A Robust Interpolation Algorithm for Spectral Analysis

Kaushik Mahata and Minyue Fu, *Fellow, IEEE*

Abstract—We propose a *robust* interpolation algorithm for model-based spectral analysis. Instead of estimating the spectral model directly, the so-called *half spectrum*, which has a one-to-one relationship with the spectrum through standard spectral decomposition, is estimated using an interpolation approach. The interpolation data consists of the values and the derivatives of the half spectrum function at a set of user-specified points, and can be easily estimated using an input-to-state filter. Our algorithm allows a large number of noisy interpolation data to be used to optimally fit a half spectrum function of a fixed order. The capability of handling large number of interpolation data makes our algorithm *robust* to the inherent finite sample noise in the interpolation data. The algorithm involves solving some least-squares problems and semidefinite programming problems, and is thus numerically efficient. Numerical tests show that our algorithm gives very reliable spectral estimates.

Index Terms—Autoregressive moving average (ARMA) modeling, input-to-state filtering, Nevanlinna–Pick interpolation, spectral analysis.

I. INTRODUCTION

THIS paper is concerned with the following standard spectral estimation problem: Given N samples of a real-valued signal $u(t), t = 0, 1, 2, \ldots, N$, find an nth-order discrete-time autoregressive moving average (ARMA) model with transfer function $g(z)$ such that the spectrum of $u(t)$ is *best* approximated by $|g(e^{i\omega})|^2$ in some measure. To make this problem more tractable and more meaningful technically, $u(t)$ is often assumed to be generated by an ARMA process but possibly corrupted by noises, and N is not necessarily large.

Estimation of ARMA models is a classical problem in time-series analysis and signal processing. The maximum-likelihood method (MLM) [1] is the most accurate among the available algorithms. However, MLM requires to solve a nonlinear and nonconvex optimization problem, which is not tractable for models of large order. An alternative is the instrumental variable method (IVM); see [2] and references therein. IVM is computationally efficient and yields accurate estimates of the autoregressive (AR) part of the ARMA model. However, for the moving average (MA) part, the estimation performance is often poor. In addition, the estimated model is not guaranteed to be stable. The approximate maximum-likelihood (AML) approach [3] resolves the accuracy issue, but the estimated spectrum may fail to be positive. The AML approach is extended in [4] to address the positivity issue, where a semidefinite programming approach is presented. However, the semidefinite programming problem in [4] may not admit a solution.

In this paper, we consider an interpolation approach to spectral estimation initially proposed in [5]. The basic idea is as follows. Instead of estimating a model of $g(z)$ directly, we compute a model of the so-called *half spectrum* $f(z)$ defined as

$$f(z) = r_0/2 + \sum_{k=1}^{\infty} r_k z^{-k} \qquad (1)$$

where $r_k = \mathsf{E}\{u(t+k)u(t)\}$. Here, $\mathsf{E}\{\cdot\}$ denotes the mathematical expectation operator. By definition of the spectrum, it follows that

$$\sum_{k=-\infty}^{\infty} r_k z^{-k} = f(z) + f(z^{-1}) = g(z)g(z^{-1}).$$

It is well known that $f(z)$ is an nth-order strictly positive-real (SPR) rational function (i.e., $f(z)$ is real-valued and stable and $f(z) + f(z^{-1}) > 0$ for all $|z| = 1$). We first estimate the values of $f(z)$ and possibly its derivatives at a set of user-specified locations $z = \xi_i$, with $i = 0, 1, \ldots, m$. This can be done using an *input-to-state filter* [6], [7]. These values are then used to estimate the parameters of $f(z)$ by solving an interpolation problem. The advantage of this framework is the tunability [5]. By choosing the interpolation points appropriately, it is possible to tune the estimation algorithm to achieve superior resolution performance in a specified frequency interval.

The aim of this paper is to propose a robust interpolation algorithm for spectral estimation. This is motivated by the following observation. The interpolation approach [5], [8] computes an order n rational function by interpolating the half spectrum estimates at n different user chosen points located outside the unit circle in the complex plane. Clearly, if the solution exists, it is nonunique, and the set of all solutions are parameterized by the spectral zeros [5]. In order to use this approach in practical problems, it is required to estimate the spectral zeros. In addition, the interpolation approach in [5] and [8] does not account for noise in the interpolation data. The half spectrum values are estimated from a finite number of samples of $u(t)$, thus are typically noisy. One way of coping with this problem would be to use a large number of interpolation data, where the number of interpolation conditions exceeds the number of free parameters in the interpolating function, and subsequently *fit* a positive-real rational function to the interpolation data by minimizing an appropriate objective function. Strictly speaking, this is no longer an interpolation problem. However, this is a standard approach to incorporate some degree of robustness to the inherent noise in the interpolation data. We will call this approach *robust interpolation*.

In the proposed robust interpolation algorithm, we allow a large set of noisy interpolation data to improve the estimation of the half spectrum. Our algorithm also allows interpolation data to include higher order derivatives of the half spectrum. This feature is important for estimating the model order, although this topic is not studied in this paper. Apart from being statistically accurate, the robust interpolation algorithm is numerically efficient and it always guarantees a positive-definite spectrum estimate.

II. ESTIMATION OF HALF SPECTRUM

The first step in our robust interpolation algorithm is to estimate the values of the half spectrum $f(z)$ and possibly its derivatives at a set of user-specified points $\xi_i, i = 0, 1, 2, \ldots, m$. The points ξ_i are chosen to be *self-conjugate* with $|\xi_i| > 1$. Here, we propose a variation of the approach outlined in [6] and [7]. The proposed approach is relatively simpler to understand and implement.

Let us define

$$
\mathbf{A} = \begin{bmatrix} -a_1 & \cdots & -a_{\nu-1} & -a_\nu \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (2)
$$

where a_i are obtained from

$$
a(z) := z^\nu + a_1 z^{\nu-1} + \cdots + a_\nu := \prod_{i=0}^m \left\{ z - \xi_i^{-1} \right\}^{\nu_i}. \quad (3)
$$

In the above, ν_i is the multiplicity of ξ_i^{-1} and $\nu = \sum_{i=0}^m \nu_i$. For convenience, we assume that $\xi_0 = \infty$ with $\nu_0 = 1$. Let q be the unit advance operator, i.e., $qx(t) = x(t+1)$. Consider

$$
x(t) = \left[q(qI - \mathbf{A})^{-1} - \frac{1}{2} I \right] \mathbf{b} u(t)
$$
$$
= \frac{1}{2} \mathbf{b} u(t) + \sum_{k=1}^\infty \mathbf{A}^k \mathbf{b} u(t-k). \quad (4)
$$

Then, it is straightforward that

$$
\mathbf{r} := \mathsf{E}\{x(t)u(t)\} = \frac{1}{2} r_0 \mathbf{b} + \sum_{k=1}^\infty r_k \mathbf{A}^k \mathbf{b}. \quad (5)
$$

We choose $\mathbf{A}$ and $\mathbf{b}$ in a controllable canonical form to avoid complex numbers in the filtering operation and the covariance computation, which are the most computationally intensive steps in our algorithm.

Define $f_k(z) := (1/k!)\frac{\mathrm{d}^k f(z)}{\mathrm{d}z^k}$, and

$$
\mathbf{f}_i(z) := [f_{\nu_i - 1}(z) \quad \cdots \quad f_1(z) \quad f(z)]^\top. \quad (6)
$$

It is shown in Appendix A that

$$
\mathbf{f}_i(\xi_i) = \mathbf{T}_i \Gamma^{-1} \mathbf{r}, \quad i = 0, 1, \ldots, m \quad (7)
$$

where $\Gamma \in \mathbb{C}^{\nu \times \nu}$ is such that the pair $(\Gamma^{-1}\mathbf{A}\Gamma, \Gamma^{-1}\mathbf{b})$ is in the Jordan canonical form, see (37). The expressions for the matrices $\mathbf{T}_i, i = 0, 1, \ldots, m$ are also derived in Appendix A.

Using (7) we can estimate $\mathbf{f}_i(\xi_i), i = 0, 1, \ldots, m$, from an estimate of $\mathbf{r}$. However, if $|\xi_i|$ is very close to unity for some i, then the corresponding pole ξ_i^{-1} of the filter in [4] is located very close to the unit circle causing a very long impulse response. Hence, the transient error in $x(t)$ due the unknown initial conditions decay slowly. This may induce a significant error in the estimate of $\mathbf{r}$. In order to keep the error within an acceptable limit it is recommended to keep $|\xi_i| > 1.25$ when $N < 1000$. One can, however, allow ξ_i close to unity when N is large enough.

Once the estimates $\hat{\mathbf{f}}_i$ of $\mathbf{f}_i(\xi_i), i = 0, 1, \ldots, m$ are made available, we can address the robust Nevanlinna–Pick interpolation problem [9], [10] below: *Find an SPR rational function $f(z)$ of degree n, such that*

$$
\mathbf{f}_i(\xi_i) = \hat{\mathbf{f}}_i \quad (8)
$$

for all $i = 0, 1, \ldots, m$ or that $\mathbf{f}_i(\xi_i)$ are as close to $\hat{\mathbf{f}}_i$ as possible in some measure.

We require $\nu \geq 2n + 1$ in the sequel in order to produce a reliable spectral estimate. The technical difficulty with this setting is that, on one hand, we tend to have an overdetermined problem because $f(z)$ has only $2n + 1$ free parameters, and on the other hand, we may not necessarily have an SPR solution that satisfies all the constraints [11].

III. INTERPOLATION CONDITIONS FOR HALF SPECTRUM

We assume in the sequel that $n = \sum_{i=1}^p \nu_i$ for some $0 < p < m$ and that the set $\{\xi_i\}_{i=1}^p$ is self conjugate. In this section, we assume that the interpolation data $\hat{\mathbf{f}}_i$ for $\mathbf{f}_i(\xi_i), i = 0, 1, \ldots, m$ are perfect. That is, there exists a nth-order rational SPR function $f(z)$ that interpolates the given data. We want to know how these interpolation conditions can be satisfied. This is done in the following three steps:
1) we first parameterize all $f(z)$ so that (8) holds for $i = 0, 1, \ldots, p$;
2) we then give the constraints on the free parameters of $f(z)$ such that (8) is satisfied for $i = p + 1, \ldots, m$;
3) the final step is to consider the SPR requirement.

Some notation is in order:

$$
\begin{aligned}
X_i &:= \begin{bmatrix} \xi_i & 1 & \cdots & 0 \\ 0 & \xi_i & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \xi_i \end{bmatrix} \in \mathbb{R}^{\nu_i \times \nu_i}; \\
\varepsilon_i &= [0 \quad \cdots \quad 0 \quad 1]^\top \in \mathbb{R}^{\nu_i \times 1}; \\
X_\star &= \mathrm{diag}\{X_1, \ldots, X_p\}; \\
\bar{e} &= [\varepsilon_1^\top \quad \cdots \quad \varepsilon_p^\top]^\top; \\
\bar{c} &= [\hat{\mathbf{f}}_1^\top \quad \cdots \quad \hat{\mathbf{f}}_p^\top]^\top; \\
T &= [\bar{e} \quad X_\star \bar{e} \quad \cdots \quad X_\star^{n-1}\bar{e}]; \\
X &= T^{-1} X_\star T; \quad c = T^{-1}\bar{c}; \\
e &= T^{-1}\bar{e}; \\
\Omega(z) &:= [X - zI]^{-1}. \quad (9)
\end{aligned}
$$

The following is a generalization of a result in [11], where we parameterize all $f(z)$ such that the interpolation conditions (8) at $\xi_0, \xi_1, \ldots, \xi_p$ are satisfied.

Lemma 1: All nth-order rational transfer functions $f(z)$ satisfying the interpolation conditions (8) for $i = 0, \ldots, p$, are parameterized in terms of a free parameter $b \in \mathbb{R}^{1 \times n}$ by

$$f(z) = \frac{\hat{\mathbf{f}}_0 - b\Omega(z)c}{1 - b\Omega(z)e} \tag{10}$$

where , and c are given by (9), and are real-valued. In particular

$$e = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^\top.$$

See Appendix B for proof.

Our next result explores the relationship between b and the remaining interpolation data at $\{\xi_i\}_{i=p+1}^m$. We define

$$C_i(z) := \begin{bmatrix} \Omega^{\nu_i}(z)c & \Omega^{\nu_i - 1}(z)c & \cdots & \Omega(z)c \end{bmatrix};$$
$$E_i(z) := \begin{bmatrix} \Omega^{\nu_i}(z)e & \Omega^{\nu_i - 1}(z)e & \cdots & \Omega(z)e \end{bmatrix}. \tag{11}$$

For any vector $\boldsymbol{v} = \begin{bmatrix} v_1 & \cdots & v_\ell \end{bmatrix}$, we denote

$$\mathcal{L}(\boldsymbol{v}) := \begin{bmatrix} v_\ell & 0 & \cdots & 0 \\ v_{\ell-1} & v_\ell & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ v_1 & \cdots & v_{\ell-1} & v_\ell \end{bmatrix}. \tag{12}$$

Lemma 2: Let

$$L_i := E_i(\xi_i)\,\mathcal{L}(\hat{\mathbf{f}}_i) - C_i(\xi_i); \tag{13}$$
$$h_i := (\hat{\mathbf{f}}_i - \varepsilon_i \hat{\mathbf{f}}_0)^\top \tag{14}$$

for $i = p+1, \ldots, m$. Then, $f(z)$ in (10) satisfies the interpolation conditions (8) for $i = p+1, \ldots, m$, if and only if

$$h_i = bL_i, \quad \forall i = p+1, \ldots, m. \tag{15}$$

See Appendix C for proof.

Since the set

$$\mathbb{X} := \{\xi_i : i = p+1, p+2, \ldots, m\}$$

is self conjugate, we do not need to consider those ξ_i with a negative imaginary part. Without loss of generality, we assume that $\mathbb{X}$ is organized such that the first p_r of the $m-p$ members in $\mathbb{X}$ are real-valued, the next p_c elements are complex-valued with a positive imaginary part, and the last p_c elements are their complex conjugates. We define

$$h_r := \begin{bmatrix} h_{p+1} & \cdots & h_{p+p_r} \end{bmatrix};$$
$$h_c := \begin{bmatrix} h_{p+p_r+1} & \cdots & h_{p+p_r+p_c} \end{bmatrix};$$
$$h := \begin{bmatrix} h_r & \mathrm{Re}\{h_c\} & \mathrm{Im}\{h_c\} \end{bmatrix}. \tag{16}$$

We also define L_r, L_c, and L in the same fashion. Using (15), the interpolation conditions for b can be written as

$$h = bL. \tag{17}$$

Finally, we consider the SPR constraint on $f(z)$.

Lemma 3: The function $f(z)$ in (10) is SPR if and only if there exists $Q = Q^\top > 0$ such that

$$\begin{bmatrix} ce^\top + ec^\top + Q & c + e\hat{\mathbf{f}}_0 & XQ \\ c^\top + \hat{\mathbf{f}}_0 e^\top & 2\hat{\mathbf{f}}_0 & \beta \\ QX^\top & \beta^\top & Q \end{bmatrix} > 0 \tag{18}$$

where $\beta = bQ$. If b is a free parameter, then there exists b such that $f(z)$ in (10) is SPR if and only if

$$ce^\top + ec^\top > XQX^\top - Q; \tag{19}$$
$$\begin{bmatrix} ce^\top + ec^\top + Q & c + e\hat{\mathbf{f}}_0 \\ c^\top + \hat{\mathbf{f}}_0 e^\top & 2\hat{\mathbf{f}}_0 \end{bmatrix} > 0 \tag{20}$$

for some $Q = Q^\top > 0$. Moreover, (19) and (20) holds for some $Q = Q^\top > 0$ if and only if the following *generalized Pick matrix*

$$P := \begin{bmatrix} XEX^\top & c + e\hat{\mathbf{f}}_0 \\ c^\top + \hat{\mathbf{f}}_0 e^\top & 2\hat{\mathbf{f}}_0 \end{bmatrix} > 0 \tag{21}$$

where $E = E^\top > 0$ is the unique solution to the Lypunov equation

$$XEX^\top - E = ce^\top + ec^\top. \tag{22}$$

See Appendix D for proof.

Remark 1: It is a classical result [9], [10] that $P > 0$ is the necessary and sufficient condition for the existence of an admissible $f(z)$ satisfying the interpolation conditions (8) for $i = 1, 2, \ldots, p$.

IV. ROBUST INTERPOLATION

When the interpolation data are noisy, the interpolation requirements given in the previous section typically fail. One may face three potential problems, as follows:

- for the parameterization of $f(z)$ in (10), there may not exist any solution for b such that $f(z)$ is SPR;
- the same parameterization of $f(z)$ may make the (17) for b unsolvable;
- even if (17) has a solution, the resulting $f(z)$ may not be SPR.

In this section, we propose a robust interpolation algorithm which essentially relaxes the interpolation conditions but guarantees a SPR solution to $f(z)$.

Our algorithm involves four steps. The first three steps aim at producing an $f(z)$ that optimally fits the given data, and this is done in a bootstrapping fashion. The fourth step addresses the SPR requirement. More specifically, the first step computes an initial estimate of $f(z)$ where the parameter c is chosen to satisfy the first $n + 1$ interpolation conditions and the parameter b is chosen to meet the remaining interpolation conditions in some least-squares sense, but the SPR requirement is not considered. This estimate of $f(z)$ is used in the second step to produce a more reliable estimate of b. This is in turn used in step three to improve the estimate of c. By now, the new estimate of $f(z)$ typically fits the given data well, but still has the potential to be not SPR. This is fixed in the final step where c and b are tuned

again by solving some semidefinite programs to meet the SPR requirement. The details are given below.

Assume that r is estimated using

$$\hat{\boldsymbol{r}} = \frac{1}{N} \sum_{t=1}^{N} x(t)u(t)$$

and this is used in (44) to produce the estimates of the interpolation data $\{\hat{\mathbf{f}}_i\}_{i=0}^{m}$. The corresponding h and L [see (16)] are denoted by $\hat{h}$ and $\hat{L}$, respectively.

The asymptotic covariance matrix of $\hat{\boldsymbol{r}}$ is defined as

$$\Psi := \lim_{N \to \infty} N\text{Cov}\{\hat{\boldsymbol{r}}\}$$

which is characterized below.

Lemma 4: Let $u(t)$ be a Gaussian random process with a power spectral density $|g(\mathrm{e}^{\mathrm{i}\omega})|^2$, and

$$\alpha(q^{-1}) := [q(qI - \mathbf{A})^{-1} + q^{-1}(q^{-1}I - \mathbf{A})^{-1} - I]\mathbf{b} \quad (23)$$

then

$$\Psi = \frac{1}{4\pi} \int_{-\pi}^{\pi} \alpha(\mathrm{e}^{-\mathrm{i}\omega})|g(\mathrm{e}^{\mathrm{i}\omega})|^4 \alpha^{\top}(\mathrm{e}^{-\mathrm{i}\omega})\,\mathrm{d}\omega. \quad (24)$$

See Appendix F for a proof.

A. Step 1: Initial Estimation of $f(z)$

The initial $f(z)$ takes the form of (10) with the estimates of $\hat{\mathbf{f}}_0$, c and b given, respectively, by $\hat{\mathbf{f}}_0$

$$\hat{c} = T^{-1}[\hat{\mathbf{f}}_1^{\top} \dots \hat{\mathbf{f}}_p^{\top}]^{\top} \quad (25)$$

$$\hat{b} = \arg \min_{\bar{b}} \|\hat{h} - \bar{b}\hat{L}\|^2. \quad (26)$$

B. Step 2: Approximate Gauss–Markov Estimation of b

In this step, the same parameterization of $f(z)$ is used, but b is estimated by solving

$$\overset{\circ}{b} = \arg \min_{\bar{b}} (\hat{h} - \bar{b}\hat{L})\Sigma^{-1}(\hat{h} - \bar{b}\hat{L})^{\top} \quad (27)$$

where $\Sigma = \Sigma^{\top} > 0$ is a weighting matrix to be chosen. We use the so-called *Gauss–Markov estimation* approach [12] to determine Σ, where the aim is to minimize the asymptotic covariance matrix of the estimation error.

Denoting

$$\hat{\epsilon} := (\hat{h} - b\hat{L})^{\top}$$

and defining the asymptotic covariance matrix of $\hat{\epsilon}$ as

$$\Psi_2 := \lim_{N \to \infty} N\text{cov}\{\hat{\epsilon}\}$$

the best linear estimation requires $\Sigma_1 = \Psi_2$ (see [12]). The difficulty, however, is that Ψ_2 depends on the knowledge of the true $f(z)$. In the following, we explain how this can be approximated using the initial estimate of $f(z)$.

We note that $\hat{\epsilon}$ is a linear function of $\hat{\boldsymbol{r}}$ for a fixed b because both h and L are linear functions of $\hat{\boldsymbol{r}}$, i.e.,

$$\hat{\epsilon} = G\hat{\boldsymbol{r}} \quad (28)$$

for some matrix G depending on b. The expression for G is given in Appendix E. The above implies that

$$\Psi_2 = G\Psi G^{\top}. \quad (29)$$

Since Ψ is not available, we use its approximation

$$\hat{\Psi} = \frac{2}{K} \sum_{j=0}^{K-1} \alpha(\mathrm{e}^{-\mathrm{i}w_j})[\text{Re}\{\hat{f}(\mathrm{e}^{\mathrm{i}w_j})\}]^2 \alpha^{\top}(\mathrm{e}^{-\mathrm{i}w_j})\,\mathrm{e}^{\mathrm{i}w_j k}$$

where $w_j = 2\pi j/K$ and K is sufficiently large to avoid any time domain aliasing ($K = 1024$) is usually sufficient). In addition, $\hat{b}$ is used in lieu of b in computing G. Naturally, the resulting G and Ψ_2 in (29) are denoted by $\hat{G}$ and $\hat{\Psi}_2$, respectively.

Setting $\Sigma_1 = \hat{\Psi}_2$, we obtain

$$\overset{\circ}{b} = \hat{h}\hat{\Psi}_2^{-1}\hat{L}^{\top} \left[\hat{L}\hat{\Psi}_2^{-1}\hat{L}^{\top}\right]^{-1}. \quad (30)$$

C. Step 3: Minimum Variance Estimation of $\hat{\mathbf{f}}_0$ and c

Using $\overset{\circ}{b}$, we obtain

$$\overset{\circ}{\epsilon} := \hat{h} - \overset{\circ}{b}\hat{L} = \hat{\Pi}\hat{G}\hat{\boldsymbol{r}}$$

where the last equality is valid for large N with

$$\hat{\Pi} = \left\{I - \hat{L}^{\top}\left[\hat{L}\hat{\Psi}_2^{-1}\hat{L}^{\top}\right]^{-1}\hat{L}\hat{\Psi}_2^{-1}\right\}.$$

Denoting $\mathbf{c} = [c^{\top}\ \hat{\mathbf{f}}_0]^{\top}$ and its initial estimate in Step 1 by $\hat{\mathbf{c}}$, it is clear that

$$\hat{\mathbf{c}} = H\hat{\boldsymbol{r}} \quad (31)$$

for some H (also to be derived in Appendix E). It follows that, when N is sufficiently large

$$N\text{cov}\{[\hat{\mathbf{c}}^{\top}\quad \overset{\circ}{\epsilon}^{\top}]\} \approx \begin{bmatrix} H \\ \hat{\Pi}\hat{G} \end{bmatrix} \hat{\Psi} \begin{bmatrix} H \\ \hat{\Pi}\hat{G} \end{bmatrix}^{\top}.$$

In particular, $\hat{\mathbf{c}}$ and $\overset{\circ}{\epsilon}$ are correlated. Therefore, we can improve the estimate of $\mathbf{c}$ by minimizing its covariance. Since $\overset{\circ}{\epsilon}$ has zero mean as $N \to \infty$, it follows that the *minimum variance estimate* of $\mathbf{c}$ is given by [13, pp. 26]

$$\overset{\circ}{\mathbf{c}} = \hat{\mathbf{c}} - \hat{\Psi}_3\hat{\Pi}^{\top}[\hat{\Pi}\hat{\Psi}_2\hat{\Pi}^{\top}]^{\dagger}\overset{\circ}{\epsilon} \quad (32)$$

where $\dagger$ denotes the Moore–Penrose pseudoinverse operator, and

$$\hat{\Psi}_3 = \hat{H}\hat{\Psi}\hat{G}^{\top}.$$

The normalized covariance matrix of $\overset{\circ}{c}$, when N is sufficiently large, can be estimated as

$$
\begin{aligned}
\hat{\Psi}_4 &:= N\mathrm{cov}\{\overset{\circ}{c}\} \\
&\approx H\hat{\Psi}H^\top - \hat{\Psi}_3\hat{\Pi}^\top[\hat{\Pi}\hat{\Psi}_2\hat{\Pi}^\top]^\dagger\hat{\Pi}\hat{\Psi}_3.
\end{aligned}
$$

D. Step 4: SPR Regularization

We denote the estimate of $f(z)$ so far by

$$
\overset{\circ}{f}(z) := \frac{\overset{\circ}{\mathbf{f}}_0 + \overset{\circ}{b}(zI - X)^{-1}\overset{\circ}{c}}{1 + \overset{\circ}{b}(zI - X)^{-1}e}.
$$

There is a possibility that $\overset{\circ}{f}(z)$ is not SPR. This may occur in either of the following cases:

1) for $\boldsymbol{c} = \overset{\circ}{\boldsymbol{c}}$, the solution set $\mathbb{B}$ of all b such that $f(z)$ is SPR may be empty;

2) even if $\mathbb{B}$ is nonempty, $\overset{\circ}{b} \notin \mathbb{B}$.

In either case, the estimate of $f(z)$ needs to be modified. This can be done using Lemma 3.

Checking whether $\mathbb{B}$ is nonempty can be easily done using (21)–(22). If it fails, $\overset{\circ}{c}$ needs to be modified. One simple modification is to use the following:

$$
\check{\boldsymbol{c}} := [\check{c}^\top \quad \check{\mathbf{f}}_0]^\top := \arg\min_{\boldsymbol{c}}(\boldsymbol{c} - \overset{\circ}{\boldsymbol{c}})^\top\hat{\Psi}_4^{-1}(\boldsymbol{c} - \overset{\circ}{\boldsymbol{c}})
$$

subject to (19) and (20). This is equivalent to solving the following semidefinite program:

$$
\check{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}}\ell
$$

$$
\text{subject to } Q > 0, \quad ce^\top + ec^\top > XQX^\top - Q
$$

$$
\begin{bmatrix} \ell & (\boldsymbol{c} - \overset{\circ}{\boldsymbol{c}})^\top \\ (\boldsymbol{c} - \overset{\circ}{\boldsymbol{c}}) & \hat{\Psi}_4 \end{bmatrix} > 0
$$

$$
\begin{bmatrix} ce^\top + ec^\top + Q & c + e\mathbf{f}_0 \\ c^\top + \mathbf{f}_0e^\top & 2\mathbf{f}_0 \end{bmatrix} > 0. \tag{33}
$$

The second step is to re-estimate b. Instead of doing it directly, we estimate β and Q (18) by solving the following optimization problem:

$$
\min_{\beta,Q}(\beta - \overset{\circ}{b}Q)(\beta - \overset{\circ}{b}Q)^\top
$$

subject to (18) with c and $\mathbf{f}_0$ replaced by $\check{c}$ and $\check{\mathbf{f}}_0$, respectively. This is equivalent to solving the following semidefinite program:

$$
[\check{\beta}, \quad \check{Q}] = \arg\min_{\beta,Q}\ell
$$

$$
\text{subject to } \begin{bmatrix} \ell & \beta - \overset{\circ}{b}Q \\ \beta^\top - Q\overset{\circ}{b}^\top & I \end{bmatrix} > 0
$$

$$
\begin{bmatrix} \check{c}e^\top + e\check{c}^\top + Q & \check{c} + e\check{\mathbf{f}}_0 & XQ \\ \check{c}^\top + \check{\mathbf{f}}_0e^\top & 2\check{\mathbf{f}}_0 & \beta \\ QX^\top & \beta^\top & Q \end{bmatrix} > 0.
$$

The resulting estimate of b is given by

$$
\check{b} = \check{\beta}\check{Q}^{-1}
$$

which gives the final SPR estimate of $f(z)$ as

$$
\check{f}(z) = \frac{\check{\mathbf{f}}_0 + \check{b}(zI - X)^{-1}\check{c}}{1 + \check{b}(zI - X)^{-1}e}.
$$

E. Asymptotic Accuracy

When N is large, the interpolation data $\hat{\mathbf{f}}_k$ is accurate enough to ensure $\overset{\circ}{f}(z)$ is SPR, and hence the SPR regularization step is not needed. Therefore, the asymptotic covariance matrix of $[\overset{\circ}{c}^\top \quad \overset{\circ}{b}]$ serves as a measure of the asymptotic accuracy of the above estimation algorithm. In Appendix G it is shown that

$$
\lim_{N\to\infty} N\mathrm{cov}\{[\overset{\circ}{c}^\top \quad \overset{\circ}{b}]\}
$$

$$
= \left\{\begin{bmatrix} I & 0 \\ 0 & L \end{bmatrix}\left(\begin{bmatrix} H \\ G \end{bmatrix}\Psi\begin{bmatrix} H \\ G \end{bmatrix}^\top\right)^{-1}\begin{bmatrix} I & 0 \\ 0 & L \end{bmatrix}^\top\right\}^{-1}. \tag{34}
$$

Let

$$
\Phi_\nu = \lim_{N\to\infty} N\mathrm{cov}\{[\overset{\circ}{c}^\top \quad \overset{\circ}{b}]\}
$$

with ν interpolation conditions. Now, suppose we add an additional interpolation condition and the resulting normalized covariance matrix be $\Phi_{\nu+1}$. Then, (34) can be used to show that $\Phi_\nu - \Phi_{\nu+1}$ is a positive-definite matrix. See [2, Lemma 2] for a similar proof.

V. ILLUSTRATIVE EXAMPLE

To show how the proposed algorithm works, we test it on an ARMA process $u(t)$ generated by filtering a Gaussian white noise through

$$
g(z) = \frac{z^4 - 0.4789z^3 + 0.3790z^2 - 0.3879z + 0.6561}{z^4 - 0.5414z^3 + 0.9077z^2 - 0.4385z + 0.6561}. \tag{35}
$$

It can be verified that the poles and zeros of $g(z)$ are very close to each other. In addition, the zeros are located very close to the unit circle. This makes it a difficult problem to identify $g(z)$ accurately. In addition, the algorithms where positivity of the estimated spectrum is not guaranteed may often give a nonpositive spectrum estimate or an unstable estimate of $g(z)$.

We take $m = 20$, $\xi_0 = \infty$, $\xi_1 = 1.5$, $\xi_2 = -1.5$, and

$$
\left.\begin{aligned} \xi_{2k-1} &= 1.5e^{i(k-1)\pi/10} \\ \xi_{2k} &= 1.5e^{-i(k-1)\pi/10} \end{aligned}\right\}, \quad k = 2,\ldots,10
$$

with $\nu_i = 1$ for all i. The performance of the proposed algorithm is compared with the maximum-likelihood approach. Simulation results based on 100 independent runs are shown in Fig. 1, where we plot the true spectrum (solid) with the mean (dashed) $\pm$ the standard deviation (dashed–dotted) of the estimated spectrum for each approach. In each run, 1000 points of
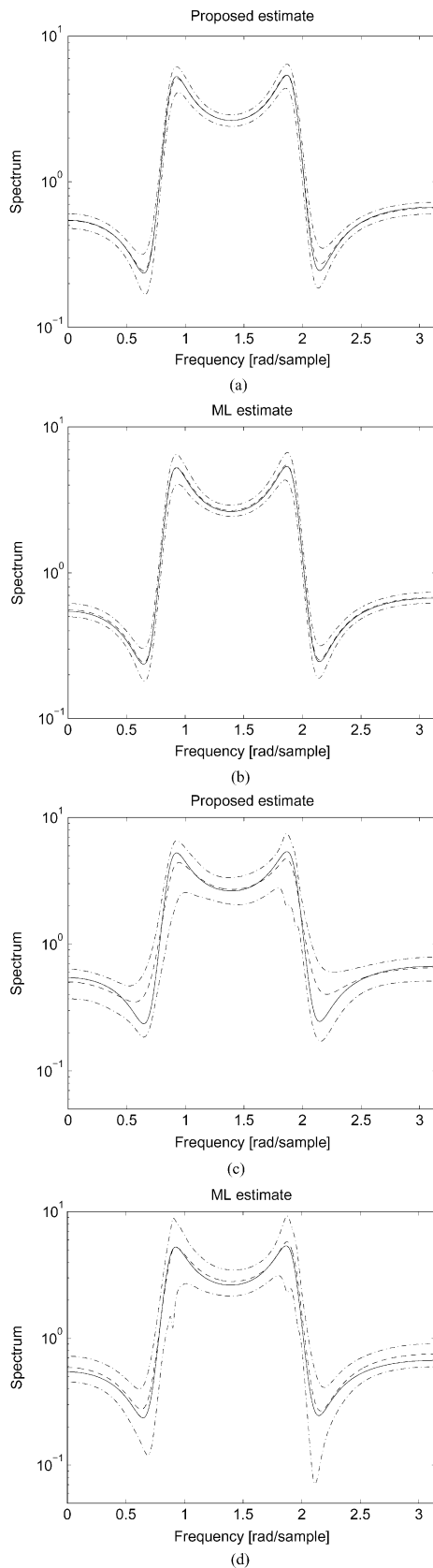
Fig. 1. Comparison between the proposed approach and maximum-likelihood method: (a) Proposed approach, $N = 1000$, (b) maximum-likelihood method, $N = 1000$, (c) proposed approach, $N = 200$, and (d) maximum-likelihood method, $N = 200$.

$u(t)$ are used. Simulation results for $N = 1000$ are shown in Fig. 1(a) and (b), where we notice that the performance of the proposed estimator is very close to that of the maximum-likelihood approach. Fig. 1(c) and (d) depicts the simulation results for $N = 200$. In this case, the proposed estimate has larger bias but smaller variance than the maximum-likelihood approach. The small sample bias effect is mainly because the initial conditions in the filtering operation (4) are unknown. This transient effects disappear as we increase N. In fact, the performance of the interpolation approach improves with the increase in the interpolation data, and may achieve the optimal performance limit asymptotically (however, we do not yet have a theoretical justification supporting this statement). This makes our algorithm an attractive alternative for spectral analysis because we solve a convex problem to arrive at the final estimates but, unlike the maximum-likelihood approach, do not encounter problems due to nonlinear optimization.

VI. CONCLUSION

In this paper we have proposed a robust interpolation algorithm for spectral estimation. The problem considered here can be seen as a generalization of the classical Nevanlinna-Pick interpolation problem. Our algorithm has three main features. Firstly, we allow noise-corrupted interpolation data. Secondly, we can incorporate a large number of interpolation data points to reduce the estimation error. Finally, an SPR solution is always guaranteed. Since SPR functions are stable, the stability of the estimated model is also automatically guaranteed. In addition, our algorithm employs semidefinite programming and is thus computationally efficient in the sense that the solution is obtained in polynomial time.

The robust spectral estimation approach can be easily generalized to multivariable processes. It is also possible to estimate the model order by using high order derivatives of $f(z)$ at appropriate interpolation points. A similar robust algorithm for covariance extension can also be derived.

APPENDIX A
PROOF OF (7)

There exists $\Gamma \in \mathbb{C}^{\nu \times \nu}$ such that

$$\Gamma^{-1}\mathbf{A}\Gamma = \mathbf{A}_*, \quad \Gamma^{-1}\mathbf{b} = \mathbf{b}_* \tag{36}$$

with $(\mathbf{A}_*, \mathbf{b}_*)$ in Jordan canonical form. Thus

$$\mathbf{A}_* = \begin{bmatrix} \mathbf{A}_0 & 0 & \cdots & 0 \\ 0 & \mathbf{A}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}_m \end{bmatrix}, \quad \mathbf{b}_* = \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_m \end{bmatrix} \tag{37}$$

where $\mathbf{A}_k \in \mathbb{C}^{\nu_k \times \nu_k}$ and $\mathbf{b}_k \in \mathbb{R}^{\nu_k \times 1}$ are given by

$$\mathbf{A}_k = \begin{bmatrix} \xi_k^{-1} & 1 & \cdots & 0 \\ 0 & \xi_k^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \xi_k^{-1} \end{bmatrix}, \quad \mathbf{b}_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{38}$$

Using (5) and (36), it follows that

$$\boldsymbol{\gamma} := \Gamma^{-1}\boldsymbol{r} = \frac{1}{2}r_0\mathbf{b}_* + \sum_{k=1}^{\infty} r_k\mathbf{A}_*^k\mathbf{b}_*. \tag{39}$$

Now, partition

$$\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\gamma}_0^\top & \boldsymbol{\gamma}_1^\top \cdots \boldsymbol{\gamma}_m^\top \end{bmatrix}^\top, \quad \boldsymbol{\gamma}_k \in \mathbb{C}^{\nu_k \times 1}. \tag{40}$$

Since $\xi_0 = \infty$ and $\nu_0 = 1$, we have $\mathbf{A}_0 = 0$. Hence, (37) and (39) gives

$$\boldsymbol{\gamma}_0 = \frac{1}{2}r_0 = f(\infty).$$

Also for $i > 0$ using (1), (37), and (39), it follows that

$$\boldsymbol{\gamma}_i = \frac{1}{2}r_0\mathbf{b}_i + \sum_{k=1}^{\infty} r_k\mathbf{A}_i^k\mathbf{b}_i. \tag{41}$$

Consequently, for $i > 0$, we have [14, pp. 557 and 565]

$$[\boldsymbol{\gamma}_i]_\ell = \frac{1}{(\nu_i - \ell)!} \frac{\mathrm{d}^{\nu_i - \ell}}{\mathrm{d}z^{\nu_i - \ell}} f\left\{\frac{1}{z}\right\}\bigg|_{z = \xi_i^{-1}},$$
$$\ell = 1, \dots, \nu_i. \tag{42}$$

Defining $^{\ell-1}C_{k-1} = (n!)/(r!(n-r)!)$, it is also straightforward to show that

$$\frac{1}{\ell!} \frac{\mathrm{d}^\ell f(z^{-1})}{\mathrm{d}z^\ell} = (-1)^\ell \sum_{k=1}^{\ell} {}^{\ell-1}C_{k-1} f_k(z) z^{-(\ell+k)} \tag{43}$$

for all $\ell > 0$. Now, setting $z = \xi_k^{-1}$ in (43) and using (42), we get

$$S_k J_k \boldsymbol{f}_k(\xi_k) = \boldsymbol{\gamma}_k$$

where $J_k \in \mathbb{R}^{\nu_k \times \nu_k}$ is the permutation matrix having *ones* along the antidiagonal and *zeros* elsewhere, $S_i = \mathrm{diag}\{1, \bar{S}_i\}$, and $\bar{S}_i \in \mathbb{C}^{(\nu_i-1)\times(\nu_i-1)}$ is given by

$$[\bar{S}_k]_{j\ell} = \begin{cases} (-1)^j {}^{j-1}C_{\ell-1}\xi_k^{j+\ell}, & \text{if } j \geq \ell \\ 0, & \text{otherwise.} \end{cases}$$

Now, S_k is invertible since it is lower triangular with nonzero diagonal entries. Hence, we get

$$\boldsymbol{f}_k(\xi_k) = J_k S_k^{-1}\boldsymbol{\gamma}_k = \mathbf{T}_k\Gamma^{-1}\boldsymbol{r} \tag{44}$$

where

$$\begin{bmatrix} \mathbf{T}_0 \\ \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_m \end{bmatrix} = \begin{bmatrix} J_0 S_0^{-1} & 0 & \cdots & 0 \\ 0 & J_1 S_1^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_m S_m^{-1} \end{bmatrix}. \tag{45}$$

APPENDIX B
PROOF OF LEMMA 1

It is clear that $f(z)$ in (10) can be rewritten (through the state transformation T) as

$$f(z) = \frac{\hat{\mathbf{f}}_0 + \bar{b}[zI_n - X_\star]^{-1}\bar{c}}{1 + \bar{b}[zI_n - X_\star]^{-1}\bar{e}} \tag{46}$$

where $\bar{b} = bT^{-1}$. It suffices to show that the $f(z)$ in (46) interpolates the given data for $i = 0, 1, \dots, p$.

It is easy to see that $f(\infty) = \hat{\mathbf{f}}_0$, which satisfies the interpolation condition for ξ_0. Consider the partial fraction expansions

$$N(z) := \hat{\mathbf{f}}_0 + \bar{b}(zI_n - X_\star)^{-1}\bar{c} = \hat{\mathbf{f}}_0 + \sum_{k=1}^{m}\sum_{j=1}^{\nu_k} \frac{n_{k,j}}{(z-\xi_k)^j};$$

$$D(z) := 1 + \bar{b}(zI_n - X_\star)^{-1}\bar{e} = 1 + \sum_{k=1}^{m}\sum_{j=1}^{\nu_k} \frac{d_{k,j}}{(z-\xi_k)^j}.$$

Since $(z - \xi_k)^{\nu_k}D(z)f(z) = (z - \xi_k)^{\nu_k}N(z)$, applying Leibnitz's theorem of successive differentiation yields

$$\frac{1}{j!}\frac{\mathrm{d}^j}{\mathrm{d}z^j}\{(z-\xi_k)^{\nu_k}N(z)\}$$
$$= \sum_{\ell=0}^{j}\frac{1}{\ell!}\frac{\mathrm{d}^\ell}{\mathrm{d}z^\ell}\{(z-\xi_k)^{\nu_k}D(z)\}f_{j-\ell}(z).$$

Evaluating at $z = \xi_k$, we have

$$n_{k,\nu_k-j} = \sum_{\ell=0}^{j} d_{k,\nu_k-\ell}f_{j-\ell}(\xi_k), \quad j = 0, 1, \dots, \nu_k - 1 \tag{47}$$

where we use the standard formula

$$n_{k,\nu_k-\ell} = \frac{1}{\ell!}\frac{\mathrm{d}^\ell}{\mathrm{d}z^\ell}\{(z-\xi_k)^{\nu_k}N(z)\}\bigg|_{z=\xi_k}$$

$$d_{k,\nu_k-\ell} = \frac{1}{\ell!}\frac{\mathrm{d}^\ell}{\mathrm{d}z^\ell}\{(z-\xi_k)^{\nu_k}D(z)\}\bigg|_{z=\xi_k}$$

for $\ell = 0, 1, \dots, \nu_k - 1$. Denoting

$$\boldsymbol{n}_k := [n_{k,\nu_k} \cdots n_{k,1}];$$
$$\boldsymbol{d}_k := [d_{k,\nu_k} \cdots d_{k,1}]$$

we can rewrite (47) as

$$\boldsymbol{n}_k = \boldsymbol{d}_k\mathcal{L}^\top\{\boldsymbol{f}_k(\xi_k)\} = \boldsymbol{f}_k^\top(\xi_k)J_k\mathcal{L}^\top(\boldsymbol{d}_kJ_k). \tag{48}$$

Defining $\boldsymbol{u}_k = [(z-\xi_k)^{-\nu_k} \cdots (z-\xi_k)^{-1}]^\top$, we can write

$$[zI - X_k]^{-1} = \mathcal{L}^\top(\boldsymbol{u}_k).$$

We also partition b as

$$\bar{b} = [b_1 \quad \cdots \quad b_p], \quad b_k \in \mathbb{C}^{1\times\nu_k}.$$

It follows that

$$\boldsymbol{n}_k \boldsymbol{u}_k = b_k [zI - X_k]^{-1} \hat{\mathbf{f}}_k = b_k \mathcal{L}^\top (\boldsymbol{u}_k) \hat{\mathbf{f}}_k = \hat{\mathbf{f}}_k^\top J_k \mathcal{L}^\top (b_k J_k) \boldsymbol{u}_k;$$
$$\boldsymbol{d}_k \boldsymbol{u}_k = b_k [zI - X_k]^{-1} \varepsilon_k = b_k \mathcal{L}^\top (\varepsilon_k) \boldsymbol{u}_k = b_k \boldsymbol{u}_k.$$

In the above, we have used $\mathcal{L}(\varepsilon_k) = I$. Now, equating coefficients of $(z - \xi_k)^{-j}$, and using (48), we get

$$\boldsymbol{d}_k = b_k, \quad \boldsymbol{n}_k = \boldsymbol{f}_k^\top (\xi_k) J_k \mathcal{L}^\top (\boldsymbol{d}_k J_k) = \hat{\mathbf{f}}_k^\top J_k \mathcal{L}^\top (\boldsymbol{d}_k J_k).$$

If $\mathcal{L}(\boldsymbol{d}_k J_k)$ is nonsingular, (8) holds. It remains to exclude the possibility that $\mathcal{L}(\boldsymbol{d}_k J_k)$ is singular. If this happens, then

$$d_{k,1} = 0, \quad n_{k,1} = 0$$

[see (48)], implying that both $N(z)$ and $D(z)$ have at least one common zero at ξ_k. This is not because $f(z)$ is of order n. Hence, $\mathcal{L}(\boldsymbol{d}_k J_k)$ must be nonsingular.

It remains to show the properties of X, b, c and e. It is known [15, pp. 109] that X is real-valued and $e = [1 \ 0 \cdots 0]$. To see c is real-valued, let us partition $T = [T_1^\top \cdots T_p^\top]^\top$ with $T_k \in \mathbb{C}^{\nu_k \times n}$. Clearly, $\hat{\mathbf{f}}_k = T_k c$. In addition, due to block diagonal structure of $X_\star$ it follows that

$$T_k = [\varepsilon_k \quad X_k \varepsilon_k \quad \cdots \quad X_k^{n-1} \varepsilon_k]. \tag{49}$$

Express $c = c_R + i c_I$. If ξ_k is real-valued, then $\hat{\mathbf{f}}_k$ is real-valued. Also by (49) T_k is real-valued. This implies that $T_k c_I = 0$.

Now, assume ξ_k is complex-valued and express $T_k = T_{Rk} + i T_{Ik}$. Then, there exists $1 \leq j \leq p$ such that ξ_j is the complex conjugate of ξ_k. Consequently, $\hat{\mathbf{f}}_j$ is the complex conjugate of $\hat{\mathbf{f}}_k$. Also by (49) T_j is the complex conjugate of T_k. Hence

$$\text{Re}\{T_k c\} = \text{Re}\{T_j c\}, \quad \text{Im}\{T_k c\} = -\text{Im}\{T_j c\}$$

which implies

$$T_{Ik} c_I = 0, \quad T_{Rk} c_I = 0$$

or $T_k c_I = 0$. It follows that $T c_I = 0$. Since ξ_k are distinct, T is nonsingular. Hence, $c_I = 0$.

Whenever ξ_j is complex conjugate of ξ_k, $f(\xi_j)$ is complex conjugate of $f(\xi_k)$. Hence, the coefficients in $f(z)$ must be real-valued. Since X, c and e are real-valued, the above implies that b is real-valued.

APPENDIX C
PROOF OF LEMMA 2

Denoting

$$\alpha_k(z) = \Omega^k(z) c, \quad \beta_k(z) = \Omega^k(z) e \quad k = p+1, \ldots, m$$

then (10) implies

$$f(z) - \hat{\mathbf{f}}_0 = b \{\beta_1(z) f(z) - \alpha_1(z)\}. \tag{50}$$

Note that

$$\frac{1}{k!} \frac{d^k \beta_1(z)}{dz^k} = \beta_{k+1}(z),$$
$$\frac{1}{k!} \frac{d^k \alpha_1(z)}{dz^k} = \alpha_{k+1}(z), \quad k > 0. \tag{51}$$

Using (51), Leibnitz's theorem of successive differentiation in (50), we get

$$f_k(z) = b \{-\alpha_{k+1} + \beta_1 f_k(z) + \beta_2 f_{k-1}(z) + \cdots \\ + \beta_{k+1} f(z)\}, \quad k > 0. \tag{52}$$

Rearranging (50) and (52) in a matrix form, we obtain

$$\boldsymbol{f}_i(z) - \varepsilon_i \hat{\mathbf{f}}_0 = b[E_i(z) \quad \mathcal{L}\{\boldsymbol{f}_i(z)\} - C_i(z)].$$

Evaluating the above at ξ_i for $i = p+1, \ldots, m$ yields (15).

APPENDIX D
PROOF OF LEMMA 3

Let us define

$$N(z) = \hat{\mathbf{f}}_0 + b(zI - X)^{-1} c,$$
$$D(z) = 1 + b(zI - X)^{-1} e. \tag{53}$$

Note that $f(z)$ is SPR if and only if $f(z)$ is stable and

$$N(z)D(z^{-1}) + N(z^{-1})D(z) = [b(zI - X)^{-1} \quad 1] \\ \cdot \begin{bmatrix} ce^\top + ec^\top & c + e\hat{\mathbf{f}}_0 \\ \hat{\mathbf{f}}_0 e^\top + c^\top & 2\hat{\mathbf{f}}_0 \end{bmatrix} \begin{bmatrix} (z^{-1}I - X^\top)^{-1} b^\top \\ 1 \end{bmatrix} > 0 \tag{54}$$

for all $|z| = 1$. By the KYP lemma [16], the above holds if and only if there exists $Q = Q^\top > 0$ such that

$$\begin{bmatrix} ce^\top + ec^\top & c + e\hat{\mathbf{f}}_0 \\ \hat{\mathbf{f}}_0 e^\top + c^\top & 2\hat{\mathbf{f}}_0 \end{bmatrix} + \begin{bmatrix} Q - XQX^\top & -XQb^\top \\ -bQX^\top & -bQb^\top \end{bmatrix} > 0. \tag{55}$$

Rewriting (55) as

$$\begin{bmatrix} ce^\top + ec^\top + Q & c + e\hat{\mathbf{f}}_0 \\ \hat{\mathbf{f}}_0 e^\top + c^\top & 2\hat{\mathbf{f}}_0 \end{bmatrix} - \begin{bmatrix} XQ \\ \beta \end{bmatrix} Q^{-1} \begin{bmatrix} XQ \\ \beta \end{bmatrix}^\top > 0$$

and applying Schur complement, (55) is equivalent to (18).

We can rewrite (18) as

$$\Theta + U\beta V^\top + V\beta^\top U^\top > 0$$

where

$$\Theta = \begin{bmatrix} ce^\top + ec^\top + Q & c + e\hat{\mathbf{f}}_0 & XQ \\ c^\top + \hat{\mathbf{f}}_0 e^\top & 2\hat{\mathbf{f}}_0 & 0 \\ QX^\top & 0 & Q \end{bmatrix}$$
$$U = \begin{bmatrix} 0_{n \times n} \\ 0_{1 \times n} \\ I_n \end{bmatrix}, V = \begin{bmatrix} 0_{n \times 1} \\ 1 \\ 0_{n \times 1} \end{bmatrix}.$$

It is straightforward to construct full-rank matrices $U_\perp$ and $V_\perp$ such that $U_\perp^\top U = 0$ and $V_\perp^\top V = 0$. One possible choice is

$$V_\perp = \begin{bmatrix} I_n & 0_{n \times n} \\ 0_{1 \times n} & 0_{1 \times n} \\ 0_{n \times n} & I_n \end{bmatrix}, \quad U_\perp = \begin{bmatrix} I_{n+1} \\ 0_{n \times 2n} \end{bmatrix}.$$

Using the elimination lemma ([17, pp. 22]), there exists β such that (18) holds if and only if $U_\perp^\top \Theta U_\perp > 0$ and $V_\perp^\top \Theta V_\perp > 0$. It can be verified that $U_\perp^\top \Theta U_\perp > 0$ is equivalent to (20), while $V_\perp^\top \Theta V_\perp > 0$ is equivalent to

$$\begin{bmatrix} ce^\top + ec^\top + Q & XQ \\ QX^\top & Q \end{bmatrix} > 0$$

which can be transformed via Schur complement into

$$ce^\top + ec^\top + Q - XQX^\top > 0, \quad Q > 0.$$

Next, we show that (19), (20), and $Q > 0$ are equivalent to (21). We first show the sufficiency. Suppose that P is positive definite. Then, $XEX^\top > 0$, implying $E > 0$ since X is stable and thus invertible. Set $Q = E$ in (19) and (20). Comparing with (22), we see that (19) becomes an equality, while the inequality (20) is satisfied. Therefore, it is possible to find $Q > 0$ (by perturbing from E slightly) such that both the inequalities (19) and (20) hold.

Conversely, suppose that there exists $Q > 0$ such that (19) and (20) hold. Then, using (22) and (19), we get

$$X(E - Q)X^\top > E - Q \quad \Rightarrow \quad E > Q. \tag{56}$$

However, from (21) and (22)

$$P - \begin{bmatrix} ce^\top + ec^\top + Q & c + e\hat{\mathbf{f}}_0 \\ c^\top + \hat{\mathbf{f}}_0 e^\top & 2\hat{\mathbf{f}}_0 \end{bmatrix} = \begin{bmatrix} E - Q & 0 \\ 0 & 0 \end{bmatrix} \geq 0.$$

Hence, (56) and (20) imply $P > 0$.

APPENDIX E
DERIVATION OF G IN (28) AND H IN (31)

Using (44), (25) and denoting $\hat{\boldsymbol{c}} = [\hat{c}^\top \ \ \mathbf{f}_0]^\top$, we have

$$H = \begin{bmatrix} T^{-1}[\mathbf{T}_1^\top \cdots \mathbf{T}_p^\top]^\top \\ \mathbf{T}_0 \end{bmatrix} \Gamma^{-1}.$$

Defining $\hat{\epsilon}_k = \hat{h}_k - b\hat{L}_k$ for $k = p_1, \ldots, m$, then (13) and (14) becomes

$$\hat{\epsilon}_k = (\hat{\mathbf{f}}_k - \varepsilon_k \hat{\mathbf{f}}_0)^\top + b\{C_k(\xi_k) - E_k(\xi_k)\mathcal{L}(\hat{\mathbf{f}}_k)\}. \tag{57}$$

Using (11) and (12), the identity $\boldsymbol{w}\mathcal{L}(\boldsymbol{v}) = \boldsymbol{v}^\top \mathcal{L}(\boldsymbol{w})$ for $\boldsymbol{v}$ and $\boldsymbol{w}$ with compatible dimensions, and

$$O_k := \begin{bmatrix} [b\Omega^{\nu_k}(\xi_k)]^\top & \cdots & [b\Omega(\xi_k)]^\top \end{bmatrix}^\top$$

we can rewrite (57) as

$$\begin{aligned} \hat{\epsilon}_k^\top &= (\hat{\mathbf{f}}_k - \varepsilon_k \hat{\mathbf{f}}_0)^\top + (O_k \hat{c})^\top - \hat{\mathbf{f}}_k^\top \mathcal{L}\{bE_k(\xi_k)\} \\ &= [I - \mathcal{L}\{bE_k(\xi_k)\}]\hat{\mathbf{f}}_k + [O_k \quad -\varepsilon_k]\hat{c} \\ &= B_k \hat{\boldsymbol{r}} \end{aligned} \tag{58}$$

where B_k is obtained using (44) and (31) and is given by

$$B_k = [O_k \quad -\varepsilon_k]H + [I - \mathcal{L}\{bE_k(\xi_k)\}]\mathbf{T}_k \Gamma^{-1}.$$

Now, using the definition for h in (16) and a similar definition for L, we get

$$\hat{\epsilon} = (\hat{h} - b\hat{L})^\top = \begin{bmatrix} \hat{\epsilon}_r^\top & \text{Re}\{\hat{\epsilon}_c^\top\} & \text{Im}\{\hat{\epsilon}_c^\top\} \end{bmatrix}^\top$$

where

$$\hat{\epsilon}_r := \begin{bmatrix} \hat{\epsilon}_{p+1} & \cdots & \hat{\epsilon}_{p+p_r} \end{bmatrix}^\top,$$
$$\hat{\epsilon}_c := \begin{bmatrix} \hat{\epsilon}_{p+p_r+1} & \cdots & \hat{\epsilon}_{p+p_r+p_c} \end{bmatrix}^\top.$$

Hence, from (58), we get (28) with

$$G_1 = \begin{bmatrix} B_{p+1} \\ \vdots \\ B_{p+p_r} \end{bmatrix}, \ G_2 = \begin{bmatrix} B_{p+p_r+1} \\ \vdots \\ B_{p+p_r+p_c} \end{bmatrix}, G = \begin{bmatrix} G_1 \\ \text{Re}\{G_2\} \\ \text{Im}\{G_2\} \end{bmatrix}.$$

APPENDIX F
PROOF OF LEMMA 4

In this Appendix, we write

$$G(q^{-1}) = [0.5I + \mathbf{A}(qI - \mathbf{A})^{-1}]\mathbf{b}.$$

Note that $\alpha(q^{-1}) = G(q^{-1}) + G(q)$. For two stationary vector valued processes $x(t)$ and $y(t)$, we denote

$$\rho_{xy}(t_1 - t_2) = \mathsf{E}\{x(t_1)y^\top(t_2)\}.$$

Using a standard formula for the fourth-order moments of Gaussian random vectors [18], we get

$$\begin{aligned} &\mathsf{E}\{\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}^\top\} \\ &= \frac{1}{N^2} \sum_{t_1=1}^N \sum_{t_2=1}^N \mathsf{E}\{x(t_1)u(t_1)u(t_2)x^\top(t_2)\} \\ &= \frac{1}{N^2} \sum_{t_1=1}^N \sum_{t_2=1}^N [\rho_{xu}(0)\rho_{ux}(0) + \rho_{xu}(t_1-t_2)\rho_{ux}(t_1-t_2) \\ &\quad + \rho_{xx}(t_1-t_2)\rho_{uu}(t_1-t_2)] \\ &= \boldsymbol{r}\boldsymbol{r}^\top + \frac{1}{N^2} \sum_{t_1=1}^N \sum_{\tau=-N+1}^{N-1} |N - \tau|[\rho_{xu}(\tau)\rho_{ux}(\tau) \\ &\quad + \rho_{xx}(\tau)\rho_{uu}(\tau)]. \end{aligned} \tag{59}$$

Note that the inner summation in the right-hand side of (59) is independent of t_1. In addition, the magnitudes of the correla-

tion functions $\rho_{ux}(\tau)$, $\rho_{xx}(\tau)$, etc., decay exponentially with τ. Hence, for large N, we can simplify (59) as

$$\hat{\boldsymbol{r}}\hat{\boldsymbol{r}}^\top - \boldsymbol{r}\boldsymbol{r}^\top = \frac{1}{N}\sum_{\tau=-N+1}^{N-1}[\rho_{xu}(\tau)\rho_{ux}(\tau) + \rho_{xx}(\tau)\rho_{uu}(\tau)]. \tag{60}$$

We know that

$$\rho_{xu}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi}G(\mathrm{e}^{-\mathrm{i}\omega})|g(\mathrm{e}^{\mathrm{i}\omega})|^2\mathrm{e}^{\mathrm{i}\omega\tau}\,\mathrm{d}\omega.$$

Hence, using Poisson's summation formula, it follows that

$$\lim_{N\to\infty}\sum_{\tau=-N+1}^{N-1}\rho_{xu}(\tau)\rho_{ux}(\tau)$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}G(\mathrm{e}^{-\mathrm{i}\omega})|g(\mathrm{e}^{\mathrm{i}\omega})|^4 G^\top(\mathrm{e}^{\mathrm{i}\omega})\,\mathrm{d}\omega. \tag{61}$$

Also, a similar calculation gives

$$\lim_{N\to\infty}\sum_{\tau=-N+1}^{N-1}\rho_{xx}(\tau)\rho_{uu}(\tau)$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}G(\mathrm{e}^{-\mathrm{i}\omega})|g(\mathrm{e}^{\mathrm{i}\omega})|^4 G^\top(\mathrm{e}^{-\mathrm{i}\omega})\,\mathrm{d}\omega. \tag{62}$$

Hence, by combining (60), (61), and (62), we get

$$N\mathrm{cov}\{\hat{\boldsymbol{r}}\} = \frac{1}{2\pi}\int_{-\pi}^{\pi}G(\mathrm{e}^{-\mathrm{i}\omega})|g(\mathrm{e}^{\mathrm{i}\omega})|^4\alpha^\top(\mathrm{e}^{-\mathrm{i}\omega})\,\mathrm{d}\omega. \tag{63}$$

Note that $\alpha(\mathrm{e}^{-\mathrm{i}\omega}) = \mathrm{Re}\{G(\mathrm{e}^{-\mathrm{i}\omega})\}$, and $\mathrm{Im}[G(\mathrm{e}^{-\mathrm{i}\omega})]$ is an odd function. Using these observations, it is straightforward to derive (24) from (63).

APPENDIX G
PROOF OF (34)

In this Appendix, we denote

$$\Psi_1 = H\Psi H^\top, \quad \Psi_3 = H\Psi G^\top,$$
$$\Pi = \left\{I - L^\top\left[L\Psi_2^{-1}L^\top\right]^{-1}L\Psi_2^{-1}\right\}.$$

It can be verified that

$$\Pi L^\top = 0, \tag{64}$$

and

$$\mathbf{Q} := \Pi\Psi_2 = \Psi_2 - L^\top\left[L\Psi_2^{-1}L^\top\right]^{-1}L = \Pi\Psi_2\Pi^\top.$$

Hence

$$\Pi^\top[\Pi\Psi_2\Pi^\top]^\dagger\Pi = \Psi_2^{-1}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{Q}\Psi_2^{-1} = \Psi_2^{-1}\mathbf{Q}\Psi_2^{-1}. \tag{65}$$

For large N, it holds that $\hat{L}\approx L$, $\hat{G}\approx G$, $\hat{\Pi}\approx\Pi$, etc. Consequently, it follows from (27) that

$$(\overset{\circ}{b}-b)^\top = \left[L\Psi_2^{-1}L^\top\right]^{-1}L\Psi_2^{-1}\hat{\boldsymbol{r}}. \tag{66}$$

Similarly, using (31) and (32), it follows that

$$\overset{\circ}{\mathbf{c}}-\mathbf{c} = \left\{H - \Psi_3\Pi^\top\left[\Pi\Psi_2\Pi^\top\right]^\dagger\Pi G\right\}\boldsymbol{r}. \tag{67}$$

Consequently, by (66), we have

$$\mathsf{E}\{(\overset{\circ}{b}-b)^\top(\overset{\circ}{b}-b)\} = \left[L\Psi_2^{-1}L^\top\right]^{-1} =: \mathbf{X}^{-1} \tag{68}$$

while by (67) and (65), we get

$$\mathsf{E}\{(\overset{\circ}{\mathbf{c}}-\mathbf{c})(\overset{\circ}{\mathbf{c}}-\mathbf{c})^\top\} = \Psi_1 - \Psi_3\Pi^\top[\Pi\Psi_2\Pi^\top]^\dagger\Pi\Psi_3^\top$$
$$= \Psi_1 - \Psi_3\Psi_2^{-1}\mathbf{Q}\Psi_2^{-1}\Psi_3^\top = \mathbf{Y} - \mathbf{B}^\top\mathbf{X}^{-1}\mathbf{B} \tag{69}$$

where

$$\mathbf{B} = L\Psi_2^{-1}\Psi_3^\top, \quad \mathbf{Y} = \Psi_1 - \Psi_3\Psi_2^{-1}\Psi_3^\top. \tag{70}$$

In addition, using (64)–(67), it follows that

$$\mathsf{E}\{(\overset{\circ}{\mathbf{c}}-\mathbf{c})(\overset{\circ}{b}-b)\} = \Psi_3\Psi_2^{-1}L^\top\mathbf{X}^{-1} = \mathbf{B}\mathbf{X}^{-1}. \tag{71}$$

Combining (68), (69), (70), and (71), and using a standard block matrix inversion formula [12, pp. 512], we get

$$\mathbf{R} = \mathsf{E}\left\{\begin{bmatrix}\overset{\circ}{\mathbf{c}}-\mathbf{c}\\(\overset{\circ}{b}-b)^\top\end{bmatrix}\begin{bmatrix}\overset{\circ}{\mathbf{c}}-\mathbf{c}\\(\overset{\circ}{b}-b)^\top\end{bmatrix}^\top\right\}$$
$$= \begin{bmatrix}\mathbf{Y}^{-1}+\mathbf{B}^\top\mathbf{X}^{-1}\mathbf{B} & \mathbf{B}^\top\mathbf{X}^{-1}\\\mathbf{X}^{-1}\mathbf{B} & \mathbf{X}^{-1}\end{bmatrix}$$
$$= \begin{bmatrix}\mathbf{Y}^{-1} & -\mathbf{Y}^{-1}\mathbf{B}^\top\\-\mathbf{B}\mathbf{Y}^{-1} & \mathbf{X}+\mathbf{B}\mathbf{Y}^{-1}\mathbf{B}^\top\end{bmatrix}^{-1}$$
$$= \left\{\begin{bmatrix}I & 0\\0 & L\end{bmatrix}\begin{bmatrix}\Psi_1 & \Psi_3\\\Psi_3^\top & \Psi_2\end{bmatrix}^{-1}\begin{bmatrix}I & 0\\0 & L\end{bmatrix}^\top\right\}^{-1}$$

which is same as (34).

REFERENCES

[1] S. M. Kay, *Modern Spectral Estimation*. Upper Saddle River, NJ: Prentice-Hall, 1988.

[2] P. Stoica, T. Söderström, and B. Friedlander, "Optimal instrumental variable estimates of the AR parameters of an ARMA process," *IEEE Trans. Autom. Control*, vol. AC-30, no. 11, pp. 1066–1074, 1985.

[3] P. Stoica, B. Friedlander, and T. Söderström, "Approximate maximum-likelihood approach for ARMA spectral estimation," *Int. J. Control*, vol. 45, no. 4, pp. 1281–1310, 1987.

[4] J. Mari, P. Stoica, and T. McKelvey, "Vector ARMA estimation: A reliable subspace approach," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 2092–2104, Jul. 2000.

[5] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 3189–3205, Nov. 2000.

[6] T. T. Georgiou, "Signal estimation via selective harmonic amplification: MUSIC, redux," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 780–790, Mar. 2000.

[7] T. T. Georgiou, "The structure of state covariances and its relation to the power spectrum of the input," *IEEE Trans. Autom. Control*, vol. 47, no. 7, pp. 1056–1066, Jul. 2002.

[8] C. I. Byrnes, P. Enqvist, and A. Lindquist, "Cepstral coefficients, covariance lags and pole-zero models for finite data strings," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 677–693, Apr. 2001.

[9] N. I. Akhiezer, *The Classical Moment Problem*. London, U.K.: Oliver and Boyd, 1965.

[10] D. Sarason, "Generalized interpolation in H-infinity," *Trans. Amer. Math. Soc.*, vol. 127, pp. 179–203, 1967.

[11] M. Fu and K. Mahata, "Generalizations of the Nevanlinna–Pick interpolation problem," presented at the 44th IEEE Conf. Decision Control, Seville, Spain, Dec. 2005.

[12] T. Söderström and P. Stoica, *System Identification*. Hemel, Hempstead, U.K.: Prentice-Hall Int., 1989.

[13] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[14] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The John Hopkins Univ. Press, 1996.

[15] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[16] A. Rantzer, "On the Kalman–Yakubovitz–Popov lemma," *Syst. Control Lett.*, vol. 28, pp. 7–10, 1996.

[17] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.

[18] P. Janssen and P. Stoica, "On the expectation of the product of four matrix-valued Gaussian random variables," *IEEE Trans. Autom. Control*, vol. AC-33, pp. 867–870, 1988.

Kaushik Mahata received the M.E. degree in signal processing from the Indian Institute of Science, Bangalore, in 2000 and the Ph.D. degree in signal processing from Uppsala University, Uppsala, Sweden, in 2003.

Currently, he is a Lecturer in the University of Newcastle, Australia. His research interest includes estimation identification, spectrum analysis, and machine learning.

Minyue Fu (S'84–M'87–SM'94–F'04) received the Bachelor's degree in electrical engineering from the University of Science and Technology of China, Hefei, in 1982 and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison in 1983 and 1987, respectively.

From 1983 to 1987, he held a teaching assistantship and a research assistantship at the University of Wisconsin-Madison. He worked as a Computer Engineering Consultant at Nicolet Instruments, Inc., Madison, WI, during 1987. From 1987 to 1989, he served as an Assistant Professor in the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI. In summer 1989, he was employed by the Universite Catholoque de Louvain, Belgium, as a Maitre de Conferences Invited. He joined the Department of Electrical and Computer Engineering, the University of Newcastle, Australia, in 1989, where he is currently a Chair Professor in Electrical Engineering and Head of School of Electrical Engineering and Computer Science. In addition, he was a Visiting Associate Professor at University of Iowa during 1995–1996, and a Senior Fellow/Visiting Professor at Nanyang Technological University, Singapore, 2002. His main research interests include control systems, signal processing, and communications.

Dr. Fu has been an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and the *Journal of Optimization and Engineering*.