
Distances between time-series and their autocorrelation statistics

Tryphon T. Georgiou

Department of Electrical and Computer Engineering, University of Minnesota
Minneapolis, MN 55455; email: tryphon@umn.edu

This paper is dedicated to Giorgio Picci on the occasion of his sixty-fifth birthday

Summary. We begin with an interpretation of the L_1 -distance between two power spectral densities and then, following an analogous rationale, we develop a natural metric for quantifying distance between respective covariance matrices.

1 Introduction

Consider two discrete-time, stationary, zero-mean, (real-valued for notational convenience) random processes \mathbf{y}_k and $\hat{\mathbf{y}}_k$ ($k \in \mathbb{Z}$) having power spectral densities $f_{\mathbf{y}}(\theta)$ and $f_{\hat{\mathbf{y}}}(\theta)$ ($\theta \in [-\pi, \pi]$), and autocorrelation functions R_ℓ and \hat{R}_ℓ ($\ell \in \mathbb{Z}$), respectively, i.e.,

$$R_\ell = E\{\mathbf{y}_k \mathbf{y}_{k+\ell}\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-j\ell\theta} d\theta,$$

and similarly for the “hatted” quantities. When the power spectrum contains a singular part, then $f(\theta)d\theta$ needs to be replaced by a non-negative finite spectral measure $d\mu(\theta)$.

We are interested in quantifying the distance between respective spectra and statistics for two such random process \mathbf{y}_k and $\hat{\mathbf{y}}_k$. When two vectors

$$\begin{aligned} \mathbf{R}_n &:= [R_0 \ R_1 \ \dots \ R_{n-1}], \text{ and} \\ \hat{\mathbf{R}}_n &:= [\hat{R}_0 \ \hat{R}_1 \ \dots \ \hat{R}_{n-1}] \end{aligned}$$

of autocorrelation samples are available and need to be compared, one may use any metric in \mathbb{R}^n for that purpose, as for instance $\|\mathbf{R}_n - \hat{\mathbf{R}}_n\|_2 = \sqrt{\sum_k (R_k - \hat{R}_k)^2}$. However, we are not aware of any significance that can be attached to such a distance other than the fact that it is a metric in \mathbb{R}^n . Our goal in this paper, is to seek a metric which can be physically motivated.

Similarly, if we are to compare $f_{\mathbf{y}}(\theta)$ and $f_{\hat{\mathbf{y}}}(\theta)$, it appears difficult to motivate the use of an L_2 -distance $\|f_{\mathbf{y}}(\theta) - f_{\hat{\mathbf{y}}}(\theta)\|_2$. For one thing, the L_2 -distance cannot be generalized to deal with spectral measures when singular parts are present. There are certainly other alternatives. In the speech processing literature in particular there is a plethora of distances that, however, are not metrics [6] but have been motivated by specific needs. Function theoretic alternatives that one can use (e.g., L_p -norms, etc.) including Wasserstein-like transportation measures typically lack a physical interpretation. In a recent study [4, 5] a pseudometric was constructed as a geodesic between spectral densities/measures with respect to a rather natural Riemannian metric —this metric quantifies the degradation of predictive-error variance when the predictor is designed based on the wrong choice between two alternatives and the geometry is, in essence, Euclidean but only after we transform spectral densities using the logarithmic map.

In the current paper we focus on the L_1 distance

$$\|f_{\mathbf{y}}(\theta) - f_{\hat{\mathbf{y}}}(\theta)\|_1 := \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_{\mathbf{y}}(\theta) - f_{\hat{\mathbf{y}}}(\theta)| d\theta$$

which has also a rather natural interpretation. After a brief discussion of the relevance of the L_1 -distance, following a similar rationale, we will develop an analogous metric between finite partial covariance data of the corresponding random processes.

2 Interpretation of the L_1 distance

Given \mathbf{y}_k and $\hat{\mathbf{y}}_k$ we postulate that there exist two random processes ψ_k and $\hat{\psi}_k$ so that

$$\mathbf{y}_k + \psi_k = \hat{\mathbf{y}}_k + \hat{\psi}_k. \quad (1)$$

Alternatively, we postulate that there exists a random process \mathbf{z}_k and that the two original random processes relate to \mathbf{z}_k via

$$\begin{aligned} \mathbf{y}_k &= \mathbf{z}_k - \psi_k \\ \hat{\mathbf{y}}_k &= \mathbf{z}_k - \hat{\psi}_k. \end{aligned}$$

It is natural to seek such perturbations of minimal total combined variance

$$E\{\psi_k^2\} + E\{\hat{\psi}_k^2\} \quad (2)$$

that is sufficient to “reconcile” the two processes. The combined variance $E\{\psi_k^2\} + E\{\hat{\psi}_k^2\}$ represents the minimal amount of “energy” of perturbations in the two time-series that is needed to render the two indistinguishable. Intuitively, the minimal combined variance which is consistent with the available data quantifies the distance between the two.

Given $f_{\mathbf{y}}, f_{\hat{\mathbf{y}}}$, the optimal choice consists of random processes ψ_k and $\hat{\psi}_k$ such that ψ_k and $\hat{\psi}_k$ are independent, $\hat{\mathbf{y}}_k$ and $\hat{\psi}_k$ are also independent, and

$$f_{\psi}(\theta) = \begin{cases} f_{\hat{\mathbf{y}}}(\theta) - f_{\mathbf{y}}(\theta) & \text{if } f_{\hat{\mathbf{y}}}(\theta) - f_{\mathbf{y}}(\theta) \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (3a)$$

$$f_{\hat{\psi}}(\theta) = \begin{cases} f_{\mathbf{y}}(\theta) - f_{\hat{\mathbf{y}}}(\theta) & \text{if } f_{\hat{\mathbf{y}}}(\theta) - f_{\mathbf{y}}(\theta) \leq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3b)$$

Then, the power spectrum of the “sum”

$$\mathbf{z}_k := \mathbf{y}_k + \psi_k = \hat{\mathbf{y}}_k + \hat{\psi}_k$$

is simply

$$f_z(\theta) := \max\{f_{\mathbf{y}}(\theta), f_{\hat{\mathbf{y}}}(\theta)\}, \theta \in [-\pi, \pi],$$

and

$$\begin{aligned} d(f_{\mathbf{y}}, f_{\hat{\mathbf{y}}}) &:= E\{\psi_k^2\} + E\{\hat{\psi}_k^2\} & (4) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (f_{\psi}(\theta) + f_{\hat{\psi}}(\theta)) d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_{\mathbf{y}}(\theta) - f_{\hat{\mathbf{y}}}(\theta)| d\theta \\ &= \|f_{\mathbf{y}} - f_{\hat{\mathbf{y}}}\|_1. & (5) \end{aligned}$$

Obviously, this construction extends in the obvious way to the case of not-necessarily absolutely continuous power spectra as well, and the metric includes the measure of any discrepancies between the singular parts of the two spectral measures.¹ Clearly, $d(f_{\mathbf{y}}, f_{\hat{\mathbf{y}}})$ is a metric as seen from (5). Building on a similar rationale, in the next section, we develop a metric for covariance matrices.

3 A distance between covariance matrices

It is often the case that only a finite segment of the autocorrelation function of time-series \mathbf{y}_k and $\hat{\mathbf{y}}_k$ is available (and even then, possibly uncertain). Thus, it is of interest to consider distances between the partial autocorrelation statistics \mathbf{R} and $\hat{\mathbf{R}}$. To this end, we follow the dictum of the previous section and define as a distance measure the minimal combined variance of

¹ It will be interesting to explore the practical significance of other possibilities for quantifying distance such as

$$\frac{\int_{-\pi}^{\pi} (f_{\psi}(\theta) + f_{\hat{\psi}}(\theta)) d\theta}{\int_{-\pi}^{\pi} f_{\mathbf{z}}(\theta) d\theta} \quad \text{or} \quad \int_{-\pi}^{\pi} \frac{f_{\psi}(\theta) + f_{\hat{\psi}}(\theta)}{f_{\mathbf{z}}(\theta)} \frac{d\theta}{2\pi}.$$

random processes ψ_k and $\hat{\psi}_k$ for which (1) holds. Naturally, since only partial covariance samples are available, the random processes $\psi_k, \hat{\psi}_k$ and (1) need to be consistent with these data.

First denote by

$$\mathbf{R}_n := \begin{bmatrix} R_0 & R_1 & \dots & R_{n-1} \\ R_{-1} & R_0 & \dots & R_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ R_{-(n-1)} & R_{-(n-2)} & \dots & R_0 \end{bmatrix}$$

the $n \times n$ covariance matrix corresponding to \mathbf{y}_k and the covariance samples in \mathbf{R}_n and, similarly, $\hat{\mathbf{R}}_n$ for the Toeplitz matrix based on $\hat{\mathbf{R}}_n$. If $\mathbf{Q}_n, \hat{\mathbf{Q}}_n$ denote the corresponding finite Toeplitz covariances of the random processes ψ_k and $\hat{\psi}_k$, respectively, for which (1) holds, then

$$\mathbf{R}_n + \mathbf{Q}_n = \hat{\mathbf{R}}_n + \hat{\mathbf{Q}}_n \quad (6)$$

and the minimal sum $Q_0 + \hat{Q}_0$ of the respective variances can serve as a metric quantifying the distance between \mathbf{R} and $\hat{\mathbf{R}}$.

The computation of $\mathbf{Q}_n, \hat{\mathbf{Q}}_n$ minimizing the sum $Q_0 + \hat{Q}_0$, or equivalently minimizing

$$\frac{1}{n} \text{trace}(\mathbf{Q}_n + \hat{\mathbf{Q}}_n),$$

is a convex problem –since the positivity constraints are convex. The Toeplitz structure is peripheral, and the idea of defining such metrics extends equally well to non-negative definite Hermitian matrices and to more general positive operators. For notational convenience we develop the framework in the context of real symmetric matrices.

So, we let

$$\mathcal{M}_n := \{M \in \mathbb{R}^{n \times n} \mid M = M' \geq 0\}$$

be the cone of non-negative symmetric $n \times n$ -matrices and

$$\mathcal{T}_n := \{\mathbf{R} \in \mathcal{M}_n \mid \mathbf{R} \text{ is a Toeplitz matrix}\}$$

be the cone of non-negative Toeplitz matrices in \mathcal{M}_n . We address the case of matrices in \mathcal{M}_n and define a suitable metric, which is then specialized to \mathcal{T}_n .

Proposition 1. Let $M_1, M_2 \in \mathcal{M}_n$ and

$$\tau(M_1, M_2) := \min \left\{ \frac{1}{n} \text{trace}(M) \mid M \in \mathcal{M}_n, \right. \\ \left. M \geq M_1 \text{ and } M \geq M_2 \right\}.$$

Then

$$\delta(M_1, M_2) := 2\tau(M_1, M_2) - \text{trace}(M_1) - \text{trace}(M_2) \quad (7)$$

defines a metric on \mathcal{M}_n .

Proof. Given $M_1, M_2 \in \mathcal{M}_n$,

$$\mathcal{C}(M_1, M_2) := \{M \mid M \geq M_1 \text{ and } M \geq M_2\}$$

is a (convex) cone of non-negative definite matrices. It follows that there is an element $M_{12} \in \mathcal{C}(M_1, M_2)$ having minimal trace.

Clearly $\delta(M_1, M_2)$ is symmetric in its arguments and takes positive values unless $M_1 = M_2$, in which case $\delta(M_1, M_2) = 0$. Thus, we only need to prove the triangle inequality. Given $M_i \in \mathcal{M}_n$ for $i \in \{1, 2, 3\}$, we denote by M_{ik} corresponding minimal elements as above for $i, k \in \{1, 2, 3\}$, and we let

$$\Delta_{ik} := M_{ik} - M_k.$$

These matrices are non-negative by construction, the identities

$$M_i + \Delta_{ki} = M_k + \Delta_{ik}$$

hold, and

$$\delta(M_i, M_k) = \frac{1}{n} \text{trace}(\Delta_{ik} + \Delta_{ki})$$

for $i, k \in \{1, 2, 3\}$. But then,

$$\begin{aligned} M_1 + \Delta_{21} - \Delta_{12} &= M_2 \\ &= M_3 + \Delta_{23} - \Delta_{32}, \end{aligned}$$

and hence,

$$M_1 + \Delta_{21} + \Delta_{32} = M_3 + \Delta_{23} + \Delta_{12}.$$

From the minimal property of Δ_{13} and of Δ_{31} with regard to having the least value for the combined trace so that $M_1 + \Delta_{31} = M_3 + \Delta_{13}$, it follows that

$$\text{trace}(\Delta_{13} + \Delta_{31}) \leq \text{trace}(\Delta_{21} + \Delta_{32} + \Delta_{23} + \Delta_{12}).$$

Therefore,

$$\delta(M_1, M_2) + \delta(M_2, M_3) = \delta(M_1, M_3),$$

which completes the proof. \square

We now observe that the steps of the proof of Proposition 1 permit incorporating linear constraints on the structure of elements of \mathcal{M}_n , such as the constraint of all matrices being Toeplitz. Hence, whereas $\delta(\cdot, \cdot)$ may be used directly as a distance measure between elements of \mathcal{T}_n , the corresponding minimal-trace perturbations Δ_{ik} may not belong to \mathcal{T}_n in general. But, since the Toeplitz property is a linear constraint, we may define a completely analogous distance measure enforcing such perturbations (if so desired) to be Toeplitz.

Proposition 2. Let $M_1, M_2 \in \mathcal{T}_n$ and

$$\tau_{\mathcal{T}}(M_1, M_2) := \min \left\{ \frac{1}{n} \text{trace}(M) \mid M \in \mathcal{T}_n, \text{ and } M \geq M_1, M \geq M_2 \right\}.$$

Then

$$\delta_{\mathcal{T}}(M_1, M_2) := 2\tau_{\mathcal{T}}(M_1, M_2) - \text{trace}(M_1) - \text{trace}(M_2) \quad (8)$$

defines a metric on \mathcal{T}_n .

Proof. The proof follows the steps of the proof of Proposition 1 verbatim, except for the fact that we now constraint all matrices to belong to \mathcal{T}_n . \square

Proposition 3. Let $f_{\mathbf{y}}, f_{\hat{\mathbf{y}}}$ be power spectral densities, i.e., nonnegative and integrable on $[-\pi, \pi]$. Let as before $\mathbf{R}_n, \hat{\mathbf{R}}_n$ denote the corresponding Toeplitz covariance matrices, and let $n \in \{1, 2, \dots\}$. Then

$$\lim_{n \rightarrow \infty} \delta_{\mathcal{T}}(\mathbf{R}_n, \hat{\mathbf{R}}_n) = \|f_{\mathbf{y}} - f_{\hat{\mathbf{y}}}\|_1.$$

Proof. Clearly

$$\lim_{n \rightarrow \infty} \delta_{\mathcal{T}}(\mathbf{R}_n, \hat{\mathbf{R}}_n) \leq \|f_{\mathbf{y}} - f_{\hat{\mathbf{y}}}\|_1$$

since a choice of ψ_k and $\hat{\psi}_k$ with power spectra as in (3a-3b) gives rise to partial covariance matrices $\mathbf{Q}_n, \hat{\mathbf{Q}}_n$, for all n , for which (6) holds. The respective 0th elements Q_0 and \hat{Q}_0 remain the same for all n and the left hand side is

$$\|f_{\mathbf{y}} - f_{\hat{\mathbf{y}}}\|_1 = Q_0 + \hat{Q}_0$$

since the power spectra in (3a-3b) have no overlap in their support.

To show the converse inequality, consider the sequence of minimizing $\mathbf{Q}_n, \hat{\mathbf{Q}}_n$. These are Toeplitz matrices with bounded entries (since their corresponding 0th element is bounded by $\|f_{\mathbf{y}} - f_{\hat{\mathbf{y}}}\|_1$). Each can be extended to an infinite Toeplitz matrix, and thereby, gives rise to power spectral densities q_n and \hat{q}_n such that the first n Fourier coefficients of $f_{\mathbf{y}} + q_n$ and $f_{\hat{\mathbf{y}}} + \hat{q}_n$ coincide. The spectral densities q_n and \hat{q}_n can be obtained from $\mathbf{Q}_n, \hat{\mathbf{Q}}_n$ by any particular positive extension, for instance a “maximum entropy” one. We can take those as pairs, and since they are bounded there exists a subsequence weakly convergent to possibly non-negative measures, $d\mu$ and $d\hat{\mu}$, such that

$$f_{\mathbf{y}}d\theta + d\mu = f_{\hat{\mathbf{y}}}d\theta + d\hat{\mu}$$

since their Fourier coefficients must coincide. If $d\mu, d\hat{\mu}$ do have singular parts then these should be identical and the absolutely continuous parts must balance as well, so there exist power spectral densities q and \hat{q} such that

$$f_{\mathbf{y}} + q = f_{\hat{\mathbf{y}}} + \hat{q}. \quad (9)$$

But then,

$$\begin{aligned} \lim_{n \rightarrow \infty} \delta_T(\mathbf{R}_n, \hat{\mathbf{R}}_n) &\geq \lim_{n \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} (q_n(\theta) + \hat{q}_n(\theta)) d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (d\mu + d\hat{\mu}) \\ &\geq \frac{1}{2\pi} \int_{-\pi}^{\pi} (q + \hat{q}) d\theta \\ &\geq \|f_{\mathbf{y}} - f_{\hat{\mathbf{y}}}\|_1, \end{aligned}$$

the last inequality from (9). □

3.1 An example

The metric $\delta_T(\mathbf{R}_n, \hat{\mathbf{R}}_n)$ of the previous section admits no simple expression in terms of the respective eigenvalues. This should be contrasted with its limiting value $d(f_{\mathbf{y}}, f_{\hat{\mathbf{y}}})$ which is the L_1 distance between the corresponding power spectral densities. We highlight this with an example.

Let

$$\mathbf{R}_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

and

$$\hat{\mathbf{R}}_3 = \begin{bmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{bmatrix}.$$

Then, clearly,

$$\mathbf{Q}_3 = \begin{bmatrix} x & y & y \\ y & x & y \\ y & y & x \end{bmatrix}$$

and

$$\hat{\mathbf{Q}}_3 = \begin{bmatrix} x & v & v \\ v & x & v \\ v & v & x \end{bmatrix}$$

where

$$1 + y = 1/2 + v$$

and x is minimal subject to $\mathbf{Q}_3 \geq 0$ as well as $\hat{\mathbf{Q}}_3 \geq 0$. The last two inequalities imply that

$$\begin{aligned} 0 &\leq x \quad \text{as well as} \\ -\frac{1}{2}x &\leq y, v \leq x. \end{aligned}$$

It follows that the optimal choice (minimal x) is

$$\begin{aligned}x &= 1/3 \\y &= -1/6 \\v &= 1/3.\end{aligned}$$

Then

$$\delta_{\mathcal{T}}(\mathbf{R}_3, \hat{\mathbf{R}}_3) = 2x = 2/3,$$

while the respective eigenvalues are

$$\begin{aligned}\text{spec}(\mathbf{R}_3) &= \{3, 0, 0\} \text{ and} \\ \text{spec}(\hat{\mathbf{R}}_3) &= \{1, 1, 1/2\}.\end{aligned}$$

It appears that there is no simple expression for $\delta_{\mathcal{T}}(\mathbf{R}_3, \hat{\mathbf{R}}_3)$ based solely on knowledge of $\text{spec}(\mathbf{R}_3)$ and $\text{spec}(\hat{\mathbf{R}}_3)$.

The covariance \mathbf{R}_3 has a unique extension and corresponds to a measure with unit weight at $\theta = 0$, i.e., a spectral line (Dirac delta) at $\theta = 0$. Assuming that $\hat{\mathbf{R}}_3$ originates from a spectral measure which has a similar weight of amplitude $1/2$ at $\theta = 0$ and a uniform absolutely continuous part of amplitude $1/2$, then

$$\|d\mu - d\hat{\mu}\|_1 = 1/2 + 1/2 = 1$$

adding the L_1 -norm of the difference of the absolutely continuous parts with the absolute integral of the discrepancy between the two measures. We leave it as an exercise to the reader to verify that if $\hat{\mathbf{R}}_n$ is as we just assumed, namely $\hat{R}_k = 1/2$ for $k \geq 1$, and similarly, $R_k = 1$ for all k , then $\delta_{\mathcal{T}}(\mathbf{R}_n, \hat{\mathbf{R}}_n) \rightarrow 1$ as $n \rightarrow \infty$.

4 Approximating sample covariances

It is often the case that the autocovariance matrix \mathbf{R}_n of a random process \mathbf{y}_k is estimated in a way that does not guarantee this to be Toeplitz. For instance, it is quite common for \mathbf{R}_n to be estimated by averaging observation samples

$$\hat{\mathbf{R}}_n = \frac{1}{N+1} \sum_{\ell=0}^N \begin{bmatrix} y_{1+\ell} \\ \vdots \\ y_{n+\ell} \end{bmatrix} [y_{1+\ell} \cdots y_{n+\ell}]$$

The estimate $\hat{\mathbf{R}}_n$ is non-negative definite by construction, but may not be Toeplitz. Yet, for purposes of analysis it is often beneficial to approximate \mathbf{R}_n by a Toeplitz one, or possibly, by one with additional structure (e.g., corresponding to a moving average process or, more generally, to the state of a known dynamical system). The problem of seeking such an approximant which is closest to \mathbf{R}_n in $\delta(\cdot, \cdot)$, is readily solvable via convex optimization.

4.1 Comparison with the von Neumann entropy

In [1], the question was raised as to what are appropriate ways to approximate a given sample covariance with one that abides by a known linear structure. It was proposed that the Kullback-Leibler-von Neuman distance

$$\mathbb{S}(\hat{\mathbf{R}}\|\mathbf{R}) := \text{trace} \left(\hat{\mathbf{R}} \left(\log \hat{\mathbf{R}} - \log \mathbf{R} \right) \right)$$

provides a convenient convex functional for which the optimal approximant is uniquely defined. An academic example was presented in [1] which is recapitulated here as it helps underscore differences with approximation in the sense of minimizing $\delta(\hat{\mathbf{R}}, \mathbf{R})$.

Consider the positive-definite matrix below as the estimated value for a covariance matrix

$$\hat{\mathbf{R}}_3 = \frac{1}{3} \begin{bmatrix} 1.1 & .9 & 1.05 \\ .9 & .8 & .9 \\ 1.05 & .9 & 1.1 \end{bmatrix}.$$

The minimizer of

$$\{\mathbb{S}(\hat{\mathbf{R}}, \mathbf{R}) \mid \mathbf{R} \text{ being Toeplitz, } \mathbf{R} > 0, \text{trace}(\mathbf{R}) = \text{trace}(\hat{\mathbf{R}})\}$$

is unique (see [1]) and given by

$$\mathbf{R}_{3,\text{vN}} = \frac{1}{3} \begin{bmatrix} 1 & .942 & .957 \\ .942 & 1 & .942 \\ .957 & .942 & 1 \end{bmatrix}.$$

It is interesting to point out the the closest Toeplitz matrix to $\hat{\mathbf{R}}$ in the least-squares sense fails to be positive-definite ([1], cf. [2]). On the other hand, the optimal approximant in $\delta(\cdot, \cdot)$ -sense can be obtained by observation and is equal to

$$\mathbf{R}_{3,\delta} = \frac{1}{3} \begin{bmatrix} 1.1 & .9 & 1.05 \\ .9 & 1.1 & .9 \\ 1.05 & .9 & 1.1 \end{bmatrix}.$$

In the above, a second subscript indicates the sense in which the matrix approximates $\hat{\mathbf{R}}_3$. Obviously the traces of $\mathbf{R}_{3,\delta}$ and $\hat{\mathbf{R}}_3$ are not the same, in general. However, equality of the traces can be easily imposed as an added linear constraint.

4.2 Structured covariances

For purposes of illustration, consider a moving average process

$$\mathbf{y}_k = \mathbf{w}_k + \mathbf{w}_{k-1} + \mathbf{w}_{k-2}$$

where \mathbf{w}_k is a zero-mean, unit-variance, Gaussian white noise process. The autocorrelation sequence of \mathbf{y}_k is

$$[R_0 \ R_1 \ R_2 \ R_3 \ 0 \ \dots] = [3 \ 2 \ 1 \ 0 \ 0 \ \dots].$$

Simulating \mathbf{y}_k over a window $k \in \{0, 1, \dots, 100\}$, and based on a particular such realization, the corresponding $n \times n$ sample covariance matrix, for $n = 5$, was computed to be

$$\hat{\mathbf{R}}_5 = \begin{bmatrix} 4.0362 & 2.9053 & 1.8043 & 0.4042 & 0.1718 \\ 2.9053 & 4.0547 & 2.9268 & 1.7945 & 0.3800 \\ 1.8043 & 2.9268 & 4.0792 & 2.9143 & 1.7733 \\ 0.4042 & 1.7945 & 2.9143 & 4.0819 & 2.9421 \\ 0.1718 & 0.3800 & 1.7733 & 2.9421 & 4.0237 \end{bmatrix}.$$

Obviously, this matrix is not Toeplitz due to the finiteness of the observation record. The closest Toeplitz approximant to $\hat{\mathbf{R}}_5$, in the sense of the metric $\delta(\cdot, \cdot)$, turns out to be

$$\mathbf{R}_{5, \text{Toeplitz}} = \begin{bmatrix} 4.0677 & 2.9237 & 1.7912 & 0.3979 & 0.1822 \\ 2.9237 & 4.0677 & 2.9237 & 1.7912 & 0.3979 \\ 1.7912 & 2.9237 & 4.0677 & 2.9237 & 1.7912 \\ 0.3979 & 1.7912 & 2.9237 & 4.0677 & 2.9237 \\ 0.1822 & 0.3979 & 1.7912 & 2.9237 & 4.0677 \end{bmatrix}$$

for which

$$\delta(\hat{\mathbf{R}}_5, \mathbf{R}_{5, \text{Toeplitz}}) = 0.0308.$$

Interestingly, $\mathbf{R}_{5, \text{Toeplitz}}$ does not correspond to a moving average process of order 2 (or even, of order 3, 4) as it can be readily verified by the fact that the trigonometric polynomials, e.g.,

$$\sum_{k=-4}^4 R_k e^{jk\theta}$$

takes negative values.

The set of covariance matrices which are generated by moving average processes of a given order, is convex and admits a characterization via a set of linear matrix inequalities ([8, 3]). Thus, the closest approximant to $\hat{\mathbf{R}}$ which corresponds to a moving average process of any given order can be readily computed. In particular, if we specify the order to be 2, then the optimal approximant to $\hat{\mathbf{R}}_5$ becomes

$$\mathbf{R}_{5, \text{MA}(2)} = \begin{bmatrix} 3.9945 & 2.1588 & 0.5693 & 0 & 0 \\ 2.1588 & 3.9945 & 2.1588 & 0.5693 & 0 \\ 0.5693 & 2.1588 & 3.9945 & 2.1588 & 0.5693 \\ 0 & 0.5693 & 2.1588 & 3.9945 & 2.1588 \\ 0 & 0 & 0.5693 & 2.1588 & 3.9945 \end{bmatrix}$$

for which

$$\delta(\hat{\mathbf{R}}_5, \mathbf{R}_{5, \text{MA}(2)}) = 1.2161.$$

References

1. T.T. Georgiou (2003) Toeplitz covariance matrices and the von Neumann relative entropy. In: K. Hashimoto, Y. Oishi, and Y. Yamamoto (eds) Control and Modeling of Complex Systems: Cybernetics in the 21st Century. MA: Birkhauser, Boston
2. T.T. Georgiou (2006) Relative Entropy and the multi-variable multi-dimensional Moment Problem, IEEE Trans. on Information Theory, 52(3): 1052-1066.
3. T.T. Georgiou (2006) Decomposition of Toeplitz matrices via convex optimization, IEEE Signal Processing Letters, 13: 537-540.
4. T.T. Georgiou (2007) Distances and Riemannian metrics for spectral density functions, IEEE Trans. on Signal Processing, to appear. Preprint available at: <http://arxiv.org/abs/math/0608486/>
5. T.T. Georgiou (2007) An intrinsic metric for power spectral density functions, IEEE Signal Processing Letters, to appear. Preprint available at: <http://arxiv.org/abs/math/0607026/>
6. R. Gray, A. Buzo, A. Gray, and Matsuyama (1980) Distortion measures for speech processing, IEEE Trans. on Acoustics, Speech, and Signal Proc., 28(4): 367-376.
7. P. Stoica and R. Moses (2005) Introduction to Spectral Analysis. Prentice Hall
8. P. Stoica, T. McKelvey, and J. Mari (2000) MA estimation in polynomial time, IEEE Trans. on Signal Processing, 48: 1999-2012.