Multivariate spectral approximation in the Hellinger distance

Federico Ramponi, Augusto Ferrante ∗ and Michele Pavon †

Abstract
We first describe a globally convergent matricial Newton-type algorithm designed to solve the multivariable spectrum approximation problem. Then, we apply this approximation procedure to the estimation of multivariate spectral densities, and test its effectiveness through simulation.

Keywords: spectrum approximation, Hellinger distance, multivariable spectrum estimation, Newton algorithm.

1 Introduction
The aim of this paper is to introduce a globally convergent matricial Newton-type algorithm, designed to solve the problem of multivariable spectrum approximation in the Hellinger distance. The latter problem has been investigated in [6], and the algorithm to solve its dual problem, presented here, is thoroughly described in [12].

It is well known that ARMA(\textit{m, n}) identification procedures usually lead to nonconvex optimization problems for which global convergence is not guaranteed. A new approach to identification, based on \textit{convex} optimization over different subclasses of ARMA models, has been recently introduced by Byrnes, Georgiou, Lindquist and their collaborators, in the frame of their research on analytic inter-
polation with degree constraint.

In particular, their paper [2] describes a new setting for spectral estimation. The THREE algorithm, is shown to allow for higher resolution in prescribed frequency bands and to be particularly suitable in the case of few data. An outline of this method is as follows. A finite collection of data \( y_1 \ldots y_N \) is fed to a suitably structured bank of filters, and the steady-state covariance matrix of the resulting output is estimated by statistical methods. Finding an input process whose rational spectrum is compatible with the estimated covariance poses naturally a Nevanlinna-Pick interpolation problem with bounded degree: The solution of this interpolation problem is now considered as a mean of estimating the spectrum.

A (very) particular case described in the paper is the spectrum with maximum entropy rate compatible with the covariance matrix, which amounts to the so-called central solution in the Nevanlinna-Pick theory. In this case, an expression for the analytical solution of the problem is known and readily computable without need for further numerical analysis, even in the multivariate case [8]. But more generally, the scheme allows to exploit a non constant a priori estimate \( \Psi \) of the spectrum. The Byrnes-Georgiou-Lindquist school has shown how this and other important problems of control theory may be advantageously casted in terms of the minimization of a logarithmic distance from \( \Psi \), subject to second-order moment constraints (logarithmic functionals such as the Kullback-Leibler distance arise as distances between spectral density measures in a natural way, due to their well-known connection with prediction error methods). The duals of these convex problems can be shown to admit a solution: The existence result, due to Byrnes and Lindquist [1] (see also [5]) is, however, nontrivial since the optimization occurs on an open, unbounded set of Hermitian matrices. In the multivariable case, a Kullback-Leibler pseudodistance may also be readily defined, but the related spectrum approximation problem does not lead to a computable solution unless \( \Psi \) is the identity matrix (maximum entropy solution).

This paper adds to this effort in that we consider estimation of a multivariate spectral density in the spirit of THREE, but employing a different metric for the optimization, namely an extension of the Hellinger distance which was introduced in [6]. With this metric, the approximation problem generalizes nicely to the multivariable case for any prior estimate \( \Psi \) of the spectrum.

An outline of the paper follows: After some background material in Sections 2—4, in Section 5 we present a Newton-type matricial iteration designed to numerically solve the dual of the multivariable spectrum approximation problem. It had originally been sketched in [6], and it is treated systematically in [12]. Finally, in Section 6 we present guidelines for its application to spectral estimation and present some simulations.

2 Constrained spectrum approximation

Paper [7] introduces and solves the following generalized moment problem: Given an input-to-state stable transfer function \( G(z) = (zI - A)^{-1}B \) (which is a generalization of the above bank of filters) and a state covariance matrix \( \Sigma \), give necessary and
sufficient conditions for the existence of input spectra Φ(e^{iθ}) such that the steady state output has variance Σ, that is,

\[ \int G\Phi G^* = \Sigma. \] (1)

(here, and in the sequel, integration takes place on the unit circle \( T \) with respect to normalized Lebesgue measure \( d\theta/2\pi \)). Throughout this paper we use the notations \( A^* = A^\top \) for matrices and \( G^* = G^\top (z^{-1}) \) for spectra and transfer functions. The scalar product between square matrices is defined as \( \langle A, B \rangle = \text{tr} AB^* \).

Let \( S = S_m^{n \times m}(T) \) be the family of \( \mathbb{C}^{m \times m} \)-valued functions defined on the unit circle which are Hermitian, positive-definite, bounded and coercive. We have the following existence result [7]: There exists \( \Phi \in S \) satisfying (1) if and only if there exists \( H \in \mathbb{C}^{m \times n} \) such that

\[ \Sigma - A\Sigma A^* = BH + H^*B^* \] (2)

Paper [11] deals with the following (scalar) spectrum approximation problem: When constraint (1) is feasible, find the spectrum \( \Phi \) which minimizes the Kullback-Leibler pseudo distance

\[ d_{KL}(\Psi, \Phi) = \int \Psi \log \frac{\Psi}{\Phi} \]

from an “a priori” spectrum \( \Psi \), subject to the constraint (1). It turns out that, when the prior \( \Psi \) is rational, the solution is also rational, and with degree that can be bounded in terms of the degrees of \( G(z) \) and \( \Psi \). This problem again admits the maximum differential entropy spectrum, compatible with the constraint, as a particular case (\( \Psi(e^{iθ}) \equiv 1 \)). The above minimization poses naturally a variational problem, which can be solved using Lagrange theory. Its dual problem admits a maximum and can be solved exploiting numerical algorithms. In [4] we restated and solved a similar variational problem with respect to a different metric, namely the Hellinger distance:

\[ d_H(\Psi, \Phi) = \sqrt{\int (\sqrt{\Psi} - \sqrt{\Phi})^2} \] (3)

Equation (3) defines a bona fide distance, well-known in mathematical statistics. The main advantage of this approach is that it easily generalizes to the multivariable case, whereas log-like functionals do not enjoy this property.

3 Feasibility and the operator \( \Gamma \)

Following [11], [6] and [12], let \( \mathcal{H}(n) = \{ M \in \mathbb{C}^{n \times n} : M = M^* \} \), let \( \mathcal{C}(T; \mathcal{H}(m)) \) be the space of \( \mathcal{H}(m) \)-valued continuous functions defined on the unit circle, and let the operator \( \Gamma : \mathcal{C}(T; \mathcal{H}(m)) \to \mathcal{H}(n) \) be defined as follows:

\[ \Gamma(\Phi) := \int G\Phi G^* \] (4)
Consider the range of the operator $\Gamma$ (which is understood as a vector space over the reals, since we must deal with Hermitian matrices):

**Proposition 1.** ([12]) The following facts hold:

1. Let $\Sigma = \Sigma^* > 0$. The following are equivalent:
   - There exists $H \in \mathbb{C}^{m \times n}$ such that identity (2) holds.
   - There exists $\Phi \in S_+^{m \times m}(\mathbb{T})$ such that $\int G\Phi G^* = \Sigma$.
   - There exists $\Phi \in \mathcal{C}(\mathbb{T} ; \mathcal{H}(m))$, $\Phi > 0$ such that $\Gamma(\Phi) = \Sigma$.

2. Let $\Sigma = \Sigma^*$ (not necessarily positive definite). There exists $H \in \mathbb{C}^{m \times n}$ such that identity (2) holds if and only if $\Sigma \in \text{Range } \Gamma$.

3. $X \in \text{Range } \Gamma^\perp$ if and only if $G^*(e^{j\theta})XG(e^{j\theta}) = 0 \ \forall \theta \in [0, 2\pi]$.

**Remark.** Proposition 1 shows that $\text{Range } \Gamma$ is the set of all the Hermitian matrices $\Sigma$ for which there exists $H$ such that (2) holds. This fact is useful in numerical computations. Indeed, $\text{Range } \Gamma$ is obviously finite-dimensional, and if $\{H_1, \ldots, H_N\}$ is a base of $\mathbb{C}^{m \times n}$, then the corresponding solutions $\{\Sigma_1, \ldots, \Sigma_N\}$ of (2), considered as a discrete-time Lyapunov equation in the unknown $\Sigma$, generate $\text{Range } \Gamma$. Note that $\{\Sigma_1, \ldots, \Sigma_N\}$ are not necessarily linearly independent.

4 **Multivariable spectrum approximation in the Hellinger distance**

Let the function $d_H : S_+^{m \times m}(\mathbb{T}) \times S_+^{m \times m}(\mathbb{T}) \to \mathbb{R}^+$ be defined as follows:

$$d_H(\Psi, \Phi)^2 := \inf_{W_\Psi, W_\Phi} \text{tr} \int (W_\Psi - W_\Phi)(W_\Psi - W_\Phi)^*$$

such that $W_\Psi W_\Psi^* = \Psi$ and $W_\Phi W_\Phi^* = \Phi$  \hspace{1cm} (5)

that is, $d_H(\Psi, \Phi)$ is the $L^2$ distance between the sets of spectral factors of the two spectra. It was shown in [6] that the infimum in (5) is actually a minimum and that $d_H$ is a bona fide distance between spectral densities. Moreover, $d_H$ reduces to the ordinary Hellinger distance in the scalar case. We call it “the” multivariable Hellinger distance, despite not being the obvious, naive generalization of the Hellinger distance to multivariable densities (which simply stems from the use of matrix square roots and the adjunction of a trace operator to the integral in (3)).

The multivariable spectrum approximation problem addressed in [6] is the following. Let $G(z) = (zI - A)^{-1}B$, with $A$ stable, $B$ of full rank and $(A, B)$ reachable. Given $\Psi \in \mathcal{S}$, find

$$\arg \min_{\Phi \in \mathcal{S}} d_H(\Psi, \Phi) \text{ such that } \int G\Phi G^* = \Sigma$$  \hspace{1cm} (6)
The key property of (5) is that once we fix a factor $W_\Psi$ of $\Psi$, we can simply minimize over the spectral factors of $\Phi$:

$$d_H(\Psi, \Phi)^2 \equiv \min_W \text{tr} \int (W_\Psi - W) (W_\Psi - W)^+ \text{ such that } WW^* = \Phi$$

(7)

It’s not difficult to see that there is no loss of generality in taking $\Sigma = I$. Thus, once a factor $W_\Psi$ of $\Psi$ is fixed, (6) reduces to:

$$\arg\min_W \text{tr} \int (W_\Psi - W) (W_\Psi - W)^+ \text{ such that } \int GWW^* G^* = I$$

(8)

which is an $L^2$ constrained minimization. Now let us assume that the problem is feasible, i.e., condition (2) holds. To solve (8), form the Lagrangian:

$$L(W, \Lambda) = \text{tr} \int (W_\Psi - W) (W_\Psi - W)^+ + \langle \Lambda, \int GWW^* G^* - I \rangle$$

(9)

where $\Lambda \in \mathcal{H}(n)$. Since $\int GWW^* G^* \in \text{Range } \Gamma$ by construction, and $I \in \text{Range } \Gamma$ by the feasibility assumption, it is natural to restrict a priori the Lagrange parameter $\Lambda$ to $\text{Range } \Gamma$ (a $\Lambda \in \text{Range } \Gamma^\perp$ would not play any role in the above Lagrangian).

We proceed with unconstrained minimization of (9). The functional (9) is convex and differentiable in $W$. Thus, to find the unique minimizing solution we impose that the first variation of (9) is zero in each direction $\delta W$. We easily find the following condition for $W$ (see [6] for the details):

$$W - W_\Psi + G^* \Lambda G W = 0$$

To carry on with the computations, and to ensure that the resulting optimum spectrum is integrable over the unit circle, we require a posteriori that $\Lambda$ belongs to the set $\mathcal{L}^H := \{ \Lambda \in \mathcal{H}(n), \ I + G^* \Lambda G > 0 \ \forall \ \omega^0 \in \mathbb{T} \}$ that is, $\Lambda \in \mathcal{L}^H$ is an $H$ matrix.

Remark. Observe that, when $\Psi$ is rational, (10) yields a rational spectrum with McMillan degree that can be bounded. The same applies to scalar spectrum approximation problem in a Kullback-Leibler type distance where the degree of the optimal approximant is actually lower [11]. In the multivariable case, however, the Kullback-Leibler solution is computable and of bounded McMillan degree only when $\Psi = I$. Consider now the issue of existence of a matrix $\Lambda \in \mathcal{L}^H$ such that

$$\int G(I + G^* \Lambda G)^{-1} \Psi(I + G^* \Lambda G)^{-1} G^* = I$$

(11)

that is, such that the corresponding optimal spectrum satisfies the constraint (1).
states that such a $\Lambda$ always exists. In order to find the optimal $\Lambda$, we form the dual functional:

$$L_d(\Lambda) = L(\hat{W}, \Lambda) = \text{tr} \int (\Psi - (I + G^* \Lambda G)^{-1} \Psi) - \text{tr} \Lambda$$  \hfill (12)

but instead of maximizing (12), we consider the equivalent problem of minimizing the following functional:

$$J_\Psi(\Lambda) = -L_d(\Lambda) + \text{tr} \int \Psi = \text{tr} \int (I + G^* \Lambda G)^{-1} \Psi + \text{tr} \Lambda$$  \hfill (13)

Besides being convex, and therefore continuous, $J_\Psi$ has very strong regularity properties:

**Theorem 2.** Consider $J_\Psi : \mathcal{L}_H^H \subset \text{Range} \Gamma \to \mathbb{R}$. Then

1. $J_\Psi \in C^\infty(\mathcal{L}_H^H)$.
2. $J_\Psi$ is strictly convex on $\mathcal{L}_H^H$.

**Theorem 3.** $J_\Psi$ has an unique minimum point in $\mathcal{L}_H^H$.

Theorems 2 and 3 are proven in [12] and [6], respectively. The minimization of $J_\Psi$ over $\mathcal{L}_H^H$ is the main subject of the following section.

## 5 A matricial Newton algorithm

The Newton algorithm is an iterative procedure for the search of roots of a function or the minimization of a functional. With respect to the latter objective, it can be formulated as follows. Let $f : S \to \mathbb{R}$ be a functional defined over $S \subset \mathbb{R}^n$. In order to find an estimate $\hat{x}$ of a minimum point $x^*$ of $f$,

1. Make an initial guess $x_0$, possibly near the minimum point.
2. At each iteration, compute the Newton step

$$\Delta x_i = -H^{-1}_{x_i} \nabla f_{x_i}$$ \hfill (14)

where $H_{x_i}$ is the Hessian of $f$ at $x_i$ and $\nabla f_{x_i}$ is the gradient of $f$ at $x_i$ (understood as a column vector).
3. Set $t_i^0 = 1$, and let $t_i^{k+1} = t_i^k / 2$ until both of the following conditions hold:

$$x_i + t_i^k \Delta x_i \in S$$ \hfill (15)

$$f(x_i + t_i^k \Delta x_i) < f(x_i) + \alpha t_i^k \nabla f_{x_i}^T \Delta x_i$$ \hfill (16)

where $\alpha$ is a real constant, $0 < \alpha < 1/2$. 
4. Set $x_{i+1} = x_i + t_i \Delta x_i$.

5. Repeat steps 2, 3 and 4 until $|\nabla f_{x_i}| < \varepsilon$, where $\varepsilon$ is a (small) tolerance threshold, then set $\hat{x} = x_i$.

In its “pure” form, the iteration of the Newton algorithm only consists in step 2, which is indeed its essential part. Step 3 is the so-called backtracking procedure. In essence, the “pure” Newton algorithm works very well when the starting point happens to be near the minimum and the function $f$ is there effectively approximated by a quadratic form, but it can suffer from numerical problems when this is not the case. The backtracking line search is a remedy to this drawback; moreover it can be shown that, under certain regularity assumptions on $f$, which hold in our case, after a finite number of iterations step 3 always selects the multiplier $t = 1$, that is, the full step, and converges quadratically to the minimum point.

We must minimize the functional $J_\Psi(\Lambda)$ over the set $\mathcal{L}$. As initial condition, we can safely choose $\Lambda_0 = 0$. It turns out that, although the problem is finite-dimensional, the inversion of the Hessian is more demanding than inverting a matrix. In order to compute the Newton step $\Delta \Lambda_i$, we must solve at $\Lambda_i$ the following linear equation:

$$H_{\Lambda_i}(\Delta \Lambda_i, \cdot) = -\nabla J_\Psi(\cdot)$$

where $\nabla J_\Psi(\cdot)$ and $H_{\Lambda_i}(\Delta \Lambda_i, \cdot)$ must be understood as a linear and a bilinear form, defined respectively by the first second variations of $J_\Psi$ taken at $\Lambda_i$ in the direction $\Delta \Lambda_i$ and in the arbitrary direction “ $\cdot$ ”. Comparing with these variations, (17) reduces to:

$$\int GQ^{-1}_{\Lambda_i} \left[ (G^*(\Delta \Lambda_i)GQ^{-1}_{\Lambda_i}\Psi) + (G^*(\Delta \Lambda_i)GQ^{-1}_{\Lambda_i}\Psi)^* \right] Q^{-1}_{\Lambda_i}G^*$$

$$= -\left( I - \int GQ^{-1}_{\Lambda_i}\Psi Q^{-1}_{\Lambda_i}G^* \right)$$

(18)

In principle, equation (18) is not difficult to solve. We suggest the following procedure. At the beginning, take a base $\{H_1, ..., H_k, ..., H_N\}$ of $\mathbb{C}^{m \times n}$ and compute the solutions $\{\Sigma_1, ..., \Sigma_k, ..., \Sigma_N\}$ of the following discrete-time Lyapunov equations:

$$\Sigma_k - A \Sigma_k A^* = B H_k + H_k^* B^*$$

As shown before, these solutions generate Range $\Gamma$. To compute $\Delta \Lambda_i$ at each step,

1. Compute the integral

$$Y = \int GQ^{-1}_{\Lambda_i}\Psi Q^{-1}_{\Lambda_i}G^* - I$$

(19)

2. For each $\Sigma_k$ in the precomputed generators, compute the following integral:

$$Y_k = \int GQ^{-1}_{\Lambda_i} \left[ (G^* \Sigma_k GQ^{-1}_{\Lambda_i}\Psi) + (G^* \Sigma_k GQ^{-1}_{\Lambda_i}\Psi)^* \right] Q^{-1}_{\Lambda_i}G^*$$

(20)
3. Solve, by means of linear algebraic methods (the Moore-Penrose pseudoinverse), the equation
\[ \sum_k \alpha_k Y_k = Y \]  

(21)

By linearity, the solution to (18) is now \( \Delta \Lambda = \sum_k \alpha_k \Sigma_k \).

It is clear that the real difficulty here is the computation of the integrals (19) and (20). This task requires extensive use of manifold results of linear stochastic systems theory. The details are carefully explained in [12].

**Theorem 4.** ([12]) The sequence \( \{\Lambda_i ; i \geq 0\} \) generated by the Newton algorithm of Section 5 converges to the unique minimum point of \( J_\Psi \) in \( L^H \).

In essence, the proof exploits some of the regularity properties of \( J_\Psi \) (Theorem 2) to apply a convergence result which is well-known in the literature on the Newton algorithm.

### 6 Application to spectrum estimation

#### 6.1 A spectral estimation procedure

Now we describe an application of the above approximation algorithm to the estimation of spectral densities. Consider for instance the scalar case, and suppose that the finite sequence \( y_1, ..., y_N \) is extracted from a realization of a zero-mean, weakly stationary discrete-time process \( \{y_t\}_{t=-\infty}^{+\infty} \). We want to estimate the spectral density \( \Phi_y(e^{j\theta}) \) of \( y \). The idea is the following:

- Fix a transfer function \( G(z) = (zI - A)^{-1}B \), feed the data \( \{y_i\} \) to it, and collect the output data \( \{x_i\} \).

- Compute a consistent, and possibly unbiased, estimate \( \tilde{\Sigma} \) of the covariance matrix of the outputs \( \{x_i\} \). Note that some output samples \( x_1, ..., x_M \) should be discarded so that the filter can be considered to operate in steady state.

- Choose as “prior” spectrum \( \hat{\Phi}_y \), a coarse, low-order, estimate of the true spectrum of \( y \) obtained by means of another (simple) identification method.

- “Refine” the estimate \( \hat{\Phi}_y \) by solving the approximation problem (6) with respect to \( G(z) \), \( \Sigma = \tilde{\Sigma} \), and \( \Psi = \hat{\Phi}_y \).

The result of the above procedure is the only spectrum, compatible with the output variance \( \tilde{\Sigma} \), which is closest to the rough estimate \( \hat{\Phi}_y \) in the \( d_H \) distance.

Note that we are left with significant degrees of freedom in applying the above procedure: The method for estimating \( \Phi_y \), in particular its degree, and the whole structure of \( G(z) = (zI - A)^{-1}B \), which has no constraints other than \( A \) being a stability matrix and \( (A, B) \) being reachable. The coarsest possible estimate of \( \Phi_y \) is the constant spectrum equal to the sample variance of the \( \{y_i\} \), i.e. \( \hat{\Phi}_y(e^{j\theta}) \equiv \hat{\sigma}_y^2 \).
where $\hat{\sigma}^2_y = \frac{1}{N-1} \sum_{i=1}^{N} |y_i|^2$. Another simple choice is $\hat{\Phi}_y = W(z)W^*(z)$, where $W(z)$ is a low-order AR, MA or ARMA model estimated from $y_1, ..., y_N$ by means of predictive error minimization methods or the like. The flexibility in the choice of $G(z)$ is more essential, and has both theoretical and practical implications. In general, as described in [2], the magnitude of the poles of $G(z)$ has implications on the variance of the sample covariance $\hat{\Sigma}$: The closer the eigenvalues to the origin, the smaller that variance. Moreover, at least as far as THREE is concerned, the phase of the eigenvalues influences resolution capability: More precisely, the spectrum estimation procedure has higher resolution in those sectors of the unit circle where more eigenvalues are located. According to simulations, the latter statement appears to be true also in our setting.

**Remark.** In the above setting $\hat{\Sigma}$ is a consistent estimate of the true steady-state variance. Although $\hat{\Sigma}$ must belong to $\text{Range } \Gamma$ in its ergodic limit as $N \to +\infty$, it is almost certainly not the case that $\hat{\Sigma} \in \text{Range } \Gamma$ when we have available only the finitely many data $x_{M+1}, ..., x_N$. Strictly speaking, this implies that the constraint (1) with $\Sigma = \hat{\Sigma}$ is almost always not feasible. The solution we implement is to enforce well-posedness by projecting $\hat{\Sigma}$ onto $\text{Range } \Gamma$ (and to check positivity of the result) before starting the optimization procedure, but other solutions can be adopted. This approach and the considerations on the positivity issue should be compared to [2, Section II.D], which deals with the particular case when $\text{Range } \Gamma$ is the space of Toeplitz matrices, and with [10], where this very issue is solved minimizing the Kullback-Leibler-Von Neumann distance of the sample variance from the space of Toeplitz matrices.

### 6.2 Simulation results

![Simulation results](image)

**Figure 1.** Estimation of an ARMA(6,4) spectrum by means of Hellinger-distance spectrum approximation, constant prior and AR(3) prior.

Figure 1 shows the results of the above estimation procedure with $G(z)$ structured...
according to the covariance extension setting (i.e. all poles in 0) with 6 covariance lags (i.e. \( n = 6 \), \( A \) is \( 6 \times 6 \)), run over 500 samples of the following ARMA process:

\[
y(t) = 0.5y(t - 1) - 0.42y(t - 2) + 0.602y(t - 3) - 0.0425y(t - 4) + 0.1192y(t - 5) + e(t) + 1.1e(t - 1) + 0.08e(t - 2) - 0.15e(t - 3)
\]

(poles in 0.9, \(-0.2 \pm 0.7j\), \(\pm 0.5j\)) where \( e(t) \) is a zero-mean Gaussian white noise with unit variance. Two priors, both estimated from data, have been considered: the constant spectrum \( \hat{\Phi}_y(\omega^0) \equiv \hat{\sigma}_y^2 \) and the spectrum \( \hat{\Phi}_y = W_{AR}(z)W^*_{AR}(z) \), where \( W_{AR}(z) = \frac{\hat{\sigma}_e}{\hat{\sigma}_y} \) is an AR model of order 3 obtained from the data by means of the predictive error method procedure in MATLAB’s System Identification toolbox. Figure 2 shows the performance of the above procedure in a setting that resembles that of [2, Section IV.B, Example 1]. The estimation procedure was run on 300 samples of a superposition of two sinusoids in colored noise:

\[
y(t) = 0.5 \sin(\omega_1 t + \phi_1) + 0.5 \sin(\omega_2 t + \phi_2) + z(t)
\]

\[
z(t) = 0.8z(t - 1) + 0.5\nu(t) + 0.25\nu(t - 1)
\]

with \( \phi_1, \phi_2 \) and \( \nu(t) \) independent normal random variables with zero mean and unit variance, \( \omega_1 = 0.42 \) and \( \omega_2 = 0.53 \). The prior here considered is the constant spectrum equal to the sample variance of the \( \{y_i\} \) data. Following [2], \( A \) was chosen.
real block-diagonal with the following poles (equispaced in a narrow range where the frequencies of the two sinusoids lie, to increase resolution in that region):

\[ 0, 0.85, -0.85, 0.9e^{\pm j0.42}, 0.9e^{\pm j0.44}, 0.9e^{\pm j0.46}, 0.9e^{\pm j0.48}, 0.9e^{\pm j0.50} \]

(and \( B \) a column of ones). It can be seen that Hellinger-distance based approximation does a good job, as does the THREE algorithm, at detecting the spectral lines at frequencies \( \omega_1 \) and \( \omega_2 \).

The reader can find in [12] an example of estimation of a multivariate spectrum compared with existing methods. Simulation in the multivariable case shows that, at the price of some moderate extra complexity in the model, our method may perform better than MATLAB’s PEM and MATLAB’s N4SID in the case of a short observation record.
Bibliography


