

## A subspace method for frequency selective identification of stochastic systems<sup>\*</sup>

T. McKelvey<sup>\*</sup> H. Atala<sup>\*\*</sup> D.O. Blanco Parada<sup>\*\*</sup>

<sup>\*</sup> Chalmers University of Technology, Gothenburg, Sweden (Tel: +46 31 7728061; e-mail: mckelvey@chalmers.se).

<sup>\*\*</sup> Ericsson AB, Lund, Sweden  
(e-mail:{hector.atala,david.blanco.parada}@ericsson.com)

---

**Abstract:** A parametric method for the estimation of vector valued discrete-time stochastic systems or equivalently the spectrum of a stochastic process is presented. The key feature is that the method can be used to frequency selectively fit the model to the data. This means that parts of the spectrum can be modeled with a lower model order than otherwise would be necessary if the entire spectrum would be modeled. The method is based on a frequency domain subspace method which delivers a state-space model. It explicitly takes into account that the frequency domain data is derived from finite data and hence suppresses the leakage effects. Furthermore the method employs convex optimization to guarantee that the estimated parametric model represents a non-negative spectrum.

Keywords: subspace methods; frequency domain identification; Time series modelling

---

### 1. INTRODUCTION

Identification of discrete time stochastic systems or, equivalently, parametric spectral estimation is used in numerous signal processing and control applications. This paper deals with estimation of finite order discrete-time stochastic systems from time domain output samples. A wide range of techniques are available ranging from simple least-squares estimation of AR models to more complex techniques such as subspace and maximum-likelihood methods, Stoica and Moses (1997); Van Overschee and De Moor (1996); Marple (1987). A common denominator for most methods is that the parametric model sought should describe the second order properties of the sampled signal as well as possible. Hence, the entire frequency spectrum is captured by the model. In some applications the measured data is a mix between important and unimportant parts. Particularly, here we consider the case when it is desired to only model a frequency sub-band of the spectrum and disregard the spectrum outside this band. From the application point of view only the interesting sub-band needs to be modeled with a parametric model. One benefit of only partially model the spectrum is that a lower order model can be used to accurately model the sub-band. This becomes particularly important if the spectrum outside the sub-band of interest has a complex shape and thus would require a very high model order and consequently a large amount of data. Of course, from an optimality point of view, deliberate under-modeling will often have a price in terms of reduced performance and this effect has to balance the above mentioned advantages.

The method introduced in this paper is frequency selective in the user can select in which frequency bands the

parametric model should fit the spectrum. It is based on a subspace method coupled with convex optimization to guarantee that the estimated model correspond to a valid non-negative spectrum. Similar combinations has been described in earlier work but only for a full model of the spectrum by Stoica et al. (2000); Mari et al. (2000) Byrnes et al. (2001). Similar techniques has also been applied for identification of positive real systems, e.g. McKelvey and Moheimani (2005); Hoagg et al. (2004). Also a related frequency domain subspace method has been described earlier by Van Overschee et al. (1997). However that method departs from samples of a discrete time spectrum which were assumed to be available.

The technique we present below will pre-process the time-domain samples by usage of the discrete Fourier transform (DFT) to filter out the sub-band of interest. The estimation of the parametric model is then performed using a subspace based method based on techniques in McKelvey et al. (1996).

The paper is organized as follows. In the following section we introduce the stochastic model, the associated autocorrelation function and the important parametric model of the truncated half spectrum. In Section 3 we illustrate how the stochastic system can be realized from a covariance sequence. In Section 4 we present the estimation algorithm. In Section 5 a small example is presented and in the last section the conclusions are presented.

### 2. PRELIMINARIES

For a discrete time, zero-mean, *wide sense stationary* random process  $\mathbf{y}(t) \in \mathbb{R}^n$ , the autocorrelation sequence  $\mathbf{R}_{\mathbf{y}\mathbf{y}}(\tau)$  is defined as

$$\mathbf{R}_{\mathbf{y}\mathbf{y}}(\tau) \triangleq E\{\mathbf{y}(t)\mathbf{y}^T(t-\tau)\} \quad (1)$$

---

<sup>\*</sup> This work was supported in part by the Swedish Research Council (VR)

where  $\mathbf{R}_{yy}(\tau) \in \mathbb{R}^{n \times n}$ ,  $E\{\cdot\}$  denotes the expectation,  $t$  is the sample index of the process and  $\tau$  is the lag length. Directly from the definition we have the property  $\mathbf{R}_{yy}(-\tau) = \mathbf{R}_{yy}^T(\tau)$ .

### 2.1 Power Spectrum

Based on the auto-correlation function, the Power Spectrum is defined as the Fourier transformation of the auto-correlation function,

$$\Phi(\omega) \triangleq \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{yy}(\tau) e^{-j\omega\tau}, \quad \omega \in \mathbb{R} \quad (2)$$

which is a *Hermitian positive semi definite*, periodic function with period  $2\pi$  where, for all  $\omega$

$$\mathbf{v}^* \Phi(\omega) \mathbf{v} \geq 0; \quad \forall \mathbf{v} \in \mathbb{C}^n, \mathbf{v} \neq \mathbf{0}. \quad (3)$$

Here  $\mathbf{v}^*$  denotes transpose and complex conjugation.

By considering only the positive lags of the autocorrelation function and half the zero lag we introduce the *half spectrum*,

$$\phi(\omega) \triangleq \frac{1}{2} \mathbf{R}_{yy}(0) + \sum_{\tau=1}^{\infty} \mathbf{R}_{yy}(\tau) e^{-j\omega\tau} \quad (4)$$

and consequently  $\Phi(\omega) = \phi(\omega) + \phi(\omega)^*$ . In the estimation we will only deal with a finite number of samples of the auto-correlation function and it is convenient to introduce the *truncated half spectrum* which is DFT of the  $N$  samples long truncated auto-correlation function,

$$\varphi(k) = \frac{1}{2} \mathbf{R}_{yy}(0) + \sum_{\tau=1}^{N-1} \mathbf{R}_{yy}(\tau) W_N^{-k\tau} \quad (5)$$

where  $W_N = e^{j2\pi/N}$ .

### 2.2 Stochastic system model class

This paper deals with the identification of finite dimensional vector valued stochastic systems. A state-space model class is used in the following form

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}\mathbf{x}(t) + \mathbf{w}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{v}(t) \end{aligned} \quad (6)$$

where  $\mathbf{x}(t) \in \mathbb{R}^n$  is the state vector,  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is a matrix with all eigenvalues inside the unit disc and  $\mathbf{C} \in \mathbb{R}^{p \times n}$ . The signals  $\mathbf{w}(t) \in \mathbb{R}^n$  and  $\mathbf{v}(t) \in \mathbb{R}^p$  are zero mean stochastic processes which are temporally white and with a joint covariance matrix

$$E \left\{ \begin{bmatrix} \mathbf{w}(t) \\ \mathbf{v}(t) \end{bmatrix} \begin{bmatrix} \mathbf{w}(t) \\ \mathbf{v}(t) \end{bmatrix}^T \right\} \triangleq \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{12}^T & \mathbf{Q}_{22} \end{bmatrix} \triangleq \mathbf{Q} \quad (7)$$

We also assume that the model order  $n$ , i.e. the length of the state vector  $\mathbf{x}(t)$  is minimal. The auto-correlation function for the stochastic state-space model can easily be derived by using the equations (6) and (7). We obtain the relations

$$\begin{aligned} \mathbf{R}_{xx}(\tau) &= \mathbf{A}\mathbf{R}_{xx}(\tau-1), \quad \tau > 0 \\ \mathbf{R}_{yy}(\tau) &= \mathbf{C}\mathbf{A}^T \mathbf{R}_{xx}(0) \mathbf{C}^* + \mathbf{C}\mathbf{A}^{\tau-1} \mathbf{Q}_{12}, \quad \tau > 0 \end{aligned} \quad (8)$$

To simplify the notation, define

$$\begin{aligned} \mathbf{P} &\triangleq \mathbf{R}_{xx}(0) \\ \mathbf{G} &\triangleq \mathbf{A}\mathbf{R}_{xx}(0) \mathbf{C}^* + \mathbf{Q}_{12}. \end{aligned} \quad (9)$$

The state-covariance matrix  $\mathbf{P}$  is defined by the Lyapunov equation

$$\mathbf{P} = \mathbf{A}\mathbf{P}\mathbf{A}^T + \mathbf{Q}_{11} \quad (10)$$

and for the output covariance matrix we have at lag zero.

$$\mathbf{R}_0 \triangleq \mathbf{R}_{yy}(0) = \mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{Q}_{22}. \quad (11)$$

The eigenvalue assumption of the  $\mathbf{A}$  matrix guarantee that a solution  $\mathbf{P} > 0$  exists. In summary we have

$$\mathbf{R}_{yy}(\tau) = \begin{cases} \mathbf{R}_0 = \mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{Q}_{22}, & \tau = 0 \\ \mathbf{C}\mathbf{A}^{\tau-1} \mathbf{G}, & \tau > 0 \\ \mathbf{G}^T (\mathbf{A}^{-\tau-1})^T \mathbf{C}^T, & \tau < 0 \end{cases} \quad (12)$$

### 2.3 Frequency domain model

The truncated half spectrum for the stochastic state-space model then follows as

$$\begin{aligned} \varphi(k) &= \frac{1}{2} \mathbf{R}_0 + \sum_{\tau=1}^{N-1} \mathbf{R}_{yy}(\tau) W_N^{-k\tau} \\ &= \frac{1}{2} \mathbf{R}_0 + \sum_{\tau=1}^{N-1} \mathbf{C}\mathbf{A}^{\tau-1} \mathbf{G} W_N^{-k\tau} \\ &= \frac{1}{2} \mathbf{R}_0 + \mathbf{C} (W_N^k \mathbf{I} - \mathbf{A})^{-1} (\mathbf{I} - W_N^k \mathbf{A}^{N-1}) \mathbf{G} \end{aligned} \quad (13)$$

where the last equality is due to the properties of a geometric series. Recall that  $k$  is the frequency index (DFT bin). Finally, by introducing an auto-correlation frequency domain state-variable matrix  $\mathbf{Z}_k \in \mathbb{C}^{n \times p}$

$$\mathbf{Z}_k \triangleq (W_N^k \mathbf{I} - \mathbf{A})^{-1} (\mathbf{I} - W_N^k \mathbf{A}^{N-1}) \mathbf{G} \quad (14)$$

we obtain the alternative recursive description of the truncated half spectrum.

$$\begin{aligned} \mathbf{Z}_k W_N^k &= \mathbf{A}\mathbf{Z}_k + (\mathbf{I} - W_N^k \mathbf{A}^{N-1}) \mathbf{G} \\ \varphi(k) &= \frac{1}{2} \mathbf{R}_0 + \mathbf{C}\mathbf{Z}_k, \quad k = 0, \dots, N-1 \end{aligned} \quad (15)$$

The factor  $(\mathbf{I} - W_N^k \mathbf{A}^{N-1})$  in (13) and (14) explicitly accounts for the inherent leakage effect which always occur for finite data.

The particular model description (15) also appears in the F-Esprit algorithm McKelvey and Viberg (2001); Gunnarson and McKelvey (2004) and in frequency domain subspace based algorithms McKelvey et al. (1996). Associated with the model in (15) are the extended observability matrix with  $s$  block rows

$$\mathcal{O}_s \triangleq \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^{s-1} \end{bmatrix} \quad (16)$$

and the lower triangular block matrix

$$\Gamma_s \triangleq \begin{bmatrix} \frac{\mathbf{R}_0}{2} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{C}\mathbf{G} & \frac{\mathbf{R}_0}{2} - \mathbf{C}\mathbf{A}^{N-1} \mathbf{G} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{C}\mathbf{A}\mathbf{G} & \mathbf{C}\mathbf{G} - \mathbf{C}\mathbf{A}^N \mathbf{G} & \frac{\mathbf{R}_0}{2} - \mathbf{C}\mathbf{A}^{N-1} \mathbf{G} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \quad (17)$$

We note that since the state-space model is minimal  $(\mathbf{A}, \mathbf{C})$  is an observable matrix pair and the extended observability matrix in (16) has full rank whenever  $s \geq n$ . The number of block rows must be selected to satisfy  $s > n$ .

Phase shifted versions of  $\varphi(k)$  (multiplications with  $W_N^k$ ) can easily be derived via (15) and by stacking them in a long vector we obtain

$$\mathbf{Y}_k = \mathcal{O}_s \mathbf{Z}_k + \Gamma_s \mathbf{U}_k \quad (18)$$

where

$$\mathbf{Y}_k \triangleq \mathbf{W}_{k,s} \otimes \varphi(k), \quad \mathbf{U}_k \triangleq \mathbf{W}_{k,s} \otimes \mathbf{I}_p \quad (19)$$

and

$$\mathbf{W}_{k,s} \triangleq \begin{bmatrix} 1 & W_N^k & W_N^{2k} & \dots & W_N^{(s-1)k} \end{bmatrix}^T \quad (20)$$

Here  $\otimes$  denotes the Kronecker matrix product, see Graham (1981).

### 3. A FREQUENCY DOMAIN STOCHASTIC REALIZATION ALGORITHM

In this section we will introduce a subspace based realization algorithm which will recover a state-space model from a finite set of auto-correlation coefficients. As stated in the introduction we are primarily concerned with the problem of only recover parts of the spectrum and hence we will focus on an algorithm which uses samples of the truncated half spectrum, i.e. a subset of  $\{\varphi(k)\}_{k=0}^{N-1}$ . In this section we assume the covariance sequence is known without errors and hence the algorithm is a realization algorithm. In the next section we will assume only signal samples and modify the algorithm to provide an approximation instead.

Assume the set  $\{\mathbf{R}_{yy}(\tau)\}_{\tau=0}^{N-1}$  is known and calculate  $\varphi(k)$  for a suitable set of frequency indices  $\{k_i\}_{i=1}^M$ . The size of the index set must satisfy  $M > s + n$ . For each frequency index the vector  $\mathbf{Y}_k$  from (19) is constructed and assembled into a matrix as

$$\mathbf{Y} \triangleq [\mathbf{Y}_{k_1} \quad \mathbf{Y}_{k_2} \quad \dots \quad \mathbf{Y}_{k_M}] \quad (21)$$

From (18) it directly then follows

$$\mathbf{Y} = \mathcal{O}_s \mathbf{Z} + \Gamma_s \mathbf{U} \quad (22)$$

where

$$\begin{aligned} \mathbf{Z} &\triangleq [\mathbf{Z}_{k_1} \quad \mathbf{Z}_{k_2} \quad \dots \quad \mathbf{Z}_{k_M}] \\ \mathbf{U} &\triangleq [\mathbf{U}_{k_1} \quad \mathbf{U}_{k_2} \quad \dots \quad \mathbf{U}_{k_M}] \end{aligned} \quad (23)$$

From (22) we see that the known matrix  $\mathbf{Y}$  is the sum of two matrices where the product  $\mathcal{O}_s \mathbf{Z}$  by construction has rank of  $n$  or less. However, following the technique in Gunnarson and McKelvey (2004) it can be shown it has rank  $n$ . The second matrix product  $\Gamma_s \mathbf{U}$  has higher rank but can be removed by using a projection matrix which projects onto the nullspace of  $\mathbf{U}$ . Algebraically this projection can be formulated as

$$\mathbf{\Pi}^\perp = \mathbf{I} - \mathbf{U}^*(\mathbf{U}\mathbf{U}^*)^{-1}\mathbf{U}. \quad (24)$$

Multiplying this matrix from the left then leads to

$$\mathbf{Y}\mathbf{\Pi}^\perp = \mathcal{O}_s \mathbf{Z}\mathbf{\Pi}^\perp \quad (25)$$

since  $\mathbf{U}\mathbf{\Pi}^\perp = 0$ . It has been established that the rank of the product  $\mathcal{O}_s \mathbf{Z}\mathbf{\Pi}^\perp$  remains as  $n$  and hence the range space of  $\mathbf{Y}\mathbf{\Pi}^\perp$  coincide with the range space of  $\mathcal{O}_s$ , see Gunnarson and McKelvey (2004). As the rank is  $n$  we can factor  $\mathbf{Y}\mathbf{\Pi}^\perp$  into two matrices where the left matrix factor can be selected with a size identical to  $\mathcal{O}_s$ . A numerically convenient way to obtain the desired factor is by using the singular value decomposition (SVD). As we originally assumed the model matrices in (6) are real valued we can enforce the matrix to be real by

$$[\text{Re}\{\mathbf{Y}\mathbf{\Pi}^\perp\} \quad \text{Im}\{\mathbf{Y}\mathbf{\Pi}^\perp\}] = [\hat{\mathbf{Z}}_s \quad \hat{\mathbf{Z}}_o] \begin{bmatrix} \hat{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \hat{\Sigma}_o \end{bmatrix} \begin{bmatrix} \hat{\mathbf{V}}_s^T \\ \hat{\mathbf{V}}_o^T \end{bmatrix}. \quad (26)$$

Now, as  $\mathbf{Y}\mathbf{\Pi}^\perp$  has rank  $n$  and has a range space which is similar to the real matrix  $\mathcal{O}_s$  the compound matrix of the real and imaginary parts also have rank  $n$ . Hence, the singular value decomposition has  $n$  positive singular values which are collected on the diagonal in the  $\hat{\Sigma}_s$  matrix. The remaining singular values are all zero and hence  $\hat{\Sigma}_o$  is a zero matrix. As a realization of the observability matrix we thus simply select  $\hat{\mathbf{Z}}_s$ . Due to the relation (25) there exists a non-singular matrix  $\mathbf{T}$  such that

$$\hat{\mathbf{Z}}_s = \mathcal{O}_s \mathbf{T} \quad (27)$$

which means that  $\hat{\mathbf{Z}}_s$  is the extended observability matrix of a state-space realization *similar* to the realization defined by the matrices  $\mathbf{A}$  and  $\mathbf{C}$ . From the knowledge of the observability matrix we directly obtain the  $\mathbf{C}$  in this alternative realization as the first  $p$  rows in  $\hat{\mathbf{Z}}_s$ . With MATLAB matrix notation we have

$$\bar{\mathbf{C}} = \hat{\mathbf{Z}}_s(1:p, :) \quad (28)$$

i.e.  $\bar{\mathbf{C}}$  is the first  $p$  rows of the observability matrix. The structure of the observability matrix implies that the following block row can be obtained as the present block row multiplied by the  $\mathbf{A}$  from the right. This leads to the linear relation

$$\mathbf{J}_1 \hat{\mathbf{Z}}_s \bar{\mathbf{A}} = \mathbf{J}_2 \hat{\mathbf{Z}}_s \quad (29)$$

where  $\mathbf{J}_1$  and  $\mathbf{J}_2$  are defined as

$$\mathbf{J}_1 = [\mathbf{I}_{(s-1)m} \quad \mathbf{0}_{(s-1)m \times m}] \quad (30)$$

$$\mathbf{J}_2 = [\mathbf{0}_{(s-1)m \times m} \quad \mathbf{I}_{(s-1)m}]. \quad (31)$$

The linear equations (29) can be solved using the pseudo-inverse as  $\mathbf{J}_1 \hat{\mathbf{Z}}_s$  corresponds to an extended observability matrix with  $s-1 \geq n$  block rows and consequently has full rank  $n$ . Hence, we have

$$\bar{\mathbf{A}} = (\mathbf{J}_1 \hat{\mathbf{Z}}_s)^\dagger \mathbf{J}_2 \hat{\mathbf{Z}}_s. \quad (32)$$

where  $(\cdot)^\dagger$  denotes the pseudo-inverse operator.

Now we turn to the remaining parameter  $\mathbf{G}$ . Recall the relation (13). Given that  $\bar{\mathbf{A}}$  and  $\bar{\mathbf{C}}$  have been determined the  $\mathbf{G}$  matrix appears linearly in the equation. Hence, we can form the linear equation

$$\begin{bmatrix} \bar{\mathbf{C}}(\mathbf{I}W_N^{k_1} - \bar{\mathbf{A}})^{-1}(\mathbf{I} - W_N^{k_1} \bar{\mathbf{A}}^{N-1}) & \mathbf{I} \\ \bar{\mathbf{C}}(\mathbf{I}W_N^{k_2} - \bar{\mathbf{A}})^{-1}(\mathbf{I} - W_N^{k_2} \bar{\mathbf{A}}^{N-1}) & \mathbf{I} \\ \vdots & \vdots \\ \bar{\mathbf{C}}(\mathbf{I}W_N^{k_M} - \bar{\mathbf{A}})^{-1}(\mathbf{I} - W_N^{k_M} \bar{\mathbf{A}}^{N-1}) & \mathbf{I} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{G}} \\ \bar{\mathbf{R}}_0 \\ 2 \end{bmatrix} = \begin{bmatrix} \varphi(k_1) \\ \varphi(k_2) \\ \vdots \\ \varphi(k_M) \end{bmatrix} \quad (33)$$

and solve for the unknown  $\bar{\mathbf{G}}$  and  $\bar{\mathbf{R}}_0$  using the pseudo-inverse. Note that  $\bar{\mathbf{R}}_0$  is independent of the state-space basis. Hence, we obtain in the noise free case  $\bar{\mathbf{R}}_0 = \mathbf{R}_0$ .

To this end we have demonstrated how the system matrices  $\bar{\mathbf{C}}$ ,  $\bar{\mathbf{A}}$  and the related matrix  $\bar{\mathbf{G}}$  can be explicitly calculated. Together with  $\bar{\mathbf{R}}_0$  these matrices represents the half spectrum, (see (4))

$$\phi(\omega) = \frac{1}{2} \bar{\mathbf{R}}_0 + \bar{\mathbf{C}}(\mathbf{I}e^{j\omega} - \bar{\mathbf{A}})^{-1} \bar{\mathbf{G}} \quad (34)$$

The parametric model of the half spectrum in (34) is also known as a spectral summand model of the spectrum. However, for most filtering applications a more useful

representation of a stochastic system is by the spectral factor model in (6). By using samples of the spectrum derived from the spectral summand model we can retrieve a covariance matrix  $\mathbf{Q}$  for the spectral factor model. The spectrum of the spectral factor model is given by

$$\bar{\Phi}(\omega, \bar{\mathbf{Q}}) = \left[ \bar{\mathbf{C}}(\mathbf{I}e^{j\omega} - \bar{\mathbf{A}})^{-1} \mathbf{I} \right] \underbrace{\begin{bmatrix} \bar{\mathbf{Q}}_{11} & \bar{\mathbf{Q}}_{12} \\ \bar{\mathbf{Q}}_{12}^T & \bar{\mathbf{Q}}_{22} \end{bmatrix}}_{\bar{\mathbf{Q}}} \left[ \begin{matrix} (\bar{\mathbf{C}}(\mathbf{I}e^{j\omega} - \bar{\mathbf{A}})^{-1})^* \\ \mathbf{I} \end{matrix} \right] \quad (35)$$

where we explicitly have marked the dependence of the covariance matrix  $\bar{\mathbf{Q}}$ . Note that the covariance matrix  $\bar{\mathbf{Q}}$  appears linear in the spectrum. The spectral factor model is now easily calculated by forming a set of linear equations where each equation equates the spectrum from the spectral summand model with the spectrum from the factor model. We also need to add a constraint to guarantee that  $\mathbf{Q} \geq 0$ , i.e. it is a proper covariance matrix. It is also worth pointing out that the set defined by

$$\{\bar{\mathbf{Q}} \mid \bar{\mathbf{Q}} \geq 0 \text{ and } \Phi(\omega) = \bar{\Phi}(\omega, \bar{\mathbf{Q}})\} \quad (36)$$

is an infinite set. Hence, the spectral factor model is over-parametrized in terms of the elements of the matrix  $\bar{\mathbf{Q}}$  and infinitely many solutions exists. However, any of them represents the same spectrum which is of course the entity we have set out to model from the beginning. Finding a point in the set (36) can be achieved by semidefinite programming, see e.g. Vandenberghe and Boyd (1996).

#### 4. ESTIMATION FROM TIME DOMAIN DATA

In a practical application we are normally faced with samples of the stochastic process and not samples of the autocorrelation function. To make the procedure outlined above useful we have to provide estimates of the autocorrelation function and derive a state-space *approximation* of the spectrum. Since the first step will only provide an estimate of the autocorrelation function the second step has to take into account that the estimated autocorrelation function has errors.

From samples of the stochastic process,  $\{y(t)\}_{t=0}^{N_t}$  we form the unbiased sample estimate of the correlation function as

$$\hat{\mathbf{R}}_{yy}(\tau) \triangleq \sum_{t=\tau}^{N_t-\tau-1} \frac{y(t)y(t-\tau)^T}{N_t-\tau}, \quad \tau = 0, \dots, N_t - 1 \quad (37)$$

Then samples of the truncated half spectrum is calculated as, c.f. (13),

$$\hat{\varphi}(k) \triangleq \frac{1}{2} \hat{\mathbf{R}}_0 + \sum_{\tau=1}^{N-1} \hat{\mathbf{R}}_{yy}(\tau) W_N^{-k\tau} \quad (38)$$

Recall that the indices  $k$  corresponds to the DFT frequency bins and bin  $k$  corresponds to frequency  $2\pi k/N$  [rad/sample]. The index set  $\{k_i\}_{i=1}^M$  is then selected to correspond to the sub-band(s) where the spectrum is to be well approximated by the model. The choice of  $N$  is a trade-off between the variance of the last lag and  $M$ , the size of the index set used in the subspace steps.

The next step is to construct the matrix  $\hat{\mathbf{Y}}$  according to (19)-(21) and use the projection matrix as in (25). In the

noise free case  $\mathbf{Y}\mathbf{\Pi}^\perp$  has rank  $n$  but in the case with estimates from data it has normally full rank. The SVD step in (26) will then provide a rank  $n$  estimate of the range space of the observability matrix. Indeed  $\hat{\mathbf{Z}}_s$  in (26) is a solution to the following problem Golub and van Loan (1980)

$$\hat{\mathbf{Z}}_s, \hat{\mathbf{T}} = \arg \min_{\mathbf{Z} \in \mathbb{R}^{sp \times n}, \mathbf{T}} \|\mathbf{Z}\mathbf{T} - \hat{\mathbf{Y}}\mathbf{\Pi}^\perp\|_F, \quad \text{s.t. rank } \mathbf{Z} = n \quad (39)$$

Again, due to the noise, the linear relation (29) does not hold so using (32) provides the solution to

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \|\mathbf{J}_1 \hat{\mathbf{Z}}_s \mathbf{A} - \mathbf{J}_2 \hat{\mathbf{Z}}_s\|_F. \quad (40)$$

The estimate of the  $\mathbf{C}$  matrix,  $\hat{\mathbf{C}}$  is taken according to (28). To proceed we need to ensure that  $\hat{\mathbf{A}}$  is stable. This can be checked for example by an eigenvalue decomposition. If any eigenvalues  $\lambda$  are outside the unit circle we modify them to  $1/\lambda$  as this does not change the spectrum. If any eigenvalues are on the unit circle, a small perturbation is employed to move them inside the unit disc.

The aim of the final step is to provide estimates of the  $\mathbf{G}$  matrix and the covariance matrix  $\mathbf{R}_0$ . A straightforward technique would be to solve (33) in a least-squares sense. However, the least-squares solution might lead to a spectral summand model which does not satisfy the constraint that for all  $\omega$ ,  $\hat{\varphi}(\omega) + \hat{\varphi}(\omega)^* \geq 0$ , i.e. that the spectrum is non-negative or equivalently that  $\hat{\varphi}(\omega)$  is non-negative real. To enforce this inequality constraint to hold for all  $\omega$  is, in a numerical method, of course intractable. However, by using the Positive Real Lemma, the non-negative real condition can be recast to a finite dimensional matrix inequality Andersson (1967). See Appendix A for a statement of the lemma.

Define the quadratic loss function with  $\mathbf{G}$  and  $\mathbf{R}_0$  as free parameters as

$$J(\mathbf{G}, \mathbf{R}_0) = \sum_{i=1}^M \left\| \begin{bmatrix} \text{Re } \mathbf{S}(k_i) \\ \text{Im } \mathbf{S}(k_i) \end{bmatrix} \begin{bmatrix} \mathbf{G} \\ \mathbf{R}_0 \end{bmatrix} - \begin{bmatrix} \text{Re } \varphi(k_i) \\ \text{Im } \varphi(k_i) \end{bmatrix} \right\|_F^2 \quad (41)$$

where

$$\mathbf{S}(k) = [\hat{\mathbf{C}}(\mathbf{I}W_N^k - \hat{\mathbf{A}})^{-1}(\mathbf{I} - W_N^k \hat{\mathbf{A}}^{N-1}) \mathbf{I} / 2] \quad (42)$$

The estimate is then defined as the solution to the following constrained optimization problem

$$\begin{aligned} \hat{\mathbf{G}}, \hat{\mathbf{R}}_0, \hat{\mathbf{P}} &= \arg \min_{\mathbf{G}, \mathbf{R}_0, \mathbf{P}} J(\mathbf{G}, \mathbf{R}_0, \mathbf{P}) \\ \text{subject to } &\begin{cases} \mathbf{P} - \mathbf{A}^T \mathbf{P} \mathbf{A} & \mathbf{G} - \mathbf{A} \mathbf{P} \mathbf{C}^T \\ \mathbf{G}^T - \mathbf{C} \mathbf{P} \mathbf{A}^T & \mathbf{R}_0 - \mathbf{C} \mathbf{P} \mathbf{C}^T \end{cases} \geq 0, \\ &\mathbf{P} = \mathbf{P}^T \text{ and } \mathbf{R}_0 = \mathbf{R}_0^T \end{aligned} \quad (43)$$

The constraint in (43) in the form of the linear matrix inequality will by Lemma 2 guarantee that the half spectrum is non-negative real. This constrained optimization problem is a version of a second order cone program with linear matrix equality constraints and can be solved efficiently with interior point methods, see e.g. Vandenberghe and Boyd (1996); Boyd and Vandenberghe (2004). The software package SeDuMi by Sturm (1999) support this class of problems. The final step to produce a spectral factor model is identical as in the realization case outlined in the previous section and involves no approximations.

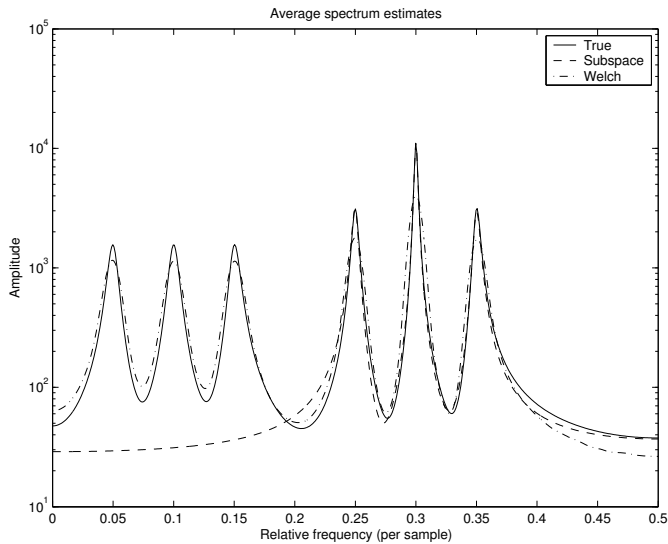


Fig. 1. Average power spectra based on 100 Monte Carlo simulations

### 5. NUMERICAL EXAMPLE

This section shows a comparison of the derived subspace based estimation algorithm with Welch's estimation method which is a non-parametric estimator.

A stochastic processes of model order 12 is considered with a spectrum shown in Figure 1 as a solid line. The frequency region of interest ( $0.2 \geq f \geq 0.4$ ) where  $f = 1$  is the sampling frequency. The number of time domain data samples used for the spectrum estimation was  $N_t = 10000$ .

Welch's averaged, modified periodogram method with 50% overlap and window size of 128 was used to calculate the spectrum in 1024 frequency points (i.e. a 1024 points DFT).

The presented subspace algorithm first produced sample estimates of the autocorrelation function for a total of  $N = 128$  non-negative lags, see (37). Then, using the DFT, samples of the truncated positive spectrum was derived. Within the frequency region of interest a total number of  $M = 26$  samples were retained in the estimation and the size of the phase shifted vector (19)-(20) was selected as  $s = 20$ .

A Monte Carlo simulation with  $N_{MC} = 100$  trials was performed. The average spectrum estimate based on the 100 Monte Carlo runs are shown Figure 1. The dashed line represent the parametric method based on the presented subspace algorithm while the dash dotted line represents Welch's method. The figure show that both methods resolve the peaks but the parametric method has less average error. Furthermore the average mean square spectrum error was calculated as

$$e_{est}^2 = \frac{2\pi/N}{N_{MC}} \sum_{t=1}^{N_{MC}} \sum_{i=1}^M (\Phi(2\pi k_i/N) - \hat{\Phi}_{est}(2\pi k_i/N))^2. \quad (44)$$

The results of the Monte Carlo simulations for both estimation methods are shown in Table 1. For this example the parametric method has the lowest average mean square error in the selected sub-band.

	$e_{Welch}^2$	$e_{F-Esprit}^2$	$\frac{e_{Welch}^2}{e_{F-Esprit}^2}$
MSE error	$7.5 \times 10^5$	$1.0 \times 10^5$	7.5

Table 1. Sample based estimation error based on 100 Monte Carlo simulations.

### 6. CONCLUSIONS

This paper has shown a parametric method for the estimation of stochastic systems or equivalently the spectrum of a stochastic process. The key feature is that the method can be used to frequency selectively fit the model to the data. This means that parts of the spectrum can be modeled with a lower model order than otherwise would be necessary if the entire spectrum would be modeled. The method is based on a frequency domain subspace method that explicitly take into account that the frequency domain data is derived from finite data and hence suppress the leakage effects. Furthermore the method employs convex optimization to guarantee that the estimated parametric model represents a non-negative spectrum.

### ACKNOWLEDGEMENTS

Partially supported by the Swedish Research Council (VR)

### REFERENCES

- B.D.O. Anderson. A system theory criterion for positive real matrices. *SIAM J. Control*, 5:171-182, 1967.
- S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- C.I. Byrnes, S.G. Gusev, and A. Lindquist. From finite covariance windows to modeling filters: A convex optimization approach. *SIAM Review*, 43(4):645-675, 2001.
- G. H. Golub and C. F. van Loan. An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, 17(6):883-893, 1980.
- A. Graham. *Kronecker Products and Matrix Calculus With Applications*. Ellis Horwood Limited, Chichester, England, 1981.
- J. Gunnarson and T. McKelvey. Consistency analysis of a frequency domain subspace algorithm for multi-component harmonic retrieval. In *Proc. of the International Conference on Acoustics, Speech and Signal Processing*, Montreal, Canada, 2004.
- J.B. Hoagg, S.L. Lacy, Erwin R.S., and D.S. Bernstein. First-order-hold sampling of positive real systems and subspace identification of positive real models. In *Proc. 2004 American Control Conference*, pages 861-866, June 2004.
- J. Mari, P. Stoica, and T. McKelvey. Vector ARMA estimation: A reliable subspace approach. *IEEE Trans. on Signal Processing*, 48(7):2092-2104, July 2000.
- L Marple. *Digital Spectral Analysis with Applications*. Prentice Hall, Englewood Cliffs, NJ, 1987.
- T. McKelvey and S. O. R. Moheimani. Estimation of phase constrained mimo transfer functions with application to flexible structures with mixed collocated and non-collocated actuators and sensors. In *Proc. 16th IFAC World Congress*, pages Mo-MO2-TO/2, Prag, Czech Republic, July 2005.
- T. McKelvey and M. Viberg. A robust frequency domain subspace algorithm for multi-component harmonic retrieval. In *Proc. of 35th Asilomar Conference on Signals*,

- Systems and Computers*, volume 2, pages 1288–1292, 2001.
- T. McKelvey, H. Akçay, and L. Ljung. Subspace-based multivariable system identification from frequency response data. *IEEE Trans. on Automatic Control*, 41(7): 960–979, July 1996.
- A. Rantzer. On the Kalman-Yakubovich-Popov lemma. *Systems Control Lett.*, 28(1):7–10, 1996. ISSN 0167-6911.
- P. Stoica and R. Moses. *Introduction to Spectral Analysis*. Prentice-Hall, Englewood Cliffs, USA, 1997.
- P. Stoica, T. McKelvey, and J. Mari. MA estimation in polynomial time. *IEEE Trans. on Signal Processing*, 48(7):1999–2012, July 2000.
- J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11-12:625–653, 1999. Special issue on Interior Point Methods.
- P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems - Theory Implementation Application*. Kluwer, Dordrecht, The Netherlands, 1996.
- P. Van Overschee, B. De Moor, W. Dehandschutter, and J. Swevers. A subspace algorithm for the identification of discrete time frequency domain power spectra. *Automatica*, 33(12):2147–2157, Dec 1997.
- L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, 38(1):49–95, 1996. ISSN 0036-1445.

#### Appendix A. POSITIVE REAL LEMMA

*Lemma 1.* (PR). Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{B}, \mathbf{C}^T \in \mathbb{R}^{n \times m}$ ,  $\mathbf{D} \in \mathbb{R}^{m \times m}$ , with  $\det(e^{j\omega} \mathbf{I} - \mathbf{A}) \neq 0$  for  $\omega \in \mathbb{R}$ ,  $(\mathbf{A}, \mathbf{B})$  controllable and  $\Psi(\omega) = \mathbf{D} + \mathbf{C}(e^{j\omega} \mathbf{I} - \mathbf{A})^{-1} \mathbf{B}$ , the following two statements are equivalent:

i For  $\omega \in \mathbb{R}$

$$\Psi^*(\omega) + \Psi(\omega) \geq 0 \quad (\text{A.1})$$

ii There exists  $P = P^T \in \mathbb{R}^{n \times n}$  such that

$$\begin{bmatrix} \mathbf{P} - \mathbf{A}^T \mathbf{P} \mathbf{A} & \mathbf{C}^T - \mathbf{A}^T \mathbf{P} \mathbf{B} \\ \mathbf{C} - \mathbf{B}^T \mathbf{P} \mathbf{A} & \mathbf{D} + \mathbf{D}^T - \mathbf{B}^T \mathbf{P} \mathbf{B} \end{bmatrix} \geq 0 \quad (\text{A.2})$$

**Proof.** Based on the Kalman- Yakobovich - Popov lemma see Rantzer (1996) the Positive Real Lemma follows immediately.

By a hermitian transpose of the transfer function it is easy to see that the dual result exists.

*Lemma 2.* (PR dual). Given  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{B}, \mathbf{C}^T \in \mathbb{R}^{n \times m}$ ,  $\mathbf{D} \in \mathbb{R}^{m \times m}$ , with  $\det(e^{j\omega} \mathbf{I} - \mathbf{A}) \neq 0$  for  $\omega \in \mathbb{R}$ ,  $(\mathbf{A}, \mathbf{C})$  observable and  $\Psi(\omega) = \mathbf{D} + \mathbf{C}(e^{j\omega} \mathbf{I} - \mathbf{A})^{-1} \mathbf{B}$ , the following two statements are equivalent:

i For  $\omega \in \mathbb{R}$

$$\Psi^*(\omega) + \Psi(\omega) \geq 0 \quad (\text{A.3})$$

ii There exists  $P = P^T \in \mathbb{R}^{n \times n}$  such that

$$\begin{bmatrix} \mathbf{P} - \mathbf{A}^T \mathbf{P} \mathbf{A} & \mathbf{B} - \mathbf{A} \mathbf{P} \mathbf{C}^T \\ \mathbf{B}^T - \mathbf{C} \mathbf{P} \mathbf{A}^T & \mathbf{D} + \mathbf{D}^T - \mathbf{C} \mathbf{P} \mathbf{C}^T \end{bmatrix} \geq 0 \quad (\text{A.4})$$