Decomposition of Toeplitz matrices via convex optimization

Tryphon T. Georgiou *Fellow, IEEE*

Abstract— **We point out that autocovariance functions of moving-average processes of any given order m, can be characterized via a linear matrix inequality (LMI). This LMI-condition can be used to decompose any Toeplitz autococovariance matrix into a sum of a singular Toeplitz covariance plus the autocovariance matrix of a moving average process of order m and of *maximal variance*. The decomposition is unique and subsumes the Pisarenko harmonic decomposition which corresponds to $m = 0$. It can be used to account for mutual couplings between elements in linear antenna arrays, or identify colored noise consistent with the covariance data. The same LMI-condition leads to an efficient computation of the least order of a MA-spectrum which agrees with covariance moments.**

Keywords—**Spectral analysis, Pisarenko harmonic decomposition, moving average processes, convex optimization; EDICS: DSP-TFSR.**

I. Introduction

THE observation that singularities in covariance matrices reveal a deterministic linear dependence between observed quantities, forms the basis of a wide range of techniques, from Gauss' least squares to modern subspace methods in time-series analysis. In particular, the so-called Pisarenko harmonic decomposition [3], [7], relies on a decomposition of a Toeplitz covariance matrix into a singular one and a positive scalar multiple of the identity—the first corresponding to a deterministic component and the second to a white noise process. The present work suggests a more versatile framework where a given Toeplitz covariance is decomposed according to the hypothesis that the underlying process consists of a deterministic component and a moving average one. The moving average component of a prespecified order m can be selected to have maximal variance. This allows a unique decomposition which generalizes the spirit of the Pisarenko dictum.

A decomposition into a low order moving average part plus a deterministic component seems appropriate in modeling mutual coupling between elements in antenna arrays in the context of direction of arrival (DOA) estimation where spectral lines are also present. It may also provide a convenient framework for modeling continuous spectra and colored noise. Such a decomposition was introduced in [2] in the process of seeking suitable multivariable generalizations of the Pisarenko harmonic decomposition for general state-covariance matrices. In the present note we specialize such a decomposition in the context of Toeplitz matrices. We also introduce and study a sequence of parameters g_m for $m = 0, 1, \ldots$ representing the fraction of the energy/variance of the underlying process which can be accounted for by a moving average one of order m. The behaviour of g_m as a function of m may suggest suitable choices for the order of the continuous part of the spectrum (e.g., m selected at a point of "discontinuity"). Further, the smallest value m for which $g_m = 1$ is guaranteed to occur after a finite number of steps and is the least order of a MA-spectrum that can account for the complete set of covariance data.

Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455; tryphon@ece.umn.edu, (612) 625-3303, Fax: (612) 625-4583; Research partially supported by the NSF and the AFOSR.

Section II discusses theoretical and computational issues of the proposed decomposition, while Section III presents an academic example that elucidates the key ideas. It should be emphasized that the proposed decomposition leads to alternatives of the Pisarenko paradigm which are similarly consistent with given covariance data. The practical significance of the methodology appears considerable in cases of low SNR, and is currently under investigation.

II. Theoretical considerations

Throughout we consider a scalar, discrete-time, zero-mean, stationary stochastic process $\{u_k : k \in \mathbb{Z}\}$ taking values in $\mathbb{C}$ with $m \in \mathbb{N}$. We denote by

$$R_k := \mathcal{E}\{u_\ell u_{\ell-k}^*\},$$

for $k, \ell \in \mathbb{Z}$, the sequence of covariances and by $d\mu(\theta)$ the corresponding power spectrum for which

$$R_k = \int_0^{2\pi} e^{-jk\theta} d\mu(\theta),$$

where "$*$" denotes the "complex-conjugate", $j := \sqrt{-1}$, and $\mathcal{E}\{\cdot\}$ is the expectation operator.

It is well-known that a covariance sequence $\{R_\ell : \ell \in \mathbb{Z} \text{ and } R_{-\ell} = R_\ell^*\}$ is completely characterized by the non-negativity of the Toeplitz matrices

$$\mathbf{R}_\ell := \begin{bmatrix} R_0 & R_1 & \ldots & R_\ell \\ R_{-1} & R_0 & \ldots & R_{\ell-1} \\ \vdots & \vdots & \ddots & \vdots \\ R_{-\ell} & R_{-\ell+1} & \ldots & R_0 \end{bmatrix} \tag{1}$$

for all ℓ. At the same time, the infinite sequence of R_ℓ's specifies the power spectrum $d\mu$ uniquely. However, it is often the case that only a finite set of second-order statistics is available, and then, it is of interest to characterize possible extensions of the finite covariance sequence $\{R_0, R_1, \ldots, R_n\}$, or equivalently, characterize the totality of consistent power spectra. When only $\{R_0, R_1, \ldots, R_n\}$ is known and $\mathbf{R}_n > 0$, the family of power spectra is infinite, whereas if $\mathbf{R}_n$ is singular, then R_ℓ is specified uniquely for $\ell > n$ by the requirement that $\mathbf{R}_\ell \geq 0$. In this case the underlying process is deterministic (see e.g., [7, Section 3.9.2]) and the unique power spectrum consists of pure spectral lines (see e.g., [7, Section 4.5 and Exercise 4.5], or [2, Theorem 5] for a considerably more general setting).

If $\{v_k,\ k \in \mathbb{Z}\}$ is a moving average process of order m, then the corresponding covariance sequence

$$Q_\ell := E\{v_k v_{k+\ell}^*\}$$

vanishes for $\ell > m$. In this case, all successive principle minors of the infinite banded Toeplitz matrix

$$\mathbf{Q} := \begin{bmatrix} Q_0 & \dots & Q_m & 0 & \dots \\ \vdots & \ddots & & \ddots & \ddots \\ Q_{-m} & & & & \\ 0 & \ddots & & & \\ \vdots & \ddots & & & \end{bmatrix} \quad (2)$$

are positive. Consistently with (1), we denote by $\mathbf{Q}_\ell$ the $(\ell+1)\times(\ell+1)$ principle submatrix of (2). Although $\mathbf{Q}$ is made up of finitely many parameters, testing whether $\mathbf{Q}_\ell > 0$ for all ℓ is not a "finite" test[1]. Alternatively, the spectral density

$$q(e^{j\theta}) := Q_m^* e^{-jm\theta} + \ldots + Q_0 + \ldots Q_m e^{jm\theta}, \quad (3)$$

is a nonnegative trigonometric polynomial which is also $2 \times \Re e(F(e^{j\theta}))$ where

$$\begin{aligned} F(z) &:= \frac{1}{2}Q_0 + Q_1 z + \ldots + Q_m z^m \\ &= \frac{1}{2}Q_0 + z\mathbf{C}(I - z\mathbf{A})^{-1}\mathbf{B} \end{aligned}$$

for $z = e^{j\theta}$. Here $\mathbf{A}$ is a $m \times m$ companion-shift matrix and $\mathbf{B}, \mathbf{C}$ compatible column and row vectors of length m given by

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix},\ \mathbf{B} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix},\ \text{and}$$
$$\mathbf{C} = \begin{bmatrix} Q_1 & \dots & Q_m \end{bmatrix}. \quad (4)$$

Since $F(z)$ is analytic inside the unit disc, then the condition that $q(e^{j\theta}) \geq 0$ for all $\theta \in [-\pi, \pi]$ is equivalent to $F(z)$ being a "positive-real" function, i.e., having nonnegative real part inside the unit disc. The positive realness property can be most efficiently characterized via a linear matrix inequality:

Proposition 1: The finite sequence $\{Q_0, \ldots, Q_m\}$ is a partial covariance sequence of a moving average process of order m if and only if there exists an $m \times m$ matrix $\mathbf{P} \geq 0$ such that, with $\mathbf{A}, \mathbf{B}, \mathbf{C}$ as in (4),

$$\begin{bmatrix} \mathbf{P} - \mathbf{A}^*\mathbf{P}\mathbf{A} & \mathbf{C}^* - \mathbf{A}^*\mathbf{P}\mathbf{B} \\ \mathbf{C} - \mathbf{B}^*\mathbf{P}\mathbf{A} & Q_0 - \mathbf{B}^*\mathbf{P}\mathbf{B} \end{bmatrix} \geq 0. \quad (5)$$

Proof: In essence, the statement of the proposition is a special case of the well-known as the "positive-real lemma" ([4], see e.g. [1, Théorème 3.1]) which applies to more general $(\mathbf{A}, \mathbf{B})$ pairs. ■

Testing condition (5) amounts to a convex optimization problem, and so is the following:

Problem 1: Given a finite non-negative definite Toeplitz matrix $\mathbf{R}_n$ for $n \in \mathbb{N}$, and given $m \in \mathbb{N}$, determine the maximal value for $Q_0 \geq 0$ for which there exist $Q_1, \ldots, Q_m \in \mathbb{C}$ and an $m \times m$ matrix $\mathbf{P}$ such that with $\mathbf{Q}$ as in (2) and $\mathbf{Q}_n$ defined accordingly,

$$\begin{aligned} &(6a)\quad \mathbf{P} \geq 0 \\ &(6b)\quad \textit{condition } (5) \textit{ holds, and} \\ &(6c)\quad \mathbf{R}_n \geq \mathbf{Q}_n. \end{aligned} \quad (6)$$

Conditions (6a-c) are standard linear matrix inequalities and Problem 1, which asks for the minimum of a linear functional subject to (6a-c), is readily solvable numerically using software which are standard in e.g., the Matlab LMI toolbox. A routine for solving Problem 1 which uses this Matlab LMI-software is provided at the author's website [8].

Proposition 2: If $Q_0, \ldots, Q_m$ and $\mathbf{P}$ represent the solution to Problem 1, and accordingly $\mathbf{R}_n, \mathbf{Q}_n$ are as stated before, then their difference given below is nonnegative and singular:

$$\mathbf{S}_n := \mathbf{R}_n - \mathbf{Q}_n. \quad (7)$$

Proof: $\mathbf{S}_n$ is clearly nonnegative from (6c). To show that it is singular, assume the contrary. Then, if Q_0 is replaced by $Q_0 + \epsilon$ and the same values for $Q_1, \ldots, Q_m$ and $\mathbf{P}$ are used, conditions (6a-c) are still valid for a sufficiently small $\epsilon > 0$. This is true because a small increase in Q_0 so that $\mathbf{S}_n$ is still positive, does not invalidate (5). This contradicts the assumption that $\mathbf{S}_n$ originates from solving Problem 1 since Q_0 is not maximal. ■

We introduce the sequence of parameters g_m, for $m = 0, 1, \ldots$, via

$$g_m = Q_{0,m}/R_0$$

where the expanded indexing $Q_{0,m}$ indicates that the value for Q_0 is obtained as in Problem 1 for a particular choice of m. These parameters represent the fraction of the energy/variance in the data $\{R_0, R_1, \ldots, R_n\}$ that can be attributed to a moving average component of order m.

Proposition 3: If $\mathbf{R}_n > 0$, then for $m = 0, 1, 2, \ldots$, g_m is non-decreasing, $0 < g_m \leq 1$ and there is a finite m_1 such that $g_m = 1$ for all $m \geq m_1$. If on the other hand $\mathbf{R}_n$ is singular, then $g_m = 0$ for all $m \geq 0$.

Proof: Assume first that $\mathbf{R}_n > 0$. Clearly, $Q_{0,m} > 0$. Also, from (6c), we deduce that $Q_{0,m} \leq R_0$. Hence, $0 < g_m \leq 1$. The sequence is nondecreasing because moving

[1] This issue underlies basic questions in linear estimation to which the periodogram is an easy way out, since it provides an approximate spectral density to a given set of covariances in the form of a moving average spectrum.

average processes of a given order trivially include those of lower order.

We now sketch the proof that $g_m \to 1$ in a finite number of steps. Since $\mathbf{R}_n > 0$, there exists a covariance extension $R_{n+1}, R_{n+2}, \ldots$ which corresponds to a maximal entropy power spectrum (e.g., see [3], [7]) and is consistent with the first $n+1$ covariance "samples" that make up $\mathbf{R}_n$. The corresponding spectral density $\sum_{k=-\infty}^{\infty} R_k e^{jk\theta}$ is strictly positive and continuous (hence uniformly approximable by trigonometric polynomials which already shows that $g_m \to 1$ as $m \to \infty$). But furthermore, the maximum entropy extension $R_{n+1}, R_{n+2}, \ldots$ decays geometrically. This is due to the fact that it corresponds to an all-pole model with poles that do not lie on the circle. Therefore, there is a finite value m_1 for which the "tail" $\sum_{|k|>m_1} R_k e^{jk\theta}$ is sufficiently small to ensure positivity of the sum over the remaining k's (i.e. over $|k| \leq m_1$) —but this remaining terms correspond to a moving average spectral density which is consistent with the data. This proves the claim.

Now consider the case where $\mathbf{R}_n$ is singular. Since $\mathbf{Q}_n$ corresponds to a moving average process, it is necessarily positive definite. Hence, $\mathbf{R}_n - \mathbf{Q}_n$ must be negative when projected on the null space of $\mathbf{R}_n$. Therefore $\mathbf{Q}_n \equiv 0$, and so is $Q_{m,0}$ and g_m for any m. ∎

III. An example

We consider as our data a Toeplitz matrix $\mathbf{R}_4$ made up of the following partial covariance sequence

$$R_\ell = \sqrt{5 - |\ell|} \text{ for } \ell = 0, \pm 1, \pm 2, \pm 3, \pm 4. \qquad (8)$$

The fraction g_m of the energy/variance that can be accounted for by a moving average process of order m, as a function of m, is shown in Figure 1 for $m = 0, 1, \ldots, 10$. Note that $g_8 = 1$ and hence, that $g_m = 1$ for $m > 8$ as well.

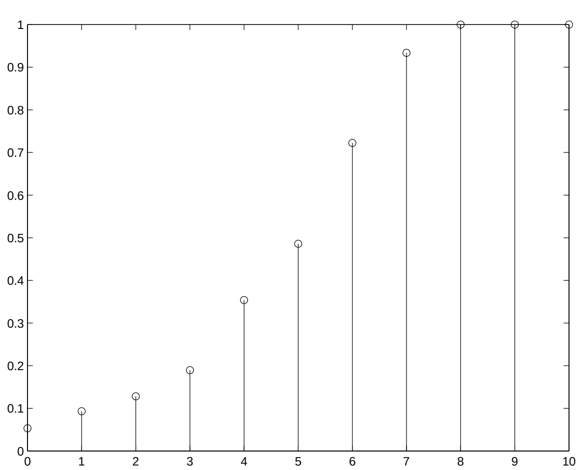

Fig. 1. g_m vs. m

Two alternative spectra for (8) are shown in Figures 2 and 3 for comparison. The first is based on interpreting the data as originating from a white background noise on top of a deterministic component consisting of sinusoidal signals. The maximal energy level for the white noise (according to the Pisarenko dictum) is shown by the horizontal line in Figure 2 while the power spectrum of the deterministic component corresponds to the spectral lines shown. Similarly, Figure 3 shows the power spectral density of a moving average component of order 3 which is consistent with data and has maximal variance, along with spectral lines that account for the remaining energy. *Both spectra shown in Figures 2 and 3 are consistent with the data.* Yet the latter adjusts for colored MA(3)-noise. If such noise is indeed present in the underlying process, and our intention is to identify spectral lines, accounting for the color of the noise may reduce bias and other effects.

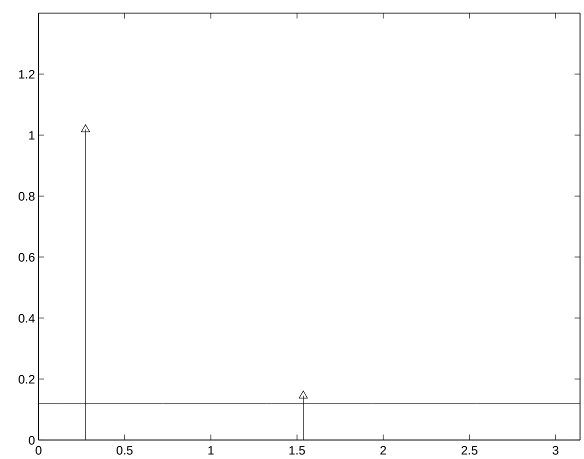

Fig. 2. Power spectrum corresponding to white noise + sinusoids

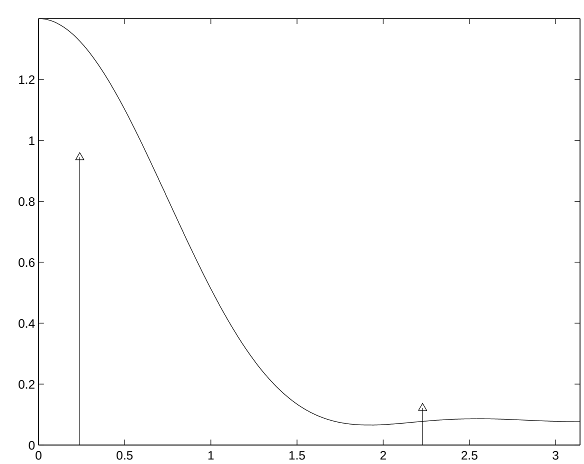

Fig. 3. Power spectrum corresponding to a MA(3)-component + sinusoids

Remark 1: It is interesting to point out that there is not a uniform bound on m_1 in Proposition 3 (i.e., a bound that depends only on n and not on the particular values making up $\mathbf{R}_n$). To see this, consider $\{R_0 = 1, R_{1,k} = (k-1)/k\}$ and let $k \to \infty$. If there was a fixed value m_1 and a set $\{R_{2,k}, \ldots, R_{m_1,k}\}$ consistent with the hypothesis of a moving average model of order m_1 or less, then as $k \to \infty$, convergent subsequences for the $R_{\ell,k}$'s having limits R_ℓ for $2 \leq \ell \leq m$ would give rise to a nonnegative trigonometric polynomial $1 + 2\cos(\theta) + 2R_2\cos(2\theta) + \ldots + 2R_m\cos(m\theta)$. But this is impossible because the limit partial sequence $\{R_0 = 1, R_1 = 1\}$ is already singular (since $\det(\mathbf{R}_1) = 0$) and has the unique nonnegative extension $R_\ell = 1$ for *all* ℓ.

IV. Concluding remarks

It is pointed out that the positive-real lemma provides an effective way to characterize autocovariance functions of moving-average processes. Utilizing this fact and in a way analogous to the well-known Pisarenko dictum, any given Toeplitz covariance matrix $\mathbf{R}_n$ can be decomposed into a sum $\mathbf{S}_n + \mathbf{Q}_n$ consistent with the decomposition of the underlying process into a deterministic component plus a MA(m) component of maximal variance. The Pisarenko harmonic decomposition corresponds to the special case where $m = 0$. In applications, the choice of m may be dictated by prior information on the nature of the component that corresponds to $\mathbf{Q}_n$ (e.g., its order), or by a more detailed analysis of the effect of m on the variability of a certain number of sinusoids that we anticipate in the component corresponding to $\mathbf{S}_n$.

The characterization of MA-autocovariances in Proposition 1 also allows an efficient solution to a long standing problem, the computation of the least order MA-spectrum which fully accounts for the given moments (see [5], [6] and the references therein). Indeed, provided $\mathbf{R}_n > 0$ (see Proposition 3) there is a finite minimal value m_1 for which $g_{m_1} = 1$. This is determined after successive evaluation of the increasing sequence g_m for $m = 1, 2, \ldots, m_1$. Clearly the sought least MA-order is m_1 and the parameters of the corresponding MA(m_1) model are obtained by the LMI-solver (cf. [8]). This is an efficient alternative to the *expanding hull algorithm* proposed in [5].

References

[1] P. Faurre, M. Clerget, and F. Germain, **Opérateurs Rationnels Positifs: Application à l'hyperstabilité et aux processus aléatoires**, Paris, France: Dunod, 293 pages, 1979.

[2] T.T. Georgiou, "The Caratheéodory-Fejér-Pisarenko decomposition and its multivariable counterpart," submitted to the *IEEE Trans. on Automatic Control*, September 2005; preprint: arXiv:math.OC/0509225v1 9 Sep 2005.

[3] S. Haykin, **Nonlinear Methods of Spectral Analysis,** New York, Springer-Verlag, 247 pages, 1979.

[4] R.E. Kalman, "Lyapunov functions for the problem of Lur'e in Automatic Control," Proc. of the National Academy of Sciences, **49(2)**: 201-205, 1963.

[5] A. O. Steinhardt, "Correlation matching by finite length sequences," *IEEE Trans. on ASSP*, **36(4)**: 545-559, 1988.

[6] A. O. Steinhardt and J. Makhoul, "On the autocorrelation of finite length sequences," *IEEE Trans. on ASSP*, **33(6)**: 1516-1520, 1985.

[7] P. Stoica and R. Moses, **Introduction to Spectral Analysis**, Prentice Hall, 2005.

[8] www.ece.umn.edu/users/georgiou/files/MA_decomposition.m