Multivariate Itakura-Saito distance for spectral estimation: Relation between time and spectral domain relative entropy rates*

Augusto Ferrante, Chiara Masiero

Department of Information Engineering, via Gradenigo 6/B, University of Padova, 35131
Padova, Italy

E-mail: augusto@dei.unipd.it, masieroc@dei.unipd.it

Michele Pavon

Department of Mathematics, via Trieste 63, University of Padova, 35131 Padova, Italy

E-mail: pavon@math.unipd.it

Abstract

The notion of *spectral relative entropy rate* is defined for jointly stationary Gaussian processes. Using classical information-theoretic results, we establish a remarkable connection between time and spectral domain relative entropy rates, and therefore with a multivariate version of the classical *Itakura-Saito divergence*. This information-theoretic result appears promising for applications where spectral entropy already plays an important role such as EEG analysis. It also lends support to a new spectral estimation technique recently developed by the authors where a multivariate version of the Itakura-Saito distance is employed as a spectrum divergence. A minimum complexity spectrum is provided by this new approach. Simulations suggest the effectiveness of the new technique in tackling multivariate spectral estimation tasks, especially in the case of short data records.

1 Introduction

Information theoretic indexes have played a key role in a widespread variety of statistical, signal processing and identification problems, see e.g. [15, 18, 2, 3, 4, 12]. On the other hand, distances

between multivariable power spectra have only recently received any attention. In this direction, we mention the generalization of the Hellinger distance introduced in [6], a version of the quantum-mechanical Umegaki-von Neumann relative entropy [9] and Riemannian metrics [13].

In our recent paper [7], a multivariate version of the *Itakura-Saito divergence* was employed to solve the state-covariance matching problem. The choice of this *Bregman divergence* was there motivated by the connection between such a distance and the *relative entropy rate* for stationary, multidimensional, Gaussian processes [18], [11, p.371]. Here, we show that it is possible to introduce a *spectral relative entropy rate*, namely a relative entropy rate between the corresponding orthogonal increments processes of the spectral representation. It is then possible to prove that the relative entropy rates in the time and spectral domains are in fact *equal*! On the one hand, this information theoretic result lends further support to the choice of the multivariate Itakura-Saito distance for spectrum approximation. On the other hand, it opens up a new perspective in applications involving spectral entropy such as EEG analysis, see e.g. [20].

2 Multivariate Itakura-Saito Distance

The *relative entropy* or *Kullback-Leibler* pseudo-distance or *divergence* between two probability den-

*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”

sities p and q , with the support of p contained in the support of q , is defined by

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

In the case of two zero mean, Gaussian densities p and q , with positive definite covariance matrices P and Q , respectively, it is given by

$$\mathbb{D}(p||q) := \frac{1}{2} \{ \log \det (P^{-1}Q) + \text{tr} [Q^{-1} (P - Q)] \}.$$

Consider now two wide-sense stationary, purely nondeterministic, zero mean, jointly Gaussian stochastic processes $y = \{y_k; k \in \mathbb{Z}\}$ and $z = \{z_k; k \in \mathbb{Z}\}$, with values in $\mathbb{R}^m$. Let $Y_{[-n,n]}$ be the Gaussian random vector whose elements are given by the window $y_{-n}, y_{-n+1}, \dots, y_0, \dots, y_{n-1}, y_n$ and let $Z_{[-n,n]}$ be defined in the same fashion. Let $p_{Y_{[-n,n]}}$ and $p_{Z_{[-n,n]}}$ denote the corresponding joint densities, respectively. Then, the *relative entropy rate* between z and y is defined by

$$\mathbb{D}_r(y||z) := \lim_{n \rightarrow \infty} \frac{1}{2n+1} \mathbb{D}(p_{Y_{[-n,n]}} || p_{Z_{[-n,n]}}),$$

if the limit exists. Let Φ_y and Φ_z be the spectral density functions of y and z , respectively. Assume that at least one of the following conditions is satisfied:

1. $\Phi_y \Phi_z^{-1}$ is bounded;
2. $\Phi_y \in L^2(-\pi, \pi)$ and Φ_z is *coercive* (i.e. $\exists \alpha > 0$ s.t. $\Phi_z(e^{j\vartheta}) - \alpha I_m > 0$ a.e. on $\mathbb{T}$),

where $\mathbb{T}$ denotes the unit circle. Then, the following classical result holds [18, 12, 22]:

$$\begin{aligned} \mathbb{D}_r(y||z) = & \frac{1}{4\pi} \int_{-\pi}^{\pi} \{ \log \det (\Phi_y^{-1}(e^{j\vartheta}) \Phi_z(e^{j\vartheta})) \\ & + \text{tr} [\Phi_z^{-1}(e^{j\vartheta}) (\Phi_y(e^{j\vartheta}) - \Phi_z(e^{j\vartheta}))] \} d\vartheta. \end{aligned} \quad (1)$$

The right-hand side of (1) represents a multivariable *Itakura-Saito distance*. Indeed, in the case of scalar spectra, $\mathbb{D}_r(y||z) = 1/2 d_{IS}(\Phi, \Psi)$, where

$$d_{IS}(\Phi, \Psi) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \frac{\Phi(e^{j\vartheta})}{\Psi(e^{j\vartheta})} - \log \frac{\Phi(e^{j\vartheta})}{\Psi(e^{j\vartheta})} - 1 \right\} d\vartheta$$

is the classical Itakura-Saito distance of maximum likelihood estimation for speech processing [4].

3 Circularly-symmetric complex Gaussian random vectors

We shall need to consider zero-mean, n -dimensional, complex-valued Gaussian random vectors $z = \alpha + j\beta$, where the real and imaginary parts

are jointly Gaussian. The covariance matrix is then given by $\mathbb{E}[zz^*]$, where $*$ denotes transposition plus conjugation. The corresponding density function is the joint probability density of the $2n$ -dimensional compound vector $\gamma = [\alpha^\top \beta^\top]^\top$. The relative entropy between two zero-mean n -dimensional complex Gaussian densities p and q is given by

$$\mathbb{D}(p||q) := \frac{1}{2} [\log \det(R_p^{-1}R_q) + \text{tr}(R_q^{-1}R_p) - 2n],$$

where R_p and R_q are the covariance matrices of the $2n$ -dimensional vectors γ_p and γ_q corresponding to the densities p and q , respectively. If the zero-mean, $\mathbb{C}^n$ -valued Gaussian random vector z has the property that $\mathbb{E}[zz^\top] = 0$, then we say that z is a *circular symmetric* normally distributed random vector [17, 8]. This implies that $\mathbb{E}[\alpha\alpha^\top] = \mathbb{E}[\beta\beta^\top]$. If p and q are two n -dimensional complex Gaussian distribution with circular symmetry, the expression of the relative entropy simplifies to the formula

$$\mathbb{D}(p||q) = \log \det(P^{-1}Q) + \text{tr}(Q^{-1}P) - n,$$

where P and Q are the covariance matrices of $z_p := \alpha_p + j\beta_p$ and $z_q := \alpha_q + j\beta_q$, respectively.

4 Spectral relative entropy rate

The stationary, purely nondeterministic, Gaussian process y with spectrum Φ_y admits the *spectral representation*

$$y_k = \int_{\mathbb{T}} e^{jk\vartheta} d\hat{y}(e^{j\vartheta}), \quad (2)$$

where $\hat{y}$ is an m -dimensional stochastic process with orthogonal increments [14, 21, 16]. Moreover,

$$\mathbb{E} \{ d\hat{y}(e^{j\vartheta}) d\hat{y}(e^{j\vartheta})^* \} = \Phi_y(e^{j\vartheta}) d\vartheta. \quad (3)$$

Notice that, given $\vartheta_1, \vartheta_2 \in (-\pi, \pi]$, $\hat{y}(\vartheta_1, \vartheta_2)$ is a complex-valued Gaussian vector, i.e. a complex vector for which the real and imaginary parts are jointly Gaussian. Recall that the corresponding density function is the joint probability density of the $2m$ -dimensional vector γ obtained by stacking the real part of the complex vector over its imaginary part [23, 17]. The first contribution of this paper is to introduce a natural concept of *spectral relative entropy rate*. Let $\vartheta_k = \frac{\pi k}{n}$, $k = 0, 1, \dots, 2n$ and consider the complex Gaussian random vectors

$$\begin{aligned} \hat{Y}_{2n} &:= [\hat{y}(e^{j\vartheta_0}, e^{j\vartheta_1})^\top, \dots, \hat{y}(e^{j\vartheta_{2n-1}}, e^{j\vartheta_{2n}})^\top]^\top, \\ \hat{Z}_{2n} &:= [\hat{z}(e^{j\vartheta_0}, e^{j\vartheta_1})^\top, \dots, \hat{z}(e^{j\vartheta_{2n-1}}, e^{j\vartheta_{2n}})^\top]^\top. \end{aligned}$$

Let $p(\hat{Y}_{2n})$ and $p(\hat{Z}_{2n})$ be their joint probability density, respectively. We define the spectral relative entropy rate between y and z as the following limit, provided it exists:

$$\mathbb{D}_r(d\hat{y}||d\hat{z}) := \lim_{n \rightarrow \infty} \frac{1}{2n} \mathbb{D} \left(p(\hat{Y}_{2n}) || p(\hat{Z}_{2n}) \right). \quad (4)$$

Our main contribution is the following thought-provoking result:

Theorem 1. *Let y and z be as above. Assume that both Φ_y and Φ_z are piecewise continuous and coercive spectral densities. Then, the following equality holds:*

$$\mathbb{D}_r(y||z) = \mathbb{D}_r(d\hat{y}||d\hat{z}). \quad (5)$$

The proof involves a number of nontrivial steps, due to the fact that the spectral processes are complex-valued. The two following lemmas, whose proof may be found in [7], are crucial to this end. The first result establishes a *circular symmetry property* of the increments of the process $\hat{y}$ occurring in the spectral representation.

Lemma 1. *Suppose $\vartheta_1, \vartheta_2 \in (-\pi, \pi]$, then $\hat{y}(e^{-j\vartheta_2}, e^{-j\vartheta_1}) = \overline{\hat{y}(e^{j\vartheta_1}, e^{j\vartheta_2})}$. If, moreover, ϑ_1, ϑ_2 have the same sign, then, $\hat{y}(e^{j\vartheta_1}, e^{j\vartheta_2})$ is a circularly symmetric, normally distributed random vector.*

The second lemma proves an independence of the increments result.

Lemma 2. *Let $\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4$ be such that $[\vartheta_1, \vartheta_2] \cap [\vartheta_3, \vartheta_4] = [\vartheta_1, \vartheta_2] \cap [-\vartheta_4, -\vartheta_3] = \emptyset$. Then, $\hat{y}(e^{j\vartheta_1}, e^{j\vartheta_2})$ and $\hat{y}(e^{j\vartheta_3}, e^{j\vartheta_4})$ are independent random vectors.*

The last preparatory result allows to restrict our attention to increments on the interval $[0, \pi]$.

Lemma 3.

$$\mathbb{D} \left(p(\hat{Y}_{2n}) || p(\hat{Z}_{2n}) \right) = \mathbb{D} \left(p(\hat{Y}_n) || p(\hat{Z}_n) \right).$$

Proof. (Theorem 1) In view of Lemma 1, the last n components of $\hat{Y}_{2n}$ are functions (the complex conjugate) of the first n and the same holds for $\hat{Z}_{2n}$. Hence, in view of Lemma 3, we have $\mathbb{D} \left(p(\hat{Y}_{2n}) || p(\hat{Z}_{2n}) \right) = \mathbb{D} \left(p(\hat{Y}_n) || p(\hat{Z}_n) \right)$. Using

Lemma 2, we have that the elements of $\hat{Y}_n$ are independent random vectors and the same holds for the elements of $\hat{Z}_n$. Hence, we have the following additive decomposition:

$$\begin{aligned} \mathbb{D} \left(p(\hat{Y}_{2n}) || p(\hat{Z}_{2n}) \right) &= \mathbb{D} \left(p(\hat{Y}_n) || p(\hat{Z}_n) \right) \\ &= \sum_{k=0}^{n-1} \mathbb{D} (p(\delta \hat{y}_k) || p(\delta \hat{z}_k)), \end{aligned} \quad (6)$$

with $p(\delta \hat{y}_k)$ and $p(\delta \hat{z}_k)$ being the probability densities of the random vector $\delta \hat{y}_k = \hat{y}(e^{j\vartheta_k}, e^{j\vartheta_{k+1}})$ and $\delta \hat{z}_k = \hat{z}(e^{j\vartheta_k}, e^{j\vartheta_{k+1}})$, respectively. Since $\delta \hat{y}_k$ and $\delta \hat{z}_k$ are jointly Gaussian and circularly symmetric, by (2) and (3), we get,

$$\begin{aligned} &\mathbb{D} (p(\delta \hat{y}_k) || p(\delta \hat{z}_k)) \\ &= \log \det [Q_y^{-1}(\vartheta_k, \vartheta_{k+1}) Q_z(\vartheta_k, \vartheta_{k+1})] \\ &\quad + \text{tr} [Q_z^{-1}(\vartheta_k, \vartheta_{k+1}) Q_y(\vartheta_k, \vartheta_{k+1})] - m, \end{aligned}$$

where, by virtue of the orthogonality of the increments

$$Q_y(\vartheta_k, \vartheta_{k+1}) := \int_{\vartheta_k}^{\vartheta_{k+1}} \Phi_y(e^{j\xi}) d\xi,$$

and similarly for z . By piecewise continuity and the mean value theorem, we have that, except for a finite number of k 's,

$$\begin{aligned} &\mathbb{D} (p(\delta \hat{y}_k) || p(\delta \hat{z}_k)) \\ &= \log \det \left[\left(\Phi_y(e^{j\bar{\vartheta}_k}) \frac{\pi}{n} \right)^{-1} \Phi_z(e^{j\bar{\vartheta}_k}) \frac{\pi}{n} \right] \\ &\quad + \text{tr} \left[\left(\Phi_z(e^{j\bar{\vartheta}_k}) \frac{\pi}{n} \right)^{-1} \Phi_y(e^{j\bar{\vartheta}_k}) \frac{\pi}{n} \right] - m \\ &= \log \det [\Phi_y(e^{j\bar{\vartheta}_k})^{-1} \Phi_z(e^{j\bar{\vartheta}_k})] \\ &\quad + \text{tr} [\Phi_z(e^{j\bar{\vartheta}_k})^{-1} \Phi_y(e^{j\bar{\vartheta}_k})] - m, \end{aligned}$$

where $\vartheta_k \leq \bar{\vartheta}_k < \vartheta_{k+1}$. By employing the latter expression together with (6) and (4), we get

$$\begin{aligned}
\mathbb{D}_r(d\hat{y}||d\hat{z}) &= \lim_{n \rightarrow \infty} \frac{1}{2n} \mathbb{D} \left(p(\hat{Y}_n) || p(\hat{Z}_n) \right) \\
&= \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_{k=0}^{n-1} \mathbb{D} (p(\delta\hat{y}_k) || p(\delta\hat{z}_k)) \\
&= \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_{k=0}^{n-1} \log \det \Phi_y(e^{j\bar{\vartheta}_k})^{-1} \Phi_z(e^{j\bar{\vartheta}_k}) \\
&\quad + \text{tr} \left[\Phi_z^{-1}(e^{j\bar{\vartheta}_k}) \left(\Phi_y(e^{j\bar{\vartheta}_k}) - \Phi_z(e^{j\bar{\vartheta}_k}) \right) \right] \\
&= \lim_{n \rightarrow \infty} \frac{1}{2\pi} \sum_{k=0}^{n-1} \left\{ \log \det \Phi_y(e^{j\bar{\vartheta}_k})^{-1} \Phi_z(e^{j\bar{\vartheta}_k}) \right. \\
&\quad \left. + \text{tr} \left[\Phi_z^{-1}(e^{j\bar{\vartheta}_k}) \left(\Phi_y(e^{j\bar{\vartheta}_k}) - \Phi_z(e^{j\bar{\vartheta}_k}) \right) \right] \right\} \frac{\pi}{n} \\
&= \frac{1}{2\pi} \int_0^\pi \left\{ \log \det (\Phi_y^{-1}(e^{j\vartheta}) \Phi_z(e^{j\vartheta})) \right. \\
&\quad \left. + \text{tr} \left[\Phi_z^{-1}(e^{j\vartheta}) (\Phi_y(e^{j\vartheta}) - \Phi_z(e^{j\vartheta})) \right] \right\} d\vartheta \\
&= \frac{1}{4\pi} \int_{-\pi}^\pi \left\{ \log \det (\Phi_y^{-1}(e^{j\vartheta}) \Phi_z(e^{j\vartheta})) \right. \\
&\quad \left. + \text{tr} \left[\Phi_z^{-1}(e^{j\vartheta}) (\Phi_y(e^{j\vartheta}) - \Phi_z(e^{j\vartheta})) \right] \right\} d\vartheta,
\end{aligned}$$

which, by (1), is $\mathbb{D}_r(y||z)$. $\blacksquare$

5 Approximation in the Itakura-Saito distance

Let $\mathcal{S}_+^{m \times m}$ denote the family of bounded and coercive spectral densities on $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ of $\mathbb{R}^m$ -valued processes. Suppose that the data $\{y_i\}_{i=1}^N$ are generated by an unknown, zero-mean, m -dimensional, $\mathbb{R}^m$ -valued, purely nondeterministic, stationary, Gaussian process $y = \{y_k; k \in \mathbb{Z}\}$. We wish to estimate the spectral density $\Phi \in \mathcal{S}_+^{m \times m}$ of y from $\{y_i\}_{i=1}^N$.

A THREE-like approach [5, 19] generalizes Burg-like methods in several ways. The second order statistics that are estimated from the data $\{y_i\}_{i=1}^N$ are not necessarily the covariance lags $C_l := E\{y_{k+l}y_k^\top\}$ of y . Moreover, a prior estimate of Φ may be included in the estimation procedure. More explicitly these methods hinge on the following four elements:

1. A rational filter to process the data. The filter has transfer function

$$G(z) = (zI - A)^{-1}B, \quad (7)$$

where $A \in \mathbb{R}^{n \times n}$ has all its eigenvalues inside the unit circle, $B \in \mathbb{R}^{n \times m}$ 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

$$x(k+1) = Ax(k) + By(k);$$

3. a *prior* spectral density $\Psi \in \mathcal{S}_+^{m \times m}$;
4. an index that measures the distance between two spectral densities.

The filterbank (8) provides Carathéodory or, more generally, Nevanlinna-Pick interpolation data for the positive real part Φ_+ of Φ , see [5, Section II]. This occurs through the constraint

$$\int G\Phi G^* = \Sigma \quad (8)$$

which must be satisfied by the spectrum of y (here and in the sequel integration occurs on the unit circle with respect to normalized Lebesgue measure). Since, in general, the prior estimate Ψ is not consistent with the interpolation conditions, an approximation problem arises. It is then necessary to introduce an adequate distance index. This crucial choice is dictated by 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, typically by solving the dual optimization problem. In [7], the following approximation problem was considered:

Problem 1. Let $\Psi \in \mathcal{S}_+^{m \times m}$, $G(z)$ as in (7) and $\Sigma = \Sigma^\top > 0$. Find Φ° that solves:

$$\begin{aligned}
&\text{minimize} \quad d_{\text{IS}}(\Phi, \Psi) \\
&\text{over} \quad \left\{ \Phi \in \mathcal{S}_+^{m \times m} \mid \int G\Phi G^* = \Sigma \right\}.
\end{aligned}$$

The first issue to be addressed is *feasibility* of the optimization problem, namely existence of a $\Phi \in \mathcal{S}_+^{m \times m}$ that satisfies (8). This problem was solved in [10] (see also [19]) where an algebraic condition that is necessary and sufficient for feasibility was derived. In [7], it was shown that, if the problem is feasible, then the optimal spectrum has the form

$$\Phi^\circ(\Lambda) := [\Psi^{-1} + G^* \Lambda G]^{-1}. \quad (9)$$

The Hermitian matrix Λ playing the role of a multiplier must be such as to satisfy the constraint, namely

$$\int G \left\{ [\Psi^{-1} + G^* \Lambda G]^{-1} \right\} G^* = \Sigma$$

Existence of a Λ° satisfying this constraint can be established proving that the dual problem has a solution (this is a highly nontrivial task since the multipliers set is open and unbounded). The multiplier can then be computed through a globally convergent, matricial, Newton-like method, see [7] for the details. The most interesting feature of the expression (9) is that it provides an upper bound for the McMillan degree of the optimal solution $\Phi^\circ(\Lambda^\circ)$ (which is the most natural index of complexity). This upper bound is given by $2n + \deg(\Psi)$ and is equal to the best one so far established in the scalar case. Thus, this result represents a significant improvement in the frame of multivariable spectral estimation. Indeed, the best so far available upper bound was $\deg[\Psi] + 4n$.

Following [5], in [7] it is shown that Problem 1 may be used as a basis for a spectral estimation technique that features high resolution in a desired range of frequencies. Moreover, simulations show that in the case of a very short data record this method is robust with regard to artifacts, as opposed to standard identification techniques like MATLAB's PEM and N4SID. We refer the reader to [7] for further details and more comprehensive simulations.

6 Conclusions

In conclusion, the main contribution of the paper is to provide a novel, precise connection between spectral relative entropy and its time domain counterpart via a multivariate Itakura-Saito divergence. This information-theoretic result seems to suggest that the pseudo distance induced by the spectral entropy rate is quite natural in dealing with many optimization problems in science and information engineering. It also appears promising for applications where spectral entropy already plays an important role such as EEG analysis.

Acknowledgments

We wish to thank an anonymous reviewer for a careful reading of the paper.

References

- [1] M. Basseville. Distance Measures for Signal Processing and Pattern Recognition. *Signal Processing*, 18:349–369, 1989.
- [2] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Aut. Contr.*, AC-19:716–723, 1974.
- [3] H. Akaike. Canonical correlation analysis of time series and the use of an information criterion. In R.K. Mehra and D.G. Lainiotis, editors, *System Identification: Advances and Case Studies*, pages 27–96, 1976.
- [4] M. Basseville. Distance Measures for Signal Processing and Pattern Recognition. *Signal Processing*, 18:349–369, 1989.
- [5] C. I. Byrnes, T. Georgiou, and A. Lindquist. A new approach to spectral estimation: A tunable high-resolution spectral estimator. *IEEE Trans. Sig. Proc.*, 48:3189–3205, 2000.
- [6] A. Ferrante, M. Pavon, and F. Ramponi. Hellinger vs. Kullback-Leibler multivariable spectrum approximation. *IEEE Trans. Aut. Control*, 53:954–967, 2008.
- [7] A. Ferrante, C. Masiero, and M. Pavon. Time and spectral domain relative entropy: A new approach to multivariate spectral estimation. March 2011. *IEEE Trans. Aut. Contr.*, to appear, preprint arXiv:1103.5602v2.
- [8] R. Gallager. *Circularly-Symmetric Gaussian random vectors*. M.I.T. notes, Jan. 2008.
- [9] T. Georgiou. Relative entropy and the multivariable multidimensional moment problem. *IEEE Trans. Inform. Theory*, 52:1052–1066, 2006.
- [10] 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.
- [11] R. Gray, A. Buzo, A. Jr Gray, and Y. Matsuyama. Distortion measures for speech processing. *IEEE Trans. Acoustics, Speech and Signal Proc.*, 28:367–376, 1980.
- [12] S. Ihara. *Information Theory for Continuous Systems*. World Scientific, Singapore, 1993.
- [13] X. Jiang, L. Ning, and T. Georgiou. Distances and Riemannian metrics for multivariate spectral densities. *preprint*, June 2011.
- [14] H. Kramer and M. R. Leadbetter. *Stationary and Related Stochastic Processes*. Wiley, New York, 1966.

- [15] S. Kullback. *Information Theory and Statistics* 2nd ed. Dover, Mineola NY, 1968.
- [16] A. Lindquist and G. Picci. *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification*. In preparation: preprint available in <http://www.math.kth.se/~alq/LPbook>.
- [17] K. S. Miller. *Complex Stochastic Processes*. Addison Wesley, Reading, MA, 1974.
- [18] M. S. Pinsker. *Information and information stability of random variables and processes*. Holden-Day, San Francisco, 1964. Translated by A. Feinstein.
- [19] 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.
- [20] I. Rezek, and S. Roberts. Stochastic Complexity Measures for Physiological Signal Analysis. *IEEE Transactions on Biomedical Engineering* Vol. 45 (9):1186–1191, Sept. 1998.
- [21] Yu. A. Rozanov. *Stationary Random Processes*. Holden-Day, San Francisco, 1967.
- [22] A. A. Stoorvogel and J. H. Van Schuppen. System identification with information theoretic criteria. In S. Bittanti and G. Picci, editors, *Identification, Adaptation, Learning: The Science of Learning Models from Data*. Springer, Berlin-Heidelberg, 1996.
- [23] A. Van Den Bos. The Multivariate Complex Normal Distribution - A Generalization. *IEEE Trans. Inform. Theory*, 41:537–539, 1995.