where $I$ is the $N \times N$ identity matrix, and $D_u = \text{diag}[\sigma_{u_1}^2, \sigma_{u_2}^2, \cdots, \sigma_{u_N}^2]$ is a real diagonal matrix. Because $H$ has zero principal diagonal, then

$$\sigma_{u_i}^2 = \frac{1}{\gamma_i}, \qquad i = 1, 2, \cdots, N \tag{2}$$

where $\gamma_i$ is the $i$th element on the principal diagonal of $R_x^{-1}$. Define the random vector $\boldsymbol{u} = [u_1 u_2 \cdots u_N]^T$ by

$$\boldsymbol{u} = [I - H]\boldsymbol{x} \tag{3}$$

so that $\boldsymbol{x}$ can be written as

$$\boldsymbol{x} = H\boldsymbol{x} + \boldsymbol{u}. \tag{4}$$

Because $H$ has zero principal diagonal, (4) represents each observation $x_i$ in terms of a weighted sum of the observations $x_j$ for all $j \neq i$ plus an interpolation error $u_i$. Using (3) and (1) one finds that the covariance matrix of observations and interpolation errors $R_{xu} = E\{\boldsymbol{x}\boldsymbol{u}^*\}$ is the diagonal matrix $D_u$

$$R_{xu} = D_u. \tag{5}$$

Consequently, $u_i$ is orthogonal to $x_j$ for all $j \neq i$. Therefore, the $i$th row of $H$ is the LMMSE interpolator of $x_i$ from the $x_j$, $j \neq i$, and $u_i$ is the interpolation error. The covariance matrix of interpolation errors $R_u = E\{\boldsymbol{u}\boldsymbol{u}^*\}$ is, from (3) and (1)

$$R_u = D_u R_x^{-1} D_u. \tag{6}$$

The $i$th element on the principal diagonal of $D_u R_x^{-1} D_u$ is, by (2), $\sigma_{u_i}^2$. Therefore, $\sigma_{u_i}^2$ is the LMMSE interpolation error variance.

As described by Therrien [1], the covariance matrix $R_x$ can be factored as

$$R_x = L D_w L^* \tag{7}$$

where $D_w = \text{diag}[\sigma_{w_1}^2, \sigma_{w_2}^2, \cdots, \sigma_{w_N}^2]$ is a real diagonal matrix, $L$ is lower triangular with 1's on the diagonal, and * represents conjugate transposition. Therrien shows that the rows of $L_p = L^{-1}$ are the coefficients of LMMSE prediction for orders 0 through $N-1$ and the elements of $D_w$ are the corresponding prediction error variances. More specifically, he shows that $L_p$ has the form

$$L_p = \begin{bmatrix} 1 & & & & \\ -\tilde{\boldsymbol{a}}_1^* & 1 & & 0 & \\ -\tilde{\boldsymbol{a}}_2^* & \rightarrow & 1 & & \\ \vdots & & & & \\ \leftarrow & & & -\tilde{\boldsymbol{a}}_{N-1}^* \rightarrow & 1 \end{bmatrix} \tag{8}$$

where $\tilde{\boldsymbol{a}}_k$ represents the reversal of $\boldsymbol{a}_k$ (coefficients written in reverse order), and $\boldsymbol{a}_k$ is the vector of linear prediction coefficients for a predictor of order $k$.

The relation between triangular matrix decomposition and linear interpolation follows directly from (7) and (1). Because

$$R_x^{-1} = L^{*-1} D_w^{-1} L^{-1} = L_p^* D_w^{-1} L_p \tag{9}$$

then

$$H = I - D_u L_p^* D_w^{-1} L_p. \tag{10}$$

The extension of these results to the two-dimensional interpolation problem is straightforward. The special case arising for wide sense Markov processes is the topic of [2], [3].

### References


[1] C. W. Therrien, "On the relation between triangular matrix decomposition and linear prediction," *Proc. IEEE*, vol. 71, no. 12, pp. 1459–1460, Dec. 1983.

[2] J. A. Stuller and B. Kurz, "Two-dimensional Markov representations of sampled images," *IEEE Tran. Commun.*, vol. COM-24, pp. 1148–1152, Oct. 1976.

[3] J. A. Stuller, "Correction to two-dimensional Markov representations of samples images," *IEEE Trans. Commun.*, vol. COM-32, p. 744, June 1984.


# An Information-Theoretic Proof of Burg's Maximum Entropy Spectrum


B. S. CHOI AND THOMAS M. COVER



*It is known that the maximum entropy stationary Gaussian stochastic process, subject to a finite number of autocorrelation constraints, is the Gauss–Markov process of appropriate order. The associated spectrum is Burg's maximum entropy spectral density. We pose a somewhat broader entropy maximization problem, in which stationarity and normality are not assumed, and shift the burden of proof from the previous focus on the calculus of variations and time series techniques to a string of information-theoretic inequalities. This results in an elementary proof of greater generality.*


## I. Preliminaries

Let $\{X_i\}_{i=1}^{\infty}$ be a stochastic process specified by its marginal probability density functions $f(x_1, x_2, \cdots, x_n)$, $n = 1, 2, \cdots$. Then the (*differential*) *entropy* of the $n$-sequence $X_1, X_2, \cdots, X_n$ is defined by

$$h(X_1, X_2, \cdots, X_n) = -\int f(x_1, \cdots, x_n) \cdot \ln f(x_1, \cdots, x_n)\, dx_1\, dx_2 \cdots dx_n = h(f). \tag{1}$$

The stochastic process $\{X_i\}$ will be said to have an *entropy rate*

$$h = \lim_{n \to \infty} \frac{h(X_1, X_2, \cdots, X_n)}{n} \tag{2}$$

if the limit exists. It is known that the limit always exists for stationary processes.

## II. History

Previous characterizations of the maximum entropy spectral density assume that the process is stationary and Gaussian. For such a process, the entropy rate is given [1] by

$$h = \frac{1}{2}\ln(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln(2\pi S(\lambda))\, d\lambda \tag{3}$$

where the spectral density $S(\lambda)$ is given by

$$S(\lambda) = \frac{1}{2\pi}\sum_{\ell=-\infty}^{\infty} \sigma(\ell) e^{-i\lambda\ell} \tag{4}$$

and $\{\sigma(\ell)\}_{\ell=-\infty}^{\infty}$ is an arbitrary autocovariance function subject to the constraints $\sigma(0) = \alpha_0, \cdots, \sigma(p) = \alpha_p$. Burg [2], [3] was first to find the maximum of (3). Subsequent proofs were exhibited in [4]–[13]. These proofs are deeper than the proof of Theorem 1 that we give in the next section, perhaps because of the complexity of the functional in (3). Such techniques as the calculus of variations, complex integration, and linear prediction theory are used in the proofs. A fuller set of references and further details can be found in [14].

We shall see that the entropy maximization can be captured in the information-theoretic string of inequalities in (6) of the next section.

## III. Theorem and Proof

We prove the following theorem.

*Theorem 1:* The stochastic process $\{X_i\}_{i=1}^{\infty}$ that maximizes the differential entropy rate $h$ subject to the autocorrelation constraints


Manuscript received October 27, 1983. This work was partially supported by the National Science Foundation under Grant ECS82-11568 and by the Joint Services Electronics Program DAAG29-81-K-0057.



B. S. Choi was with the Department of Statistics, Stanford University, Stanford, CA 94305, USA. He is now with Yonsei University, Seoul, Korea.

T. M. Cover is with the Department of Statistics and the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA.



0018-9219/84/0800-1094$01.00 ©1984 IEEE

$$EX_iX_{i+k} = \alpha_k, \qquad \begin{array}{l} k = 0,1,2,\cdots,p \\ i = 1,2,\cdots \end{array} \tag{5}$$

is the $p$th order Gauss–Markov process satisfying these constraints.

*Remark:* We do not assume that $\{X_i\}$ is a) zero mean, b) Gaussian, or c) wide-sense stationary. Nonetheless, the maximum entropy process over the larger set of stochastic processes remains the same. Apparently this wider maximization forces a natural proof, while an (unnecessary) assumption of Gaussianity leads to the maximization of (3), which is a difficult result to prove rigorously.

*Proof:* Let $X_1, X_2, \cdots, X_n$ be any collection of random variables satisfying (5). Let $Z_1, Z_2, \cdots, Z_n$ be zero mean multivariate normal with a covariance matrix given by (5), and let $Z_1', Z_2', \cdots, Z_n'$ be the $p$th-order zero-mean Gauss–Markov process satisfying (5). Then, for $n \geqslant p$

$$\begin{aligned}
&h(X_1, X_2, \cdots, X_n) \\
&\overset{(a)}{\leqslant} h(Z_1, Z_2, \cdots, Z_n) \\
&\overset{(b)}{=} h(Z_1, Z_2, \cdots, Z_p) + \sum_{k=p+1}^{n} h(Z_k | Z_{k-1}, \cdots, Z_1) \\
&\overset{(c)}{\leqslant} h(Z_1, Z_2, \cdots, Z_p) + \sum_{k=p+1}^{n} h(Z_k | Z_{k-1}, Z_{k-2}, \cdots, Z_{k-p}) \\
&= h(Z_1', Z_2', \cdots, Z_p') + \sum_{k=p+1}^{n} h(Z_k' | Z_{k-1}', \cdots, Z_{k-p}') \\
&\overset{(d)}{=} h(Z_1', Z_2', \cdots, Z_p') + \sum_{k=p+1}^{n} h(Z_k' | Z_{k-1}', \cdots, Z_1') \\
&= h(Z_1', Z_2', \cdots, Z_n').
\end{aligned} \tag{6}$$

Here (b) is the chain rule for entropy, and inequality (c) follows from the conditional entropy inequality $h(A|B,C) \leqslant h(A|B)$. Equality (d) follows from the Markovity of $\{Z_i'\}$. Inequality (a) follows from the information inequality

$$D(f \| \phi) \triangleq \int f \ln \frac{f}{\phi} \geqslant 0$$

where $f(\boldsymbol{x})$ is the density of $\boldsymbol{x} \in \boldsymbol{R}^n$, and $\phi(\boldsymbol{x}) = (1/(2\pi)^{n/2}|K|^{1/2}) e^{-(1/2)\boldsymbol{x}^t K^{-1} \boldsymbol{x}}$ is the $n$-variate normal density with mean 0 and covariance $K = \int \boldsymbol{x}\boldsymbol{x}^t f(\boldsymbol{x})\, d\boldsymbol{x}$ given by the *autocorrelation* matrix for $f$. (See Berger [1, p. 108, Theorem 4.5.1].)

Thus the $p$th-order Gauss–Markov process $Z_1', Z_2', \cdots, Z_n'$ with covariances $\alpha_0, \alpha_1, \cdots, \alpha_p$ has higher entropy $h(Z_1', Z_2', \cdots, Z_n')$ than any other process satisfying the autocorrelation constraints. Consequently

$$\lim_{n \to \infty} \frac{1}{n} h(X_1, \cdots, X_n) \leqslant \lim_{n \to \infty} \frac{1}{n} h(Z_1', \cdots, Z_n') = h \tag{7}$$

for all stochastic processes $\{X_i\}$ satisfying the autocorrelation constraints, thus proving the theorem.

## IV. Equivalent Characterizations of the Solution

It is straightforward to provide an explicit characterization of the maximum entropy process specified in the theorem. The $p$th-order Gauss–Markov process $\{X_i\}$ satisfying the correlation constraints is given by

$$X_n = -\sum_{i=1}^{p} a_i X_{n-i} + Z_n \tag{8}$$

where $Z_1, Z_2, \cdots$ are i.i.d. normal random variables with mean 0 and variance $\sigma^2$, and $a_1, a_2, \cdots, a_p$ and $\sigma^2$ are determined by the Yule–Walker equations

$$\sum_{j=0}^{p} a_j \alpha_{\ell-j} = 0, \qquad \ell = 1, 2, \cdots, p \tag{9}$$

where

$$a_0 = 1 \qquad \sigma^2 = \sum_{j=0}^{p} a_j \alpha_j.$$

Then $\{X_i\}_{-\infty}^{\infty}$ is the desired maximum entropy process. Equations (8) and (9) force the remaining autocovariance values to satisfy

$$\alpha_\ell = \sum_{j=1}^{p} a_j \alpha_{\ell-j}, \qquad \ell \geqslant p+1. \tag{10}$$

Thus as was observed by Burg, the maximum entropy stochastic process is not obtained by setting the unspecified covariance terms equal to zero, but instead is given by letting the $p$th-order autoregressive process "run" according to the Yule–Walker equations.

The Fourier transform of $\alpha_0, \alpha_1, \cdots$ given in (5) and (10) yields

$$S(\lambda) = \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} \alpha_\ell e^{-i\lambda\ell} = \frac{\sigma^2}{2\pi} \frac{1}{\left| \sum_{j=0}^{p} a_j e^{i\lambda j} \right|^2}. \tag{11}$$

This is Burg's maximum entropy spectral density subject to the covariance constraints $\alpha_0, \alpha_1, \cdots, \alpha_p$.

## V. Conclusion

A bare bones summary of the proof is that the entropy of a finite segment of a stochastic process is bounded above by the entropy of a segment of a Gaussian random process with the same covariance structure. This entropy is in turn bounded above by the entropy of the minimal order Gauss–Markov process satisfying the given covariance constraints. Such a process exists and has a convenient characterization via the Yule–Walker equations. Thus the maximum entropy stochastic process is obtained.

## References


[1] T. Berger, *Rate Distortion Theory, A Mathematical Basis for Data Compression.* Englewood Cliffs, NJ: Prentice-Hall, 1971.

[2] J. P. Burg, "Maximum entropy spectral analysis," in *Proc. 37th Meet. Society of Exploration Geophysicists*, 1967. Reprinted in *Modern Spectrum Analysis*, D. G. Childers, Ed. New York: IEEE Press, 1978, pp. 34–41.

[3] ____, "Maximum entropy spectral analysis," Ph.D. dissertation, Dept. of Geophysics, Stanford University, Stanford, CA, 1975.

[4] D. G. Smylie, G. K. C. Clarke, and T. J. Welch, "Analysis of irregularities in the Earth's relation," *Methods Comput. Phys.*, vol. 13, pp. 391–430 (New York: Academic Press, 1973).

[5] H. Akaike, "An entropy maximization principle," in *Proc. Symp. on Applied Statistics*, P. Krishnaiah, Ed. Amsterdam, The Netherlands: North-Holland, 1977.

[6] J. Grandell, M. Hamrud, and P. Toll, "A remark on the correspondence between the maximum entropy method and the autoregressive model," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 6, pp. 750–751, 1980.

[7] E. A. Robinson, "A historical perspective of spectrum estimation," *Proc. IEEE*, vol. 70, no. 9, pp. 885–907, Sept. 1982.

[8] J. A. Edward and M. M. Fitelson, "Notes on maximum-entropy processing," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 232–234, 1973. Reprinted in *Modern Spectrum Analysis*, D. G. Childers, Ed. New York: IEEE Press, 1978, pp. 94–96.

[9] T. Ulrych and T. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," *Rev. Geophys. Space Phys.*, vol. 13, pp. 183–200, 1975. Reprinted in *Modern Spectrum Analysis*, D. G. Childers, Ed. New York: IEEE Press, 1978, pp. 54–71.

[10] S. Haykin and S. Kesler, "Prediction-error filtering and maximum entropy spectral estimation," in *Nonlinear Methods of Spectral Analysis*, S. Haykin, Ed. New York: Springer, 1979, pp. 9–72.

[11] T. Ulrych and M. Ooe, "Autoregressive and mixed autoregressive-moving average models and spectra," in *Nonlinear Method of Spectral Analysis*, S. Haykin, Ed. New York: Springer, 1979, pp. 73–126.

[12] R. N. McDonough, "Application of the maximum-likelihood method and the maximum entropy method to array processing," in *Nonlinear Methods of Spectral Analysis*, S. Haykin, Ed. New York: Springer, 1979, pp. 181–244.

[13] A. Van den Bos, "Alternative interpretation of maximum entropy spectral analysis," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 493–494, 1971.

[14] B. S. Choi and T. Cover, "An information theoretic proof of Burg's maximum entropy spectrum," Tech. Rep. 49, Statistics Dept., Stanford Univ., Stanford, CA, Oct. 1983.

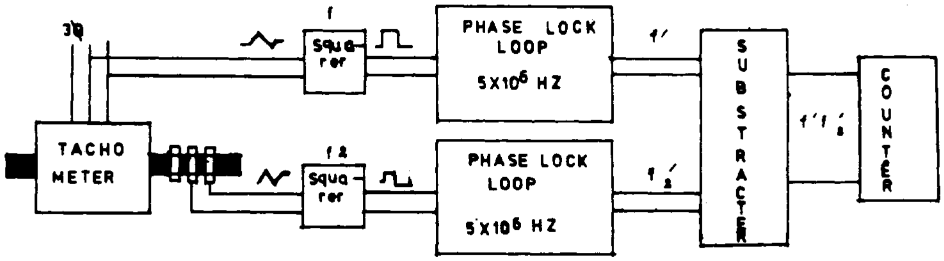


**Fig. 1.** Block diagram of the measuring system.

# A Digital Tachometer for Measurement of Very Low Speeds

MUKHTAR AHMAD

*An induction-type digital tachometer in which the number of pulses is proportional to the speed is described. Even for very low speeds the number of pulses is high, making it very suitable for extremely low speed measurement.*

## I. Introduction

There are many industries which require drives running on very low speeds (in the range of 1–5 r/min). The measurement of such a low speed particularly for the purpose of control is still one of the major problem. A simple method based on the principle of induction is suggested here.

This tachometer works on the principle of an induction machine. It must have windings on the stator as well as rotor. One of the winding is connected to the supply. When the rotor of this tachometer is coupled to the drive, the frequency of EMF induced in the rotor winding will depend on the speed of the rotor. Since the speed of the drive is very low, the frequency of the induced EMF is in the same range as that of applied voltage. This frequency can be easily measured using counters.

## II. Basic Theory

The basic theory underlying the operation of this tachometer is the same as that of a doubly fed induction motor. If a supply is given to the stator of a slip-ring motor and the rotor is rotated with the help of a prime mover, then the frequency of the induced EMF in the rotor circuit is a function of rotor speed [1]. For example, if the machine is wound for $P$ number of poles, its stator is connected to a supply of frequency $f$ hertz and its rotor is rotating at the speed of $N$ revolutions per minute, then frequency of the induced EMF in the rotor winding is given by

$$f_2 = \left( f \pm \frac{PN}{120} \right) \tag{1}$$

or

$$f_2 - f = \frac{PN}{120} \tag{2}$$

or the difference in the two frequencies is proportional to the speed of the drive. Since the supply frequency is substracted, and its effect on the rotor frequency is nullified, any variation in the supply frequency will not affect the operation of this tachometer.

## III. Implementation

The block diagram of the components used for the measurement of low speed is shown in Fig. 1. The construction of the tachometer is similar to that of a slip-ring induction motor. The stator must have three-phase windings and can be connected to a three-phase supply. The rotor has only a one-phase winding. The number of turns in the stator and rotor winding can be made equal, so that the magnitude of the EMF induced is the same.

The signals obtained from the stator and rotor windings of the tachometer are first converted into pulses. Since the speed of the drive is low, the frequency of these pulses will be in the range of supply frequency. Using a source of high frequency, these frequencies are multiplied by a factor of $10^5$ and then locked to a reference signal of the same order using phase-lock loops [2]. Using a substracter the difference between the number of pulses in the rotor and stator circuit is obtained. If the rotor is stationary, the difference will be zero and a counter is used to count the number of pulses during a particular interval of time.

*Example*

Suppose $P = 6$, $F = 50$ Hz, and $N = 1$ r/min, then $f_2 = 50 \pm 0.05$ (the sign depends on the phase sequence of stator supply).

After multiplication by a factor of $10^5$, the two frequencies become as given below

$$f_2' = f_2 \times 10^5 = 50.05 \times 10^5 \text{ (taking the +sign)}$$

and

$$f' = f \times 10^5 = 50 \times 10^5.$$

The difference $f_2' - f' = 5050$ pulses/s. However, if the counter counts the pulses for a time of 20 ms only, it will count 100 pulses and indicates the speed of 1 r/min. Now suppose the speed is changed by 1 percent or the new speed is 1.01 r/min. Then

$$f_2 = 50 + \frac{6 \times 1.01}{120} = 50.0505$$

and $f_2' - f' = 5050$ pulses/s or 101 pulses in 20 ms. Therefore, this tachometer can measure the speed of 1/100 r/min accurately.

## IV. Response to Speed Variation

The response of this speed measuring system to the variation in speed is quite fast as will be shown here. This is an important property for adoption if this system is in a closed-loop condition. As previously discussed, the tachometer is essentially a generator, however, its field system is being rotated at constant speed, depending on the supply frequency and the number of poles. For the above example the speed of the rotating field is $(120 \times f)/P = 1000$ r/min = 16.66 r/s. Since the rotor is rotating at a very low speed, we can neglect its speed in comparison with the speed of a rotating field. Therefore, the time taken up by 1 revolution of the field = 60 ms. Hence, any change in speed will be detected in less than 60 ms.

## V. Conclusion

A very simple type of digital tachometer is described. It is shown by an example that the system is quite accurate and fast. It will be a very useful device for the measurement and control of low-speed drives.

## References

[1] B. M. Bird, and R. F. Burbidge, "Analysis of doubly fed slip ring machines," *Proc. Inst. Elec. Eng.*, vol. 113, pp. 1016–1021, June 1966.
[2] F. Gardner, *Phase Lock Techniques.* New York: Wiley, 1966.

Manuscript received September 15, 1983; revised December 21, 1983.
The author was with the School of Engineering, The Polytechnic, Owo, Nigeria. He is now with the Department of Electrical Engineering, Aligarh Muslim University, Aligarh, India.