on the covariance model and to ensure a solution to (8).

III. IMPOSING POSITIVE REALNESS BY USING REGULARIZATION

A. Main Idea

The estimation problem that we consider is the following: given matrices $\hat{X}_{i+1}, Y_{i|i}$ and \hat{X}_i and given initial estimates $\hat{A}, \hat{C}, \hat{G}$, and $\hat{\Lambda}_0$, estimate new model matrices \tilde{A}, \tilde{C} such that the resulting model $\tilde{A}, \tilde{C}, \hat{\Lambda}_0$ is positive real. To impose positive realness, we will add a regularization term to the cost function $J_1(A, C)$ from (4)

$$(\tilde{A}_c, \tilde{C}_c) = \arg \min_{A, C} J_1(A, C) + cJ_2(A, C) \quad (10)$$

with

$$J_2(A, C) = \text{Tr} \left(\begin{bmatrix} A \\ C \end{bmatrix} W \begin{bmatrix} A \\ C \end{bmatrix}^T \right) \quad (11)$$

where $c \geq 0$ is a positive real scalar and W a positive definite matrix of appropriate dimensions that satisfies $W - \hat{G}\hat{\Lambda}_0^{-1}\hat{G}^T > 0$ and is typically chosen to be the unity matrix, which is motivated by [7].

A similar regularization term $\text{Tr}(AWA^T)$, involving only the system matrix A was described in [8], and was shown to impose stability on a model. We will show that by the choice of the regularization term $J_2(A, C)$ the covariance model cannot only be made stable, but also positive real, provided the regularization coefficient c is chosen sufficiently large. A further advantage of the regularization term is that the problem (10) remains quadratic and that the optimal solution follows from a linear set of equations

$$\begin{aligned} \begin{bmatrix} \tilde{A}_c \\ \tilde{C}_c \end{bmatrix} &= \begin{bmatrix} \hat{X}_{i+1} \\ Y_{i|i} \end{bmatrix} \cdot \hat{X}_i^T \cdot \left[\hat{X}_i \hat{X}_i^T + cW \right]^{-1} \\ &= \begin{bmatrix} \hat{A} \\ \hat{C} \end{bmatrix} \hat{X}_i \hat{X}_i^T \left[\hat{X}_i \hat{X}_i^T + cW \right]^{-1}. \end{aligned} \quad (12)$$

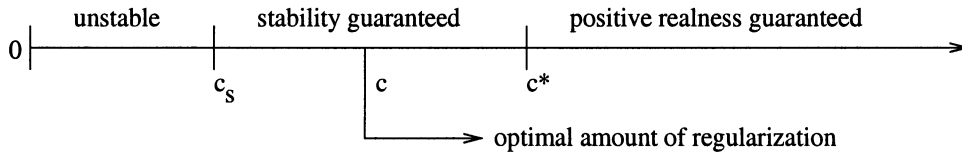
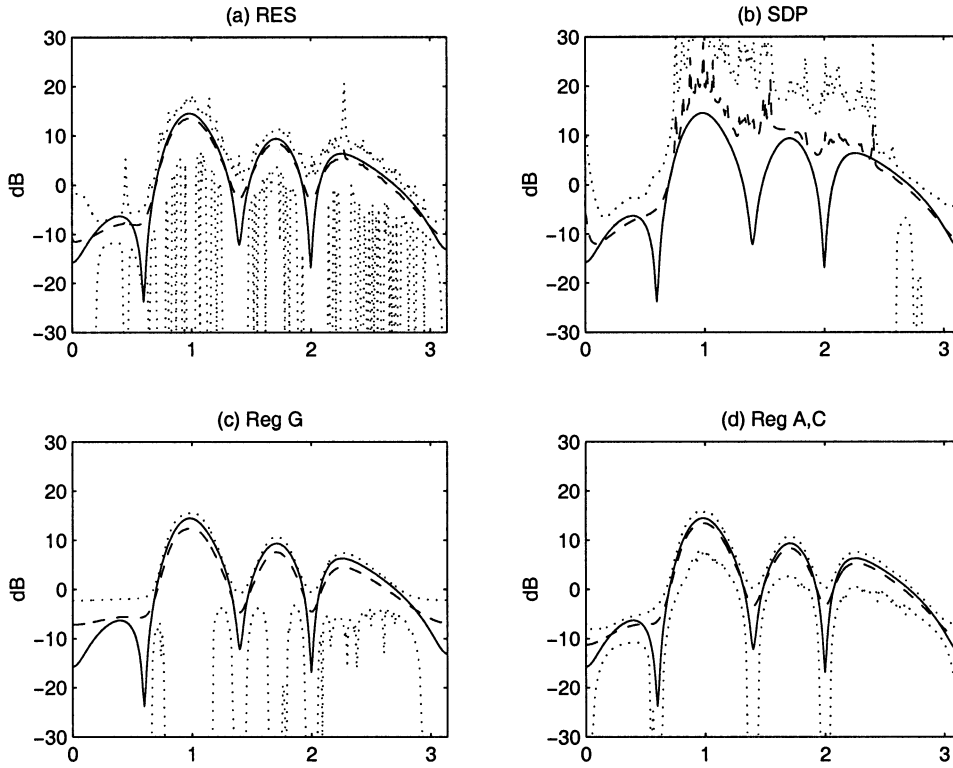
From the optimality of the least-squares estimate (12), it follows that the regularization term $J_2(\tilde{A}_c, \tilde{C}_c)$ is a nonincreasing function of c .

The idea of using regularization to deal with undesirable properties of an estimator is by no means new. In general, regularization amounts to reducing the variance of an estimator at the expense of introducing a hopefully small bias, the so-called bias-variance tradeoff. In function approximation, for instance, regularization is used to impose a certain amount of smoothness and deal with the well known problem of over fitting [9]. Other applications are found in such areas as neural networks [10] support vector machines [11], and system identification [12]. Furthermore, some known techniques can be rewritten in a regularization context. The technique described in [8] to impose stability on a model using regularization, for instance, is essentially equivalent to a technique described in [13], provided a certain choice for the weighting matrices is made in the former reference.

B. Choosing the Regularization Parameter

It will be shown in the following lemma that, by using the regularization term introduced in (10), positive realness can always be imposed provided the regularization coefficient c is chosen sufficiently large.

Lemma 1: Let $\hat{G}, \hat{\Lambda}_0$ be given. Let $W = Q_W Q_W^T > 0$, $W - \hat{G}\hat{\Lambda}_0^{-1}\hat{G}^T > 0$, and define $\hat{\Sigma} = X_i X_i^T$, $\mathcal{L} = \hat{\Sigma} \begin{bmatrix} \hat{A}^T & \hat{C}^T \end{bmatrix} \begin{bmatrix} W & \hat{G} \\ \hat{G}^T & \hat{\Lambda}_0 \end{bmatrix}^{-1} \begin{bmatrix} \hat{A} \\ \hat{C} \end{bmatrix} \hat{\Sigma}$, $P_0 = \hat{\Sigma} W^{-1} \hat{\Sigma} - \mathcal{L}$. Suppose the covariance model $(\hat{A}, \hat{G}, \hat{C}, \hat{\Lambda}_0)$ is not positive real. Then, there exists a c^* such that the system $\tilde{A}_c, \tilde{C}_c, \hat{\Lambda}_0$, with \tilde{A}_c and \tilde{C}_c as in (12), is positive real for $c \geq c^*$, with $c^* = \max_{i|\theta_i \in \mathbb{R}^+} \theta_i$,

Fig. 1. Finding the optimal amount of regularization c .Fig. 2. Averaged spectral density over 1000 runs for the example $(H(z))$ with $\hat{n} = 10$, $i = 16$, $N = 500$ (dashed line) with 95% error region (dotted line). The solid line is the spectral density of the original model used for simulation.

and θ the set of generalized eigenvalues of the following eigenvalue problem:

$$\theta = \lambda \left(\begin{bmatrix} 0_{\hat{n}} & -I_{\hat{n}} \\ P_0 & 2\hat{\Sigma} \end{bmatrix}, - \begin{bmatrix} I_{\hat{n}} & 0_{\hat{n}} \\ 0_{\hat{n}} & W \end{bmatrix} \right). \quad (13)$$

Proof: We will show that (9), with A, C, G and P replaced by $\hat{A}_c, \hat{C}_c, \hat{G}$, and \hat{P} , holds under the assumptions of the lemma for $\hat{P} = W$. This means that

$$\begin{bmatrix} W & \hat{G} \\ \hat{G}^T & \hat{\Lambda}_0 \end{bmatrix} - \begin{bmatrix} \hat{A}_c W \hat{A}_c^T & \hat{A}_c W \hat{C}_c^T \\ \hat{C}_c W \hat{A}_c^T & \hat{C}_c W \hat{C}_c^T \end{bmatrix} \geq 0 \quad (14)$$

where the first term is positive semidefinite since $W > 0$, $W - \hat{G}\hat{\Lambda}_0^{-1}\hat{G}^T > 0$, and \hat{A}_c and \hat{C}_c are as defined in (12). Taking the Schur complement and defining $\hat{\Sigma} = X_i X_i^T$ leads to

$$\left[\begin{array}{c|c} W & \hat{G} \\ \hline \hat{G}^T & \hat{\Lambda}_0 \end{array} \middle| \begin{array}{c} \hat{A}\hat{\Sigma} \\ \hat{C}\hat{\Sigma} \end{array} \right] \geq 0 \quad (15)$$

$$\left[\begin{array}{c|c} \hat{\Sigma}\hat{A}^T & \hat{\Sigma}\hat{C}^T \\ \hline (\hat{\Sigma} + cW)W^{-1}(\hat{\Sigma} + cW) \end{array} \right]$$

and again taking the Schur complement, with $W - \hat{G}\hat{\Lambda}_0^{-1}\hat{G}^T > 0$

$$(\hat{\Sigma} + cW)W^{-1}(\hat{\Sigma} + cW) - \mathcal{L} \geq 0 \quad (16)$$

TABLE I
PERFORMANCE ON SIMULATED DATA

	REG $_{\hat{A},\hat{C}}$	REG $_{\hat{G}}$	RES	SDP
Mean(d_∞)	1.7	2.45	2.19	18.3
Var(d_∞)	0.488	1.57	6.02	4e+03
Mean(d_2)	0.579	0.784	0.591	2.46
Var(d_2)	0.0172	0.0907	0.0294	8.59
Mean(d_1)	1.36	1.98	1.35	4.08
Var(d_1)	0.0709	0.643	0.0655	6.7

which can also be written as

$$c^2 W + 2c\hat{\Sigma} + \hat{\Sigma}W^{-1}\hat{\Sigma} - \mathcal{L} \geq 0, \quad (17)$$

Equation (17) is clearly satisfied for $c \rightarrow \infty$. The exact lower bound c^* for c in (17) is given by the largest positive root of

$$\det(c^2 W + 2c\hat{\Sigma} + \hat{\Sigma}W^{-1}\hat{\Sigma} - \mathcal{L}) = 0. \quad (18)$$

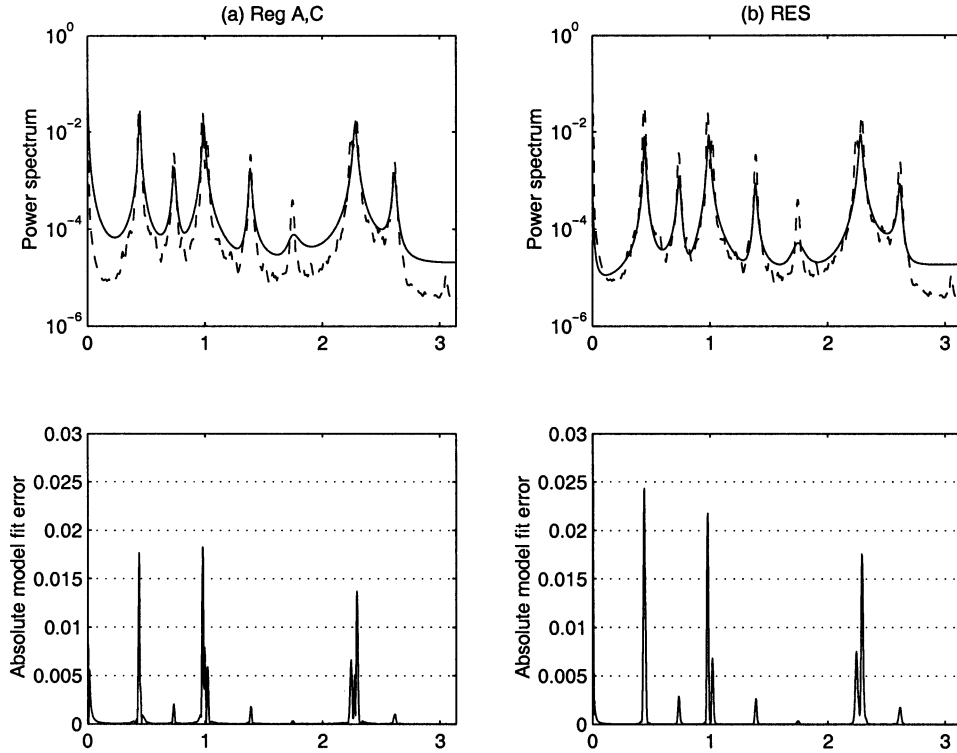


Fig. 3. Output spectra of one of the accelerometers on a steel mast (dashed lines), together with the estimated spectra using $\text{REG}_{\hat{A}, \hat{C}}$ and RES (full line). The absolute differences between the spectra in the uppermost two figures are depicted in the figures in the bottom row. The variances of the differences are $3.71 \cdot 10^{-6}$ for the $\text{REG}_{\hat{A}, \hat{C}}$ case and $11.05 \cdot 10^{-6}$ for the RES case. In similar experiments, the variances for the $\text{REG}_{\hat{C}}$ and SDP techniques were found to be $10.47 \cdot 10^{-6}$ and $15.05 \cdot 10^{-6}$, respectively.

Using the definition of P_0 , this reduces to

$$\begin{aligned} \det(c^2W + 2c\hat{\Sigma} + P_0) &= 0 \\ \iff \det(c(cW + 2\hat{\Sigma}) + P_0) &= 0 \\ \iff \det\left(c \begin{bmatrix} I_{\hat{n}} & 0_{\hat{n}} \\ 0_{\hat{n}} & W \end{bmatrix} + \begin{bmatrix} 0_{\hat{n}} & -I_{\hat{n}} \\ P_0 & 2\hat{\Sigma} \end{bmatrix}\right) &= 0. \end{aligned} \quad (19)$$

Hence, a positive real model is always obtained for $c \geq c^*$, and in particular for $c = c^*$. Furthermore, since any positive real model is necessarily stable [which follows immediately from the upper left part of (9)], stability is automatically guaranteed. However, c^* can be a too conservative estimate. In general it seems reasonable to keep the amount of regularization as low as possible. Hence, one should search for the smallest possible $c \leq c^*$ for which a positive real model is found. A lower bound c_s for c can be found from a theorem presented in [8], where c_s follows from a generalized eigenvalue problem and is shown to be the smallest c imposing stability on the estimated covariance model. As shown in Fig. 1, a minimal c imposing positive realness will always satisfy $c_s \leq c \leq c^*$. When the realization $(\hat{A}_{c_s}, \hat{G}, \hat{C}_{c_s}, \hat{\Lambda}_0)$ is not yet positive real, i.e., $\Phi(z) < 0$ for a certain $z = e^{j\theta}$, we can find a $c \geq c_s$ imposing positive realness, for instance by applying a bisection algorithm on the interval $c_s \leq c \leq c^*$.

IV. RELATION TO OTHER ALGORITHMS

As mentioned in Section I, some alternative techniques have been reported in the literature in order to impose positive realness on a covariance model [1], [14]–[17], many of whom rely on regularization principles. Apart from changing \hat{A} and \hat{C} in the initial covariance model, regularization could also be applied to \hat{G} , $\hat{\Lambda}_0$, or a combination of both. A common problem with many of these alternatives, in contrast to the

method presented here, is that they cannot be used if the initial covariance model is unstable. A situation which is not uncommon in many practical situations. Furthermore, the approach proposed in this note is seen to outperform existing techniques in simulations. As an example, we take a known single-input–single-output (SISO) system with transfer function

$$H(z) = \frac{(z - 0.99e^{\pm 2j})(z - 0.98e^{\pm 1.4j})}{(z - 0.8e^{\pm 2.1j})(z - 0.8e^{\pm j})} \cdot \frac{(z - 0.99e^{\pm 0.6j})(z - 0.9)(z + 0.9)}{(z - 0.8e^{\pm 1.7j})(z - 0.8e^{\pm 0.8j})} \quad (20)$$

of which the spectral density is displayed as the solid curves in Fig. 2(a)–(d). 1000 sequences of Gaussian, zero mean, unit variance white input noise with length $N = 500$ were filtered through (20) and used as input for the stochastic subspace identification algorithm with $\hat{n} = 10$ and $i = 16$. As a total of 727 of the obtained covariance models turned out not to be positive real of whom 182 were unstable, regularization was applied where necessary. The average of the obtained spectra over all 1000 runs are depicted under the title $\text{REG}_{\hat{A}, \hat{C}}$ in Fig. 2(d), together with those of the best-performing methods described in the literature, namely the following.

- RES: A method described in [1] and [16], using an algorithm based on the residuals of the least squares problem (4) displayed in Fig. 2(a). The method described in [16] also deals with the related problem of systems of the form (1) for which an innovation model simply does not exist, e.g., if the output noise in (1) is zero.
- SDP: A method described in [17] which obtains positive real covariance models by solving an SDP-problem, displayed in Fig. 2(b).
- $\text{REG}_{\hat{C}}$: A method described in [14] using regularization on \hat{G} , displayed in Fig. 2(c).

As RES and $\text{REG}_{\hat{C}}$ were unable to deal with unstable covariance models, the covariance model was stabilized for RES and $\text{REG}_{\hat{C}}$ where necessary by using the techniques described in [8]. For a fair comparison, the numerical results for all methods, which are given in Table I, are limited to the set of 545 models that were stable but not positive real. The table lists the average performance of each method on these covariance models and the variance upon this performance. The performance is measured as the average distance d_x between the transfer function of the original model $H(z)$ and the estimated transfer function in x -norm, where x is chosen from the set $\{1, 2, \infty\}$.

From the figures and the table it is clear that the regularization technique described in this note outperforms the others. As for the complexity, all algorithms are roughly $\mathcal{O}(qn^3)$, with q the number of iterations in a regularization approach or an SDP problem. For RES, $q = 1$ as no optimization is performed. These complexity results have been found to be consistent with the required computation times for each iteration in our simulations which were comparable for all methods discussed.

V. PRACTICAL APPLICATION

The regularization procedure described in this note was used to identify a stochastic subspace model from measurements on a steel transmitter mast for cellular phone networks [18]. Nine accelerometers were placed on the mast and the mast's response on the wind turbulence was measured. A 16th-order stochastic SISO subspace model was thereafter created for one of the accelerometers using subspace identification with i , the number of block rows, set to 32. For this set of parameters, a stable but nonpositive real covariance model was obtained, whereafter the different regularization techniques described in this note were used to obtain positive real models. The original measurement spectrum and the modeled spectra resulting from the two best performing techniques in the simulations of Section IV, namely RES and $\text{REG}_{\hat{A}, \hat{C}}$ are displayed in Fig. 3, together with the absolute values of the differences between them. Note that all the spectra are strictly positive. Also, note that the RES technique performs better in the regions between the peaks, while $\text{REG}_{\hat{A}, \hat{C}}$ is seen to fit the peaks themselves better. For comparison, the variances of the model fit errors for $\text{REG}_{\hat{C}}$ and SDP are given in below the figure.

VI. CONCLUSION

Stochastic subspace methods for the identification of linear time-invariant systems are known to be asymptotically unbiased [3]. However, if a finite amount of data is used, the procedure might break down due to positive realness problems. In this note, a regularization approach was proposed to impose positive realness on a formerly identified covariance model. It was shown that, if an adequate amount of regularization is used, a positive real model can always be obtained. The simulation results indicate that this new approach yields better models than other existing techniques.

REFERENCES

- [1] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory—Implementation—Applications*. Boston, MA: Kluwer, 1996.
- [2] A. Dahlén, A. Lindquist, and J. Mari, "Experimental evidence showing that stochastic subspace identification methods may fail," *Syst. Control Lett.*, vol. 34, pp. 303–312, 1998.
- [3] P. Van Overschee and B. De Moor, "Subspace algorithms for the stochastic identification problem," *Automatica*, vol. 29, no. 3, pp. 649–660, Mar. 1993.

- [4] B. L. Ho and R. E. Kalman, "Efficient construction of linear state variable models from input/output functions," *Regelungstechnik*, vol. 14, pp. 545–548, 1966.
- [5] H. Zeiger and A. McEwen, "Approximate linear realizations of given dimension via Ho's algorithm," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 390–396, Apr. 1974.
- [6] P. Faurre, M. Clerget, and F. German, "Opérateurs rationnels positifs, application à l'hyperstabilité et aux processus aléatoires," *Méth. Math. l'Informat.*, vol. 8, 1978.
- [7] S. F. Gull, "Bayesian inductive inference and maximum entropy," in *Maximum-Entropy and Bayesian Methods in Science and Engineering*, G. J. Erickson and R. Smith, Eds. Dordrecht, The Netherlands: Kluwer, 1988, vol. 1, pp. 53–74.
- [8] T. Van Gestel, J. Suykens, P. Van Dooren, and B. De Moor, "Identification of stable models in subspace identification by using regularization," *IEEE Trans. Automat. Contr.*, vol. 46, pp. 1416–1420, Sept. 2001.
- [9] F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural Comput.*, vol. 7, no. 2, pp. 219–269, 1995.
- [10] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.
- [11] A. Smola, B. Schölkopf, and K. R. Müller, "The connection between regularization operators and support vector kernels," *Neural Networks*, vol. 11, pp. 637–649, 1998.
- [12] J. Sjöberg, T. McKelvey, and L. Ljung, "On the use of regularization in system identification," in *Proc. 12th IFAC World Congr.*, vol. 7, Sydney, Australia, 1993, pp. 381–386.
- [13] N. L. C. Chui and M. Maciejowski, "Realization of stable models with subspace methods," *Automatica*, no. 32, pp. 1587–1595, 1996.
- [14] R. J. Vaccaro and T. Vukina, "A solution to the positivity problem in the state-space approach to modeling vector-valued time series," *J. Econ. Dyna. Control*, vol. 17, pp. 401–421, 1993.
- [15] K. Peternell, "Identification of linear dynamic systems by subspace and realization-based algorithms," Ph.D. dissertation, T.U. Wien, Vienna, Austria, 1995.
- [16] R. Johansson, M. Verhaegen, C. T. Chun Tung Chou, and A. Robertsson, "Residual models and stochastic realization in state-space identification," *Int. J. Control*, vol. 74, no. 10, pp. 988–995, 2001.
- [17] J. Mari, P. Stoica, and T. McKelvey, "Vector ARMA estimation: A reliable subspace approach," *IEEE Trans. Signal Processing*, vol. 48, pp. 2092–2104, July 2000.
- [18] B. Peeters, "Stochastic subspace system identification of a steel transmitter mast," in *Proc. Int. Modal Analysis Conf. (IMAC 16)*, Santa Barbara, CA, Feb. 1998, pp. 130–136.