in [7] was, unfortunately, based on the incorrect expression given in [5]. For example, the problem

$$p_1 = 0.22 \quad q_1 = 0.98$$
$$p_2 = 0.05 \quad q_2 = 0.8$$
$$p_3 = 0.01 \quad q_3 = 0.71$$

has $P_e(1,2) > P_e(2,3)$ even though the condition in [7] is satisfied.

The actual necessary and sufficient condition for $P_e(1,2) < P_e(2,3)$ is

$$|z_{23}| - |z_{12}| < P_e(3) - P_e(1).$$

An example of this condition is given by the following problem:

$$p_1 = 0.15 \quad q_1 = 0.95$$
$$p_2 = 0.5 \quad q_2 = 0.8$$
$$p_3 = 0.35 \quad q_3 = 0.6.$$

We have

$$P_e(1) < P_e(1,2) < P_e(1,3) < P_e(2,3) < P_e(2) < P_e(3).$$

The Bayes rule for this problem does not use two variables. Thus, both variables X_2 and X_3 degrade performance even though they are independent of X_1. The best single variable is no worse than the best pair. Equation (3) shows that one of the variables in a pair must be worse than the pair itself. This ordering is an example of a previous result [4] that shows that the addition of an independent binary variable need not decrease the error probability.

Cover [2] examines a repeated experiment, which is equivalent to imposing identical distributions on two variables. The error probability for two independent and identically distributed variables X_1 and X_1' is

$$P_e(1,1') = (1/2)(1 - l_1 - 2|l_1 h_1|) < P_e(1) = P_e(1').$$

The example in [2] results in

$$P_e(2,2') < P_e(1,2) < P_e(1,1') = P_e(1) < P_e(2).$$

Thus, repeating the poorer individual experiment provides more class discrimination than repeating the better one.

IV. Conclusion

We conclude that feature extraction with dichotomous features is fundamentally different from feature extraction with continuous features and must be approached with care.

References

[1] W. G. Cochran, "On the performance of the linear discriminant function," *Bull. Int. Statist. Inst.*, vol. 39, pp. 435–447, 1962.

[2] T. M. Cover, "The two best independent measurements are not the two best," *IEEE Trans. Syst., Man, Cybern*, vol. SMC-4, pp. 116–117, 1974.

[3] T. M. Cover and J. M. Van Campenhout, "On the possible orderings in the measurement selection problem," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, pp. 657–661, 1977.

[4] R. P. Duin, C. E. van Haersmabuma, and L. Roosma, "On the evaluation of independent binary features," *IEEE Trans. Inform. Theory*, vol. IT-24, p. 248, 1978.

[5] J. D. Elashoff, R. M. Elashoff, and G. E. Goldman, "On the choice of variables in classification problems with dichotomous variables," *Biometrika*, vol. 54, pp. 668–670, 1967.

[6] G. S. Fang, "A note on optimal selection of independent observables," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 309–311, 1979.

[7] G. T. Toussaint, "A note on the optimal selection of independent, binary-valued features for pattern recognition," *IEEE Trans. Inform. Theory*, vol. IT-17, p. 618, 1971.

Estimating the Frequency of a Noisy Sinusoid by Linear Regression

STEVEN A. TRETTER, SENIOR MEMBER, IEEE

Abstract—**The estimation of the parameters of a sinusoid from observations of signal samples corrupted by additive noise is investigated. At high signal-to-noise ratios the additive noise is viewed as an equivalent phase noise, suggesting frequency and phase estimation by linear regression on the signal phase. The variances of the regression estimates are shown to achieve the Cramer–Rao bounds. A formula for the variance of the regression frequency estimator is derived in terms of the noise power spectrum. A simple formula for the variance with $1/f^2$ phase noise is presented.**

I. Introduction

This correspondence examines a method for estimating the frequency and phase of a sinusoid from a sequence of uniformly spaced signal samples corrupted by additive Gaussian noise. The method is linear regression on the instantaneous signal phase. The variances of the estimates are shown to be identical to the Cramer–Rao bounds [1] at high signal-to-noise ratios (SNR's). Frequency and phase estimation has been studied extensively in the past. However, the linear regression technique presented here provides some new insight into the physical significance of the Cramer–Rao bounds. It is relatively simple computationally.

A variety of techniques have been proposed for frequency and phase estimation. The maximum-likelihood method has been examined by Slepian [2] and Kelly *et al.* [3] for continuous-time observations. More recently, Rife and Boorstyn [4], [5] have examined the single-tone and multitone estimation problem for discrete-time observations. Standard techniques for FM demodulation such as the limiter-discriminator, the zero crossing detector, and differential or product detector can be used. Lucky *et al.* [6, ch. 8] describe these methods and include a list of significant references. Lank *et al.* [7] examine a discrete-time technique for frequency estimation based on product detection. Frequency estimation from discrete-time observations using autoregressive models has been studied under a variety of names, including all-pole models, linear prediction, maximum entropy estimation, and maximum-likelihood whitening filter. Tufts and Kumaresan [8] summarize and extend this method to achive results close to the Cramer–Rao bound. They include an extensive list of references. This is just a small sample of the literature on frequency and phase estimation.

The low SNR threshold where the variances of parameter estimates increase rapidly above the Cramer–Rao bound is an important property of estimators. The threshold behavior of the linear regression frequency estimator is not investigated in this correspondence. However, this estimator has been successfully used with actual received radio signals with SNR greater than 15 dB. The threshold behavior needs to be investigated and compared with other methods.

The observed sequence is assumed to have the form

$$r(n) = Ae^{j(\omega_0 nT + \theta)} + z(n), \quad \text{for } n = n_0, n_0 + 1, \cdots, n_0 + N - 1, \quad (1)$$

Manuscript received February 4, 1985; revised April 12, 1985. The material in this correspondence was partially presented at the Conference on Information Sciences and Systems, Princeton, University, Princeton, NJ, March 1984. This work was supported by System Engineering and Development Corporation, PO Box 666, Columbia, MD 21045.

The author is with the Electrical Engineering Department, University of Maryland, College Park, MD 20742, USA.

0018-9448/85/1100-0832$01.00

where $z(n)$ is a white Gaussian complex noise sequence; $\operatorname{var} z(n) = E\{|z(n)|^2\} = \sigma_z^2$; A, ω_0, and θ are unknown constants; and T is the sampling period. The SNR is defined as

$$\mathrm{SNR}_r = A^2/\sigma_z^2. \tag{2}$$

In Section II it is shown how for large SNR_r the additive noise can be converted into an equivalent additive phase noise. This suggests estimating ω_0 and θ by linear regression on the observed signal phase. The estimator variances in the case of white noise are derived and found to be identical to the Cramer–Rao bounds. In Section III a formula for the variance of the linear regression frequency estimator of II is derived for the case where the noise has an arbitrary power spectral density. A simple formula for the variance of the linear regression frequency estimator is derived in Section IV for the case of $1/f^2$ phase noise.

II. The High SNR Approximation

The observed signal can be expressed as

$$r(n) = [1 + v(n)]Ae^{j(\omega_0 nT + \theta)}, \tag{3}$$

where

$$v(n) = \frac{1}{A} z(n) e^{-j(\omega_0 nT + \theta)} \tag{4}$$

is a complex white noise sequence with

$$\operatorname{var} v(n) = \sigma_z^2/A^2 = 1/\mathrm{SNR}_r. \tag{5}$$

Let $v(n) = v_I(n) + jv_Q(n)$. Then

$$1 + v(n) = \left\{[1 + v_I(n)]^2 + v_Q^2(n)\right\}^{1/2} \cdot \exp\left\{ j \tan^{-1} \frac{v_Q(n)}{1 + V_I(n)} \right\}. \tag{6}$$

For $\mathrm{SNR}_r \gg 1$

$$1 + v(n) \cong \exp\left[j \tan^{-1} v_Q(n)\right] \cong \exp\left[jv_Q(n)\right], \tag{7}$$

so

$$r(n) \cong Ae^{j[\omega_0 nT + \theta + v_Q(n)]}. \tag{8}$$

Thus, the additive noise has been converted into an equivalent phase noise $v_Q(n)$ with

$$\operatorname{var} v_Q(n) = 0.5 \operatorname{var} v(n) = \frac{1}{2\,\mathrm{SNR}_r}. \tag{9}$$

All the information required to estimate ω_0 and θ is contained in the phase angle

$$\phi(n) = \omega_0 nT + \theta + v_Q(n). \tag{10}$$

This angle can be computed by applying a phase unwrapping algorithm [9]–[12] to the principal value of $\arg r(n)$ obtained by using an inverse tangent.

The parameters ω_0 and θ can be estimated by the method of least squares or linear regression. The least-squares estimates are equivalent to maximum-likelihood estimates when the noise is Gaussian. The parameters that minimize the square error

$$\Lambda = \sum_{n=n_0}^{n_0+N-1} \left[\phi(nT) - \hat{\omega}_0 nT - \hat{\theta}\right]^2 \tag{11}$$

are

$$\begin{bmatrix} \hat{\omega}_0 \\ \hat{\theta} \end{bmatrix} = \frac{12}{T^2N^2(N^2-1)} \cdot \begin{bmatrix} N & -T(Nn_0 + P) \\ -T(Nn_0+P) & T^2(Nn_0^2 + 2n_0P + Q) \end{bmatrix} \cdot \begin{bmatrix} \sum_{n=n_0}^{n_0+N-1} nT\phi(n) \\ \sum_{n=n_0}^{n_0+N-1} \phi(n) \end{bmatrix}, \tag{12}$$

where

$$P = \sum_{n=0}^{N-1} n = (N-1)N/2$$

and

$$Q = \sum_{n=0}^{N-1} n^2 = (N-1)N(2N-1)/6.$$

These estimators are unbiased. The error covariance matrix is

$$\operatorname{cov}\begin{bmatrix} \hat{\omega}_0 \\ \hat{\theta} \end{bmatrix} = \frac{6}{\mathrm{SNR}_r T^2N^2(N^2-1)} \cdot \begin{bmatrix} N & -T(Nn_0 + P) \\ -T(Nn_0+P) & T^2(Nn_0^2 + 2n_0P + Q) \end{bmatrix}. \tag{13}$$

The variances of $\hat{\theta}$ and $\hat{\omega}$ are exactly the same as the Cramer–Rao bounds [1], [4].

The estimators depend on the choice of the initial time n_0. It is particularly convenient to choose n_0 to diagonalize the square matrix in (12), that is,

$$n_0 = -P/N = -(N-1)/2. \tag{14}$$

Then $r(0)$ becomes the center sample in the observed sequence. With this choice

$$\hat{\omega}_0 = \frac{12}{TN(N^2-1)} \sum_{n=-(N-1)/2}^{(N-1)/2} n\phi(n) \tag{15}$$

and

$$\hat{\theta} = \frac{1}{N} \sum_{n=-(N-1)/2}^{(N-1)/2} \phi(n). \tag{16}$$

The corresponding error covariance matrix is

$$\operatorname{cov}\begin{bmatrix} \hat{\omega}_0 \\ \hat{\theta} \end{bmatrix} = \begin{bmatrix} \dfrac{6}{\mathrm{SNR}_r T^2 N(N^2-1)} & 0 \\ 0 & \dfrac{1}{2N\mathrm{SNR}_r} \end{bmatrix}. \tag{17}$$

Now the frequency and phase estimation errors are uncorrelated. The variance of $\hat{\theta}$ is also minimized for this choice of n_0. This choice makes the basis functions 1 and n orthogonal over the interval $[n_0, n_0 + N - 1]$.

Rife and Boorstyn [4] have observed the dependence of the Cramer–Rao bounds on the initial time n_0. The derivation in this section shows that at high SNR's the problem is equivalent to estimating the slope and intercept of a linear ramp corrupted by additive noise, which explains this behavior.

The estimators (15) and (16) are simple to compute given the instantaneous signal phase. At high SNR's simple phase unwrapping algorithms can be successfully used. The maximum-likelihood estimators require finding the peak of the periodogram of the observed signal sequence. Rife and Boorstyn [4] suggest first using a fast Fourier transform to find a coarse estimate of the peak location and then refining the estimate with a numerical search procedure. The autoregressive model technique requires lag product computations, an effective matrix inversion, and root finding. The linear regression estimators do not require any of these computations, but only an arc tangent, phase unwrapping, and the simple sums (15) and (16). In some applications it is useful to plot the instantaneous signal phase to see various anomalies such as periodic variations in phase slope or sudden changes in SNR. This sequence is generated for the linear regression estimators but is hidden in the maximum-likelihood and autoregressive methods.

III. Variance of the Linear Regression Frequency Estimator with Colored Noise

The linear regression frequency estimator given by (15) can be used even when the noise is not a white noise sequence. A formula for the variance of $\hat{\omega}_0$ in terms of the power spectral density for $v_Q(n)$ will be derived in this section.

Let $v_Q(n)$ have the autocorrelation function

$$R(n) = E\{v_Q(k+n)v_Q(k)\} \tag{18}$$

and power spectral density

$$S(\omega) = \sum_{n=-\infty}^{\infty} R(n)e^{-j\omega nT}. \tag{19}$$

Also, let

$$n_0 = -\frac{N-1}{2} = -M. \tag{20}$$

Since

$$\omega_0 = \frac{12}{TN(N^2-1)} \sum_{n=-M}^{M} n(\omega_0 nT + \theta), \tag{21}$$

the estimation error can be expressed as

$$\hat{\omega}_0 - \omega_0 = \frac{12}{TN(N^2-1)} \sum_{n=-M}^{M} n v_Q(n). \tag{22}$$

So

$$\operatorname{var}\hat{\omega}_0 = E\{(\hat{\omega}_0-\omega_0)^2\} = \left[\frac{12}{TN(N^2-1)}\right]^2 \sum_{n=-M}^{M}\sum_{k=-M}^{M} nkR(n-k). \tag{23}$$

In terms of the window function

$$g(n) = \begin{cases} n & \text{for } |n| \le M \\ 0, & \text{elsewhere} \end{cases}, \tag{24}$$

(23) becomes

$$\operatorname{var}\hat{\omega}_0 = \left[\frac{12}{TN(N^2-1)}\right]^2 \sum_{k=-\infty}^{\infty} g(k) \sum_{n=-\infty}^{\infty} g(n)R(n-k). \tag{25}$$

According to Parseval's theorem

$$\sum_{n=-\infty}^{\infty} g(n)R(n-k) = \frac{1}{\omega_s}\int_{-\omega_{s/2}}^{\omega_{s/2}} S(\omega)e^{-j\omega kT}G(-\omega)\,d\omega, \tag{26}$$

where $\omega_s = 2\pi/T$. Substituting (26) into (25) yields

$$\operatorname{var}\hat{\omega}_0 = \left[\frac{12}{TN(N^2-1)}\right]^2 \frac{1}{\omega_s}\int_{-\omega_{s/2}}^{\omega_{s/2}} S(\omega)|G(\omega)|^2\,d\omega. \tag{27}$$

To compute $G(\omega)$, first observe that

$$H(\omega) = \sum_{n=-M}^{M} e^{-j\omega nT} = \frac{\sin\frac{\omega NT}{2}}{\sin\frac{\omega T}{2}}. \tag{28}$$

Then

$$G(\omega) = \sum_{n=-M}^{M} ne^{-j\omega nT} = \frac{j}{T}\frac{d}{d\omega}H(\omega) = \frac{j}{T\sin\frac{\omega T}{2}}\left[\frac{NT}{2}\cos\frac{\omega NT}{2} - \frac{T}{2}\frac{\sin\frac{\omega NT}{2}\cos\frac{\omega T}{2}}{\sin\frac{\omega T}{2}}\right]. \tag{29}$$

Substituting (29) into (27) gives the desired result,

$$\operatorname{var}\hat{\omega}_0 = \left[\frac{6N}{T(N^2-1)}\right]^2 \frac{1}{\omega_s}\int_{-\omega_{s/2}}^{\omega_{s/2}} \frac{S(\omega)}{N^2\sin^2\frac{\omega T}{2}} \cdot\left[\cos\frac{\omega NT}{2} - \cos\frac{\omega T}{2}\frac{\sin\frac{\omega NT}{2}}{N\sin\frac{\omega T}{2}}\right]^2 d\omega. \tag{30}$$

IV. A Simple Error Variance Formula for $1/f^2$ Noise

An oscillator often has a slow drift that can be modeled by a random walk. This component is known as divergent noise and has a power spectral density that behaves like $1/\omega^2$. The error variance of the linear regression frequency estimate (15) resulting from divergent noise could be calculated by (30). However, a simple formula for the variance will be derived in this section by a different method.

Suppose the observed phase has the form

$$\phi(n) = \omega_0 nT + \theta + x(n), \tag{31}$$

where $x(n)$ is the divergent phase noise with autocorrelation function $R_x(n)$ and power spectral density

$$S(\omega) = \frac{\sigma_x^2}{4\left(\sin\frac{\omega T}{2}\right)^2} \cong \frac{\sigma_x^2}{\omega^2 T^2}, \quad \text{for } |\omega T| \ll 1. \tag{32}$$

The estimation error for the linear regression frequency estimator (15) can be expressed as

$$\epsilon = \hat{\omega}_0 - \omega_0 = \frac{12}{TN(N^2-1)} \sum_{n=-M}^{M} nx(n). \tag{33}$$

This error can also be written as

$$\epsilon = \frac{6}{TN(N^2-1)} \sum_{n=-M}^{M} [(M+1)M - (n+1)n] \cdot [x(n+1) - x(n)]. \tag{34}$$

This alternative formula can be derived by using the summation-by-parts formula from the calculus of finite differences

$$\sum_{n=n_1}^{n_2} a(n)\Delta b(n) = a(n_2+1)b(n_2+1) - a(n_1)b(n_1) - \sum_{n=n_1}^{n_2} b(n+1)\Delta a(n), \quad (35)$$

where Δ is the finite difference operator defined for an arbitrary sequence $f(n)$ as

$$\Delta f(n) = f(n+1) - f(n). \quad (36)$$

The sum on the left side of (35) becomes the sum on the right side of (33) if $n_2 = -n_1 = M$, $a(n) = x(n)$, and $\Delta b(n) = b(n+1) - b(n) = n$. A solution for $b(n)$ is $b(n) = n(n-1)/2$. Evaluating the right side of (35) and substituting into (33) gives (34).

The error variance is

$$\operatorname{var}\epsilon = \left[\frac{6}{TN(N^2-1)}\right]^2 \sum_{n=-M}^{M}\sum_{k=-M}^{M}[(M+1)M-(n+1)n] \cdot[(M+1)M-(k+1)k]R_\Delta(n-k), \quad (37)$$

where

$$R_\Delta(n) = E\{\Delta x(k+n)\Delta x(k)\} \quad (38)$$

is the autocorrelation function for $\Delta x(n)$. The power spectral density for $\Delta x(n)$ is

$$S_\Delta(\omega) = |e^{j\omega T} - 1|^2 S(\omega) = \sigma_x^2. \quad (39)$$

Thus, $R_\Delta(n) = \sigma_x^2\delta_{n,0}$ and (37) reduces to

$$\operatorname{var}\epsilon \cong \left[\frac{6}{TN(N^2-1)}\right]^2 \sigma_x^2 \sum_{n=-M}^{M}[(M+1)M-(n+1)n]^2 = \frac{3\sigma_x^2}{5T^2}\frac{2M^2+2M+1}{M(M+1)(2M+1)}. \quad (40)$$

For large M

$$\operatorname{var}\epsilon \cong \frac{3}{5MT^2}\sigma_x^2. \quad (41)$$

Also,

$$M = (N-1)/2 \cong N/2, \qquad \text{for large } N,$$

so

$$\operatorname{var}\epsilon \cong \frac{6}{5NT^2}\sigma_x^2 = \frac{6}{5TT_0}\sigma_x^2, \quad (42)$$

where $T_0 = NT$ is the observation time. The estimation error variance behaves as N^{-1} for $1/f^2$ noise but as N^{-3} for a white noise sequence according to (13). This difference is caused by the different noise spectra or, equivalently, autocorrelation functions.

References

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part I. New York: Wiley, 1968.

[2] D. Slepian, "Estimation of signal parameters in the presence of noise," *IRE Trans. Inform. Theory*., vol. IT-3, pp. 68–89, Mar. 1957.

[3] E. J. Kelly, I. S. Reed, and W. L. Root, "The detection of radar echoes in noise," *J. Soc. Ind. Appl. Math.*, pt. 1, vol. 8, no. 2, pp. 309–341, June 1960; pt. 2, no. 3, pp. 481–507, Sep. 1960.

[4] D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 5, pp. 591–598, Sep. 1974.

[5] D. C. Rife and R. R. Boorstyn, "Multiple tone parameter estimation from discrete-time observations," *Bell Syst. Tech. J.*, pp. 1389–1410, Nov. 1976.

[6] R. W. Lucky, J. Salz, and E. J. Weldon, *Principles of Data Communication*. New York: McGraw-Hill, 1968.

[7] G. W. Lank, I. S. Reed, and G. E. Pollon, "A semicoherent detection and doppler estimation statistic," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-9, no. 2, pp. 151–165, Mar. 1973.

[8] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," in *Proc. IEEE*, vol. 70, no. 9, pp. 975–989, Sept. 1982. (Special issue on spectral estimation).

[9] R. W. Schafer, "Echo removal by discrete generalized linear filtering," Tech. Rep. 466, Research Lab of Electronics, Mass. Inst. Tech., Feb. 1969.

[10] J. M. Tribolet, "A new phase unwrapping algorithm," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-25, no. 2, pp. 170–177, Apr. 1977.

[11] K. Steiglitz and B. Dickinson, "Phase unwrapping by factorization," *IEEE Trans. Acoust. Speech, and Signal Processing*, vol. ASSP-30, no. 6, pp. 984–991, Dec. 1982.

[12] C. M. Loeffler and R. E. Leonard, Jr., "Phase unwrapping via median filtering," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, Mar. 1984, pp. 46.8.1–46.8.3.

Approximation of Locally Optimum Detector Nonlinearities

STEVEN V. CZARNECKI, MEMBER, IEEE, AND KENNETH S. VASTOLA, MEMBER, IEEE

***Abstract*—The problem of finding a detector nonlinearity that maximizes the efficacy (or asymptotic processing gain) over a class of suboptimal nonlinearities is considered. It is shown that this efficacy maximization problem is essentially the same as the problem of finding the minimum-mean-square-error approximation to the locally optimal detector nonlinearity. This result is compared with some intuitive ideas about suboptimal detection.**

I. Introduction

An important practical problem arising in detection is the following. Given that the true noise density f and the true detector nonlinearity g_{opt} are known, what is the best way to approximate g_{opt} within some specified constraints? This correspondence provides a solution to this broadly posed question.

The locally optimal (LO) detector structure [1]–[4] is a useful model for the detection of a signal that is known but very small relative to the noise environment. For detecting a (constant) weak discrete-time signal in the presence of white non-Gaussian noise with first-order density f, it is well known that the LO detector consists of a memoryless nonlinearity (ZNL) of the form

$$g_{LO}(x) = -\frac{f'(x)}{f(x)} \quad (1)$$

followed by summation and comparison with a threshold.

Manuscript received June 27, 1984; revised April 29, 1984. This work was supported by the Office of Naval Research under Grant N00014-81-K-0146 and by the National Science Foundation under Grant EC5-83-17777. The material in this correspondence was presented in part at the Conference on Information Sciences, Johns Hopkins University, Baltimore MD, March 1983.

S. V. Czarnecki was with the Department of Electrical Engineering and Computer Science, Princeton University. He is now with IBM Federal Systems Division, Owego, NY 13827, USA.

K. S. Vastola was with the Department of Electrical Engineering and Computer Science, Princeton University, Princeton, NJ. He is now with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, USA.

0018-9448/85/1100-0835$01.00 ©1985 IEEE