A METHOD FOR COMPUTING THE INFORMATION MATRIX OF STATIONARY GAUSSIAN PROCESSES

José M. B. Dias and José M. N. Leitão
Instituto de Telecomunicações and
D.E.E.C., Instituto Superior Técnico
Tel: +351 1 8418466; fax: +351 1 8418472
Email: edias@beta.ist.utl.pt

ABSTRACT

This paper proposes a new method for the efficient computation of the Fisher information matrix of zero-mean complex stationary Gaussian processes. Its complexity (measured by the number of floating point operations) is smaller than the fastest previously available procedure. The key idea exploited is that the Fisher information matrix depends only on the sum of the diagonals of the inverse covariance matrix derivative (with respect to the model parameters), rather than on the whole matrix. To obtain the referred sum, a new efficient technique, built upon the Trench algorithm for computing the inverse of a Toeplitz matrix, is presented.

1 INTRODUCTION

The *Cramér-Rao bound* (CRB) [1], [2] plays an important rule in parameter estimation:

(a) The CRB is, for the class of unbiased estimators, a lower bound on the error variance;

(b) The CRB is attainable by maximum likelihood estimators, at least asymptotically;

(c) Very often, the variance of a particular estimator is not known; resorting to the CRB is, in this case, a usual procedure.

The CRB is obtained from the *Fisher information matrix*; given a finite sample size N, the computation of the Fisher information matrix of a zero-mean complex stationary Gaussian process includes matrix inversions and matrix multiplications [3] (not to mention the computation of the derivatives). Therefore, the total complexity, measured in floating point operations (flops) is $O(N^3)$, if the number of parameters is independent of N. Should the number of parameters be proportional to N, the mentioned complexity grows to $O(N^4)$.

1.1 Previous Work

The normalized Fisher information matrix of a zero-mean (or with mean independent of the parameter vector) complex stationary Gaussian process tends, with the sample size, to Whittle's formula [4]. Zeira and Nehorai [5] generalized this result for non zero-mean processes. Whittle's formula, despite leading to closed and simple expressions with light complexity (ARMA processes are a relevant example), does not fit accurately the Fisher information matrix for *small* sample sizes [6].

As it was pointed before, the complexity in computing the Fisher information matrix varies between $O(N^3)$ and $O(N^4)$. Such a complexity is not tolerable in many applications, and has fostered research towards more efficient algorithms; namely:

1. Porat and Friedlander [7], based on the Levinson-Durbin algorithm for computing the orthogonal polynomials of a Toepliz matrix, proposed an algorithm for the exact computation of the Fisher information matrix. The algorithm computes the information matrices for all orders between 1 and N in $O(N^2)$ flops;

2. Giannella [8], Porat and Friedlander [6], and Tuan [9], based on the Gohberg-Semencul decomposition, proposed methods for AR(p) processes;

3. Mélard and Klein [10] introduced an algorithm for ARMA processes; it takes advantage of the state equation associated to the rational transfer function, jointly with the Chandrasekhar [11] recursion, used for the computation of the likelihood function. The algorithm complexity is, roughly, $O(N(p+q)^2 \max(p, q+1)^2)$; despite its slow growing rate with the sample size (linear with N), the term $(p+q)^2 \max(p, q+1)^2$ is, frequently, too penalizing.

1.2 Proposed Approach

We begin by showing that in zero-mean (complex or not) stationary Gaussian processes, the Fisher information matrix depends only on the sum of the diagonals (from now on we refer to *diagonal sum*) of the inverse covariance matrix derivative (with respect to the model parameters), and not on the whole matrix. With this fact in mind, it is then proved (based on the Trench algorithm [12] for determining the inverse of a Toeplitz matrix) that the diagonal sum of the referred matrix can be computed

with $O(N \ln N)$ complexity. On computing the diagonal sum, it is necessary to solve a Toeplitz system; by using the preconditioned conjugate gradient (PCG) technique (see [13] and [14]), this has $O(N \ln N)$ complexity. The total complexity of the method is therefore $O(N \ln N)$.

2 PROBLEM FORMULATION

Let $\mathcal{Y} = \{y(t), t \in \Re\}$ be a zero-mean complex stationary Gaussian process and $\mathbf{Y} = [Y_1, \ldots, Y_N]^T$, with $Y_i = y(iT_p)$ for $i = 1, \ldots, N$, a random vector. We assume that the random variable $y(t) = y_r(t) + jy_i(t)$ verifies

$$E\{y(t)y(s)\} = 0, \tag{1}$$

for all $t, s \in \Re$. Property (1) allows writing the p.d.f. of $\mathbf{Y}$ as [15, p. 77]

$$p_y(\mathbf{Y} = \mathbf{y}; \theta) = \frac{1}{\pi^N |\mathbf{R}_y(\theta)|} \exp\left[-\mathbf{y}^H \mathbf{R}_y^{-1}(\theta)\mathbf{y}\right], \tag{2}$$

where θ stands for the parameter vector, and $\mathbf{R}_y(\theta) = E\{\mathbf{Y}\mathbf{Y}^H\}$ is a positive definite Hermitian Toeplitz and differentiable (with respect to θ) matrix, for $\theta \in \mathbf{\Theta} \subset \Re^p$, with $p \in \{1, 2, \ldots\}$. Under this conditions, the Fisher information matrix $J = [J_{kl}]$ is given by (see, e.g. [15, p. 144])

$$\begin{aligned} J_{kl}(\theta) &= \mathrm{tr}\{\mathbf{R}_y^{-1}\mathbf{R}_y^{(k)}\mathbf{R}_y^{-1}\mathbf{R}_y^{(l)}\} &(3)\\ &= -\mathrm{tr}\{\mathbf{R}_y^{(k)}\mathbf{R}_y^{-(l)}\}, &(4)\end{aligned}$$

where $\mathbf{A}^{(i)}$ e $\mathbf{A}^{-(i)}$ denote $\partial\mathbf{A}/\partial\theta_i$ and $\partial\mathbf{A}^{-1}/\partial\theta_i$, respectively.

If $J^{-1}(\theta)$ exists, the covariance matrix of any unbiased estimator $\widehat{\theta}(\mathbf{Y})$ obeys to (see e.g. [16, cap. 15.7])

$$\mathbf{V}(\theta) = E\{(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)^H\} \geq J^{-1}(\theta), \tag{5}$$

where $\mathbf{A} \geq \mathbf{B}$ means that matrix $\mathbf{A}-\mathbf{B}$ is positive semidefined.

Computing J, according to (3), requires three multiplications and one inversion involving N-matrices (square matrices of size N); both operations have $O(N^3)$ complexity. If the number of parameters is proportional to N, the total complexity in computing J grows to $O(N^4)$. Nevertheless, if the matrix to invert is Toeplitz, the application of the Trench algorithm [12] reduces the complexity of this operation to $O(N^2)$. The total complexity is, still, $O(N^3)$ if the number of parameters is independent of N, or $O(N^4)$ if the number of parameters is proportional to N.

3 EFFICIENT COMPUTATION OF $\mathrm{tr}\{\mathbf{A}\delta\mathbf{B}^{-1}\}$

The computation of the Fisher information matrix is, normally, based on expression (3). However, the results next presented take advantage of expression (4); the latter has the structure $\mathrm{tr}\{\mathbf{A}\delta\mathbf{B}^{-1}\}$ (symbol δ denotes the derivative in order to any parameter vector component), where $\mathbf{A}$ and $\mathbf{B}$ are Hermitian and Toeplitz N-matrices. As a step towards computing $\mathrm{tr}\{\mathbf{A}\delta\mathbf{B}^{-1}\}$, the operator $\mathrm{tr}\{\mathbf{A}\mathbf{B}^{-1}\}$ is considered firstly.

Denoting $\mathbf{A} = [a_{ij}] = [a_{j-i}]$ and $\mathbf{B}^{-1} = [b_{ij}]$, it follows that

$$\begin{aligned} \mathrm{tr}\{\mathbf{A}\mathbf{B}^{-1}\} &= \sum_{i,j=1}^{N} a_{ij}b_{ji} &(6)\\ &= \sum_{\tau=-N+1}^{N-1} a_\tau \sum_{i(\tau)} b_{i+\tau,i} &(7)\\ &= \sum_{\tau=-N+1}^{N-1} a_\tau \overline{b}_{-\tau}, &(8)\end{aligned}$$

where $\overline{b}_{-\tau} = \overline{b}_\tau^*$ and

$$\overline{b}_\tau = \begin{cases} \sum_{i=1}^{N-\tau} b_{i,i+\tau} & \tau = 0, \ldots, N-1 \\ \sum_{i=1-\tau}^{N} b_{i,i+\tau} & \tau = -N+1, \ldots, 0. \end{cases} \tag{9}$$

According to (8), $\mathrm{tr}\{\mathbf{A}\mathbf{B}^{-1}\}$ depends on a_τ and on $\overline{b}_\tau$ (sum of the elements of $\mathbf{B}^{-1}$ along diagonal τ).

3.1 Sum of Diagonals of $\mathbf{B}^{-1}$

Toeplitz matrices belong to the larger class of *persymmetric matrices*: the N-matrix $\mathbf{B}_N = [b_{ij}]$ is persymmetric if it is symmetric about its northeast-southwest diagonal, i.e., if $b_{ij} = b_{N-j+1,N-i+1}$ for $i, j = 1, \ldots, N$. In an equivalent form $\mathbf{B}_N = \mathbf{E}_N\mathbf{B}_N^T\mathbf{E}_N$, where $\mathbf{E}_N = [\delta_{N-i+1,j}]$ is the permutation matrix. Note that $\mathbf{E}_N^{-1} = \mathbf{E}_N$. The inverse of a persymmetric matrix is, if it exits, given by $\mathbf{B}_N^{-1} = \mathbf{E}_N\mathbf{B}_N^{-T}\mathbf{E}_N$, thus, also persymmetric.

Consider the partition

$$\mathbf{B}_N^{-1} = \begin{bmatrix} \mathbf{B}_{N-1} & \mathbf{E}\mathbf{r} \\ \mathbf{r}^H\mathbf{E} & r_0 \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{B}' & \nu \\ \nu^H & \gamma \end{bmatrix}, \tag{10}$$

with $\mathbf{E} = \mathbf{E}_{N-1}$, $\mathbf{r} = [r_{-1}, \ldots, r_{-N+1}]^T$. Vector $[\nu^T, \gamma]^T$, the last column of $\mathbf{B}_N^{-1}$, is given by (see, e.g. [17, p. 130])

$$\begin{cases} \nu &= \gamma\mathbf{E}\alpha^* \\ \gamma &= \dfrac{1}{r_0 + \mathbf{r}^H\alpha^*}, \end{cases} \tag{11}$$

where α^* is the solution of the Yule-Walker equation $\mathbf{B}_{N-1}^T\alpha^* = -\mathbf{r}$. Matrix $\mathbf{B}'$ is given by

$$\mathbf{B}' = \mathbf{B}_{N-1} + \frac{\nu\nu^H}{\gamma}. \tag{12}$$

By working out expression (12), one is led to (see, e.g. [17, p. 130])

$$b_{ij} = b_{N-j,N-i} + \frac{1}{\gamma}\left(\nu_i\nu_j^* - \nu_{N-j}\nu_{N-i}^*\right). \quad (13)$$

Matrix $\mathbf{B}'$ is persymmetric; its element b_{ij} verifies $b_{N-j,N-i} = b_{i+1,j+1}$ for $i,j = 1,\ldots,N-1$; replacing $b_{N-j,N-i}$ by $b_{i+1,j+1}$ in (13), it follows that

$$b_{i+1,j+1} = b_{ij} + \frac{1}{\gamma}\left(\nu_{N-j}\nu_{N-i}^* - \nu_i\nu_j^*\right), \quad (14)$$

for $i,j = 1,\ldots,N-1$. Setting $j = i+\tau$, defining $\nu_N = \gamma$, and noting that $b_{1,i} = \nu_{N+1-i}$ for $i,j = 1,\ldots,N$, it follows that

$$b_{i+1,i+1+\tau} = b_{i,i+\tau} + \frac{1}{\gamma}\left(\nu_{N-i-\tau}\nu_{N-i}^* - \nu_i\nu_{i+\tau}^*\right), \quad (15)$$

valid for $i = 1,\ldots,N-\tau-1$, when $\tau \geq 0$, and for $i = 1-\tau,\ldots,N-1$, when $\tau \leq 0$. Expressions (14) or (15) generate recursively, from vector ν, elements b_{ij} along each matrix diagonal. The diagonal sum $\overline{b}_\tau$ is, after simple but lengthy manipulation of (15), given by

$$\gamma\overline{b}_\tau = \sum_{i=1}^{N-\tau}\left(i\nu_i\nu_{i+\tau}^* + \nu_i\nu_{i+\tau}^*(i+\tau) - \nu_i\nu_{i+\tau}^*\right), \quad (16)$$

where $\nu_N^* = \gamma$.

The determination of $\overline{b}_\tau$, for $\tau = 1,\ldots,N-1$, using (16), has $O(N^2)$ complexity. Notice, however, that each term in the sum (16) defines a convolution which can be computed using the *fast Fourier transform* (FFT) with $O(N\ln N)$ complexity. For this purpose, define the sequences $\{u_i'\}$ e $\{\nu_i'\}$, with $i \in \{\ldots,-1,0,1,\ldots\}$, as periodic extensions (of period $2N$) of sequences $\{1,\ldots,N,\mathbf{0}_N^T\}$ and $\{\nu_1,\ldots,\nu_N,\mathbf{0}_N^T\}$, respectively. Sum (16), using entities u_i' and ν_i', assumes the form

$$\gamma\overline{b}_\tau = \sum_{i=1}^{2N}\left(u_i'\nu_i'\nu_{i+\tau}'^* + \nu_i'\nu_{i+\tau}'^*u_{i+\tau}' - \nu_i'\nu_{i+\tau}'^*\right) \quad (17)$$

$$= (u_{-i}'\nu_{-i}') \star \nu_i'^* + \nu_{-i}' \star (\nu_i'u_i')^* - N\nu_{-i}' \star \nu_i'^*, \quad (18)$$

where symbol $\star$ means convolution of length $2N$. Denoting the FFT and its inverse by $\mathcal{DF}$ e $\mathcal{DF}^{-1}$, respectively, and using FFT properties (convolution, time symmetry and conjugation), expression (18) is given by

$$\gamma\overline{b}_\tau = \mathcal{DF}^{-1}\left\{\mathrm{Re}\left(\mathcal{DF}[(u_i''\nu_i')]\mathcal{DF}^*[\nu_i']\right)\right\}\Big|_{(-\tau)}, \quad (19)$$

with $u_i'' = 2u_i' - N$.

The number of flops needed to implement (19) is, approximately, $3N\ln 2N$ (corresponding to 3 FFTs of size $2N$). The number of flops necessary to solve the system $\mathbf{B}_{N-1}^T\alpha^* = -\mathbf{r}$ is, using the PCG method, of $\max(p,q)3N\ln_2 N$ in the case of ARMA(p,q) processes and of $O(N\ln N)$ otherwise (see [13] and [14]). Consequently, the complexity in computing $\overline{b}_\tau$ for $|\tau| \leq N-1$ is, in any case, of $O(N\ln N)$.

3.2 Sum of Diagonals of $\delta\mathbf{B}_N^{-1}$

Based on the results of the previous section, a technique for efficient computation of the diagonal sum of $\delta\mathbf{B}_N^{-1}$ is now developed.

Differentiating both members of (19), one gets

$$\delta(\gamma\overline{b}_\tau) = \delta\gamma\,\overline{b}_\tau + \gamma\,\delta\overline{b}_\tau$$

$$= \mathcal{DF}^{-1}\left\{\mathrm{Re}\left(\mathcal{DF}[u_i''\,\delta\nu_i']\mathcal{DF}^*[\nu_i'] + \mathcal{DF}[u_i''\,\nu_i']\mathcal{DF}^*[\delta\nu_i']\right)\right\}\Big|_{(-\tau)}, \quad (20)$$

where the following facts were used:

(a) $\delta u_i'' = 0$;

(b) $\mathcal{DF}$ e $\mathcal{DF}^{-1}$ are linear operators;

(c) $\delta\mathrm{Re}(g) = \mathrm{Re}(\delta g)$.

Equation (20) allows obtaining $\delta\overline{b}_\tau = \gamma^{-1}(\delta(\gamma\overline{b}_\tau) - \delta\gamma\,\overline{b}_\tau)$ (note that $\gamma > 0$, since it is the principal diagonal of a positive defined matrix) from γ, $\delta\gamma$, $\overline{b}_\tau$, and from sequences $\{u_i''\nu_i'\}$, $\{\nu_i'\}$, $\{u_i''\,\delta\nu_i'\}$ and $\{\delta\nu_i'\}$. It is thus necessary to compute terms $\mathcal{DF}[u_i''\,\delta\nu_i']$ and $\mathcal{DF}[\delta\nu_i']$, in addition to $\mathcal{DF}[\nu_i']$ and $\mathcal{DF}[u_i''\,\nu_i']$, intervening in the expression of $\overline{b}_\tau$. Therefore, let alone the complexity of $(\delta\nu_1,\ldots,\delta\nu_{N-1},\delta\gamma)$, the number of flops to determine $\delta\overline{b}_\tau$, for $|\tau| \leq N-1$, is, roughly, $6N\ln_2 2N$.

Finally, it is necessary to determine the vector $\delta\,[\nu^T,\gamma]^T$. Given that

$$\delta\mathbf{B}_N\begin{bmatrix}\nu\\ \gamma\end{bmatrix} = [0,\ldots,0,1]^T, \quad (21)$$

we can write

$$\delta\mathbf{B}_N\begin{bmatrix}\nu\\ \gamma\end{bmatrix} + \mathbf{B}_N\,\delta\begin{bmatrix}\nu\\ \gamma\end{bmatrix} = \mathbf{0}_N. \quad (22)$$

Hence, the vector $\delta\,[\nu^T,\gamma]^T$ is the solution of the system

$$\mathbf{B}_N\left\{\delta\begin{bmatrix}\nu\\ \gamma\end{bmatrix}\right\} = -\delta\mathbf{B}_N\begin{bmatrix}\nu\\ \gamma\end{bmatrix}. \quad (23)$$

Matrix $\delta\mathbf{B}_N$ is Toeplitz; thus, the product at the right hand of (23) can be embedded in a circular convolution and computed by FFT; the number of flops involved in this operation is, approximately, $3N/2\ln_2 3N$. On the other hand, the solution of the linear Toeplitz system (23) has $O(N\ln N)$ complexity, in agreement to what was presented above.

3.3 Exact Computation of J

Define $\mathbf{R}_y(\theta) = [r_{i-j}] = [r_{-\tau}]$, for $i,j = 1,\ldots,N$, and $\overline{r}_\tau$ as $\overline{b}_\tau$ in (9) replacing $\mathbf{B}$ by $\mathbf{R}_y$. With this notation, the

element (k,l) of the Fisher information matrix is given by

$$J_{kl} = -\sum_{\tau=-N+1}^{N-1} r_{-\tau}^{(k)} \overline{r}_{-\tau}^{(l)} \quad (24)$$

$$= -\sum_{\tau=-N+1}^{N-1} r_{\tau}^{(k)} \overline{r}_{\tau}^{(l)}. \quad (25)$$

Determining J_{kl} given by (25) has the complexity of computing $\{\overline{r}_{\tau}^{(l)}(\theta), |\tau| \leq N-1\}$, which, according with the method presented in the previous section is of $O(N \ln N)$.

3.4 Approximate Computation of J

Given that $\mathbf{R}_y^{-1}$ and $\mathbf{R}_y^{(k)}$ exist for $\theta \in \mathbf{\Theta}$, it follows that $\mathbf{R}_y^{-(k)} = -\mathbf{R}_y^{-1}\mathbf{R}_y^{(k)}\mathbf{R}_y^{-1}$ and, consequently, the terms $\overline{r}_{\tau}(\theta))$ are differentiable in $\mathbf{\Theta}$. Define $\mathbf{u}_l$ as

$$\mathbf{u}_l = (\underbrace{0,\ldots,0}_{l-1}, 1, \underbrace{0,\ldots,0}_{p-l}).$$

Invoking the differentiability of $\overline{r}_{\tau}$, it follows that

$$J_{kl} = -\sum_{\tau=-N+1}^{N-1} r_{\tau}^{(k)} \frac{\overline{r}_{\tau}(\theta + \Delta\mathbf{u}_l) - \overline{r}_{\tau}(\theta)}{\Delta} + o(\Delta), \quad (26)$$

where $o(\Delta)$ is an infinitesimal of order greater than Δ. Sequences $\{\overline{r}_{\tau}(\theta + \Delta\mathbf{u}_l), |\tau| \leq N-1\}$ and $\{\overline{r}_{\tau}(\theta), |\tau| \leq N-1\}$ are given by the method presented above (Section 3.1) with complexity $O(N \ln N)$. Increment Δ must be carefully chosen, in order to have negligible error in the results given by (26).

4 CONCLUSION

Two new algorithms (one exact and another approximate) for computing the Fisher information matrix of zero-mean stationary complex Gaussian processes were presented. Both methods have $O(N \ln N)$ complexity being, therefore, faster than the method of Porat and Friedlander [7], which has $O(N^2)$ complexity.

The central idea exploited is that the Fisher information matrix depends on the sum of diagonals of the inverse covariance matrix derivative (in order to the parameter vector) and not on each single element. The computation of the referred sum is carried out by a technique here introduced, which has $O(N \ln N)$ complexity.

References

[1] H. Cramér, "A contribution to the theory of statistical estimation", *Aktuariestidskrift*, vol. 29, pp. 458–463, 1946.

[2] C. Rao, "Information and the accuracy attainable in the estimation of statistical parameters", *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–91, 1945.

[3] B. Porat, *Digital Processing of Random Signals*, Prentice-Hall, Englewood Cliffs, New Jersey, 1994.

[4] P. Whittle, "Estimation and information in stationary time series", *Arkiv Matematick*, vol. B-2, pp. 423–434, 1953.

[5] A. Zeira A. Nehorai, "Frequency domain Cramér-Rao for Gaussian processes", *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, pp. 1063–1066, 1990.

[6] B. Porat and B. Friedlander, "The exact Cramér-Rao bound for Gaussian autoregressive processes", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 23, pp. 537–542, 1987.

[7] B. Porat and B. Friedlander, "Computation of the exact information matrix of Gaussian time series with stationary random components", *IEEE Trans. Acoust. Speech Signal Process.*, vol. 34, pp. 118–130, 1986.

[8] F. Giannella, "C-R bounds for AR parameter estimation as a function of the data lenght", *IEEE Trans. Acoust. Speech Signal Process.*, vol. 34, pp. 994–995, 1986.

[9] P. Tuan, "Cramér-Rao bounds for AR and parameter reflection coefficient estimators", *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, pp. 769–772, 1989.

[10] G. Mélard and A. Klein, "On a fast algorithm for the exact information of a Gaussian ARMA time series", *IEEE Trans. Signal Process.*, vol. 42, pp. 2201–2203, 1994.

[11] M. Morf, G. Sidhu, and T. Kailath, "Some new algorithms for recursive estimation and on constant, linear, discrete-time systems", *IEEE Trans. Aut. Contr.*, vol. 19, pp. 315–323, 1974.

[12] W. F. Trench, "An algorithm for the inversion of finite Toeplitz matrices", *J. SIAM*, vol. 12, pp. 515–522, 1964.

[13] T. Ku and C. Kuo, "Design and analysis of optimal Toeplitz preconditioners", *IEEE Trans. Signal Process.*, vol. 40, pp. 129–141, 1992.

[14] T. Ku and C. Kuo, "Spectral properties of preconditioned rational Toeplitz matrices", *SIAM J. Matrix Anal. Appl.*, vol. 14, pp. 146–165, 1993.

[15] K. S. Miller, *Complex Stochastic Processes. An Introduction to Theory and Applications*, Addison-Wesly, Massachusetts, 1974.

[16] S. Kay, *Fundamentals of Statistical Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[17] G. H. Golub and C. F. Loan, *Matrix Computations*, Johns Hopkins University Press, 1983.