MINIMUM FISHER INFORMATION SPECTRAL ANALYSIS

Vojin Živojnović [1] *Dominikus Noll* [2]

[1] Integrated Systems for Signal Processing
Aachen University of Technology, Aachen, Germany
zivojnov@ert.rwth-aachen.de

[2] Département de Mathématiques
Université Paul Sabatier, Toulouse, France
noll@cict.fr

ABSTRACT

Minimizing the Fisher information measure over the set of power spectrum densities fitting a finite number of autocorrelation lag constraints is treated. Due to an explicit control of the derivative values of the densities, the Fisher information measure produces a useful smoothing effect. The Fisher information based estimate exhibits improved characteristics compared to the maximum entropy approach proposed by Burg. We show that the resulting power spectrum estimate is positive, and along with the autocorrelation constraints, satisfies a generalized Riccati differential equation. In general, the true estimate of the power spectrum may be obtained only by numerically integrating the corresponding boundary value problem. For real time applications, we therefore propose a fast and numerically stable approximate solution in explicit trigonometric form. Although suboptimal, the proposed approach has proven to be advantageous especially for flat spectra. The presented theory is verified on simulated examples.

1. INTRODUCTION

Among the procedures for computing the power spectrum density $S(\omega)$ of a discrete stationary process $x(t), t \in Z$ from a finite sequence of autocorrelation lags $\{r_k, k = 0, ..., N\}$, Burg's maximum entropy method is the best known and most frequently treated one [1,2,3]. Because of its theoretical foundation in terms of the entropy rate of autoregressive processes, the maximum entropy principle was described as one of the most important analytical results in statistics [4]. For a review and more details see [5] and [4].

The *maximum entropy power spectrum density* is defined as

$$\bar{S}(\omega) = \max\{ \quad \int_{-\pi}^{\pi} -\log(S(\omega))d\omega | \quad \int_{-\pi}^{\pi} \cos(k\omega)S(\omega)d\omega = r_k, \quad k = 0, ..., N; S(\omega) > 0\} \tag{1}$$

the rationale of Burg's measure $\int -\log S(\omega)d\omega$ being the minimization of spurious or artificial information introduced when extrapolating the autocorrelation coefficients under the assumption of Gaussian input data. An alternative interpretation, given by Van den Bos [6], is based on a least-square fitting of an all-pole model to the available data, and equally assumes a Gaussian input. Jaynes [4] has shown that the same result can be obtained if the hypothesis of a Gaussian input is replaced by the Greatest Number of Ways (GNW) criterion.

One of the drawbacks of Burg's measure is the introduction of false peaks in the case of flat power spectra and overparameterization [6,2]. This result is not unexpected regarding the form $(\sum c_k \exp(-k\omega))^{-1}$ of the spectrum maximizing Burg's measure, which corresponds in form to the spectra of all-pole (AR) models.

In this paper we propose the Fisher information measure $I_f(S) = \int S(\omega)'^2/S(\omega)d\omega$ as the optimization criterion. The following problem is treated:

$$\bar{S}(\omega) = \min \left\{ \quad \int_{-\pi}^{\pi} \left(\frac{S(\omega)'}{S(\omega)}\right)^2 S(\omega)\, d\omega \right| \quad \int_{-\pi}^{\pi} \cos(k\omega)S(\omega)d\omega = r_k, \quad k = 0, ..., N; S(\omega) > 0 \Bigg\} \tag{2}$$

Although the Fisher information measure is missing the strong analytical justification of Burg's entropy, it has some distinguishing features. Its main characteristic is the explicit control over the derivative of the argument, which has a useful smoothing effect on the resulting spectrum estimate. We shall show that in particular for processes with flat spectra, the Fisher information provides improved spectrum estimates as compared to the maximum entropy approach.

In general, computing the Fisher power spectrum estimate under autocorrelation constraints can be done by numerical means only. We show that for a subclass of autocorrelation sequences the solution can be obtained in a closed analytical form by solving a set of linear equations. For general autocorrelation lags, this analytic expression is only suboptimal, but provides a good approximation of the true estimate. The computational effort of our method is equal to the cost required to calculate Burg's maximum entropy solution.

2. INFORMATION MEASURES AND MOMENT CONSTRAINTS

Minimizing the Shannon information with given moments is a well treated subject in information theory, see e.g. [7,8,9]. The corresponding power spectrum estimation problem is

$$\bar{S}(\omega) = \min\{ \quad \int_{-\pi}^{\pi} S(\omega)\log(S(\omega))d\omega | \quad \int_{-\pi}^{\pi} \cos(k\omega)S(\omega)d\omega = r_k, \quad k = 0, ..., N; S(\omega) > 0\} \tag{3}$$

the solution having the explicit form

$$\bar{S}(\omega) = C \exp(-\sum_{k=1}^{N} \alpha_k \cos(k\omega)) \tag{4}$$

with the parameters $\{\alpha_k, k = 1, ..., N\}$ chosen to fit the autocorrelation constraints. Similarly, in the case of the entropy rate,

i.e. Burg's problem (1), the solution has an explicit form

$$\bar{S}(\omega) = 1 \Big/ \sum_{k=0}^{N} \gamma_k \cos(k\omega) \tag{5}$$

where $\{\gamma_k, k = 0, ..., N\}$ are chosen so that $S(\omega)$ fits the autocorrelation lags. Clearly, any other model has to be compared to these methods, and should in particular lead to a solution $S(\omega)$ with sufficiently explicit form.

The origin of the Fisher information measure dates back to the work of Fisher [10], where it is introduced in the context of maximum likelihood estimation. For a parameterized family of probability densities $p(x, \theta)$ with a scalar parameter $\theta \in R$ the Fisher information is defined as [7]

$$I(\theta) = E\left(\frac{\partial \log p(x,\theta)}{\partial \theta}\right)^2 = -E\left(\frac{\partial^2 \log p(x,\theta)}{\partial \theta^2}\right) \tag{6}$$

which measures the curvature of the log likelihood function in a neighborhood of the true parameter. In fact, the maximum likelihood estimate $\hat{\theta}_N$ based on N samples is asymptotically normally distributed around the true value θ^* with a variance $I(\theta)^{-1}$, $I(\theta)$ the Fisher information:

$$\sqrt{N}(\hat{\theta}_N - \theta^*) \sim \mathcal{N}(0, I(\theta)^{-1}), \quad N \to \infty. \tag{7}$$

If θ is a location parameter, i.e. $p(x, \theta) = p(x - \theta)$, then

$$I(\theta) = I_f(p) \equiv \int (p'/p)^2 p dx \tag{8}$$

only depends on the probability density p, measuring the asymptotic efficiency of the maximum likelihood method.

In the domain of robust statistics pioneered by Huber in 1964 [11], the problem of minimizing $I_f(p)$ over a class of possible densities p has been extensively treated [12]. However, surprisingly enough the case of moment constraints is not treated at all. [1] Only the result for constrained second order moment of the density yielding to a Gaussian density is known.

Minimization of the Fisher information measure of a probability density function with first N moments given was treated in [13,14] in the context of robust accuracy improvement of time series parameters. It was shown that the resulting minimum Fisher density may be obtained with a procedure similar to the one given below, and in particular, a good suboptimal solution in explicit polynomial form was obtained. For a complete mathematical treatment of the Fisher problem we refer to [15,16,17].

3. MINIMUM FISHER INFORMATION SPECTRAL ANALYSIS

Unlike the Shannon or Burg entropies, the Fisher information measure $I_f(S)$ measures the shape of the spectral density $S(\omega)$ *and* of its first derivative $S'(\omega)$. What is controlled is in fact the variance of S'/S with respect to measure dS. Roughly speaking, this means that fast changes of the density with sizable amplitude introduce a high information rate, and will therefore be suppressed by the minimization procedure.

Extrapolating the autocorrelation sequence $\{r_k, k = 0, ..., N\}$ for the lags higher than N with minimum amount of Fisher information of the spectrum $S(\omega)$ is a convex optimization problem:

(P) Minimize

$$I_f(S) = \int_{-\pi}^{\pi} \left(\frac{S'(\omega)}{S(\omega)}\right)^2 S(\omega) d\omega \tag{9}$$

subject to

$$\int_{-\pi}^{\pi} \cos(k\omega) S(\omega) d\omega = r_k, \quad S(\omega) \geq 0. \tag{10}$$

3.1. Existence and Uniqueness

For the finite interval $[-\pi, \pi]$, Borwein *et al.* [15,17] have shown that problem (P), if assumed to be feasible in the sense that there exists a function $\hat{S}(\omega) \geq 0$, $\hat{S} \neq 0$ fitting the constraints, has a unique optimal solution $\bar{S} > 0$.

3.2. Optimization Procedure

In order to compute the optimum $\bar{S}$, let us define the Lagrangian associated to problem (P):

$$\Lambda(S, \lambda) = I_f(S) + \sum_{k=0}^{N} \lambda_k \left(\int_{-\pi}^{\pi} \cos(k\omega) S(\omega)\, d\omega - r_k \right) \tag{11}$$

Evaluating the necessary optimality conditions $\nabla_{S,\lambda} \Lambda(\bar{S}, \bar{\lambda}) = 0$ leads to the Euler-Lagrange equation for (P). Using the change of variables

$$\psi(\omega) = -S(\omega)'/S(\omega), \qquad S(\omega) = C \exp\Big(- \int \psi(\omega)\, d\omega \Big)$$

leads to the generalized Riccati differential equation [18]

$$2\psi'(\omega) - \psi^2(\omega) + \sum_{k=0}^{N} \lambda_k \cos(k\omega) = 0 \tag{12}$$

with boundary condition

$$\psi(-\pi) = \psi(\pi) = 0 \tag{13}$$

With $\bar{\psi}$ and $\bar{\lambda}_k$ solving (12), (13), the optimal solution $\bar{S}(\omega)$ is then positive with $\bar{S}'(-\pi) = \bar{S}'(\pi) = 0$. Conversely, the optimum $\bar{S}$ is even characterized by the fact that it satisfies the constraints (10), and the corresponding $\bar{\psi}$ satisfies the Ricatti equation (12) with boundary condition (13) [15,16].

3.3. Polynomial Solution

As a heuristic, let us assume that (12), (13) admits a solution in form of a trigonometric polynomial:

$$\psi(\omega) = \sum_{k=1}^{L} \beta_k \sin(k\omega) \tag{14}$$

This implies the solution of (P) to take the form

$$S(\omega) = C \exp(-\sum_{k=1}^{L} \frac{\beta_k}{k} \cos(k\omega)) \tag{15}$$

which is obviously positive and satisfies the Neumann boundary conditions $\hat{S}'(-\pi) = \hat{S}'(\pi) = 0$.

Inserting (14) in (12) shows that (14) can be a solution only if $L = N/2$, i.e. the highest constrained autocorrelation lag has to be of even order ($N = 2L$) and the total number of constrained lags is odd.

[1] In [7] the minimization of the Fisher information under moment constraints was emphasized as one of the interesting, but unsolved problems.

3.4. Feasibility of the Polynomial Solution

In order for (15) to be the true solution of problem (S), it is necessary that it satisfies the constraints (10). While in the case of the Burg's estimate (1), the parameterized solution (5) has $N+1$ degrees of freedom, and a unique solution satisfying the given autocorrelation lags may therefore be found if the regression matrix is positive definite, the form of the approximate Fisher estimate (15) has only $L = N/2$ degrees of freedom, and an exact solution of the form (15) will only occur for a subset of the possible autocorrelation constellations. In the sequel we shall explore the dependence between the given autocorrelations $\{r_k, k = 0, ..., 2L\}$ and the coefficients $\{\beta_k, k = 0, ..., L\}$.

For a trigonometric polynomial solution, Eq. (10) has the following form

$$r_j = \int_{-\pi}^{\pi} \cos(j\omega) C \exp(-\sum_{k=1}^{L} \frac{\beta_k}{k} \cos(k\omega)) d\omega, j = 0, 1, ..., 2L \tag{16}$$

This equation defines a mapping $F : \beta \rightarrow r$ between the L-dimensional space of parameters β, and the $2L$-dimensional space of autocorrelation lags r. The task is now to determine the range of F in the higher dimensional r-space.

Using partial integration on Eq. (16) the following relation between r and β is obtained

$$2kr_k = \sum_{i=1}^{L} (r_{k-i} - r_{k+i})\beta_i, \; k = 1, ..., 2L \tag{17}$$

This set of $2L$ equations contains L unknown coefficients $\{\beta_k, k = 1, ..., L\}$ and L unknown autocorrelation lags $\{r_k, k = 2L + 1, ..., 3L\}$.

If we take only the first L equations of Eq. (17), the resulting set of linear equations is regular, and the coefficients β are uniquely determined. The obtained matrix is positive definite by

$$\begin{array}{l} \sum_{i,j} (r_{j-i} - r_{j+i})\lambda_i\lambda_j = \\ \sum_{i,j} \int 2\lambda_i\lambda_j \sin(i\omega)\sin(j\omega) S(\omega) d\omega = \\ \int \left(\sum \lambda_i \sin(i\omega)\right)^2 S(\omega) d\omega > 0 \end{array} \tag{18}$$

and the regularity follows for $\lambda \neq 0$. The coefficients found will guarantee that the first L autocorrelation lags of $S(\omega)$ are the given $\{r_k, k = 1, ...L\}$.

The remaining L equations from (17) contain unknown autocorrelation lags of order higher than $2L$. With known β, and under the assumption $\beta_L \neq 0$, the missing $\{r_k, k = 2L+1, ..., 3L\}$ are uniquely determined and can be computed from the remaining set of linear equations by consecutive substitution.

In order for the computed values to be valid autocorrelation lags it is necessary and sufficient that at the same time the determinants $D_j = D_j(r_0, r_1, ..., r_j), j = 1, ..., 3L$ of all upper left square matrices of

$$\begin{bmatrix} r_0 & r_1 & \cdots & r_{2L+1} & \cdots & r_{3L} \\ r_1 & r_0 & & & \cdots & \cdots \\ & & \ddots & & \cdots & \cdots \\ \cdots & & & & \cdots & \cdots \\ r_{2L+1} & & & r_0 & \cdots & \cdots \\ & & & & \ddots & \cdots \\ \cdots & & & & \cdots & \cdots \\ r_{3L} & \cdots & \cdots & & \cdots & r_0 \end{bmatrix} \tag{19}$$

be nonnegative.

Each D_j is a quadratic function of r_j, being nonnegative for values $r_j^{\min} \geq r_j \geq r_j^{\max}$, where $r_j^{\min}$ and $r_j^{\max}$ are the solutions of $D_j = 0$. So, every computed autocorrelation value with lag higher than $2L$ has to belong to this interval.

The constraints on the autocorrelation lags discussed so far are only necessary conditions. In order to verify whether the solution for $S(\omega)$ with a polynomial in the exponent is also the optimum solution, the autocorrelation constraints have to be checked.

3.5. Optimality of the Polynomial Solution

For ψ satisfying the Riccati equation the Fisher information can be computed as

$$\begin{array}{ll} I_f(S) = & \int_{-\pi}^{\pi} \psi^2 S(\omega) d\omega = \\ & 2\int_{-\pi}^{\pi} \psi' S(\omega) d\omega + \int_{-\pi}^{\pi} \sum_{k=0}^{2L} \lambda_k \cos(k\omega) S(\omega) d\omega = \\ & 2I_f(S) + \sum_{k=0}^{2L} \lambda_k r_k \end{array} \tag{20}$$

giving

$$I_f(S) = -\sum_{k=0}^{2L} \lambda_k r_k. \tag{21}$$

If we assume again that the solution of the Riccati equation is of polynomial form (15), then on substituting into (12) we get

$$\sum_{i,j=1}^{L} \beta_i\beta_j \sin(i\omega)\sin(j\omega) = 2\sum_{k=1}^{L} k\beta_k \cos(k\omega) + \sum_{k=0}^{2L} \lambda_k \cos(k\omega) \tag{22}$$

Equating coefficients of $\cos(k\omega)$ terms we get

$$\lambda_k = -2k\beta_k + \frac{1}{2} \sum_{\substack{i,j=1 \\ i-j=k}}^{L} \beta_i\beta_j - \frac{1}{2} \sum_{\substack{i,j=1 \\ i+j=k}}^{L} \beta_i\beta_j \tag{23}$$

which gives

$$I_f(S) = 2\sum_{k=1}^{L} k\beta_k r_k - \sum_{i,j=1}^{N} \beta_i\beta_j \left(\frac{r_{i-j} - r_{i+j}}{2}\right) \tag{24}$$

or in matrix form

$$I_f(S) = 2\tilde{r}^T\beta - \beta^T R\beta \tag{25}$$

The Fisher information $I_f(S)$ is now a quadratic function of the coefficients β and the minimum will be obtained for

$$\tilde{r} = R\beta. \tag{26}$$

This expression is identical to the first L equations of (17). It therefore turns out that the approximate $S(\omega)$ of the form (15), based on (14), is in fact the solution to minimizing Fisher information over an approximate set of test functions satisfying the Ricatti equation with boundary conditions, and the first L autocorrelation constraints (10).

4. SIMULATION EXAMPLES

The newly developed algorithm for minimum Fisher information spectral analysis, given by (15) and (26), was applied to the first 20 autocorrelation lags of an MA(1) process. Figure 1 shows the resulting spectrum together with the true spectrum and the maximum entropy spectrum computed from the same autocorrelation lags.

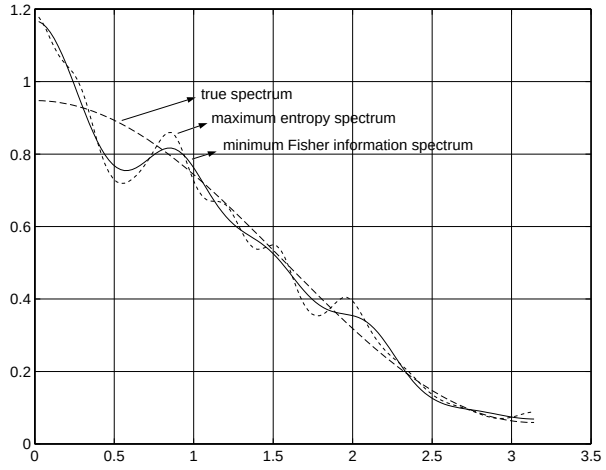

Figure 1. MA(1) Spectrum Estimate from 20 Autocorrelation Lags

Minimizing the Fisher information delivers an improved spectrum estimate compared to the maximum entropy case. Obviously it suffers from the same overparameterization problem for flat spectrum. However, because of the control of the spectrum derivative, the resulting spectrum has a less oscillatory character. Our experiments have shown that for MA models the positive effect of the derivative is stronger for low values of the spectrum.

The second example provided on Figure 2 is a spectrum estimate based on the first 100 autocorrelation lags of an AR(2) spectrum with a pair of complex poles near the unit circle. It

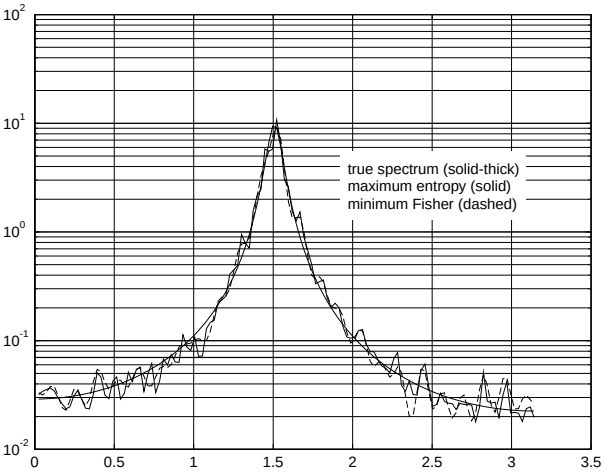

Figure 2. AR(2) Spectrum Estimate from 100 Autocorrelation Lags

is obvious that the maximum entropy and the minimum Fisher information estimates behave very similarly. The overparameterized model introduces in both cases fluctuations around the true value. Further experiments have shown that the minimum Fisher information spectrum is able to follow the peaks of the spectrum with precision comparable to the maximum entropy estimate.

REFERENCES

[1] J. P. Burg, "Maximum entropy spectral analysis," in *37th Annual Int. SEG Meeting, Oklahoma City*, 1967.

[2] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, pp. 661–675, 1971.

[3] T. Ulrych and T. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," *Rev. Geophys.*, vol. 13, no. 1, pp. 183–200, 1975.

[4] E. T. Jaynes, "On the rational of maximum entropy methods," *IEEE Proc.*, vol. 70, pp. 939–952, 1982.

[5] A. Papoulis, "Maximum entropy and spectral estimation: A review," *IEEE Trans. on ASSP*, vol. ASSP-29, pp. 1176–1186, Dec. 1981.

[6] A. van den Bos, "Alternative interpretation of maximum entropy spectral analysis," *IEEE Trans. on Inform. Theory*, vol. IT-17, pp. 193–494, July 1971.

[7] A. Kagan, Y. Linnik, and C. Rao, *Characterization Problems in Mathematical Statistics*. New York: Wiley, 1966.

[8] A. Wragg and D. Dowson, "Fitting continuous probability density functions over [0,oo) using information theory ideas," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 226–230, Mar. 1970.

[9] J. Einbu, "On the existence of a class of maximum-entropy probability density functions," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 772–775, Nov. 1977.

[10] R. Fisher, "Theory of statistical estimation," *Philosoph. Trans. of the Royal Soc. London*, vol. 222, pp. 309–368, 1930.

[11] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Stat.*, vol. 35, pp. 73–104, 1964.

[12] P. J. Huber, "Robust statistics: A review," *Ann. Math. Stat.*, vol. 43, pp. 1041–1067, 1972.

[13] V. Živojnović, "A robust accuracy improvement method for blind identification using higher order statistics," in *IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Minneapolis, USA), pp. 516–519, 1993.

[14] V. Živojnović, "Moment constrained Shannon and Fisher information: Analysis and application on blind identification," in *Proc. of the ATHOS Workshop*, (Edinburgh, UK), 1994.

[15] J. Borwein, A. Lewis, and D. Noll, "Maximum entropy reconstruction using derivative information, part I: Fisher information and convex duality," *Mathematics of Operations Research.* to appear.

[16] J. Borwein, A. Lewis, and D. Noll, "Maximum entropy reconstruction using derivative information, part II: Computational results," *Journal of Numerical Mathematics*, no. 69, pp. 243–256, 1995.

[17] J. Borwein, M. Limber, and D. Noll, "Fast heuristic methods for function reconstruction using derivative information," *Applicable Analysis.* to appear.

[18] E. Kamke, *Differentialgleichungen I.* Geest & Portig, Leipzig, 1967.