Cramér-Rao Bounds for Toeplitz Covariance Estimation

Michael J. Turmon
School of Electrical Engineering
Cornell University
Ithaca, NY 14853

June 1994

1 Introduction

Suppose we have a $G \times 1$ sample y_G from a stationary complex-valued random process and we wish to estimate its $G \times G$ covariance matrix K_G, which must of course be Toeplitz as well as positive definite and hermitian. We assume that y_G has the circularly symmetric complex Gaussian distribution

$$f(y_G; K_G) = \exp\left[g(S; K_G) - G\log(\pi)\right] \quad (1)$$
$$g(S; K_G) = -\log|K_G| - \operatorname{tr}(K_G^{-1}S) \quad (2)$$

where S is the $G \times G$ sample covariance. (If there are $P > 1$ independent copies of y_G, the above expressions remain valid if the density is raised to the power P.) Let the estimate be constrained to lie in the set $\mathcal{I}$.

In finding the MLE or the CR bound we are concerned with various choices of constraint set. Following Dembo et al. [1, 2] we define

Set	Description	Properties
Π	hermitian matrices ≥ 0	
Π^+	hermitian matrices > 0	$\Pi^+ \subset \Pi$
$\mathcal{K}_G$	$G \times G$ Toeplitz matrices $\in \Pi$	convex cone
$\mathcal{C}_N$	$N \times N$ circulant matrices $\in \Pi$	convex cone
$\mathcal{K}_G\mathcal{C}_N$	matrices $\in \mathcal{K}_G$ with extension to $\mathcal{C}_N$	convex cone, $\mathcal{K}_G\mathcal{C}_N \subseteq \mathcal{K}_G$

Additionally we define $\mathcal{K}_G^+$, $\mathcal{C}_N^+$, $\mathcal{K}_G^+\mathcal{C}_N$, $\mathcal{K}_G^+\mathcal{C}_N^+$ as the obvious positive-definite versions of the sets above. Covariance estimates will be drawn from either $\mathcal{K}_G^+$ or $\mathcal{K}_G^+\mathcal{C}_N$.

There are two motivations for the introduction of the circulant extension constraint. First, if $\operatorname{rank} S < G$ and $\mathcal{I} = \mathcal{K}_G$, Fuhrmann and Miller [3] have shown that with high probability ($\geq 1 - 2^{-G+2}$) the maximum-likelihood covariance estimate is singular, which is at odds with the problem statement assuming

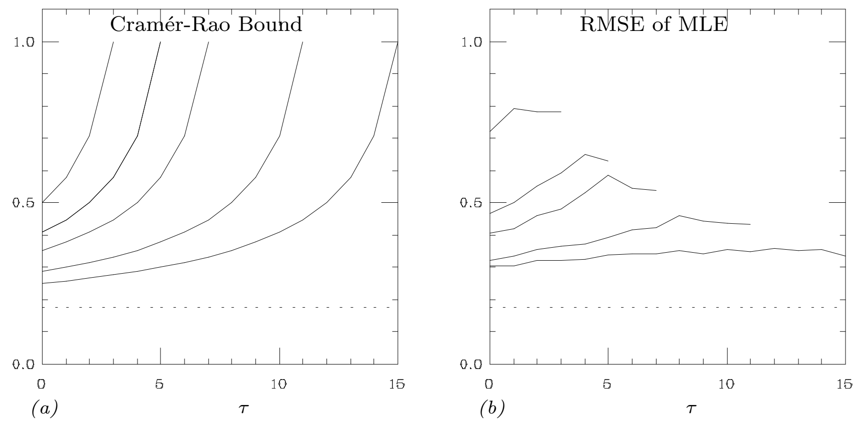

Figure 1: RMS errors for a white process for $G = 4$, 6, 8, 12, and 16, plotted versus lag number τ. The MLE and the CR bound are shown.

a nonsingular covariance. However the authors also show that if $\mathcal{I} = \mathcal{K}_G\mathcal{C}_N$, the MLE is almost surely nonsingular. The second is more pragmatic: a well-behaved, reasonably quick iterative method [4, 5] is avaliable to find the MLE in the last case. (However the seminal paper [6] and the later [7] each contain a method for finding the MLE when $\mathcal{I} = \mathcal{K}_G$.)

One open question in this framework is the effect introducing the rather extraneous circulant constraint has on the estimation process. One would expect that as long as the underlying covariance belongs to $\mathcal{K}_G\mathcal{C}_N$ that the constraint would lower the estimation error; it certainly cannot raise the error. Examples like the one in figure 1 have led us to believe that the circulant constraint does lower estimation error. The first panel shows the CR bound on standard deviation computed assuming $\mathcal{I} = \mathcal{K}_G$, for various values of G for a white process. In the second panel are the RMS errors of the MLE, which is the Miller-Snyder EM estimator [4] with $N = 32$ in each case. The RMSE is computed via upwards of 500 independent trials. It is apparent that the RMSE of the MLE is substantially lower than indicated by the CR bound, and a plausible reason for the discrepancy is that the CR bound did not incorporate the circulant extension constraint. This is the issue we wish to address in this memo, and along the way we hope to gain a better understanding of the family of constraint sets introduced above.

Finally, a remark about related problems. If the observed process is real, its covariance is real and even, and so is the corresponding spectrum. The density of the observations is just the square root of that above, and the significant part of the loglikelihood is unchanged. In this case one can use cosines alone, rather than complex exponentials, as basis vectors in representing a covariance with circulant extension. The issues are essentially the same, but we feel the complex

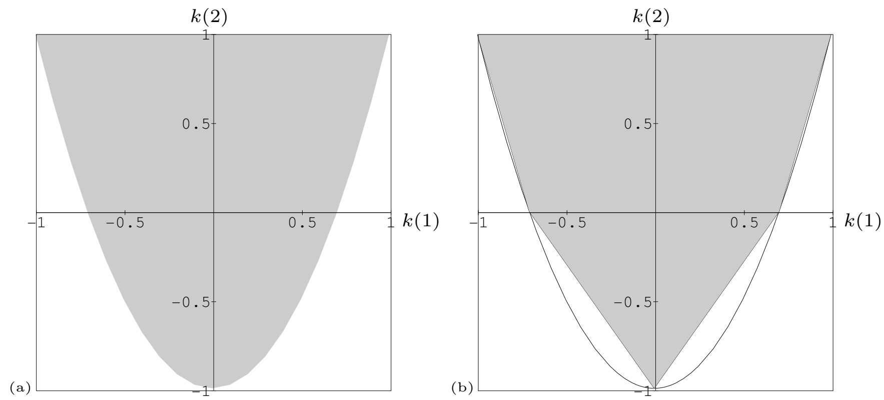

Figure 2: Normalized constraint sets for $G = 3$. In panel (b), $N = 8$ and the boundary of $\mathcal{K}_G$ is indicated.

setting is the more natural.

2 Understanding the Constraint

The stationary covariance K_G of interest to us can be represented uniquely by its top row, which we express as the $G \times 1$ complex-valued column vector k_G. Note that the last three sets in the table above, as well as their positive cohorts, are in the same way isomorphic to corresponding subsets of C^G (or C^N), all of which are again convex cones. (We will find it useful to identify a covariance matrix and the vector generating it, as well as the corresponding sets.) Due to the scaling property of cones it is useful to define the sliced versions

$$\bar{\mathcal{A}} = \mathcal{A} \cap \{z \in C^k : z(0) = 1\} \qquad (k = G \text{ or } N) \tag{3}$$

all of which are again convex. Note that a covariance vector $k_G \in \mathcal{A}$ if and only if $\alpha k_G \in \bar{\mathcal{A}}$ for some real $\alpha > 0$ (the nonnegative versions may want to include also the trivial zero covariance).

2.1 An example

It is illuminating to study the above constraint sets in the simple case where $G = 3$ and the underlying process is real. A covariance $[1\ k(1)\ k(2)]^T \in \bar{\mathcal{K}}_G$ if the determinants of its three principal submatrices are non-negative, which is equivalent to

$$-1 \le k(1) \le 1 \qquad -1 + 2k(1)^2 \le k(2) \le 1 \tag{4}$$

(For $\bar{\mathcal{K}}_G^+$, all inequalities are strict.) The region $\bar{\mathcal{K}}_G$ is depicted in figure 2a.

To find $\overline{\mathcal{K}_G\mathcal{C}_N}$, recall that, to have circulant extension to length N, a sequence must be of the form

$$k_G = W_G^\dagger \Sigma \tag{5}$$

where $\Sigma \geq 0$ (componentwise) is an $N \times 1$ vector of spectral weights, and W_G is the $N \times G$ principal submatrix of the $N \times N$ DFT matrix W_N ($W_N W_N^\dagger = I$). Moving the normalizing $N^{-1/2}$ from W_G to Σ, note that (5) is the same as saying that k_G is a nonnegative linear combination of vectors of the form $[1\ e^{i\frac{2\pi}{N}k}\ e^{i\frac{2\pi}{N}2k}]^T$, $k = 0, \ldots, N-1$. For such a superposition to have first component equal to unity, the weights must sum to unity. So for complex-valued data,

$$\overline{\mathcal{K}_G\mathcal{C}_N} = \text{conv.hull}(\{[1\ e^{i\frac{2\pi}{N}k}\ e^{i\frac{2\pi}{N}2k}]^T\}_{k=0}^{N-1})$$

This region is a linear transformation of the face of the unit simplex in R^N by the matrix $\sqrt{N}W_G^\dagger$, into some convex subset of C^G.

In this case, we are only interested in the purely real members of $\overline{\mathcal{K}_G\mathcal{C}_N}$, those having even spectral coefficients. This amounts to using cosines as basis vectors, and

$$\overline{\mathcal{K}_G\mathcal{C}_N} = \text{conv.hull}(\{[1\ \cos(\frac{2\pi}{N}k)\ \cos(\frac{2\pi}{N}2k)]^T\}_{k=0}^{N/2}) \tag{6}$$

where we assume for simplicity that N is even. The generators and the resulting convex polygon for $N = 8$ are shown in figure 2b. For larger N the polygon fills out more and more.

As for the positive relatives of $\overline{\mathcal{K}_G\mathcal{C}_N}$, note that $\overline{\mathcal{K}_G^+\mathcal{C}_N} = \overline{\mathcal{K}_G\mathcal{C}_N} \cap \overline{\mathcal{K}_G^+}$. In the case at hand, this means all but the boundary of (4); or all but the generators of the polygon. The set $\overline{\mathcal{K}_G\mathcal{C}_N^+}$ is just as in (6), except that only positive convex combinations are allowed. Thus the boundary of the polygon as well as its generators are excluded. Finally it is clear that $\overline{\mathcal{K}_G^+\mathcal{C}_N^+} = \overline{\mathcal{K}_G\mathcal{C}_N^+}$.

2.2 General description

In general we have these results, stated in the language of Dembo et al. [2], although the ideas and proofs go back at least to [8].

Theorem 1 $\overline{\mathcal{K}_G\mathcal{C}_N}$ is the convex polygon in C^G generated by $\{w(k)\}_{k=0}^{N-1}$ with $w(k) = [1\ \rho^k\ \rho^{2k}\ \cdots\ \rho^{(G-1)k}]^T$ and $\rho = e^{i\frac{2\pi}{N}}$. That is,

$$\overline{\mathcal{K}_G\mathcal{C}_N} = \text{conv.hull}(w(0), w(1), \ldots, w(N-1)) \qquad . \Box$$

Translating back to the original matrix formulation, we see that

$$K_G \in \mathcal{K}_G\mathcal{C}_N \Longleftrightarrow K_G = \sum_{i=0}^{N-1} \sigma(i)(w(i)w(i)^\dagger)^* \qquad , \tag{7}$$

where $*$ is complex conjugate. (We feel this formulation suppresses some geometrical intuition.)

Theorem 2

$$\bar{\mathcal{K}}_G = \text{conv.hull}(w(t), t \in [0,1))$$

where $w(t) = [1 \ \rho_t \ \rho_t^2 \ \cdots \ \rho_t^{G-1}]^T$ and $\rho_t = e^{i2\pi t}$. □

In this case

$$K_G \in \mathcal{K}_G \Longleftrightarrow K_G = \int_0^1 (w(t)w(t)^\dagger)^* d\sigma(t) \qquad ,$$

which is of course a classical result for Toeplitz forms.

3 Cramér-Rao Bounds

3.1 Toeplitz case

Finding a CR bound for the case $\mathcal{I} = \mathcal{K}_G$ is simple; the only complication is that we must take into account that the covariance to be estimated is complex. We might like to take derivatives of the loglikelihood $g(S; K_G)$ with respect to the (complex) entries in K_G; however g is not analytic (it is purely real) so these derivatives do not exist. This forces us to take only directional derivatives, or to put it another way to break the complex entries in K_G into real and imaginary parts.

Write K_G in terms of basis matrices B_m and real coefficients κ:

$$\begin{aligned} K_G &= \sum_{m=0}^{G-1} \kappa_m \Re\{B_m\} + j \sum_{m=G}^{2G-2} \kappa_m \Im\{B_{m-G-1}\} \\ B_m(r,s) &= \begin{cases} 1 & r-s = m = 0 \\ 1+j & r-s = m \neq 0 \\ 1-j & s-r = m \neq 0 \\ 0 & |s-r| \neq m \end{cases} \end{aligned} .$$

Each matrix B_m is Hermitian, having $1 \pm j$ along its mth sub- and super-diagonals, expressing the constraint $K_G \in \mathcal{K}_G$. The bound will be calculated using

$$Var\{\hat{K}(\tau)\} = \begin{cases} Var\{\hat{\kappa}_\tau\} & \tau = 0 \\ Var\{\hat{\kappa}_\tau\} + Var\{\hat{\kappa}_{\tau+G-1}\} & \tau \neq 0 \end{cases} \tag{8}$$

The $(2G-1) \times (2G-1)$ Fisher information matrix J_{K1} is computed from the Hessian of g. The derivatives are

$$\begin{aligned} \frac{\partial g(S; k_G)}{\partial \kappa_n} &= \text{tr}\left[\left(K_G^{-1} S K_G^{-1} - K_G^{-1}\right) \frac{\partial K_G}{\partial \kappa_n}\right] \\ \frac{\partial g(S; k_G)}{\partial \kappa_m \partial \kappa_n} &= \text{tr}\left[K_G^{-1} \frac{\partial K_G}{\partial \kappa_n} K_G^{-1}\left(K_G - 2S\right) K_G^{-1} \frac{\partial K_G}{\partial \kappa_n}\right] \end{aligned}$$

from which

$$J_{K1}(m,n) = -E\frac{\partial^2 g(s;k_G)}{\partial\kappa_m\partial\kappa_n} = \mathrm{tr}\left(K_G^{-1}\frac{\partial K_G}{\partial\kappa_m}K_G^{-1}\frac{\partial K_G}{\partial\kappa_n}\right) \qquad . \tag{9}$$

The indicated derivatives are simply expressed as

$$\frac{\partial K_G}{\partial\kappa_m} = \begin{cases} \Re\{\mathbf{B}_m\} & 0 \le m < G \\ j\Im\{\mathbf{B}_{m-G+1}\} & G \le m \le 2G-2 \end{cases} \tag{10}$$

which is sufficient to find the bound.

3.2 Circulant case

Now the covariance is parameterized in terms of the spectral coefficients as in (7):

$$K_G = \sum_{i=0}^{N-1}\sigma(i)w(i)^* w(i)^T \qquad , \tag{11}$$

which expresses both the circulant extension and the Toeplitz constraints. Since the loglikelihood remains the same, we can adapt (9):

$$\begin{aligned} J_\Sigma(m,n) &= \mathrm{tr}\left(K_G^{-1}\frac{\partial K_G}{\partial\sigma(m)}K_G^{-1}\frac{\partial K_G}{\partial\sigma(n)}\right) \\ &= \mathrm{tr}\left(K_G^{-1}w(n)^*w(n)^T K_G^{-1}w(m)^*w(m)^T\right) \\ &= \mathrm{tr}\left(w(m)^T K_G^{-1}w(n)^*w(n)^T K_G^{-1}w(m)^*\right) \\ &= |w(m)^T K_G^{-1}w(n)^*|^2 \\ J_\Sigma &= |W_G K_G^{-1}W_G^\dagger|^2 \qquad , \end{aligned}$$

where $|\cdot|^2$ is applied elementwise to a matrix argument. This result is also derived in [9], where it is also shown that this matrix has rank at most $2G-1$.

To convert J_Σ into a Fisher information for the covariance parameters, note that if an estimation problem may be parameterized by either η or θ, where $\eta = A\theta$ for $A \in R^{m\times n}$, then $J_\theta = A^T J_\eta A$, as is simple to verify. In the instant case the real covariance vector $\kappa = W_{GR}^T\Sigma$ so that

$$J_{K2} = W_{GR}|W_G K_G^{-1}W_G^\dagger|^2 W_{GR}^T \qquad , \tag{12}$$

where W_{GR} is a real $N\times(2G-1)$ matrix formed by removing the zero column from $[\Re(W_G) - \Im(W_G)]$. J_{K2} has full rank for the examples we have investigated.

We had expected to find $J_{K1} < J_{K2}$ so that the CR bound with the constraint $k_G \in \mathcal{K}_G$ would be higher than that assuming $k_G \in \mathcal{K}_G\mathcal{C}_N$. However, the remarkable thing is: for $N \ge 2G-1$, $J_{K1} = J_{K2}$. While I have not tried to prove this formally, it is true for many examples of various spectral shapes that I have tried. The circulant extension constraint yields no variance reduction. If we think about the estimation problem in the light of what we have seen

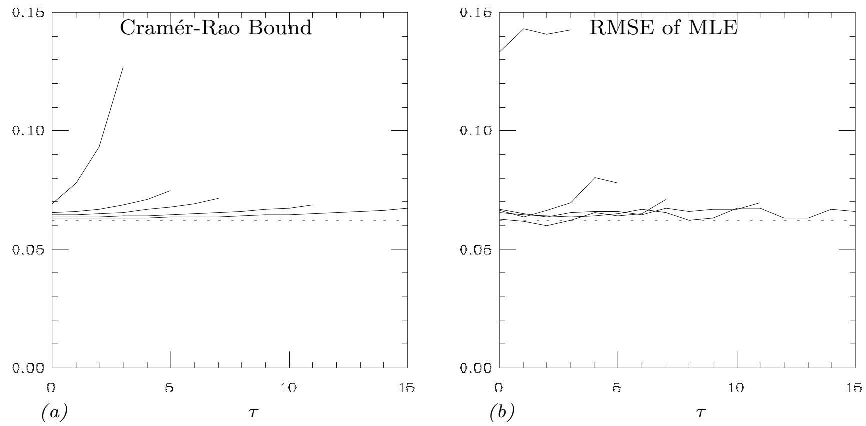

Figure 3: RMS errors for an impulsive process for $G = 4$, 6, 8, 12, and 16, plotted versus lag number τ. The MLE and the CR bound are shown.

above about the geometry of the constraint sets, this begins to make sense. In the first scheme (K1), we represent the covariance as a linear combination of $2G-1$ Toeplitz basis matrices. The CR bound is on the coefficients κ of these bases, which are the covariance parameters themselves. The circulant approach represents the covariance linearly via N basis matrices. Since there are essentially $2G-1$ parameters in the problem, only this many spectral coefficients are really estimable, which is of course why the Fisher information has rank at most $2G-1$ in this case. The Fisher information for the spectral parameters is then converted into one for the covariances. What we have is essentially two equivalent linear parameterizations of the covariance.

The only way the circulant constraint really matters is in that the spectral coefficients must be nonnegative, which is why the two parameterizations are only "essentially" equivalent. (If it were not for this, clearly any G-length vector—positive-definite or not—with purely real first element could be reached by a combination of the $N \geq 2G-1$ DFT basis vectors.) But such orthant constraints do not affect Cramér-Rao bounds ([10], lemma 4). The intuition for this is that the Fisher information is the mean curvature of the loglikelihood along directions of infinitesimal variation of the parameter. If the true parameter is in the interior of the constraint set, this rate is unaffected by the constraint. And even if the parameter is on the boundary, feasible variations (having the same curvature as before) exist in the directions *away* from the boundary.

4 Conclusions

Motivated by the fact that the mean-squared error of the MLE falls below the CR bound (figure 1), we hypothesized that the circulant extension constraint was helping the MLE to beat the Toeplitz CR bound. This motivated an examination of the difference between the constraint regions corresponding to the Toeplitz versus Toeplitz plus circulant conditions. It was seen that the former is a cone generated by the convex hull of a continuum of complex exponential basis vectors, while the latter is a cone generated by the convex hull of N such basis vectors. As N increases the second convex hull includes more and more of the first.

We derived the CR bound for covariance estimation under the Toeplitz constraint, and another CR bound for estimation under the Toeplitz plus circulant constraint. Although we had not expected it, the two bounds are identical. In retrospect it is evident that this should be the case because the two approaches essentially just parameterize the covariance in different ways. The circulant constraint only enters the parameterization as an inequality condition, which cannot affect the bound.

We are therefore left with a greater understanding of, but no solution for, the original problem presented in figure 1 of the MLE's beating the CR bound. The case of a much different spectrum, containing several impulses superimposed on low-level background noise, is presented in figure 3. In this case the MLE does not outperform the CR bound (to the accuracy of simulation): the phenomenon depends on the shape of the covariance.

Another option left explaining this phenomenon is that the MLE is biased, allowing it to perform relatively better for white spectra than for rough ones. The simulation studies so far have not shown a significant bias, however. This leaves us with the conclusion that the iterative algorithm may not be precisely computing the MLE due to incomplete convergence or multimodality.

References

[1] A. Dembo. The relation between maximum-likelihood estimation of structured covariance matrices and periodograms. *IEEE Trans. Sig. Proc.*, 34(6):1661–2, December 1986.

[2] A. Dembo, C. L. Mallows, and L. A. Shepp. Embedding nonnegative-definite Toeplitz matrices in nonnegative-definite circulant matrices, with application to covariance estimation. *IEEE Trans. Infor. Theory*, 35(6):1206–12, November 1989.

[3] D. R. Fuhrmann and M. I. Miller. On the existence of positive-definite maximum-likelihood estimates of structured covariance matrices. *IEEE Trans. Infor. Theory*, 34(4):722–9, July 1988.

[4] M. I. Miller and D. L. Snyder. The role of likelihood and entropy in incomplete-data problems. *Proc. IEEE*, 75(7):892–907, July 1987.

[5] M. J. Turmon and M. I. Miller. Maximum-likelihood estimation of constrained means and Toeplitz covariances with application to direction-finding. *IEEE Trans. Sig. Proc.*, 42(5):1074–1086, May 1994.

[6] J. P. Burg, D. G. Luenberger, and D. L. Wenger. Estimation of structured covariance matrices. *Proc. IEEE*, 70(9):963–74, September 1982.

[7] D. L. Snyder, J. A. O'Sullivan, and M. I. Miller. The use of maximum-likelihood estimation for forming images of diffuse radar-targets from delay-doppler data. *IEEE Trans. Infor. Theory*, 35(3):536–548, May 1989.

[8] U. Grenander and G. Szegő. *Toeplitz Forms and Their Applications.* Chelsea, second edition, 1984.

[9] J. A. O'Sullivan, P. Moulin, and D. L. Snyder. Cramér-Rao bounds for constrained spectrum estimation with application to a problem in radar imaging. Technical Report ESSRL-88-9, ESSRL, Washington University, 1989.

[10] J. D. Gorman and A. O. Hero. Lower bounds for parametric estimation with constraints. *IEEE Trans. Infor. Theory*, 36(6):1285–1301, November 1990.