2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

A new metric for multivariate spectral estimation leading to lowest complexity spectra

Augusto Ferrante, Chiara Masiero and Michele Pavon

***Abstract*— A new multivariate spectral estimation technique is proposed. It is based on a constrained spectrum approximation problem, where the distance between spectra is derived from the relative entropy rate between stationary Gaussian processes. This approach may be viewed as an extension of the high-resolution estimator called THREE introduced by Byrnes, Georgiou and Lindquist in 2000. The corresponding solution features a complexity upper bound which is *equal* to the one featured by THREE *in the scalar case* thereby improving on the one so far available in the multichannel framework. The solution is computed by means of a globally convergent, matricial Newton-type algorithm. Comparative simulation indicates that this new technique outperforms PEM and N4SID in the case of short data records.**

I. Introduction

Multivariate spectral estimation is a classic and difficult problem [13], [17] still generating considerable interest in the engineering sciences, see e.g. [8], [10], [16], [15]. In [2], [7], a new approach to scalar spectral estimation, called THREE, was introduced by Byrnes, Georgiou and Lindquist. It employs the output covariance of a bank of filters, representing measurement devices, to extract information on the input power spectrum and may be viewed as a (considerable) generalization of classical Burg-like maximum entropy methods. Its main features are high resolution in prescribed frequency ranges and good performances in the case of short observation records. A first attempt of generalization to the multichannel situation was made in [15], where a non entropy-like distance was employed in the optimization part of the procedure. The resulting solution, however, had higher McMillan degree than in the original scalar THREE method.

The purpose of this paper is to introduce a more natural multivariate extension of the THREE method, which hinges on the choice of a new distance index for the optimization part. Such a choice has profound motivations, as detailed in our journal paper [4]. There, the proposed metric is shown to originate from the *relative entropy rate* of two stationary Gaussian processes. Hence, in the following the new approach will be referred to as RER (Relative Entropy Rate) estimator.

Work partially supported by the Italian Ministry for Education and Research (MIUR) under PRIN grant n. 20085FFJ2Z "New Algorithms and Applications of System Identification and Adaptive Control"

A. Ferrante and C. Masiero are with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy augusto@dei.unipd.it, masiero.chiara@dei.unipd.it

M. Pavon is with the Dipartimento di Matematica Pura ed Applicata, Università di Padova, via Trieste 63, 35131 Padova, Italy pavon@math.unipd.it

This method features the same complexity upper bound as THREE in the scalar case, considerably improving on the one so far available in the multivariable setting. As for the previous THREE-like methods, RER also exhibits high resolution features and works extremely well in the case of short observation records. Indeed, scanty but encouraging simulation tests show that in such case it may outperform classical identification methods like PEM and N4SID.

The paper is outlined as follows. Section II describes THREE-like spectral estimation methods. In Section III we introduce the new approach RER via a convex optimization problem and we derive the form of the optimal spectral estimate. In Section IV, we show that there exists a unique solution for the corresponding dual problem. Such a solution is computed, in Section V, by means of a globally convergent, matricial Newton-type algorithm. Finally, in Section VI, the RER estimator performance is compared via simulation to that of the previously available methods.

II. THREE-like spectral estimation and generalized moment problems

We now briefly introduce our spectral estimation framework. Let $\mathcal{S}_+^{m\times m}(\mathbb{T})$ be the set of the bounded and coercive spectral density functions with values in $\mathcal{H}_+(m)$, the cone of positive definite, Hermitian $\mathbb{C}^{m\times m}$-valued matrices. Suppose that the data $\{y_i\}_{i=1}^N$ are generated by a zero-mean, $m-$dimensional, wide-sense stationary and purely nondeterministic process $y = \{y_k;\, k \in \mathbb{Z}\}$. Our purpose is to estimate the spectral density $\Phi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$ of y from $\{y_i\}_{i=1}^N$. We draw inspiration from THREE-like approaches [2], [11], [5], [15], which generalize Burg-like methods in several ways. Indeed, the second order statistics that are estimated from the data $\{y_i\}_{i=1}^N$ are not necessarily the covariance lags $C_l := E\{y_{k+l}y_k^*\}$ of y (here $*$ denotes transposition plus conjugation). Moreover, a prior estimate of Φ may be included in the estimation procedure. More explicitly, these methods are based on four pivotal elements:

1) A rational filter to process the data. The filter has transfer function
$$G(z) = (zI - A)^{-1}B, \tag{1}$$
where $A \in \mathbb{C}^{n\times n}$ is a stability matrix (i.e. it has all its eigenvalues inside the unit circle), $B \in \mathbb{C}^{n\times m}$ is full rank, $n \geq m$, and (A, B) is a reachable pair;
2) An estimate, based on the data $\{y_i\}_{i=1}^N$, of the steady-state covariance Σ of the state $x(k)$ of the filter
$$x(k+1) = Ax(k) + By(k); \tag{2}$$

3) A *prior* spectral density $\Psi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$;
4) An index that measures the distance between two spectral densities.

By (2), the spectrum of y must satisfy the constraint

$$\int G\Phi G^* = \Sigma. \tag{3}$$

Here and throughout the paper integration is on the unit circle with respect to normalized Lebesgue measure. Constraint (3) provides Carathèodory or, more generally, Nevanlinna-Pick interpolation data for the positive real function $Z(z)$ associated to Φ (see [2, Section II]). As for the spectral density Ψ, it allows to take into account possible *a priori* information on Φ. For example, Ψ may simply be a coarse estimate of the true spectrum.[1] Since, in general, Ψ is not consistent with the interpolation conditions, an approximation problem arises. Then, it is necessary to introduce an adequate distance index. This choice is of capital importance and should be made by considering several requirements. On the one hand, the solution should be rational of low McMillan degree at least when the prior Ψ is such. On the other hand, the variational analysis should lead to a computable solution. In the scalar case [2], [11], the choice was to minimize the Kullback-Leibler type criterion

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

which satisfies both of the above requirements. In the multivariable case, the Kullback-Leibler pseudo-distance may also be defined [10], on the basis of the *Umegaki-von Neumann's relative entropy* of statistical quantum mechanics [14]. However, the resulting spectrum approximation problem does not lead, in general, to computable solutions of bounded McMillan degree. On the contrary, with the following multivariate extension of the Hellinger distance introduced in [5]

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

$$\text{such that } W_\Psi W_\Psi^* = \Psi \text{ and } W_\Phi W_\Phi^* = \Phi,$$

which turns out to be a *bona fide* distance, the variational analysis can be carried out leading to a computable solution. An effective multivariate THREE-like spectral estimation method can therefore be based on such a distance, which leads to rational solutions when the prior Ψ is rational [15]. However, there is a drawback: The complexity of the solution is usually noticeably higher than in the original scalar THREE approach.

In this paper we propose a new distance index, which allows us to carry out the variational analysis explicitly. Moreover, such a choice entails an upper bound on the complexity of the solution which is equal to that achieved in the scalar setting by the original THREE method.

[1]When no prior information on Φ is available, Ψ is set equal either to the identity or to the sample covariance of the available data $\{y_i\}_{i=1}^N$.

III. A NEW METRIC FOR MULTIVARIATE SPECTRAL ESTIMATION

Let us briefly recall the concept of *relative entropy*, also known as *Kullback-Liebler* pseudo-distance or *divergence*, see e.g [3]. Given two probability densities p and q, it is defined by

$$\mathbb{D}(p\|q) = \int_{\mathbb{R}^n} p(x)\log\frac{p(x)}{q(x)}dx, \tag{4}$$

provided the support of p is contained in the support of q. In the case of two zero-mean Gaussian probability density functions p and q with covariance matrices P and Q, respectively, the relative entropy depends only on the covariance matrices as follows:

$$\mathbb{D}(p\|q) = \frac{1}{2}\left[\log\det(P^{-1}Q) + \operatorname{tr}(Q^{-1}P) - n\right]. \tag{5}$$

Motivated by (5), we introduce a new pseudo-distance between spectra in $\mathcal{S}_+^{m\times m}(\mathbb{T})$:

$$d_{RER}(\Phi,\Psi) := \frac{1}{2\pi}\int_{\mathbb{T}} \frac{1}{2}\left\{\log\det\left(\Phi^{-1}(e^{j\vartheta})\Psi(e^{j\vartheta})\right) + \operatorname{tr}\left[\Psi^{-1}(e^{j\vartheta})\left(\Phi(e^{j\vartheta}) - \Psi(e^{j\vartheta})\right)\right]\right\} d\vartheta. \tag{6}$$

In [4], strong motivation for employing (6) is provided by exhibiting a cogent relation between this index and time and spectral domain relative entropy rates.

We now address the following *Approximation problem*

Problem 1: Let $\Psi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$, $G(z)$ as in (1) and $\Sigma \in \mathcal{H}_+(n)$. Find Φ° that solves:

$$\text{minimize } d_{RER}(\Phi,\Psi)$$
$$\text{over } \left\{\Phi \in \mathcal{S}_+^{m\times m}(\mathbb{T}) \middle| \int G\Phi G^* = \Sigma\right\}.$$

Remark 3.1: Notice that we could also think of minimizing the distance index (6) with respect to the second argument. However, it is possible to prove that such a choice, which is related to minimum prediction error estimation [12], usually leads to a non rational approximant, even when the prior Ψ is rational. Therefore, this approach is not suitable for our purposes.

First, feasibility of Problem 1, i.e. existence of $\Phi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$ satisfying (3), needs to be considered. To this aim, let $\Gamma : C(\mathbb{T},\mathcal{H}(m)) \to \mathcal{H}(n)$ be the linear operator

$$\Gamma(\Phi) := \int G\Phi G^*, \tag{7}$$

where $C(\mathbb{T},\mathcal{H}(m))$ denotes the set of $\mathcal{H}(m)$-valued continuous functions defined on the unit circle $\mathbb{T}$. The following result was established by Georgiou in [9], see also [15].

Theorem 3.1: Consider $\Sigma \in \mathcal{H}(n)$ and a system described by (2), where A is (asymptotically) stable, B is full column rank and (A,B) is a reachable pair. Then:

1) Σ is in $\operatorname{Range}(\Gamma)$ if and only if there exists $H \in \mathbb{C}^{m\times n}$ such that
$$\Sigma - A\Sigma A^* = BH + H^*B^*. \tag{8}$$
2) Let Σ be positive definite. Then, there exists $H \in \mathbb{C}^{m\times n}$ that solves (8) if and only if there exists $\Phi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$ such that $\Gamma(\Phi) = \Sigma$.

Henceforth, we assume feasibility of Problem 1. Moreover, to simplify the exposition, we also assume $\Sigma = I$. Solving Problem 1, is equivalent to minimizing over $\mathcal{S}_+^{m\times m}(\mathbb{T})$

$$2d_{RER}(\Phi,\Psi) + m = \int \left\{\log\det\left(\Phi^{-1}\Psi\right) + \operatorname{tr}\left(\Psi^{-1}\Phi\right)\right\}.$$

subject to (3). Recall that the inner product in $\mathcal{H}(n)$ is defined by $\langle M, N\rangle = \operatorname{tr}[MN]$. Thus, we consider the Lagrangian

$$\begin{aligned} L_\Psi(\Phi,\Lambda) &= 2d_{RER}(\Phi,\Psi) + m + \langle\Lambda, \int G\Phi G^* - \Sigma\rangle \\ &= \int\left[\log\frac{\det(\Psi)}{\det(\Phi)} + \operatorname{tr}(\Psi^{-1}\Phi) + \operatorname{tr}(\Lambda G\Phi G^*)\right] - \operatorname{tr}\Lambda, \end{aligned} \tag{9}$$

where the Lagrange parameter $\Lambda \in \mathcal{H}(n)$. Each $\Lambda \in \mathcal{H}(n)$ can be uniquely decomposed as $\Lambda = \Lambda_\Gamma \oplus \Lambda_\perp$, where $\Lambda_\Gamma \in \operatorname{Range}(\Gamma)$ and $\Lambda_\perp \in (\operatorname{Range}(\Gamma))^\perp$. It can be proven that a term $\Lambda_\perp \in (\operatorname{Range}(\Gamma))^\perp$ gives no contribution to the Lagrangian (9). Therefore, we assume from now on that the Lagrange parameter Λ belongs to $\operatorname{Range}(\Gamma)$.

Next, we concentrate on the *unconstrained* minimization of the functional (9). Since $L_\Psi(\cdot,\cdot)$ in (9) is strictly convex in Φ and has continuous directional (*Gateaux*) derivatives in any direction $C(\mathbb{T},\mathcal{H}(m))$, the unconstrained minimization is realized by imposing the first variation to be zero in each direction. Recall that, for $X > 0$,

$$\delta\log\det(X;\delta X) = \operatorname{tr}(X^{-1}\delta X). \tag{10}$$

Then, it is possible to compute the first variation of the Lagrangian:

$$\begin{aligned} \delta L(\Phi,\Lambda;\delta\Phi) = \int \big[&-\operatorname{tr}(\Phi^{-1}\delta\Phi) + \operatorname{tr}(\Psi^{-1}\delta\Phi) \\ &+ \operatorname{tr}(G^*\Lambda G\delta\Phi)\big]. \end{aligned} \tag{11}$$

Therefore, we can conclude that (11) is zero $\forall\,\delta\Phi \in \mathcal{C}(\mathbb{T},\mathcal{H}(m))$ if and only if

$$\Phi = \Phi^\circ := \left[\Psi^{-1} + G^*\Lambda G\right]^{-1}. \tag{12}$$

Let W_Ψ be the stable and minimum phase spectral factor of Ψ,[2] and $G_1(e^{\mathrm{j}\vartheta})$ be defined by

$$G_1(e^{\mathrm{j}\vartheta}) := G(e^{\mathrm{j}\vartheta})W_\Psi(e^{\mathrm{j}\vartheta}). \tag{13}$$

Later, it will be useful to consider also the alternative form of (12)

$$\Phi^\circ = W_\Psi(I + G_1^*\Lambda G_1)^{-1}W_\Psi^*. \tag{14}$$

Remark 3.2: Notice that expression (12) yields an upper bound on the complexity of the approximant, whose maximum McMillan degree amounts to $\deg\Psi + 2n$, where n is the dimension of each minimal realization of $G(z)$. This result represents a significant improvement in the frame of multivariable spectral estimation. Indeed, the best so far available upper bound on the complexity of the solution was $\deg\Psi + 4n$ (see [5]).

[2]Since $\Psi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$, W_Ψ exists. It is unique up to multiplication on the right by a constant unitary matrix.

Since Φ° is required to be a bounded spectral density, we need, as indicated by (14), to restrict the Lagrange multiplier Λ to the subset $\mathcal{L}_+$, defined by

$$\mathcal{L}_+ := \{\Lambda \in \mathcal{H}(n) \,|\, I + G_1^*\Lambda G_1 > 0 \ \text{ a.e. on } \mathbb{T}\}. \tag{15}$$

Thus, the natural set for Λ is

$$\mathcal{L}_+^\Gamma := \mathcal{L}_+ \cap \operatorname{Range}(\Gamma). \tag{16}$$

To sum up, for each $\Lambda \in \mathcal{L}_+^\Gamma$ there exists a unique $\Phi^\circ \in \mathcal{S}_+^{m\times m}(\mathbb{T})$ that minimizes the Lagrangian functional. It has the form (12). If we produce a Λ° s.t. $\Phi^\circ(\Lambda^\circ)$ satisfies the constraint (3), then such a $\Phi^\circ(\Lambda^\circ)$ is the solution of Problem 1. To this aim we resort to duality.

IV. The dual problem

In order to tackle the dual problem, we equivalently Solving the dual problem is equivalent to minimizing the following functional, henceforth referred to as the *dual functional*:

$$\begin{aligned} J_\Psi(\Lambda) &:= -L(\Phi^\circ(\Lambda),\Lambda) + n \\ &= \int\left[\operatorname{tr}\Lambda - \log\det(I + G_1^*\Lambda G_1)\right]. \end{aligned} \tag{17}$$

Given $\delta\Lambda \in \mathcal{H}(n)$, by means of (11) we have that the first variation of J_Ψ is

$$\delta J_\Psi(\Lambda;\delta\Lambda) = \int\left\{\operatorname{tr}\left[\delta\Lambda\right] - \operatorname{tr}\left[(I + G_1^*\Lambda G_1)^{-1}G_1^*\delta\Lambda G_1\right]\right\} \tag{18}$$

As for the second variation: Let us denote the matrix inversion operator by $R : M \mapsto M^{-1}$ and recall that its first derivative in direction δM is given by $\delta R(M,\delta M) = -M^{-1}\delta M M^{-1}$. Then, for $\delta\Lambda_1$ and $\delta\Lambda_2$ in $\mathcal{H}(n)$, we have

$$\begin{aligned} \delta^2 J_\Psi(\Lambda;\delta\Lambda_1,\delta\Lambda_2) = \int \operatorname{tr}\Big[&(I + G_1^*\Lambda G_1)^{-1}G_1^*\delta\Lambda_2 G_1 \\ &\times(I + G_1^*\Lambda G_1)^{-1}G_1^*\delta\Lambda_1 G_1\Big]. \end{aligned} \tag{19}$$

The bilinear form $H_\Lambda(\cdot,\cdot) := \delta^2 J_\Psi(\Lambda;\cdot,\cdot)$ is the *Hessian* of J_Ψ at Λ.

The next and most delicate issue is to prove that, although the set $\mathcal{L}_+^\Gamma$ is open and unbounded, a (unique) Λ° minimizing J_Ψ over $\mathcal{L}_+^\Gamma$ *does* exist.

Theorem 4.1: The dual functional $J_\Psi(\Lambda)$ belongs to $\mathcal{C}^2(\mathcal{L}_+^\Gamma)$ and is *strictly* convex on $\mathcal{L}_+^\Gamma$. If the feasibility condition (8) holds, $J_\Psi(\Lambda)$ admits a unique minimum point Λ° in $\mathcal{L}_+^\Gamma$.

We refer the reader for both proofs to [4] for the details. Here, we just provide a sketch of the proof. As for uniqueness, it can be proven by showing that the Hessian is positive definite on $\mathcal{L}_+^\Gamma$. The proof of the existence, however, is quite nontrivial. First, we define a new function J_Ψ^∞, so that it is well defined on the closure of the set $\mathcal{L}_+^\Gamma$. Given that, for Λ belonging to the boundary $\partial\mathcal{L}_+^\Gamma$, the Hermitian matrix

$I + G_1^* \Lambda G_1$ is singular, in at least one point of $\mathbb{T}$, it is useful to introduce the following sequence of functions on $\overline{\mathcal{L}_+^\Gamma}$:

$$J_\Psi^n(\Lambda) = \int \operatorname{tr}\left[\Lambda - \log\left(I + G_1^*\Lambda G_1 + \frac{1}{n}I\right)\right], \quad n \geq 1. \tag{20}$$

Define now $J_\Psi^\infty(\Lambda)$ by considering the pointwise limit

$$J_\Psi^\infty(\Lambda) := \lim_{n\to\infty} J_\Psi^n(\Lambda). \tag{21}$$

It can be proven that such limit exists and coincides with the dual function (17) on $\mathcal{L}_+^\Gamma$. Moreover, it defines a bounded-below, convex, lower-semicontinuous and inf-compact function[3]. Therefore, by Weierstrass' Theorem, we can conclude that J_Ψ^∞ admits a minimum point in $\Lambda^\circ \in \overline{\mathcal{L}_+^\Gamma}$. Actually, more can be established: Such a minimum point belongs to $\mathcal{L}_+^\Gamma$.

In conclusion, there exists a unique $\Lambda^\circ \in \mathcal{L}_+^\Gamma$ minimizing $J_\Psi(\Lambda)$, as defined in (17). Such a Λ° annihilates the directional derivative (18) in any direction $\delta\Lambda \in \mathcal{H}(n)$. This means that the corresponding spectral density $\Phi^\circ(\Lambda^\circ) = \left[\Psi^{-1} + G^*\Lambda^\circ G\right]^{-1}$, satisfies constraint (3) and is therefore the unique solution of Problem 1. We are left with the problem of developing an efficient numerical algorithm to compute Λ°.

V. An efficient matricial Newton-like algorithm for multivariate spectral estimation

In order to compute the minimizer of the dual functional $J_\Psi(\Lambda)$, a matricial Newton-type algorithm is proposed, based on two steps:

1) Find the *search direction* $\Delta\Lambda_i$;
2) Compute the Newton *step length* t_i^k.

The starting point for the minimizing sequence $\{\Lambda_i\}_{i\in\mathbb{N}}$ can be taken to be $\Lambda_0 = 0$.

A. Search Direction

Notice that this step is rather delicate: Even though the problem is finite-dimensional, because a matricial expression of the Hessian and the gradient allowing to compute the search direction as $\Delta x = -{H_x}^{-1}\nabla f_x$ is not available. Indeed, in order to find the Newton step $\Delta\Lambda_i$, given $\Lambda_i \in \mathcal{L}_+^\Gamma$, one has to solve, for the unknown $\Delta\Lambda_i$, the equation $H_{\Lambda_i}(\Delta\Lambda_i, \cdot) = -\nabla J_{\Psi,\Lambda_i}(\cdot)$, which can be explicitly written as:

$$\int G_1(I + G_1^*\Lambda_i G_1)^{-1} G_1^* \Delta\Lambda_i G_1 (I + G_1^*\Lambda_i G_1)^{-1} G_1^* = \int G_1(I + G_1^*\Lambda_i G_1)^{-1} G_1^* - I.$$

To this aim, consider a basis of $\operatorname{Range}(\Gamma)$. Note that it can be readily obtained[4], by recalling that $\Sigma_k \in \operatorname{Range}(\Gamma)$ if and only if $\exists\, H_k \in \mathbb{C}^{m\times n}$ s.t. $\Sigma_k - A\Sigma_k A^* = BH_k + {H_k}^* B^*$.

[3]Recall that a function $f : \mathbb{C}^{n\times n} \to \mathbb{R}$ is said to be lower semicontinuous at x_0 if, $\forall\, \varepsilon > 0$, there exists a neighborhood U of x_0 such that, $\forall\, x \in U$, $f(x) \geq f(x_0) - \varepsilon$. Recall also that f is inf-compact if $\forall\, \alpha \in \mathbb{R}$, the set $\left\{x \in \mathbb{C}^{n\times n} \mid f(x) \leq \alpha\right\}$ is compact.

[4]Indeed, following the lines detailed in [6] it is possible to obtain *directly* a basis of $\operatorname{Range}(\Gamma)$ by solving only N Lyapunov equations.

Moreover, under the assumption that $I \in \operatorname{Range}\Gamma$ we can always get a basis $\{\Sigma_1, \ldots, \Sigma_N\}$ of $\operatorname{Range}(\Gamma)$ made of positive definite matrices. This result simplifies the following procedure:

1) Compute
$$Y = \int G_1(I + G_1^*\Lambda_i G_1)^{-1} G_1^* - I \tag{22}$$
2) For each generator Σ_k, compute
$$Y_k = \int G_1(I + G_1^*\Lambda_i G_1)^{-1} G_1^* \Sigma_k G_1 \times (I + G_1^*\Lambda_i G_1)^{-1} G_1^* \tag{23}$$
3) Find $\{\alpha_k\}$ s.t. $Y = \sum_k \alpha_k Y_k$;
4) Set $\Delta\Lambda_i = \sum_k \alpha_k \Sigma_k$.

Notice that the most challenging step, that is the computation of Y and Y_k, can be tackled efficiently by employing spectral factorization techniques in order to compute the integrals, along the same lines described in [15, Section VI]. Indeed, the integrand that appears in equation (22) is a coercive spectral density and the same holds for the integrand in (23), since we have chosen positive definite generators Σ_i. For the computation of Y, let us focus on $Q_{\Lambda_i}(z) = I + G_1^*(z)\Lambda_i G_1(z)$. Assume $W_\Psi(z)$ to be a realization of the stable minimum phase spectral factor of Ψ. Then, we can easily obtain a state-space realization $G_1(z) = C_1(zI - A_1)^{-1}B_1$ of (13). Since $\Lambda_i \in \mathcal{L}_+^\Gamma$, $Q_{\Lambda_i}(z)$ is positive definite on $\mathbb{T}$, so that the following ARE admits a positive definite stabilizing solution $P = P^* > 0$ (see, e.g. Lemma 6.4 in [15]):

$$P = A_1^* P A_1 - A_1^* P B_1 (B_1^* P B_1 + I)^{-1} B_1^* P A_1 + C_1^* \Lambda_i C_1. \tag{24}$$

Moreover, $Q_{\Lambda_i}(z)$ can be factorized as $Q_{\Lambda_i}(z) = \Delta_{\Lambda_i}^*(z)\Delta_{\Lambda_i}(z)$, where $\Delta_{\Lambda_i}(z)$ can be explicitly written in term of the stabilizing solution P:

$$\Delta_{\Lambda_i}(z) = (B_1^* P B_1 + I)^{-\frac{1}{2}} B_1^* P A_1 (zI - A_1)^{-1} B_1 + (B_1^* P B_1 + I)^{\frac{1}{2}}. \tag{25}$$

It is now easy to compute a state space realization of $\Delta_{\Lambda_i}^{-1}$ and then of the stable filter $W_Y := G_1\Delta_{\Lambda_i}^{-1} = C_1(zI - Z_1)^{-1}B_1(B_1^*PB_1 + I)^{-\frac{1}{2}}$, with $Z_1 := A_1 - B_1(B_1^*PB_1 + I)^{-1}B_1^*PA_1$ being the closed-loop matrix. The computation of (22) is now immediate. In fact,

$$Y + I = \int G_1(I + G_1^*\Lambda_i G_1)^{-1} G_1^* = \int G_1 \Delta_{\Lambda_i}^{-1} \Delta_{\Lambda_i}^{-*} G_1^* = \int W_Y W_Y^*. \tag{26}$$

Then, the latter integral can be evaluated by computing the unique solution of the Lyapunov equation $R - Z_1 R Z_1^* = B_1(B_1^*PB_1 + I)^{-1}B_1^*$ and setting $Y + I = C_1 R C_1^*$, so that

$$Y = C_1 R C_1^* - I. \tag{27}$$

A similar procedure may be employed to compute also the matrices Y_k.

B. Step length

The backtracking line search is implemented by halving the step t_i until both the following conditions are satisfied:

$$\Lambda_i + t_i^k \Delta\Lambda_i \in \mathcal{L}_+^\Gamma; \tag{28}$$

$$J_\Psi(\Lambda_i + t_i^k \Delta\Lambda_i) < J_\Psi(\Lambda_i) + \alpha t_i^k \nabla J_{\Psi,\Lambda_i} \Delta\Lambda_i \tag{29}$$

where $0 < \alpha < 0.5$. The first condition can be easily evaluated by testing whether $Q_{\Lambda_i + t_i^k \Delta\Lambda_i}$ admits a factorization of the kind introduced in the previous subsection or, equivalently, whether the corresponding ARE (24) admits a solution $P = P^* > 0$.

The only difficulty in checking the second condition is in computing

$$J_\Psi(\Lambda) = \operatorname{tr} \int [\Lambda - \log(I + G_1^* \Lambda G_1)] = \operatorname{tr} \Lambda - \int \log\det(I + G_1^* \Lambda G_1). \tag{30}$$

The latter integral can be evaluated in a straightforward way in the light of Wiener-Masani Theorem. It states that, if $\Xi(\mathrm{e}^{\mathrm{j}\vartheta})$ is the spectral density of a process z,

$$\exp\left\{ \frac{1}{2\pi} \int_{\mathbb{T}} \log\det[\Xi(\mathrm{e}^{\mathrm{j}\vartheta})] d\vartheta \right\} = \det R, \tag{31}$$

where R is the error covariance matrix corresponding to the one-step-ahead optimal predictor of z. In our case $Q_\Lambda(z)$ may be factorized as $Q_\Lambda = \Delta^* \Delta$, where Δ is a stable and minimum phase filter for which a minimal realization can be computed as in the previous section (see (25)). Since $\log\det Q_\Lambda = \log\det[\Delta^*\Delta] = \log\det[\Delta\Delta^*]$, $\det R$ is given by $\det[\Delta(\infty)\Delta^*(\infty)]$ which may be explicitly written in terms the solution P of the corresponding ARE as $\det[B_1^* P B_1 + I]$. Therefore,

$$\int \log\det(I + G_1^* \Lambda G_1) = \log\det\left(B_1^* P B_1 + I\right).$$

C. Convergence of the Proposed Algorithm

A sufficient condition for global convergence of the algorithm is that the following requirements are simultaneously satisfied [1, Chapter 9]:

1) $J_\Psi(\cdot)$ is twice continuously differentiable;
2) The starting point Λ_0 belongs to $\mathcal{L}_+^\Gamma$ and the sublevel set $S := \left\{ \Lambda \in \mathcal{L}_+^\Gamma | J_\Psi(\Lambda) \le J_\Psi(\Lambda_0) \right\}$ is closed;
3) The Hessian is Lipschitz continuous in S;
4) $J_\Psi(\cdot)$ is *strongly* convex, i.e. $\exists\ m$ s.t. $H_\Lambda > mI$, $\forall\ \Lambda \in S$.

In this case, after a certain number of iterations, the backtracking line search always selects the full step and the rate of convergence is quadratic, i.e. there exists a constant C such that $\|\Lambda_{i+1} - \Lambda^\circ\| \le C \|\Lambda_i - \Lambda^\circ\|^2$. It is possible to show that the above sufficient conditions for global convergence hold. While we refer the reader to [4] for a complete proof, we outline below the essential steps. The first requirement is satisfied in view of Theorem 4.1. As for the second one, it follows on the fact that the sublevel sets of the dual function J_Ψ are compact. Finally, by Weierstrass' theorem, it is possible to readily establish strong convexity of J_Ψ and Lipschitz continuity of the Hessian, so also the third and the fourth requirements are satisfied.

Thus, global convergence of the Newton algorithm is guaranteed. This result suggests the effectiveness of the proposed procedure in providing the solution of the spectral estimation Problem 1.

VI. Simulation Results

We now employ our results in a spectral estimation procedure, that may be outlined as follows.

1) Let $\{y_1, \ldots, y_N\}$ be a finite sequence extracted from a realization of the zero-mean Gaussian process $y = \{y_k; k \in \mathbb{Z}\}$, whose spectrum is $\Phi(\mathrm{e}^{\mathrm{j}\vartheta})$.
2) Design a filter $G(z)$, as described by equation (2).
3) Feed the filter with the data sequence $\{y_1, \ldots, y_N\}$, collect the output data x_i and compute a consistent estimate $\hat{\Sigma}$ of the covariance matrix.
4) In general the estimate $\hat{\Sigma}$ does not satisfy the conditions stated in Theorem 3.1. In order to guarantee feasibility, we compute a suitable matrix $\tilde{\Sigma}$ by solving an ancillary optimization problem as in [6].
5) Introduce a prior spectral density Ψ.
6) Tackle Problem 1 by means of the proposed algorithm, by fixing Ψ and $\Sigma = \tilde{\Sigma}$.

Notice that our approach provides two degrees of freedom: the choice of the prior Ψ and the design of the filter $G(z)$. In the planning stage it is important to consider their effect on the upper bound on the complexity of the approximant, as explained in Section III.

The performances of the proposed method in the multivariate setting were analyzed by considering the same estimation task as that described in [15, Section VIII.C]. The process y was obtained by filtering a bivariate Gaussian white noise process with zero mean and variance I through a square shaping filter of order 40. The filter coefficients were chosen at random, except for one fixed complex poles pair, $0.9\mathrm{e}^{\pm\mathrm{j}0.52}$ and the zeros pair $(1 - 10^{-5})\mathrm{e}^{\pm\mathrm{j}0.2}$.

We designed the filter $G(z)$ by fixing four complex poles pairs with radius 0.7 and arguments equispaced in the range $[0, \pi]$. We assumed $N = 300$ samples of the process $\{y_k\}_{k\in\mathbb{Z}}$ to be available. As for the the prior, our choice was to compute a simple PEM model of order 3, by means of the standard function `pem` provided in MATLAB's System and Identification toolbox. We compared the performance of the proposed technique to those achieved by Maximum Entropy [8] and Hellinger-distance estimators [5], which are both THREE-like approaches to multivariate spectral estimation. Aiming at making the comparison as independent as possible of the specific data set, we performed 50 trials by feeding the shaping filter with independent realizations of the input noise process $e = \{e_k; k \in \mathbb{Z}\}$. The performances of each method were evaluated by considering the average estimation error at each frequency, defined as

$$E_M(\vartheta) := \frac{1}{50} \sum_{i=1}^{50} \|\hat{\Phi}_M(\mathrm{e}^{\mathrm{j}\vartheta}) - \Phi(\mathrm{e}^{\mathrm{j}\vartheta})\|. \tag{32}$$

Here M denotes the specific algorithm and $\hat{\Phi}_M$ is the corresponding approximant. Fig. 1 allows to compare the various techniques. Our approach seems to outperform the Maximum Entropy estimator (henceforth referred to as ME). Moreover, the results achieved by Relative Entropy Rate method are slightly better than those of the Hellinger-distance approach. As for the order of the estimates, it was 19 in the Hellinger case, while it was just 11 in the case of RER. The order of the ME estimates was 8.

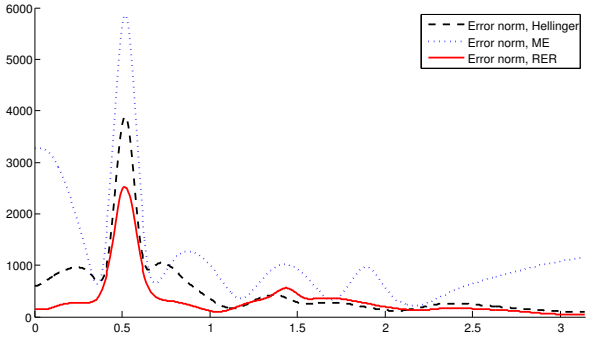

Fig. 1. Comparison of THREE-like approaches, in terms of average estimation error.

It is worthwhile to investigate what happens when only a few data of the process of interest are provided. The RER method exhibits a property that seems to characterize THREE-like approaches: It is quite robust with respect to artifacts. Such a problem, which is due to the shortness of the available data record, can heavily affect the estimates obtained by means of standard techniques such as MATLAB's PEM and MATLAB's N4SID. Fig. 2 shows the results that were obtained in a case where only $N = 100$ samples are available. Both PEM and N4SID estimates were affected by artifacts. On the contrary, the proposed approach was not. This result seems to suggest that RER estimation is suitable to tackle spectral estimation problems characterized by the presence of short data records.

VII. Conclusion

In this paper, a new THREE-like approach to multivariate spectral estimation, called RER, has been introduced. The new distance in the optimization part leads to an upper bound on the complexity of the estimate which is equal to the one achieved by THREE in the scalar context. Hence, it sensibly improves on the best one so far available in the multichannel setting with prior estimate. An efficient globally convergent Newton-type matricial algorithm has been designed and tested. Simulations suggest the effectiveness of the proposed approach in tackling multivariate spectral estimation tasks. In particular, as it is the case for THREE-like methods, RER seems to work extremely well with short observation records, often outperforming MATLAB's PEM and MATLAB's N4SID. We surmise that RER estimation may become a standard in multivariate, THREE-like spectral estimation.

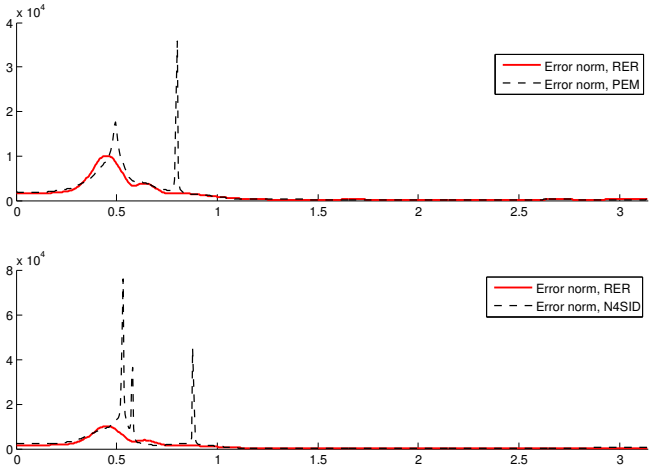

Fig. 2. Comparison of RER, PEM and N4SID in terms of average estimation error. RER is provided with a PEM(2) prior. The filter $G(z)$ has a pole in the origin and four complex conjugate poles pairs with radius 0.7.

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

[2] C. I. Byrnes, T. Georgiou, and A. Lindquist. A new approach to spectral estimation: A tunable high-resolution spectral estimator. *IEEE Trans. Sig. Proc.*, 49:3189–3205, 2000.

[3] T. M. Cover and J. A. Thomas. *Information Theory*. Wiley, New York, 1991.

[4] A. Ferrante, C. Masiero, and M. Pavon. Time and spectral domain relative entropy: A new approach to multivariate spectral estimation. March 2011. Submitted, preprint arXiv:1103.5602v1.

[5] A. Ferrante, M. Pavon, and F. Ramponi. Hellinger vs. Kullback-Leibler multivariable spectrum approximation. *IEEE Trans. Aut. Control*, 53:954–967, 2008.

[6] A. Ferrante, M. Pavon, and M. Zorzi. A maximum entropy enhancement for a family of high-resolution spectral estimators. *IEEE Trans. Aut. Control*, June 2010 (to appear).

[7] T. Georgiou. Spectral estimation by selective harmonic amplification. *IEEE Trans. Aut. Control*, 46:29–42, 2001.

[8] T. Georgiou. Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parameterization. *IEEE Trans. Aut. Control*, 47:1811–1823, 2002.

[9] T. Georgiou. The structure of state covariances and its relation to the power spectrum of the input. *IEEE Trans. Aut. Control*, 47:1056–1066, 2002.

[10] T. Georgiou. Relative entropy and the multivariable multidimensional moment problem. *IEEE Trans. Inform. Theory*, 52:1052–1066, 2006.

[11] T. Georgiou and A. Lindquist. Kullback-Leibler approximation of spectral density functions. *IEEE Trans. Inform. Theory*, 49:2910–2917, 2003.

[12] A. Lindquist. Prediction-error approximation by convex optimization. In A. Chiuso, A. Ferrante, and S. Pinzoni, editors, *Modeling, Estimation and Control: Festschrift in honor of Giorgio Picci on the occasion of his sixty-fifth birthday*, pages 265–275. Springer-Verlag, 2007.

[13] J. H. McClellan. Multidimensional spectral estimation. *Proc. IEEE*, 70:1029–1039, 1982.

[14] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge Univ. Press, Cambridge, UK, 2000.

[15] F. Ramponi, A. Ferrante, and M. Pavon. A globally convergent matricial algorithm for multivariate spectral estimation. *IEEE Transactions on Automatic Control*, 54(10):2376–2388, Oct. 2009.

[16] O. Rosen and D. Stoffer. Automatic estimation of multivariate spectra via smoothing splines. *Biometrika*, 94:335–345, 2007.

[17] P. Stoica and R. Moses. *Introduction to Spectral Analysis*. Prentice Hall, New York, 1997.