ON NONPARAMETRIC SPECTRAL ESTIMATION

Tomas Sundin and *Petre Stoica*
Systems and Control Group, Uppsala University,
P.O. Box 27, SE-75103 Uppsala, SWEDEN
Tel: +46 18 4713150, +46 18 4717619; fax: +46 18 503611
e-mail: tsu@syscon.uu.se, ps@syscon.uu.se

ABSTRACT

In this paper the Cramér-Rao bound (CRB) for a general nonparametric spectral estimation problem is derived under a local smoothness condition (more exactly, the spectrum is assumed to be well approximated by a piecewise constant function). Furthermore it is shown that under the aforementioned condition the Thomson (TM) and Daniell (DM) methods for power spectral density (PSD) estimation can be interpreted as approximations of the maximum likelihood PSD estimator. Finally the statistical efficiency of the TM and DM as nonparametric PSD estimators is examined and also compared to the CRB for ARMA-based PSD estimation. In particular for broadband signals, the TM and DM almost achieve the derived nonparametric performance bound and can therefore be considered to be nearly optimal.

1 INTRODUCTION

The *parametric approach* to spectral estimation suffers from a number of problems (such as sensitivity to mismodeling) a fact that has motivated a renewed interest in the *nonparametric approach*. For the latter approach the performance issue is an important aspect. In particular, answers to the following questions are of significant interest: (a) What is the best (achievable) statistical performance in the class of nonparametric PSD estimation methods, under some reasonable assumptions ?; (b) Is there any nonparametric PSD estimator that achieves the best statistical performance mentioned above?; (c) How do the best possible performances in the classes of parametric and nonparametric PSD estimation methods compare with one another?

Most papers in the literature do not address the above questions in any generality, but are limited to studies of specific nonparametric PSD estimators, e.g., [2] [5]. A notable exception is [4] where a fairly large class of nonparametric PSD estimators, which are quadratic functions of the data vector, were analyzed.

Our approach here is more general, although conceptually simpler, than that of [4]. Under a local smoothness condition and the Gaussian hypothesis we provide general answers to questions (a) and (b) above by making use of the Cramér-Rao bound (CRB) and the maximum likelihood (ML) estimation method, respectively. We show that two of the most successful nonparametric PSD estimators, viz. Thomson method (TM) [9] and Daniell method (DM) [3] can be interpreted as computationally convenient approximations to the nonparametric ML-based PSD estimator. This interpretation of the TM and DM provides new insights into the properties of these two methods and the relationship between them. To provide an answer to question (c) we compare the CRB for nonparametric PSD estimation derived here with the CRB for ARMA-based PSD estimation, in a number of cases.

2 THE ML APPROACH

Let $\{y(t)\}_{t=1,2,\ldots}$ denote a complex-valued stationary signal, and let $\Phi(\omega)$ denote its PSD function. Also, let N denote the number of available observations, $\{y(t)\}_{t=1}^{N}$. For the sake of convenience we assume that, for a given (positive) integer M, there exists a integer L such that

$$LM = N \tag{1}$$

We also make the following assumptions on $\{y(t)\}$ and $\{\Phi(\omega)\}$:

A1. The data vector $y = \begin{pmatrix} y(1) & \cdots & y(N) \end{pmatrix}^T$ has a circular Gaussian distribution with zero mean and covariance matrix R.

A2. The PSD function is piecewise constant on the frequency bins $[0, 2\pi\beta]$, $[2\pi\beta, 4\pi\beta]$ etc., where

$$\beta \stackrel{\triangle}{=} 1/M = L/N \tag{2}$$

Furthermore, $0 < \Phi(\omega) < \infty$ for every ω.

Let Φ_k denote the value taken by $\Phi(\omega)$ in the k:th frequency bin:

$$\Phi_k = \Phi(\omega) \text{ for } \omega \in [2\pi(k-1)\beta, 2\pi k\beta] \quad (k = 1, \ldots, M) \tag{3}$$

By invoking A2 we can reduce the problem of *nonparametric estimation* of $\Phi(\omega)$ to that of estimating the unknown constants $\{\Phi_k\}_{k=1}^{M}$. Because usually $\Phi(\omega)$ does not satisfy A2 exactly, the use of this assumption will introduce a *bias* in the so-obtained PSD estimate(s). This bias can be "controlled" by suitably choosing the *user parameter* M (or L) (see [8] for details).

We can now state the problem to be dealt with in this paper: obtain the ML estimates of $\{\Phi_k\}_{k=1}^{M}$ from $\{y(t)\}_{t=1}^{N}$, and the associated CRB, under assumptions A1 and A2.

Under assumption A1, the negative log-likelihood function of the data vector y is given (within an additive constant) by:

$$f = \log|R| + y^* R^{-1} y \tag{4}$$

where $|\cdot|$ denotes the determinant, and $*$ is the conjugate transposition symbol. The inverse matrix R^{-1} in (4) exists owing to the assumption that $\Phi(\omega) > 0$ for all ω, in A2.

Let

$$a(\omega) = \begin{pmatrix} e^{i\omega} & \ldots & e^{iN\omega} \end{pmatrix}^T \tag{5}$$

Next, we make use of A2 to rewrite R as follows:

$$R = \sum_{k=1}^{M} \Phi_k \frac{1}{2\pi} \int_{2\pi(k-1)\beta}^{2\pi k\beta} a(\omega)a^*(\omega)d\omega \triangleq \sum_{k=1}^{M} \Phi_k D_k \Gamma D_k^* \tag{6}$$

where

$$\Gamma = \frac{1}{2\pi} \int_{-\pi\beta}^{\pi\beta} a(\omega)a^*(\omega)d\omega \tag{7}$$

and

$$D_k = \begin{pmatrix} e^{i2\pi\beta(k-1/2)} & & 0 \\ & \ddots & \\ 0 & & e^{iN2\pi\beta(k-1/2)} \end{pmatrix} \tag{8}$$

The property of Γ of interest here is the fact that, for reasonably large values of N,

$$\text{rank}(\Gamma) \simeq N\beta = L \tag{9}$$

The approximate equality above should be interpreted in the sense that $\lambda_L/\lambda_{L+1} \gg 1$, where $\{\lambda_k\}$ denote the eigenvalues of Γ. Furthermore, it holds that

$$\Lambda \triangleq \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_L \end{pmatrix} \simeq I \tag{10}$$

In view of (9) we can approximately write Γ as

$$\Gamma \simeq UU^* \tag{11}$$

for an appropriately chosen $N \times L$ - matrix U. *Different choices of U* in (11) will lead to *different PSD estimation methods.* We will discuss the choice of U after completing the analysis for a generic U matrix. Inserting (11) into (6) we obtain:

$$R \simeq W \begin{pmatrix} \Phi_1 I_{L\times L} & & 0 \\ & \ddots & \\ 0 & & \Phi_M I_{L\times L} \end{pmatrix} W^* \tag{12}$$

where

$$W \triangleq \begin{pmatrix} D_1 U & \cdots & D_M U \end{pmatrix} \tag{13}$$

The approximation of R in (12) yields the following convenient approximation for the negative log-likelihood function (within an additive constant, once again):

$$f \simeq L \sum_{k=1}^{M} \log(\Phi_k) + \sum_{k=1}^{M} \|\tilde{y}_k\|^2/\Phi_k \tag{14}$$

where $\|\cdot\|$ denotes the Euclidean norm, and $\{\tilde{y}_k\}$ are the Lx1 sub-vectors of $W^{-1}y$:

$$\begin{pmatrix} \tilde{y}_1^T & \cdots & \tilde{y}_M^T \end{pmatrix}^T = W^{-1} y \tag{15}$$

The minimization of (14) with respect to $\{\Phi_k\}$ yields *the generic approximate ML estimates*:

$$\hat{\Phi}_k = \|\tilde{y}_k\|^2/L \tag{16}$$

Note that the U matrix, which enters in (16) via W, is yet to be specified.

3 APPROXIMATE ML APPROACHES

In the previous section a rank-assumption on Γ resulted in the *generic* approximate ML-estimator in (16) . In this section three different choices of U in (11) will be considered and they will lead to three different PSD-estimators. The first method is based on the following choice of U:

$$U = V\Lambda^{1/2} \tag{17}$$

where $\Lambda^{1/2}$ is the square root of the matrix Λ in (10), and V is the $N \times L$ - matrix made from the L *principal eigenvectors* of Γ. The resultant approximate ML estimator, which we will denote by AML in the following, has a large bias and is not a good PSD-estimator in finite samples. We have found no explanation for this behavior and in fact expected the AML to perform well since the choice of U in (17) seems most reasonable (note that $UU^* = V\Lambda V^*$ is the best rank-L approximation of Γ, in the Frobenius-norm metric).

Next, certain approximations leading to the Thomson method (TM) [9] will be considered. The matrix U is still chosen as in (17). Additionally, we now make use of the approximation

(a) $\Lambda \simeq I$ (cf. (10))

and of the fact that approximately ([4] [6])

(b) $D_k V$ and $D_j V$, for $k \neq j$,are orthogonal matrices.

By using approximations (a) and (b) above, we can write:

$$\begin{aligned} W^{-1} &\simeq \begin{pmatrix} D_1 V & \ldots & D_M V \end{pmatrix}^{-1} \\ &\simeq \begin{pmatrix} D_1 V & \ldots & D_M V \end{pmatrix}^* \end{aligned} \tag{18}$$

which leads to the following significantly simplified expression for the "filtered data" vectors $\{\tilde{y}_k\}$ in (15):

$$\tilde{y}_k \simeq V^*(D_k^* y) \quad (k = 1, \ldots, M) \tag{19}$$

Finally we consider the Daniell method (DM) [3]. This method turns out to use an approximate $N \times L$-square root U of Γ that is different from that used by the AML and TM. To explain how U corresponding to the DM is obtained, note that for sufficiently large values of N and L we can approximate the integral in (7) by the following sum:

$$\Gamma \simeq \frac{1}{N} \sum_{p=1}^{L} a(-\pi\beta + \frac{2\pi}{N}p) a^*(-\pi\beta + \frac{2\pi}{N}p) \tag{20}$$

(recall that $2\pi L/N = 2\pi/M = 2\pi\beta$). In view of (20), let us choose U as:

$$U = \left(a(-\pi\beta + \frac{2\pi}{N}) \ \cdots \ a(-\pi\beta + \frac{2\pi}{N}L) \right) / \sqrt{N} \tag{21}$$

Next observe that, for the U above,

$$D_k U = \left(a(\frac{2\pi}{N}(L(k-1)+1)) \cdots a(\frac{2\pi}{N}(L(k-1)+L)) \right) \sqrt{N} \tag{22}$$

It is readily verified that the vectors $a(\frac{2\pi}{N}m)/\sqrt{N}$ and $a(\frac{2\pi}{N}n)/\sqrt{N}$ are orthogonal to one another for $m \neq n$, and also that both vectors have unit (Euclidean) norm. It follows that the inverse of the matrix W corresponding to (22) is given by:

$$W^{-1} = W^* \tag{23}$$

Using (22) and (23) in (15)(16), we obtain:

$$\hat{\Phi}_k = \frac{1}{L} \sum_{p=1}^{L} |a^*(\frac{2\pi}{N}(L(k-1)+p))y|^2 / N \tag{24}$$

which is exactly the Daniell PSD estimator.

4 THE CR BOUND

Under A1, A2 *and* the approximation (11) (leading to the expression (12) for R) we can easily obtain the CRB matrix as follows. From the Bangs formula for the CRB (see [1] [7]) we have that

$$[(\mathrm{CRB})^{-1}]_{i,j} = \mathrm{tr}[R^{-1} R_i' R^{-1} R_j'] \tag{25}$$

where $\mathrm{tr}(\cdot)$ is the trace operator, and R_i' denotes the derivative of R with respect to Φ_i. A simple calculation shows that

$$R_i' = D_i U U^* D_i^* = W E_i E_i^* W^* \tag{26}$$

where

$$E_i = (0_{L\times L} \ \cdots \ 0_{L\times L} \ I_{L\times L} \ 0_{L\times L} \ \cdots \ 0_{L\times L})^* \tag{27}$$

Hence (within the approximations made):

$$\begin{aligned} &\mathrm{tr}\left[R^{-1} R_i' R^{-1} R_j'\right] = \\ &\mathrm{tr}\{ \begin{pmatrix} (1/\Phi_1) I_{L\times L} & & 0 \\ & \ddots & \\ 0 & & (1/\Phi_M) I_{L\times L} \end{pmatrix} E_i E_i^* \\ &\times \begin{pmatrix} (1/\Phi_1) I_{L\times L} & & 0 \\ & \ddots & \\ 0 & & (1/\Phi_M) I_{L\times L} \end{pmatrix} E_j E_j^* \} \\ &= (L/\Phi_i^2) \delta_{i,j} \end{aligned} \tag{28}$$

which gives the following simple expression for *the CRB matrix*:

$$\mathrm{CRB} = \frac{1}{L} \begin{pmatrix} \Phi_1^2 & & 0 \\ & \ddots & \\ 0 & & \Phi_M^2 \end{pmatrix} \tag{29}$$

5 NUMERICAL EXAMPLE

In this section a numerical example is used to study the accuracy of the discussed methods. The data are generated as an ARMA-process:

$$A(q^{-1}) y(t) = C(q^{-1}) e(t)$$

where q^{-1} denotes the unit delay operator, $A(q^{-1})$ and $C(q^{-1})$ are polynomials in q^{-1} and $e(t)$ is white Gaussian noise with unit variance. The PSD-estimators considered in what follows are: AML (given by (16) with U in (17)), TM ((16) with $\{\tilde{y}_k\}$ in (19)) and DM (given by (24)). To compare the statistical accuracy of these estimators with the derived CRB we introduce the Normalized Mean Square Error (NMSE):

$$NMSE = \mathrm{E}\left[\frac{(\hat{\Phi}_k - \Phi_k)^2}{\Phi_k^2} \right]$$

and the Integrated NMSE

$$INMSE = \frac{1}{M} \mathrm{E}\left[\sum_{k=1}^{M} \frac{(\hat{\Phi}_k - \Phi_k)^2}{\Phi_k^2} \right]$$

The CRB in (29) gives the following performance bounds:

$$NMSE \geq \frac{1}{L} \qquad INMSE \geq \frac{1}{L}$$

Let

$$A(q^{-1}) = 1 - 1.3817q^{-1} + 1.5632q^{-2} - 0.8843q^{-3} + 0.4096q^{-4}$$

$$C(q^{-1}) = 1 + 0.3544q^{-1} + 0.3508q^{-2} + 0.1736q^{-3} + 0.2401q^{-4}$$

The nature of this spectrum is such that the smoothness assumption A2 approximately holds even for relatively small values of M. The DM and TM estimates (in Fig.1(a) and (b) respectively) appear to yield almost unbiased estimates and are as expected close to achieving the derived bound (Fig.1(d)) whereas the AML has a poor behavior particularly for the areas of the spectrum with low power (Fig.1(c)) and consequently fails to achieve the CRB (see Fig.1(d)). In Fig.1(e) the estimated INMSE from 100 simulation runs is displayed as a function of M when the value of L is held constant ($L = 8$). It is seen that the TM and DM estimates approach the derived bound despite the fact that $M/N \nrightarrow 0$. The difference between the INMSE and the CRB bound for small values of M is due to the bias which decreases as the smoothness assumption becomes more and more valid with increasing M.

Finally, by comparing the nonparametric and parametric CRB bounds in Figure 1(d) we can see that there is of course some loss in performance associated with using nonparametric methods for PSD estimation, particularly so for the parts of the PSD with low power. However this performance degradation may well be balanced by the computational simplicity of the nonparametric PSD estimators.

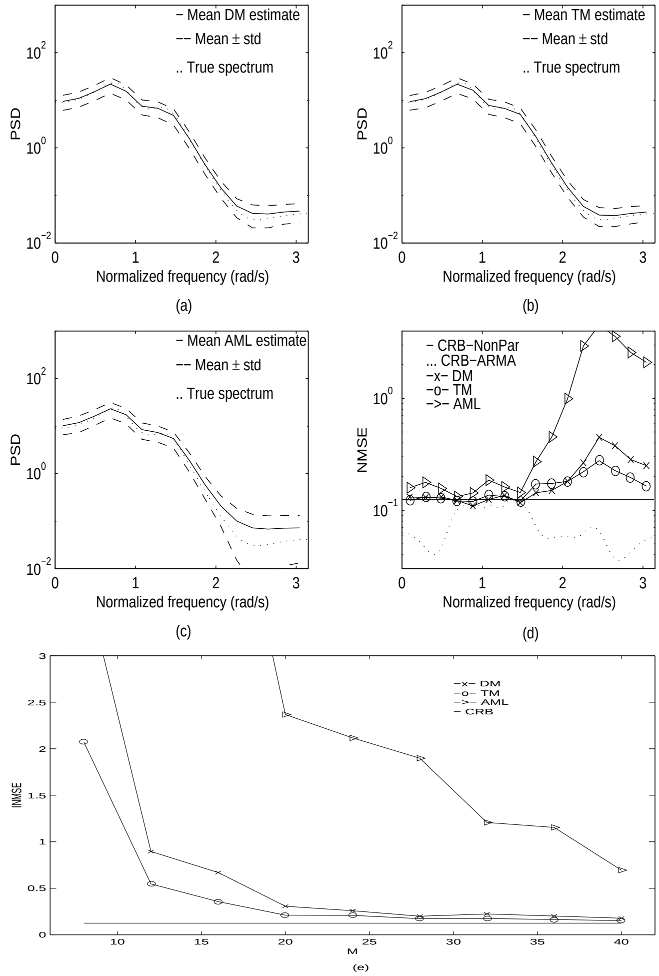

Figure 1: (a)-(d): Results obtained from 1000 simulation runs, $N = 256$, $M = 32$. (a) DM-mean ± standard deviation (std) compared with the true spectrum. (b) TM-mean ± std compared with the true spectrum. (c) AML-mean ± std compared with the true spectrum. (d) NMSE estimates compared with $\frac{1}{L}$ and the normalized CRB for ARMA-based PSD-estimation. (e) INMSE estimates obtained from 100 simulation runs, $L = 8$

References

[1] W. J. Bangs, *Array Processing with Generalized Beamformers*. PhD thesis, Yale University, New Haven, CT, 1971.

[2] T. P. Bronez, "On the performance advantage of multitaper spectral analysis". *IEEE Transactions on Signal Processing*, Vol. 40(2), pp 2941-2946, 1992.

[3] P. J. Daniell, "Discussion of "On the theoretical specification and sampling properties of autocorrelated time-series"", *J. Royal Statistical Society*, Vol. 8, pp 88-90, 1946.

[4] C.T. Mullis and L.L. Scharf, "Quadratic Estimators of the Power Spectrum" in S. Haykin, editor, *Advances in spectrum analysis and array processing*, Vol. I, pp 1-57, Prentice-Hall, Englewood Cliffs, N.J., 1991.

[5] K. Riedel and A. Sidorenko, "Minimum bias multiple taper spectral estimation", *IEEE Transactions on Signal Processing*, Vol. 43(1),pp 188-195, 1995.

[6] D. Slepian, "Prolate spherical wave functions, Fourier analysis and uncertainty-IV", *Bell System Technical Journal* , Vol. 43, pp 3009-3057, 1964.

[7] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice Hall, Upper Saddle River, NJ, 1997.

[8] P. Stoica and T. Sundin, "On Nonparametric Spectral Estimation", Submitted to *Circuits, Systems, and Signal Processing*, 1998.

[9] D. J. Thomson, "Spectrum Estimation and Harmonic Analysis", *Proceedings of the IEEE*, Vol. 70(9),pp 1055-1096, 1982.