A New Method for Moving-Average Parameter Estimation

Petre Stoica[†]   Lin Du[‡]   Jian Li[‡]   Tryphon Georgiou[§]

[†]Dept. of Information Technology, Uppsala University, SE-75105 Uppsala, Sweden
[‡]Dept. of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6130, USA
[§]Dept. of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA

Abstract— We introduce an apparently original method for moving-average parameter estimation, based on covariance fitting and convex optimization. The proposed method is shown by means of numerical simulation to provide much more accurate parameter estimates, in difficult scenarios, than a related existing method does. We derive the new method via an analogy with a covariance fitting interpretation of the Capon beamforming from array processing. In the process, we also point out some new facts on Capon beamforming.

I. INTRODUCTION AND PRELIMINARIES

Let $\{y(k)\}$ be a moving-average (MA) time series of order n:

$$y(k) = c_0 e(k) + c_1 e(k-1) + \cdots + c_n e(k-n), \quad k = 1, 2, \cdots \tag{1}$$

where $\{e(k)\}$ is the driving white noise sequence,

$$E[e(k)e(\bar{k})] = \begin{cases} 1 & \text{if } k = \bar{k} \\ 0 & \text{else} \end{cases} \tag{2}$$

and

$$c_0 + c_1 z + \cdots + c_n z^n \neq 0 \quad \text{for} \quad |z| \leq 1 \tag{3}$$

(as is well-known, the "minimum-phase" condition in (3) ensures that (1) is a unique description of the power spectrum of $\{y(k)\}$ [1]).

Our problem is to obtain estimates of the MA parameters $\{c_p\}$ from N observations of $\{y(k)\}$. For the sake of simplicity we assume that $\{y(k)\}$ is a scalar sequence; however, note that the discussion in this paper can be readily extended to vector sequences by using results from [2]. We also assume that the MA order, n, is prespecified.

The maximum likelihood estimation (MLE) of $\{c_p\}$ requires making assumptions on the distribution of $\{y(k)\}$, and even under the Gaussian assumption it leads to a highly nonlinear problem. Consequently, a host of alternative computationally simpler methods have been proposed for MA parameter estimation (see e.g., [1], [3]). Of these methods, the covariance fitting approach of [4], [5] (see also [3]) is somewhat unique in that it has, in its most refined form, an accuracy comparable with that of the MLE, and yet it obtains parameter estimates from the solution of a convex problem that can be reliably and efficiently computed in polynomial time.

In this paper we introduce a new method for MA parameter estimation that, similarly to the method of [4], [5], is based on a covariance fitting criterion whose minimization leads to a convex optimization problem. The main difference between the proposed method and that in the cited references lies in a novel form of the fitting criterion. The inspiration for using this

new form of covariance fitting criterion comes from a recent letter ([6]) as well as from one of the possible derivations of the Capon beamforming method in array processing [7], [8] (also [3]) - see the next section for details. Compared with the basic form of the method in [4], [5], the method proposed here is shown via Monte-Carlo simulations to provide much more accurate MA parameter estimates in difficult cases (where the roots of the polynomial in (3) lie close to the unit circle).

II. MAIN RESULTS

A. The basic method of [4], [5]

Let

$$\hat{r}_p = \frac{1}{N} \sum_{k=p+1}^{N} y(k)y(k-p) = \hat{r}_{-p} \quad p = 0, 1, 2, \cdots \tag{4}$$

denote the standard sample covariances of $\{y(k)\}$. Also, let $\{r_p\}$ denote the theoretical covariances of $\{y(k)\}$. Two convex parameterizations of $\{r_p\}$ have been derived in [4], [5] (see also the references therein): the "trace parameterization" and the "Kalman-Yakubovich-Popov lemma - based parameterization." These two parameterizations are equivalent in most respects, but the trace parameterization is easier to describe and thus it will be the one used in what follows. In this parameterization, the covariances $\{r_p\}$ of the MA sequence in (1) have the following expressions:

$$r_p = \text{tr}_p(\mathbf{Q}) \quad p = 0, \pm 1, \cdots, \pm n; \qquad \mathbf{Q} \geq 0 \tag{5}$$

where the $(n+1) \times (n+1)$ matrix $\mathbf{Q}$ is positive semidefinite ($\mathbf{Q} \geq 0$) but otherwise arbitrary, and $\text{tr}_p(\mathbf{Q})$ denotes the sum of the elements on the pth diagonal of $\mathbf{Q}$ (with the main diagonal corresponding to $p = 0$, and so forth).

Making use of (5), the basic form of the method of [4], [5], which we refer to as the basic method (BM), can be described by a two-step procedure:

Step 1. Solve the convex minimization problem

$$\min_{\mathbf{Q} \geq 0} \sum_{p=-n}^{n} (\hat{r}_p - r_p)^2; \qquad \{r_p = \text{tr}_p(\mathbf{Q})\} \tag{6}$$

Step 2. Obtain estimates of $\{c_p\}$ from the solution of Step 1 by using a spectral factorization algorithm.

The problem in Step 1 can be easily reformulated as a semi-definite program (SDP) that can be solved reliably and efficiently in polynomial time using public-domain software [9], [10]. The spectral factorization problem in Step 2 can also be solved efficiently by means of any of several available algorithms (see, e.g., [11] for a recent account). Therefore BM is a computationally appealing method. However, from

an estimation accuracy viewpoint, the parameter estimates obtained with BM may be statistically rather inefficient.

The approach employed in [4], [5] to improve the statistical efficiency of BM relies on a somewhat involved weighted covariance fitting criterion. Here we take a simpler route, as will be explained in what follows. Compared with the BM, or its enhanced version in [4], [5], which rely on fitting $\{r_p\}$ to $\{\hat{r}_p\}$, the new method (NM) is based on fitting the theoretical covariance matrix

$$\mathbf{R} = \begin{bmatrix} r_0 & r_1 & \cdots & r_n \\ r_1 & r_0 & \cdots & r_{n-1} \\ \vdots & \ddots & \ddots & \vdots \\ r_n & \cdots & r_1 & r_0 \end{bmatrix} \quad (7)$$

to an estimate of it, let us say $\hat{\mathbf{R}}$, obtained from the available observations. A possible expression for $\hat{\mathbf{R}}$ is the standard Toeplitz sample covariance matrix:

$$\hat{\mathbf{R}}_{\mathrm{T}} = \begin{bmatrix} \hat{r}_0 & \hat{r}_1 & \cdots & \hat{r}_n \\ \hat{r}_1 & \hat{r}_0 & \cdots & \hat{r}_{n-1} \\ \vdots & \ddots & \ddots & \vdots \\ \hat{r}_n & \cdots & \hat{r}_1 & \hat{r}_0 \end{bmatrix} \quad (8)$$

Another commonly used $\hat{\mathbf{R}}$ is the following non-Toeplitz sample covariance matrix:

$$\hat{\mathbf{R}}_{\mathrm{nT}} = \frac{1}{N-n} \sum_{k=n+1}^{N} \begin{bmatrix} y(k) \\ \vdots \\ y(k-n) \end{bmatrix} \begin{bmatrix} y(k) & \cdots & y(k-n) \end{bmatrix} \quad (9)$$

There is compelling evidence, obtained from numerical studies of a diversity of estimation problems, which suggests that in small-or-medium-sized samples the use of $\hat{\mathbf{R}}_{\mathrm{nT}}$ can lead to more accurate parameter estimates than the use of $\hat{\mathbf{R}}_{\mathrm{T}}$. We will show in the numerical example section that the same is usually true in the MA parameter estimation problem considered here, in the sense that the NM based on $\hat{\mathbf{R}}_{\mathrm{nT}}$ is more accurate than the NM that uses $\hat{\mathbf{R}}_{\mathrm{T}}$; we will also show that NM is much more accurate than BM, which can use only the Toeplitz covariances.

B. Capon and Pisarenko methods

The main inspiration for the NM has come from a covariance fitting-based derivation of the Capon beamformer in array processing. In the said derivation, the theoretical covariance matrix $\mathbf{R}$ of the observed data is modeled (using real-valued variables, for analogy with the estimation problem discussed in this paper) as:

$$\mathbf{R} = \sigma^2 \mathbf{a}\mathbf{a}^T + \mathbf{\Gamma} \quad (10)$$

where σ^2 is the signal power, which is the unknown parameter of main interest, $\mathbf{a}$ is a given vector, and $\mathbf{\Gamma}$ is an unknown residual covariance matrix (which is usually a nuisance quantity). The Capon beamforming method for determining σ^2 can be obtained as the solution to the following covariance fitting problem [7], [8]:

$$\max_{\sigma^2} \quad \sigma^2 \quad \text{subject to} \quad \mathbf{R} - \sigma^2 \mathbf{a}\mathbf{a}^T \geq 0 \quad (11)$$

where in practice $\mathbf{R}$ must be replaced by $\hat{\mathbf{R}}$.

Along similar lines to the Capon formalism, in the so-called Pisarenko harmonic analysis [3], one seeks to decompose a given Toeplitz covariance $\mathbf{R}$ into the sum of a singular Toeplitz matrix and a matrix that corresponds to the background white noise. The variance ν^2 of the white noise can be obtained as the solution to

$$\max_{\nu^2} \quad \nu^2 \quad \text{subject to} \quad \mathbf{R} - \nu^2 \mathbf{I} \geq 0 \quad (11a)$$

where $\mathbf{I}$ denotes the identity matrix. Clearly the solution of the above optimization problem coincides with the smallest eigenvalue of $\mathbf{R}$, and the residual matrix $\mathbf{\Gamma}$ in the decomposition

$$\mathbf{R} = \nu^2 \mathbf{I} + \mathbf{\Gamma} \quad (11b)$$

is a singular Toeplitz matrix that corresponds to a finite number of spectral lines.

C. The new method

It follows from the discussion in the previous sub-section that the basic rationale in Capon beamforming as well as in Pisarenko harmonic analysis can be seen as seeking a decomposition of the given or estimated covariance $\hat{\mathbf{R}}$ into the sum of a covariance matrix corresponding to a component that is postulated as being present, and a residual matrix which accounts for noise, uncertainty, estimation errors, or signals of a particular type. Reference [6] (Problem 1) explores this idea for decomposing Toeplitz matrices into one corresponding to an MA noise-component plus a singular one, for the purpose of identifying possible spectral lines in the residual. Inspired by the Capon and Pisarenko rationales and by the approach in [6], we explore a similar decomposition from which we seek to estimate the MA parameters of the underlying process. More specifically, we propose to estimate the MA covariances $\{r_p\}$ (and thus the MA parameters $\{c_p\}$, see Step 2 of BM) by solving the following covariance fitting problem:

$$\max_{\mathbf{Q} \geq 0} \quad r_0 \quad \text{subject to} \quad \hat{\mathbf{R}} - \mathbf{R} \geq 0; \quad \{r_p = \mathrm{tr}_p(\mathbf{Q})\} \quad (12)$$

where $\hat{\mathbf{R}}$ is either $\hat{\mathbf{R}}_{\mathrm{T}}$ or $\hat{\mathbf{R}}_{\mathrm{nT}}$; we denote the resulting two versions of NM as NM_{T} (based on $\hat{\mathbf{R}}_{\mathrm{T}}$) and $\mathrm{NM}_{\mathrm{nT}}$ (based on $\hat{\mathbf{R}}_{\mathrm{nT}}$). An equivalent, but perhaps intuitively more appealing, form of (12) is as follows:

$$\min_{\mathbf{Q} \geq 0} \quad \mathrm{tr}(\hat{\mathbf{R}} - \mathbf{R}) \quad \text{subject to} \quad \hat{\mathbf{R}} - \mathbf{R} \geq 0; \quad \{r_p = \mathrm{tr}_p(\mathbf{Q})\} \quad (13)$$

Similarly to (6), this is a convex problem (namely a SDP) that can be efficiently solved in polynomial time using publicly available software [9], [10].

Several remarks on the NM are in order:
(i) The trace criterion in (13) can be replaced by other related criteria, for instance by

$$\min_{\mathbf{Q} \geq 0} \quad \lambda_{\max}(\hat{\mathbf{R}} - \mathbf{R}) \quad \text{subject to} \quad \hat{\mathbf{R}} - \mathbf{R} \geq 0; \quad \{r_p = \mathrm{tr}_p(\mathbf{Q})\} \quad (14)$$

where $\lambda_{\max}$ denotes the maximum eigenvalue. The so-obtained problems, such as (14) above, are also convex and therefore they can be efficiently solved as well. However, we have observed in a number of numerical simulations that the statistical accuracies of NM and of the modified NMs (such as (14)) are quite similar to one another. Therefore, for conciseness reasons, we will focus on NM.

Interestingly, for the Capon beamforming problem in (11), we show in the Appendix that (11) (with $\mathbf{R}$ replaced by $\hat{\mathbf{R}}$),

or equivalently

$$\min_{\sigma^2} \quad \mathrm{tr}(\hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T) \qquad \text{subject to} \quad \hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T \geq 0 \quad (15)$$

has the same solution as the following alternative problems:

$$\min_{\sigma^2} \quad \lambda_{\max}(\hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T) \qquad \text{subject to} \quad \hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T \geq 0 \tag{16}$$

and

$$\min_{\sigma^2} \quad \det(\hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T) \qquad \text{subject to} \quad \hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T \geq 0 \tag{17}$$

where $\det(\cdot)$ denotes the determinant. While this apparently new result on Capon beamforming is not strictly related to the MA parameter estimation problem discussed here, at least it lends some theoretical support to our observation that NM and certain modifications of it, such as (14), have similar statistical performances.

(ii) Evidently BM cannot make use of covariances $\{\hat{r}_p\}$ for $p > n$, whereas NM can be readily extended to use longer covariance sequences. However, we have observed empirically that the so-obtained extension of NM does not necessarily have better accuracy than NM. A heuristic explanation of this behavior is as follows: an indirect effect of increasing the dimension of $\hat{\mathbf{R}}$ and $\mathbf{R}$ in (13) is to decrease the "weight" of the covariances $\{\hat{r}_p\}$ and $\{r_p\}$ for $p = 0, \cdots, n$ (which are the parameters of main interest) in the fitting criterion; furthermore, the covariance estimates with lags larger than n carry "information" on the MA parameters only via their correlation with the covariance estimates for lags $0, \cdots, n$, and this correlation is not exploited adequately in the fitting criterion used by NM.

(iii) We might think of reversing the positions of $\hat{\mathbf{R}}$ and $\mathbf{R}$ in (13) to obtain a different version of NM, viz.:

$$\min_{\mathbf{Q} \geq 0} \quad \mathrm{tr}(\mathbf{R} - \hat{\mathbf{R}}) \qquad \text{subject to} \quad \mathbf{R} - \hat{\mathbf{R}} \geq 0; \quad \{r_p = \mathrm{tr}_p(\mathbf{Q})\} \tag{18}$$

However, somewhat similarly to what we said in (ii) above, we have observed empirically that this modification of NM may provide less accurate parameter estimates than (13). Heuristically, we can try to explain this behavior in the following way: for the sake of discussion, let $\hat{\mathbf{R}} = \hat{\mathbf{R}}_T$ in (13) and (18); then (13) implies $\hat{r}_0 > r_0$, whereas for (18) we must have $r_0 > \hat{r}_0$; when $r_0 > \hat{r}_0$, the conditions for $\{r_p\}_{p=1}^n$ to belong to the set of MA(n) covariances are weaker than the corresponding conditions when $r_0 < \hat{r}_0$; consequently, for (18) the minimizing $\{r_p\}_{p=1}^n$ may be closer to $\{\hat{r}_p\}_{p=1}^n$ than for (13); this fact may explain why (18) has been observed in some of our simulations to behave somewhat similarly to BM.

Finally, we remark on the fact that while the condition $\hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T \geq 0$ in the Capon beamforming problem (see (11)) is easily motivated, the similar condition $\hat{\mathbf{R}} - \mathbf{R} \geq 0$, used in the NM for MA parameter estimation, is more intriguing. However, "the proof of the pudding is in the eating": the NM works well, and it provides more accurate parameter estimates than the BM in difficult scenarios, as we show in the next section.

III. Numerical Illustration

For MA sequences with roots well outside the unit circle (see (3)), the sample covariances $\{\hat{r}_p\}_{p=0}^n$ belong to the set of valid MA(n) covariances, with a high probability. In such cases, BM and NM give very similar results (the solution to the covariance fitting problem is likely to be $\{r_p = \hat{r}_p\}_{p=0}^n$ for both BM and NM).

In this section we consider a "difficult example" (see [4]): an MA(3) for which the polynomial in (3) has roots at $\frac{1}{0.95}$ and $\frac{1}{0.98 e^{\pm i\pi/4}}$. The corresponding MA coefficients have the following values:

$$c_0 = 1 \quad c_1 = -2.3359 \quad c_2 = 2.2770 \quad c_3 = -0.9124 \tag{19}$$

We will use BM, NM_T and NM_{nT} to estimate $\{c_p\}$ from samples of $\{y(k)\}$ of varying length: $N = 100, 200, \cdots, 1000$. For each method and each value of N, we estimate the average mean squared error (AMSE) of the parameter estimates, viz.

$$\text{AMSE} = \frac{1}{n+1} \sum_{p=0}^{n} E(\hat{c}_p - c_p)^2 \tag{20}$$

by using 1000 Monte-Carlo simulation runs. The obtained results are shown in Fig. 1.
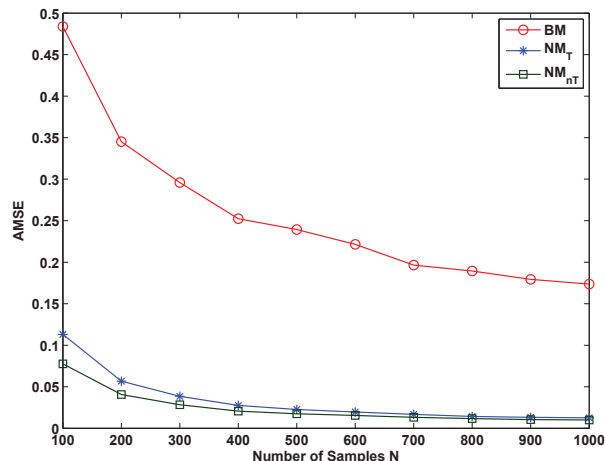

Fig. 1. The estimated AMSE of BM, NM_T and NM_{nT} versus N.

Fig. 2 shows the theoretical spectrum of the considered MA(3) sequence, viz. $|C(e^{i\omega})|^2$ (where $C(\cdot)$ is the polynomial in (3) and $\omega \in [0, 2\pi]$ is the frequency variable), along with the mean and standard deviation of the spectra estimated via BM and NM_{nT} for $N = 1000$. As can be seen from these figures, NM_{nT} provides slightly more accurate estimates than NM_T and significantly more accurate estimates than BM.

IV. Conclusions

We have proposed a computationally attractive MA parameter estimation method based on the use of convex optimization and of an original covariance fitting criterion. The new method has been shown via numerical simulations to provide more accurate parameter estimates than the basic version of an existing competitive method.

Appendix: Problems (15), (16) and (17) have the same solution

First note (assuming that $\hat{\mathbf{R}}$ is nonsingular) that:

$$\det(\hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T) = \det(\hat{\mathbf{R}}) \det(\mathbf{I} - \hat{\mathbf{R}}^{-1} \sigma^2 \mathbf{a}\mathbf{a}^T)$$
$$= \det(\hat{\mathbf{R}})(1 - \sigma^2 \mathbf{a}^T \hat{\mathbf{R}}^{-1} \mathbf{a}) \tag{A.1}$$

which implies immediately that the problems (15) and (17) are equivalent. To show the same result for (15) and (16), let

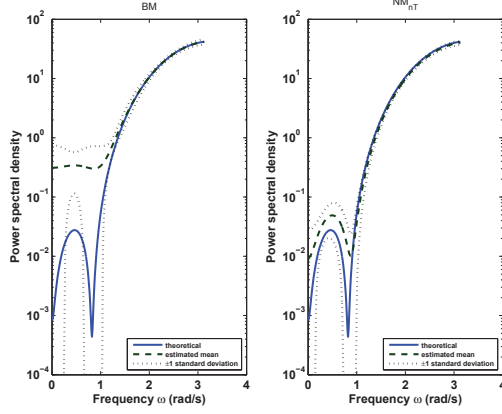Fig. 2. The theoretical spectrum along with the mean, and the mean ± 1 standard deviation curves for the spectra estimated via BM and NM_{nT} ($N = 1000$).

$\hat{\mathbf{R}}^{-\frac{1}{2}}$ denote a symmetric square-root of $\hat{\mathbf{R}}^{-1}$, and observe that:

$$\hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T \geq 0 \Leftrightarrow$$
$$\mathbf{I} - \sigma^2 \hat{\mathbf{R}}^{-\frac{1}{2}} \mathbf{a}\mathbf{a}^T \hat{\mathbf{R}}^{-\frac{1}{2}} \geq 0 \Leftrightarrow$$
$$\sigma^2 \leq \frac{1}{\mathbf{a}^T \hat{\mathbf{R}}^{-1} \mathbf{a}} \triangleq \hat{\sigma}^2 \qquad \text{(A.2)}$$

and therefore that $\hat{\sigma}^2$ above is the solution to (15). The previous calculation also implies that

$$\hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T = \hat{\mathbf{R}} - \hat{\sigma}^2 \mathbf{a}\mathbf{a}^T + (\hat{\sigma}^2 - \sigma^2)\mathbf{a}\mathbf{a}^T \geq \hat{\mathbf{R}} - \hat{\sigma}^2 \mathbf{a}\mathbf{a}^T \qquad \text{(A.3)}$$

for all values of σ^2 that satisfy the constraint in (15) or (16). It follows that $\lambda_{\max}(\hat{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^T) \geq \lambda_{\max}(\hat{\mathbf{R}} - \hat{\sigma}^2 \mathbf{a}\mathbf{a}^T)$, under the constraint in (16) on σ^2, and hence that $\hat{\sigma}^2$ is the solution to the problem (16) as well.

REFERENCES

[1] T. W. Anderson, *The Statistical Analysis of Time Series*. New York, NY: Wiley, 1994.

[2] P. Stoica, L. Xu, J. Li, and Y. Xie, "Optimal correction of an indefinite estimated MA spectral density matrix," *Statistics and Probability Letters*, vol. 77, pp. 973–980, June 2007.

[3] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice-Hall, 2005.

[4] P. Stoica, T. McKelvey, and J. Mari, "MA estimation in polynomial time," *IEEE Transactions on Signal Processing*, vol. 48, pp. 1999–2012, July 2000.

[5] B. Dumitrescu, I. Tabus, and P. Stoica, "On the parameterization of positive real sequences and MA parameter estimation," *IEEE Transactions on Signal Processing*, vol. 49, pp. 2630–2639, November 2001.

[6] T. T. Georgiou, "Decomposition of Toeplitz matrices via convex optimization," *IEEE Signal Processing Letters*, vol. 13, pp. 537–540, September 2006.

[7] T. L. Marzetta, "A new interpretation for Capon's maximum likelihood method of frequency-wavenumber spectrum estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, pp. 445–449, April 1983.

[8] P. Stoica, Z. Wang, and J. Li, "Robust Capon beamforming," *IEEE Signal Processing Letters*, vol. 10, pp. 172–175, June 2003.

[9] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, October 1999. Available: http://sedumi.ie.lehigh.edu/.

[10] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," *The 2004 IEEE International Symposium on Computer Aided Control Systems Design, Taipei, Taiwan,*, pp. pp. 284–289, September 2004. Available: http://users.isy.liu.se/johanl/yalmip/.

[11] L. M. Li, "Factorization of moving-average spectral densities by state-space representations and stacking," *Journal of Multivariate Analysis*, vol. 96, pp. 425–438, October 2005.